PRIORITY MIX PLANNING FOR CYCLE TIME-DIFFERENTIATED SEMICONDUCTOR MANUFACTURING SERVICES

Shi-Chung Chang

Shin-Shyu Su
Ke-Ju Chen

Dept. of Electrical Engineering & Graduate Institute of
Industrial Engineering
National Taiwan University
Taipei, TAIWAN, ROC, 10617

Dept. of Electrical Engineering
National Taiwan University
Taipei, TAIWAN, ROC, 10617

ABSTRACT

Semiconductor fabs often offer manufacturing service of multiple priorities in terms of cycle time-based X-factor targets (XFTs) and fab production must be planned accordingly. This paper studies a priority mix planning (PMP) problem that determines the wafer release rates of individual priorities to maximize fab profit subject to XFT and capacity constraints. It is formulated as a nonlinear programming problem, where the constraints integrate an extended M/G/m:PR queue approximation with contribution theory-based network for modeling how X-factors of individual priorities are affected by priority mix and fab capacity utilization. It is then demonstrated over a realistic fab example that PMP problem can be solved for effectively planning priority mix and machine tool capacity utilization to provide cycle-time differentiated manufacturing services.

1 INTRODUCTION

Effective provision of manufacturing services in multiple priority levels has been one critical aspect to the competitiveness of wafer fabs. A customer order with a higher priority level demands a shorter cycle time than orders of a lower priority (Narahari and Khan 1997). Wafers of lower priority orders have longer cycle times because they need to wait in line for wafers of higher priority to finish processing. Wafer release rates of individual priorities constitute the priority mix to a fab. Priority mix significantly affects fab performance such as throughput, cycle time, wafer-in-process (WIP) and bottleneck location (Kang and Lee 2007).

Among the many fab performance indices, cycle time has a significant impact on productivity learning and customer serviceability. There is a basic relationship among capacity utilization (U) and cycle time (Hopp and Separman 2008; Hu and Chang 2003). The cycle time of a fab increases exponentially with the increase of utilization when utilization goes beyond a high level, say, 90%, while it is proportional to U at a lower U-level (Yang, Ankenman, and Nelson 2006). To measure and manage cycle times, the notion of X-factor (XF), where XF = cycle time/raw processing time (RPT), has been introduced to provide a sensitive performance indicator and is standardized across different products (Hopp and Spearman 2008). It has been shown that many fab operation problems can be effectively identified through the analysis of X-factors. Customized X-factor targets can be set for short cycle time manufacturing (SCM) to allow performance differentiation among machine tool groups of different characteristics and to specify the overall fab performance as well (Hopp and Spearman 2008, Kishimoto et al. 2001, Matrin 1998).

In production planning of a fab, there are different XF target (XFT) specifications for individual priority levels of manufacturing services (Virtamo 2004). The XF of each priority (PXF) is a function of wafer release rates and processing flow requirements of individual priorities and the total utilization of the bottleneck tool group, which we shall refer to as a PXF behavior model. Note that different XFs require different levels of resource allocation and hence lead to different costs and manufacturing services of different XFTs should be priced differently. Both priority mix and product mix will effect cycle times and may cause "floating bottleneck" (Hopp and Spearman 2008). Given a pricing policy, production cost structure, and a set of XFTs, a PXFT constrained production planning decides the priority mix (or wafer release rates) of products in individual manufacturing service priorities for profit maximization subject to machine capacity and PXFT constraints. Key to this planning problem is the behavior modeling of the relationship between PXF and priority mix and capacity utilizations.

In this paper, we formulate and study priority mix planning problem of semiconductor manufacturing. The M/G/m:PR queue approximation-based model (Chang and Chen 2006) is extended and integrated with contribution

theory to form a network model that captures the relationship among profit, MG utilization (U) and priority mix (PM). The fab modeling methodology is focused on capturing how operation priority, production flow variations, and capacity utilizations may affect individual PXFs of a fab. The M/G/m:PR queue model is adopted to model the behavior of a service node (tool group). On top of the single node model, a PXF contribution theory relates PXFs of individual service nodes to the overall fab PXF and provides a novel priority network model. Model fitting is then adopted to compensate, for each priority, the errors caused by assuming Poisson arrival process. The priority mix planning problem (Figure 1) determines the wafer release rates of individual priorities to maximize the fab profits (revenue minus manufacturing and inventory costs) subject to PXFTs and capacity constraints, which is formulated as a nonlinear programming problem. After validation by simulation of a realistic example designed based on fab models in (Lu, Ramaswamy, and Kumar 1994; Lin 1996; Narahari and Khan 1997), numerical studies are performed to solve for the optimal priority mix and to investigate how PXF behavior may affect the solution.
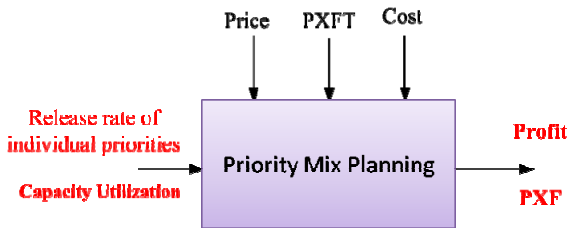

Figure 1: Priority mix planning problem

The remainder of the paper is organized as follows. Section II presents the M/G/m:PR queue approximation and priority contribution theory-based network modeling methodology. Section III then gives a problem formulation of PXFT constrained production planning based on the PXF behavior model. Model analysis and tuning are given in Section IV. In the numerical study of Section V, we obtain some insights about how to plan priority mix and determine the fab utilization to obtain the optimal profit given constraints under different scenarios. Finally, Section VI concludes the paper.

2 PRIORITY X-FACTOR APPROXIMATION MODELS FOR FABS

Cycle time-based PXFTs are the key quality of manufacturing service (QoMS) specifications for priority mix decision in this paper. To plan for the optimal priority mix of a fab, behavior models of the fab with priority differentiation are needed to characterize how cycle times

of individual priorities vary with wafer release rates of individual priorities and utilization levels of machine tools. In this section, the priority queueing network approximation proposed by Chang and Chen 2006 is adopted as the modeling approach. The PXFT models obtained then serve as the cornerstone in formulating priority mix planning problem.

Consider a fab that fabricates semiconductor wafer products in different priorities. Each product requires processing in a specific sequence of steps, named a process flow, by various machine tools. Machines of the same functionality are grouped into machine tool groups (MGs). In semiconductor manufacturing, the process flow of a product is re-entrant, namely, it requires the processing by a MG at a few different steps. The wafers of one product may be assigned different priorities and different products may have the same priority. For simplicity of later discussions in this paper, it is assumed that there is only one type of product in each priority.

To develop fab behavior models, let us first define some notations.

Notations
J: total number of priorities;
j: priority/product index, $j=1,…, J$, the smaller the number the higher the priority;
I: total number of processing steps;
i: step index, $i=1,…, I$;
K: total number of machine tool groups (MGs);
k: MG index, $k=1,…, K$;
m_k: number of heterogeneous tools in MG k;
Ω_k : {(i,j)|Step i of priority j that requires the processing by a machine of MG k};
λ_j : mean release rate of jth priority product;
τ_{ij} : mean processing time of a jth priority wafer at step i, which equals zero if jth priority wafer does not go through step i;
CT_{ij} : average cycle time of a jth priority wafer at step i;
XF_{ij} : X-Factor of jth priority wafer at step $i \equiv CT_{ij}/\tau_{ij}$;
$Var[S_{j|k}]$: variance of processing times of all jth priority steps processed by MG k, which equals zero if jth priority wafer does not go through MG k.

Let a tool group in a fab be a service node. Chang and Chen's priority queueing network approximation (Chang and Chen 2006) approach consists of a modified M/G/m:PR queueing model for PXFs of individual service nodes and a contribution theory describing how individual PXFs contribute to the overall fab XF.

2.1 Single Node Approximation Model

Single-node priority behavior modeling is the cornerstone to fab PXF modeling. The model of PXFs in terms of

wafer release rate, mean and variance of processing times and capacity utilization is based on M/G/m:PR queueing analyses of Virtamo 2004, and Lee, Kang, and Wang 2006. As the Poisson arrival assumption of the M/G/m:PR model may not be a good approximation, the model needs to be modified for a closer match to fab data. A compensation parameter is therefore added to account for the variance of a general, non-Poisson arrival process.

Consider MG k with m_k heterogeneous machines. As the process flow of priority-j wafer is reentrant and may visit a MG k at a few steps, the single node analysis in Virtamo 2004, and Lee, Kang, and Wang 2006 needs to be extended to multiple steps within one priority. Let

$$\tau_{j|k} \equiv \sum_{i \text{ with } (i,j) \in \Omega_k} \tau_{ij} \quad and \quad \rho_{j|k} \equiv \frac{1}{m_k}(\lambda_j \times \tau_{j|k}) \tag{1}$$

Our extended approximation model of XF for step i of priority j at MG k is

$$XF_{ij} \equiv CT_{ij}/\tau_{ij}$$

$$\approx \frac{1}{\tau_{ij}} \frac{\frac{1}{2}\sum_{j'=1}^{J}(\frac{Var[S_{j'|k}]}{m_k^2}+(\frac{\tau_{j|k}}{m_k})^2)\times\lambda_j+\alpha_{j|k}}{(1-\sum_{j'=1}^{j-1}\rho_{j'|k})(1-\sum_{j'=1}^{j}\rho_{j'|k})}+1 \tag{2}$$

where $\alpha_{j|k}$ is a compensation parameter to account for the variance of the general, non-Poisson arrival process to MG k. In equation (2), the re-entrant flows of priority j to MG k are considered as separated flows of priority j with the same average flow rate λ_j. The re-entrant flow effect is accounted for by the definitions of $\tau_{j|k}$ and $\rho_{j|k}$ in (1). For notational simplicity, $\{\lambda_j, j=1,...,J\}$ will be referred to as $\{\lambda_j\}$ in the later discussions.

2.2 Priority XF Contribution Theory

Network relationship among service nodes is then needed for modeling fab PXF of each priority based on nodal models. D.P. Martin (Martin 1998) derived a contribution theory that describes *the contribution by XFs of individual processing steps to fab XF*. The theory exploits the notion of weighted average and that fab cycle time is the sum of cycle times of individual steps in the process flow, namely,

$$CT_{fab} = \text{average processing time across priorities} \times XF_{fab}$$

$$= \sum_{j=1}^{J}\frac{\lambda_j}{\lambda}(\sum_{i=1}^{I}\tau_{ij})\times XF_{fab}$$

$$= \sum_{j=1}^{J}\frac{\lambda_j}{\lambda}CT_j = \sum_{j=1}^{J}\frac{\lambda_j}{\lambda}(\sum_{i=1}^{I}\tau_{ij}\times XF_{ij}), \tag{3}$$

where $\lambda = \sum_{j=1}^{I}\lambda_j$ is the total wafer release rate and

$\sum_{i=1}^{I}\tau_{ij}=\tau_j$ is the total raw processing time of priority j wafers. So, from Eq. (3),

$$XF_{fab} = \frac{\sum_{j=1}^{I}\lambda_j\sum_{i=1}^{I}(\tau_{ij}\times XF_{ij)}}{\sum_{j=1}^{J}\lambda_j\sum_{i=1}^{I}\tau_{ij}} \equiv \Phi_{s2f}(\{\lambda_j\}). \tag{4}$$

where Φ_{s2f} represents the relationship of fab XF with respect to priority mix and individual step XFs.

It is then straightforward to derive from Eq. (4) *the XF of priority j in the fab* as

$$XF_{p2f\_j} \equiv \Phi_{s2f}(\{\lambda_j\}) = \frac{CT_j}{RPT_j} = \frac{\sum_{i=1}^{I}(\tau_{ij\times XF_{ij}})}{\sum_{i=1}^{I}\lambda_j\tau_{ij}} \tag{5}$$

where Φ_{s2f_j} represents the relationship of fab XF with respect to priority mix and individual step XFs and the coefficient term for XF_{ij} can be interpreted as the relative workload incurred by step i of priority j.

The *XF* of MG k is derived by exploiting the relationship between total queue size and queue sizes of individual priorities and the relation between cycle times and wafer release rates of individual priorities according to Little's formula as follows:

$$CT_k \equiv RPT_k \times XF_k$$

$$= [\sum_{j=1}^{J}\frac{\lambda_j}{\lambda}(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij})]\times XF_k$$

$$= \sum_{j=1}^{J}\frac{\lambda_j}{\lambda}\times CT_{j|k} = \sum_{j=1}^{J}\frac{\lambda_j}{\lambda}(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij}\times XF_{ij}).$$

$$XF_k = \sum_{j=1}^{J}\lambda_j(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij}\times XF_{ij})/\sum_{j=1}^{J}\lambda_j(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij})$$

$$\tag{6}$$

The *contributions of priority j to XF of a MG k* and *to fab XF* are then

$$XF_{p2m\_jk} \equiv \frac{\lambda_j(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij}\times XF_{ij})}{\sum_{j=1}^{J}\lambda_j(\sum_{i \text{ with } (i,j)\in\Omega_k}\tau_{ij})\times XF_{ij}} , \text{ and} \tag{7a}$$

$$XFC_{p2f\_j} \equiv \sum_{i=1}^{I}\frac{\lambda_j\tau_{ij}}{\sum_{j=1}^{J}\sum_{i=1}^{I}\lambda_j\tau_{ij}}\times XF_{ij} \tag{7b}$$

3 PRIORITY MIX PLANNING PROBLEM FORMULATION

In fab operations, each priority has a pre-specified PXFT as part of QoMS; the higher the priority, the smaller value of its PXFT. There is also a pricing policy associated with the prioritized manufacturing service, the higher the priority the higher the price rate. Fab operations need to be planned to meet the PXFTs. A fab manager can control XFs of individual priorities of the fab by adjusting priority mix and machine tool utilization levels. The goal is to maximize the profit of such prioritized manufacturing service under the QoMS and machine tool capacity constraints.

Chang and Chen 2006 formulated a nonlinear programming problem of finding priority mix plan to maximize the profit rate of a fab under the differentiated price and cost structures and cycle time targets. To extend their mathematical formulation, let us further define some notations.

Notations

P_j **:** per wafer price of *j*th priority;

C_j^M **:** per wafer manufacturing cost of *j*th priority;

C_j^I **:** per wafer and per unit time WIP cost of *j*th priority.

In manufacturing make-to-order wafers, the revenue rate (revenue per unit time) equals its offered price multiplied by the wafer release rate of the priority. The cost rate (cost per unit time) consists of two sources: WIP and capacity. The WIP at a step is proportional to the release rate and cycle time at the step and the capacity use is proportional to the release rate. The profit per unit time is

$$J = \sum_j (P_j \times \lambda_j - C_j^M \times \lambda_j) - \sum_{i,j} C_j^I \times \lambda_j \times \tau_{ij} \times XF_{ij} \quad (8)$$

There are specified PXFTs for planning priority mix of a fab. In deciding on the release rates or mix of individual priorities to maximize one's manufacturing profit, a fab manager must consider constraints of capacity and PXF targets. Note that both fab XF and per priority XFs are functions of priority mix and capacity utilization. Although a high bottleneck MG utilization implies high throughputs/release rates of individual priorities and hence profit increase (Eq. (8)), it also leads to quick increase in PXFs as indicated in Eqs. (4) - (7). Under a fixed bottleneck MG utilization level, release rate increase in a high priority and the corresponding decrease in a low priority may lead to revenue gain, but PXFs of both high and low priorities may increase with a higher impact on the low priority. The tradeoff among release rates, utilization and XFs for maximum profit is nonlinear. A PXFT constrained priority mix planning problem is formulated as follows:

$$(PMP)$$

$$\underset{\{\lambda_j\}}{Max} \quad [J = \sum_j Revenue_j(\lambda_j) - Cost_j(\{XF_{ij}\}, \{\lambda_j\})] \quad (9)$$

$$= \sum_j (P_j \times \lambda_j - C_j^M \times \lambda_j) - \sum_{i,j} C_j^I \times \lambda_j \times \tau_{ij} \times XF_{ij}$$

subject to

$$\Phi_{s2f}(\{\lambda\}) \le XFT_{fab} \quad (10a)$$

$$\Phi_{s2f\_j}(\{\lambda\}) \le XFT_j; \quad (10b)$$

$$\sum_{(i,j)\in\Omega_k} \lambda_j \times \tau_{ij} \le C_k, \forall k; \quad (10c)$$

with targets XF_{fab}, and $\{XFT_j, \forall j\}$ given.

In (PMP), function Φ_{p2f} represents a model of the relationship between priority release rates and XF of the whole fab as defined in Eq. (4), while function Φ_{p2f_j} represents a model of the relationship between priority release rates and the XF of priority *j* as defined in Eq. (5). Problem (PMP) is a nonlinear programming problem with nonlinear objective and constraints.

4 PRIORITY X-FACTOR MODEL VALIDATION AND TUNING

How good the PXF approximation models are needs to be validated before solving (PMP). The validation is carried out by comparing the approximated PXFs with those obtained from discrete event simulation of an exemplary fab. The validation results then guide how the compensation parameters in Eq. (2) should be adjusted.

Exemplary Fab Model: FAB1

FAB1 model is developed by combining an aggregated full-scale production line previously studied by Lu, Ramaswamy, and Kumar 1994, with a two-product model designed by Lin 1996. In FAB1, there are two types of products and each type of products is assigned a priority, *Pi*, *i*=1,2. There are 12 MGs. The numbers of operation steps of *P1* and *P2* are 32 and 60, respectively. Their release processes are Poisson. Machine tool group service time distributions include uniform, Erlang-k and exponential distributions as designed by Hu and Chang 2003. Details of FAB1 model data are listed in *Appendix A*. At wafer release rates of 0.2525 units/hour for *P1* and 0.3788 units/ hour for *P2*, the capacity bottleneck is MG3 with utilization 91.54% (Lin 1996). This FAB1 model is implemented in eM-Plant^TM for simulation study.

Simulation Study for PXF Model Validation

Two types of X-factors (ITRS 2007) are examined for validation: (1) PXFs of individual MGs and (2) overall fab PXF. Ten simulation runs are performed in the study. Each run begins with an empty line and simulates one-year of

production for warming up and then ten years of production for PXF statistics collection.

Figures 2 and 3 show the comparisons between XFs obtained from the approximation model and simulation for the two priorities respectively. In Figure 2, the approximation results of *P1* match those of simulation quite closely but the error at MG8, of which the arrivals are from the capacity bottleneck MG3. For *P2*, there are over 10% approximation errors at MGs 1, 2, 5, 8 and 10 as shown in Figure 3. An intuitive reasoning is that as *P2* wafers need to yield to the arrivals of *P1* wafers for processing at a machine group, the compounded arrival variance of *P2* wafers is larger than that of *P2* wafer arrivals only and leads to larger errors by the Poisson arrival approximation. Approximation error analysis of fab PXFs is shown in Table 1, where *P2* also has a significantly larger error.
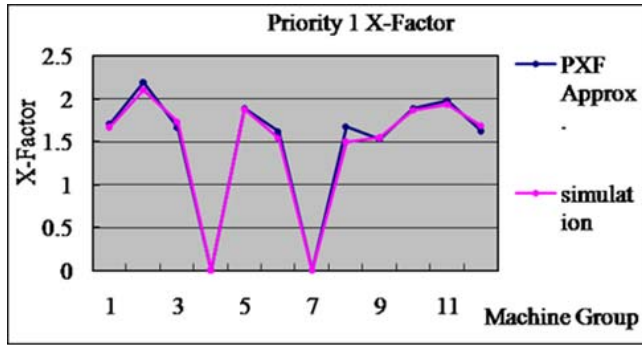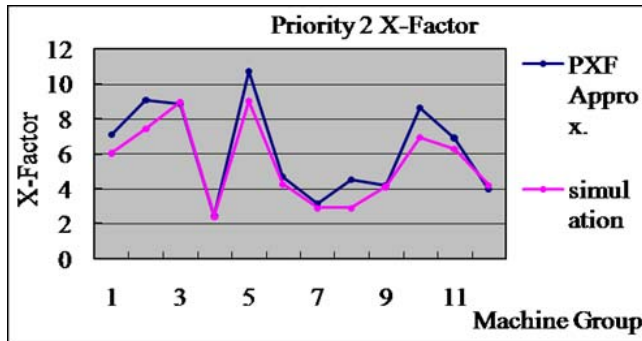

Figure 2: X-factors of *P1* over MGs

Figure 3: X-factors of *P2* over MGs

Table 1: Fab PXF errors

	Priority 1	Priority 2
PXF Model	1.789	6.189
simulation	1.767	5.468
Relative Error %	1.226	13.182

PXF model Tuning

By adjusting the arrival variance compensation parameters $\{\alpha_{jlk}\}$ in Eq. (2), the PXF approximation errors may be compensated. There are 24 compensation parameters in this case for fitting the PXFs of individual MGs and priorities obtained from simulation. By trial-and-error through Eqs. (2) and (6), a set of compensation parameters $\{\alpha_{jlk}\}$ is obtained that leads to a very close fit of PXFs to simulation results as listed in Table B1 in Appendix B.

5 APPLICATIONS TO PRIORITY MIX PLANNING

Now (PMP) is solved to determine the optimal priority mix under a set of given fab capacity and priority PXFTs. FAB1 is adopted as the baseline fab scenario. The PXF approximation models derived in Section IV are incorporated into (PMP) to explicitly express Φ_{s2f} and Φ_{s2f_j}s. Prices and costs of the baseline PMP problem are listed in Table 2. PXFT targets of *P1* and *P2* are 1.7 and 4.5 respectively.

In Table 2, the prices P_1 and P_2 are set to 980 and 700 respectively, which reflect that the average price of a finished 8-inch wafer is 840 U.S. (Chi 2006). We assume the ratios between the price and manufacturing cost are 8.13 and 2.5 for *P1* and *P2* respectively, and C_1^M =120.59 and C_2^M =280. To compute the inventory cost coefficients, we assume 5% annual interest rate and use 50% of the prices as the value of each wafer in process. C_1^I and C_2^I can then be calculated. .

Table 2: Prices and costs of baseline PMP

P_1	P_2	C_1^I	C_2^I	C_1^M	C_2^M
980.00	700.00	0.00168	0.00084	120.59	280.00

5.1 Optimal Solution and Analysis

Lingo 9.0™ is used to solve (PMP) and the optimal solution is $(\lambda_1^*, \lambda_2^*) = (0.30098, 0.27728)$ with MG 5 as the capacity bottleneck at a utilization level of 77%. As listed in Table 3, the PXF of *P1* hits its upper bound specified by PXFT1 (1.7) while PXF2 is still far away from the PXFT2 (4.5). Note that 77% of utilization at MG 5 is not very high for a bottleneck MG. Why is the maximum profit achieved at this level of utilization?

Table 3: Optimal Solution and Associated PXFs

λ_1^*	λ_2^*	PM	U	PXF1	PXF2	Profit
0.30098	0.27728	35.18	77%	1.70	2.94	445.90

In depth analysis of the objective function (Eq. (8)) shows that under the given price and cost parameters, increase in *P1* or bottleneck MG utilization will largely result in a profit rate increase. Analyses also show that the derivatives of PXF1 with respect to λ_1 and/or λ_2 at the optimum (0.30098, 0.27728) are both positive. Any increment in λ_1 and/or λ_2 will make PXF1 exceed PXFT1 and is not a feasible solution. So, the solution $(\lambda_1^*, \lambda_2^*) = (0.30098, 0.27728)$ is optimal.

5.2 Numerical Study of PXF Behavior

To obtain insights for the PMP problem, we further analyze PXF behavior which we define as the relation among profit, bottleneck MG utilization (U) and priority mix (PM). In this study, PM is formally defined as the workload percentage of P1

$$PM \equiv \frac{\lambda_1 \times \sum_i \tau_{i1}}{\lambda_1 \times \sum_i \tau_{i1} + \lambda_2 \times \sum_i \tau_{i2}} \times 100\%. \quad (11)$$

For simplicity of discussion, it is assumed that the processing times of individual steps at each MG are the same and that MG *k* is the capacity bottleneck. Under such an assumption, PM is then

$$PM = \frac{\lambda_1 \times n_{1k}}{\lambda_1 \times n_{1k} + \lambda_2 \times n_{2k}} \times 100\%, \quad (12)$$

where n_{jk} represents the number of times that a priority *j* wafer visits the capacity bottleneck MG *k*. The variation of U and PM is through changing the values of λ_1 and/or λ_2. □□□

How fab profit, *J* in Eq. (8), varies with respect to U of the bottleneck MG and PM is first studied by enumeration of the PMP model. Figure 4 shows that under a fixed PM level subject to capacity constraints but *without PXFT constraints*, the profit increases as the utilization increases except in the cases of PM=40% and 50% when the utilization reaches over 94%. The profit decrease in the latter cases is because the inventory cost incurred by the significant WIP and, equivalently, XF increase at a high bottleneck utilization and a high priority 1 release rate according to Eq. (2). The two-piece horizontal red line in the middle of the Figure 4 is obtained by *including PXFTs*. It represents the maximum profit under PXFT constraints.
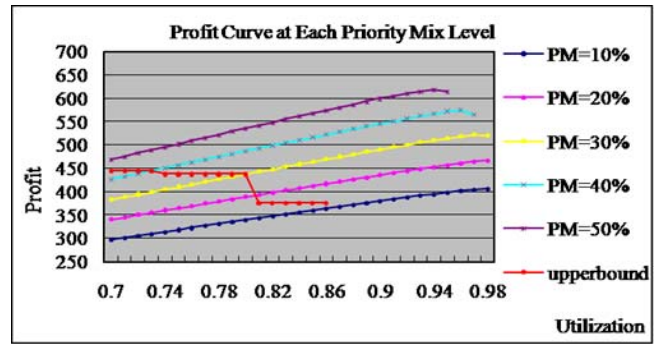

Figure 4: Profit vs priority mix and utilization

How a PXF behaves with respect to capacity bottleneck utilization and priority mix is then investigated by enumeration o7 PXF functions expressed in Section 2. *In changing PM*, the *capacity bottleneck MG may shift from one MG to another* because the difference between process flows of the two priorities. For example, the capacity bottleneck shifts to MG3 form MG5 as the PM increases beyond 20%. In our numerical study of PXF behavior, the utilization indicates the utilization of the bottleneck MG under the given PM.

Figure 5 and 6 present how PXFs under a given PM varies with respect to bottleneck MG utilization. In Figure 5, PXF1 is basically linear, of which the slope increases with PM value of *P1*. PXF2 is insensitive to PM values of *P1* when bottleneck utilization is below 90% but increases sharply otherwise; the higher the PM value of *P1*, the sharper the rise.
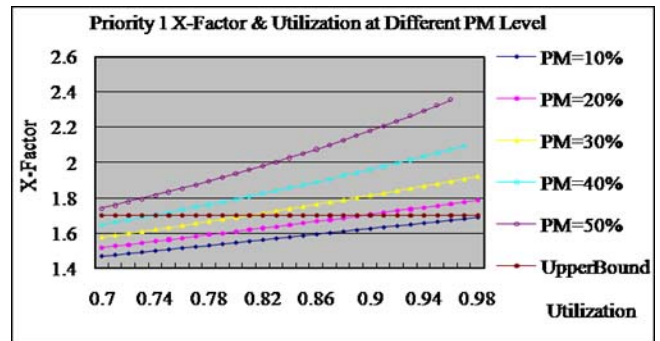

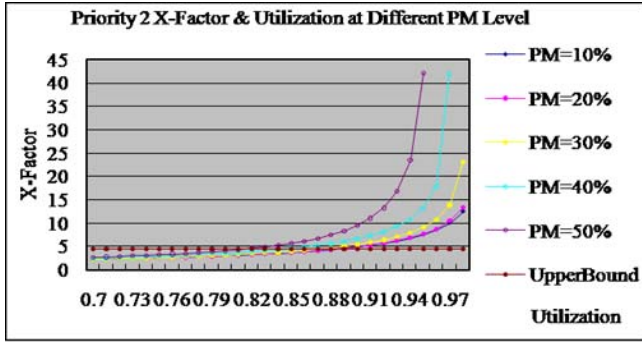Figure 5: Priority 1 X-Factor & utilization at different PM level

Figure 6: Priority 2 X-Factor & utilization at different PM level

Figure 8: Priority 2 X-Factor & PM at different utilization level

Figures 7 and 8 depict how PXF under a given bottleneck MG utilization varies with respect to PM value of *P1*. PXF1 behavior curves of the four utilization levels are quite similar, while the PXF2 behavior curve of U=95% is significantly different from the other three and it increases sharply when PM value of *P1* becomes higher than 40%. These observations provide some further understanding about the PXF behavior models of Section 2.

6 CONCLUSIONS

We have formulated and studied a priority mix planning (PMP) problem to determine the wafer release rates of individual priorities to maximize fab profits subject to XFT and capacity constraints. The nonlinear programming formulation extended a M/G/m:PR queue approximation and contribution theory-based network to model how X-factors of individual priorities are affected by priority mix and fab capacity utilization. Numerical study over a realistic fab example shows that PMP problem can be efficiently solved for planning priority mix and machine tool capacity utilization to maximize profit while satisfying differentiated XFTs.
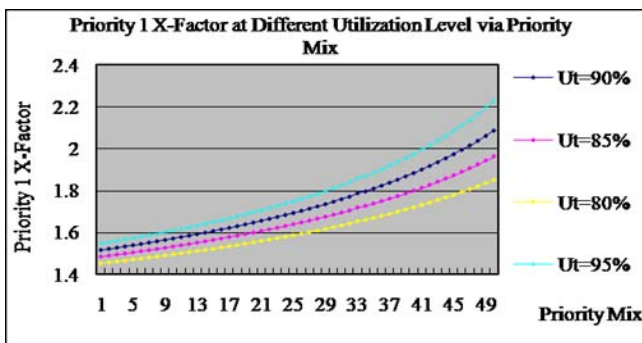
Appendix A: FAB1 Model Data

Table A1: Machine Group Data of FAB1

MG	# of tools	# of Visits (P1)	# of Visits (P2)	MPT (hr)	Service Time Distribution	Utilization %[*]
1	4	7	14	0.5	Erlang 4	88.38
2	3	9	12	0.375	Exponential	85.23
3	10	4	7	2.5	Uniform	91.54
4	1	0	1	1.8	Erlang 3	68.18
5	1	1	2	0.9	Erlang 2	90.91
6	2	1	3	1.2	Erlang 4	83.33
7	1	0	1	1.8	Exponential	68.18
8	4	4	8	0.8	Erlang 3	80.81
9	1	1	3	0.6	Uniform	83.33
10	9	3	5	3	Erlang 2	88.38
11	2	1	3	1.2	Exponential	83.33
12	2	1	1	2.5	Uniform	78.91

Appendix B: modified Priority MG X-Factor

Table B1: X-factor adjustment table

Figure 7: Priority 1 X-Factor & PM at different utilization level

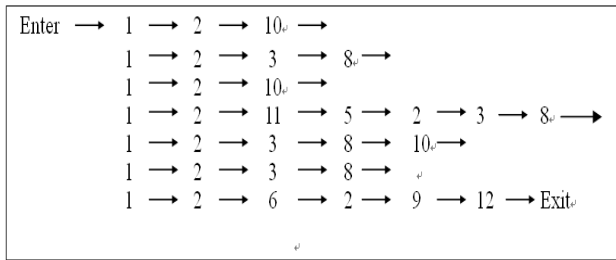| MG	Alpha1	P1 Modified PXF Approx.	Simul.	Alpha2	P2 Modified PXF Approx.	Simul.
1	0.007	1.673	1.673	0.024	6.039	6.038
2	0.015	2.107	2.105	0.043	7.431	7.430
3	0.027	1.736	1.735	0.004	8.964	8.963
4	0.000	0.000	0.000	0.009	2.421	2.421
5	0.004	1.879	1.879	0.215	9.009	9.009
6	0.070	1.545	1.545	0.068	4.282	4.284
7	0.000	0.000	0.000	0.270	2.907	2.907
8	0.055	1.503	1.502	0.100	2.890	2.889
9	0.017	1.549	1.548	0.011	4.127	4.127
10	0.009	1.869	1.868	0.099	6.925	6.925
11	0.046	1.937	1.936	0.104	6.283	6.283
12	0.107	1.687	1.687	0.091	4.214	4.211

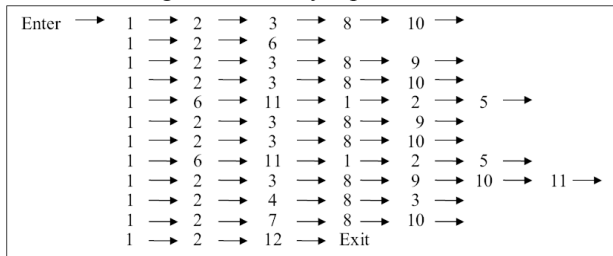Figure 9: Priority 1 process flow

Figure 10: Priority 2 process flow

References

Chang, S. C., and K. J. Chen. 2006. Priority X-Factor Modeling for Differentiated Manufacturing Service Planning. *Proceedings of 2006 International Symposium on Semiconductor Manufacturing*, 25-27.

Chung, S. H., A. H. I. Lee, C. Y. Huang, and C. C. Chuang. 2002. Capacity Pricing Mechanism for Wafer Fabrication. *In 2002 Asia-Pacific Industrial Engineering and Management Symposium*, 58-59.

China Economics. 2006. <http://big5.ce.cn/cysc/ceit/zggng/200510/28/t20051028_5043441.shtml>

Delp, D., J. Si, and J. W. Fowler. 2006. The Development of the Complete X-Factor Contribution Measurement for Improving Cycle Time and Cycle Time Variability. *IEEE Transactions on Semiconductor Manufacturing* 19:352-362.

Hu, M. D., and S. C. Chang. 2004. Translating Overall Production Goals into Distributed Flow Control Parameters for Semiconductor Manufacturing. *Journal of Manufacturing Systems* 22:46-63.

Hopp, W., and M. Spearman. 2008. *Factory Physics 3rd ed.*. MacGraw-Hill Higher Education.

International Technology Roadmap for Semiconductors 2007 edition. Factory Integration.

Kang, H. Y., and A. H. I. Lee. 2007. Priority Mix Planning for Semiconductor Fabrication by Fuzzy AHP Ranking. *Expert Systems with Applications* 32:560–570.

Kishimoto, M., K. Ozawa, K. Watanabe, and D. Martin. 2001. Optimized Operations by Extended X-Factor Theory Including Unit Hours Concept. *IEEE Transactions on Semiconductor Manufacturing* 14:187-195.

Lu, S. H., D. Ramaswamy, and P. R. Kumar. 1994. Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants. *IEEE Transactions on Semiconductor Manufacturing* 7: 374-388.

Lin, C. Y.. 1996. Shop Floor Scheduling of Semiconductor Wafer Fabrication Using Real-Time Feedback Control and Predictions. *Ph. D. Dissertation, University of California at Berkley*.

Lee, A. H. I., H. Y. Kang, and W. P. Wang. 2006. Analysis of Priority Mix Planning for the Fabrication of Semiconductors under Uncertainty. *Int. J. Adv. Manuf. Tech*. 28:351-361.

Martin, D. P.. 1998. The Advantage of Using Short Cycle Time Manufacturing (SMC) Instead of Continuous Flow Manufacturing (CFM). *Proceeding of ASMC*, 89-94.

Narahari, Y., and L. M. Khan. 1997. Modeling the Effect of Hot Lots in Semiconductor Manufacturing Systems. *IEEE Transactions on Semiconductor Manufacturing* 10:185-188.

Robinson, J., and F. Chance. 2002. Dynamic X-Factor application for Optimizing Lot Control for Agile Manufacturing. *Proceedings of ISSM*, 281-284.

Virtamo, J.. 2004. Lecture Note 38.143: Queuing Theory/Priority queues," Helsinki University of Technology. <http://www.netlab.tkk.fi/opetus//s38143/luennot/E_priority.pdf>

Yang, F., B. Ankenman, and Barry L. Nelson. 2006. Efficient Generation of Cycle Time-Throughput Curves through Simulation and Metamodeling. *Naval Research Logistics* 54:78–93.

AUTHOR BIOGRAPHIES

SHI-CHUNG CHANG received the B.S.E.E. degree from National Taiwan University, Taipei, in 1979, and the M.S. and Ph.D. degrees in electrical and systems engineering from the University of Connecticut, Storrs, in 1983 and 1986, respectively. From 1979 to 1981, he served as an Ensign in the Chinese Navy, Taiwan. He worked as a Technical Intern at the Pacific Gas and Electric Company, San Francisco, CA, in the summer of 1985. During 1987, he was a member of the Technical Staff, Decision Systems Section, ALPHATECH, Inc., Burlington, MA. He has been with the Electrical Engineering Department, National Taiwan University since 1988 and was promoted to Professor in 1994. During 2001–2002, he served as the Dean of Student Affairs and a professor of Electrical Engineering, National Chi Nan University, Pu-Li, Taiwan. He was a visiting scholar at the Electrical and Computer Engineering Department, University of Connecticut during his sabbatical leave in the 2003–2004 and 2006–2007 academic years. Besides the Electrical Engineering Department, he is now jointly appointed by the Graduate Institute of Industrial Engineering and the Graduate Institute of Communication Engineering, National Taiwan University. His research interests include optimization theory and algorithms, production scheduling and control, network management, Internet economics, and distributed decision making. He has been a principal investigator and consultant to many industry and government funded projects in the above areas, and has published more than 130 technical papers. He received the award for Outstanding Achievements in University–Industry Collaboration from the Ministry of Education for his pioneering research collaborations with the Taiwan semiconductor industry on production scheduling and control in 1996.
E-Mail: <Hscchang@cc.ee.ntu.edu.tw>

SHIN-SHYU SU received the B.S.P.M.E. degree from National Tsing Hua University, Taiwan R.O.C. in 2006. He is current working toward M.S.E.E. degree in National Taiwan University. His main research interests are in discrete-event simulation and optimization theory with application to manufacturing systems.
E-Mail: <r95921063@ntu.edu.tw>

KE-JU CHEN received the B.S.M.E degree from National Chung Hsing University, Taichung, in 2002 and received M.S.E.E. degree from National Taiwan University, Taipei, R.O.C. in 2006.
E-Mail: <Hr93921070@ntu.edu.tw>