

ECONOMIC EFFICIENCY ANALYSIS OF WAFER FABRICATION FACILITIES

Wen-Chih Chen

Dept. of Industrial Engineering and Management

National Chiao Tung University
Hsinchu, 30010, TAIWAN ROC

Chen-Fu Chien
Ming-Hsuan Chou

Dept. of Industrial Engineering and Engineering Management

National Tsing Hua University
Hsinchu, 30013, TAIWAN ROC

ABSTRACT

Semiconductor industry is capital intensive and competitive, and thus efficiently utilizing resources to provide products and services is essential for maintaining competitive advantages. Knowing whether the resource is properly utilized is the foundation for future improvements and/or decision making. This study investigates the economic efficiency of fabrication facility (fab) operations. We develop a two-stage overall efficiency model, which clearly defines and explains the “real” performance of fab production operations and the non-production issues. The model provides an overall performance index while considering different aspects. A single performance index can be used to evaluate and rank the performance for period review. Factors affecting performance can be identified. Furthermore, according to a real case, an *ex post* relative efficiency analysis is conducted and the initial results are reported. The case study can help providing diagnosis for inefficient production facilities and identifying best practices of efficient production units.

1 INTRODUCTION

Semiconductor manufacturing is a very capital intensive and competitive industry. Knowing how to utilize resources to provide products and services efficiently to maximize the profits is essential to sustain in this competitive business and towards success. Performance analysis is a foundational tool for monitoring, diagnosing and improving the business activities and processes, and can thus steer the future direction of business strategies.

There are many different methods for performance evaluation in various applications. Benchmarking, learning from the best similar practices, is one of the helpful methods. Performance benchmarking was popularized in the late 80's by the successful stories of Hewlett-Packard and Xerox, (Camp 1989). It is a set of processes and practices used to determine (i) reference values for selected performance indices, and (ii) factors for key processes affect-

ing performance. Performance indices selected should be sufficient to represent all important objectives of a firm. Generally, there are trade-offs among firm's objectives, and the trade-offs often lead to inconsistent conclusions among performance based on different indices. It is hence desired to establish a single index effectively representing the overall performance considering all aspects of a firm. In addition, understanding factors and/or practices affecting the performance is another key activity in benchmarking studies. Studying the overall performance index and possible factors diagnoses the performance and suggests the improvement direction and possible approaches.

There are several benchmarking studies focusing on semiconductor industry, particularly regarding to the manufacturing aspect. Leachman and Hodge (1996) provide the first complete benchmarking study report in the competitive semiconductor manufacturing (CSM) program, which includes several leading companies around the world. Seven key performance indices (KPIs) are proposed in their study: cycle time per wafer layer, line yield, die yield, stepper productivity, direct labor productivity, total labor productivity, and on-time delivery. Associations between practices and performance are also presented. This is the first industry wide benchmarking study with rich data in semiconductor manufacturing. The CSM program also yields other relevant studies such as (Nickerson and Sloan 1997). Extending the studies in CSM program, Leachman, Ding, and Chien (2007) use data envelopment analysis (DEA) to provide an overall relative efficiency index. They use total wafer starts, number of steppers, direct and indirect headcount, and clean room size as the inputs and die output and effective revenue as the outputs. Factors and practices are also examined and identified according to the overall single performance index. Carbone (2000) applies DEA to evaluate the efficiencies of different production areas within a semiconductor fabrication facility (fab). In his work, machine failure rate, scrap rate, cycle time, and machine downtime are the resources, and outputs are wafer move rate, overall equipment efficiency, and activity ratio (actual moves/planned moves). Chien, Hsiao, and Wang

(2004) also list a comprehensive set of performance indices used in semiconductor manufacturing practices, and they further categorize them as management indices and performance indices.

This paper aims to study economic efficiency on fab operations, particularly to evaluate the relative performance of fabrication facilities. The study is motivated by the need of performance review and performance rank for each fabrication facility. The main function of a fab is manufacturing, or named production, which utilizes resources to produce products to meet the demand. We propose a two-stage overall efficiency model, which clearly defines and classifies the “real” performance of fab production operations and the non-production issues. The evaluation model provides an overall performance index while considering different aspects by adopting DEA method. A single performance index can be used to evaluate and rank the performance for period review. Furthermore, we also report the initial results of a benchmarking study according to a real case. The proposed model and method is demonstrated, and an ex post relative efficiency analysis is conducted. The case study can help providing diagnosis for inefficient decision making units and identifying best practices for efficient decision making units.

The remainder of this paper is organized as follows. Section 2 presents a model for fabrication production process. The model gives a clearer picture of fabrication and a better reasoning of performance evaluation method proposed in Section 3. Section 3 also addresses some appealing properties of our model. Section 4 provides the initial results of a benchmarking study based on a real case. Conclusions are given in Section 5.

2 FABRICATION PRODUCTION PROCESS

This study focuses on the performance analysis on fabs, particularly the manufacturing department. Fabs are the places where all production activities happen, and can be viewed as decision units utilizing resources, e.g., labor, equipments, etc., to provide products (outputs). A good

production unit can provide more outputs by using resources as few as possible. As one can observe, the physical final product produced and delivered by a fab is wafers. In the literature and practices, typical productivity indices adopted use total wafer produced as the outputs. See (Chien, Hsiao, and Wang 2004; Leachman and Hodge, 1996; Leachman, Ding, and Chien 2007) for example. This is questionable and may underestimate the production performance as we will address. In this study, we propose a two-stage model to describe the transformation process from resources to final products in a fab.

Figure 1 presents the two-stage process which describes the activities in a fab. The first stage (Stage 1) is related to *production efficiency*, which is a process providing masking layers (outputs) by consuming labor, capital investments including equipments and space, and time. The first three resources are commonly found in other studies. Time, which is not as straightforward as the first three, is the total time used in the production. Given the same level of labor, equipment and space, more layers require longer time. Similarly, for example, less labors (equipments) generally results in longer time to generate required layers, and this reveals the substitutability among resources. To have a clear understanding, it is better to interpret the output of Stage 1, layers, as a service provided than a product. Thus the performance at this stage is directly related to production efficiency. A good performance indicated in Stage 1, namely providing more layers by fewer resources, suggests that the production process is productive.

The second stage (Stage 2) shows the process of transforming layers to wafers. Different types of wafer products require different number of layers; this is related to *manufacturability*. Design for manufacturability (DFM) is an engineering concept getting important nowadays, which includes a set of methodologies, of designing products in such a way that they are easy to manufacture. The manufacturability is resulted from both engineering supports, particularly in R&D activities of design or manufacturing such as manufacturing recipe. It is also affected by the

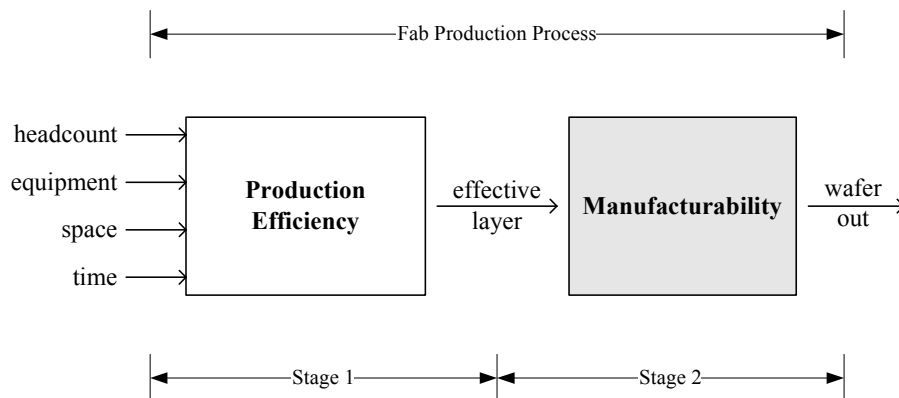


Figure 1: Fab production process

business strategies, e.g., allocating products to each fab. Some products require fewer layers per wafer and are easier to produce. A fab allocated with products easier to produce definitely will be easier to have superiors performance than others. Number of layers is the number of processes required in wafer manufacturing. Larger number of wafer outputs with fewer layers is always preferred since it means fewer efforts in manufacturing.

It is worth mentioning that the combinations of resource-output pair in both stages are identical to the most commonly used productivity indices. Index (effective layer / headcount) is identical to layer labor productivity, and (effective layer / equipment) is the layer labor productivity. Index (effective layer / space) corresponds to the effectiveness of space usage, and (effective layer / time) is the reciprocal of layer cycle time. The combination of wafer out and four resources have identical meanings as four indices but with respect to wafer. Number of layers required per wafer, of course, is one practical means to measure the product complexity.

Although number of wafer outputs is commonly used as part of a productivity index in conventional fab operation performance evaluation. Our two-stage model shows that using wafer output as the output may underestimate the performance of fabricator operations, particularly the manufacturing teams. This is because that the final outputs consist activities determined by and contributed from other departments, such as R&D (in manufacturing recipes) and central production planning department (in product allocation). An efficient manufacturing team may produce less wafers due to poor manufacturability. In fact, the “real” responsibility of a manufacturing team is to process various layers based on certain manufacturing recipes using whatever resources on hand. That is the manufacturing team consumes resources to provide services (the layer processing) so that the final product (wafer) is produced by going through the complete required processes. The proposed model provides a better picture on the process of transforming resources invested to final products. Therefore, a fair performance evaluation can be conducted with fewer arguments, a more effective diagnosis can be provided, and true responsibility for each department can be identified.

Detailed definitions of resources and outputs are listed as followed:

- Effective Layers (L): is the total number of effective masking layers produced monthly. In practices, it can be collected as (total layers produced) \times (average layer yield rate).
- Effective wafer outs (W): is the total number of effective wafers produced monthly. It can be collected as (total number of wafers produced) \times (average wafer yield rate).
- Headcount (HC): is the total labor employed in each month including direct and indirect labor. It also includes the management and assistant staff.
- Equipments (MC): is defined as the total installed capacity of steppers and scanners (exposure tools) used in the fab during the particular month. Since lithography equipment is generally the fab bottleneck among all machine types, it serves as the best proxy for the equipment capacity of the fab. The literature uses total number of steppers and scanners, e.g., Leachman, Ding, and Chien (2007). Number of tools however ignores the difference in production capability of the machines. For example, new tools may have higher throughput than old machines, and/or a more expensive machine may have better throughput as well. The difference on resulted throughput can yield significant impact on production performance. Hence, instead of tool number, weighted sum of installed tool capacity are used in our model for a better proxy.
- Space (S): is the total floor space used and available for manufacturing. It reflects the infrastructure investment, which becomes the limit of production capability.
- Time (T): is total time needed to produce total number of layers, and is computed as (total number of layers) \times (average cycle time per layer). It should be noticed that total layers, including effective layers and defects, are considered. Time is a resource and defects result in resource waste; thus defects should be taken into account.

3 MEASURING ECONOMIC EFFICIENCY

Conventional benchmarking studies in semiconductor manufacturing use different productivity indices, such as labor productivity and equipment productivity. However, trade-offs exist among productivity indices and can lead to inconsistent conclusions. DEA, introduced by Charnes, Cooper, and Rhodes in 1978, is a mathematical programming approach to compute relative efficiency. DEA considers many resources and outputs simultaneously and provides a single overall efficiency score. Leachman, Ding, and Chien (2007) investigate the fab operation efficiency using DEA as the evaluation tool.

According to the proposed two-stage production process model, the output-oriented CCR DEA model (Charnes, Cooper, and Rhodes 1978) is employed to measure the efficiency for both stages as follows:

$$\begin{aligned}
 & \max_{\phi_k^t, \lambda_j} \phi_k^t \\
 & \text{s.t. } \sum_{j \in S^t} x_{ij} \lambda_j \leq x_{ik} \quad \forall i \in I^t, \\
 & \quad \sum_{j \in S^t} y_{rj} \lambda_j \geq \phi_k^t y_{rk} \quad \forall r \in O^t, \\
 & \quad \lambda_j \geq 0 \quad \forall j \in S^t;
 \end{aligned} \tag{1}$$

where $t \in \{1, 2\}$ representing Stages 1 and 2. Sets I^t and O^t are the resource set and output set, respectively, for stage t . Namely, $I^1 = \{HC, MC, S, T\}$, $O^1 = \{L\}$, $I^2 = \{L\}$ and $O^2 = \{W\}$. Set S^t stands for the collected records for stage t . Moreover, x_{ij} and y_{rj} are the amounts of resource i and output r for record j , respectively, given corresponding stage for record set S^t . Subscript k represents the record under evaluation, and $k \in S^t$.

Let ϕ_k^{1*} and ϕ_k^{2*} be the optimal values corresponding to Stages 1 and 2 in (1), respectively. They are the efficiency scores for Stages 1 and 2. Value ϕ_k^{1*} indicates that the record k can generate ϕ_k^{1*} times of current output level while using the same level of resources. Clearly, $\phi_k^{1*} \geq 1$, $\phi_k^{1*} = 1$ indicates that record k is *CCR-efficient in production*. Larger value of ϕ_k^{1*} reveals more inefficient in production, namely, it has more significant edge losing to the best records. Similarly, if $\phi_k^{2*} = 1$, record k is said to be *CCR-efficient in manufacturability*; record k has products that are the easiest to produce. A higher ϕ_k^{2*} value shows less manufacturability, i.e., more difficult to produce.

Moreover, the dual of model (1) is as follows (Charnes, Cooper and Rhodes 1978):

$$\begin{aligned}
 & \min_{u_i, v_r} \sum_{i \in I^t} u_i x_{ik} \\
 & \text{s.t. } \sum_{r \in O^t} v_r y_{rj} \geq \sum_{i \in I^t} u_i x_{ik} \quad \forall j \in S^t \setminus \{k\}, \\
 & \quad \sum_{r \in O^t} v_r y_{rk} = 1, \\
 & \quad u_i \geq 0 \quad \forall i \in I^t, \\
 & \quad v_r \geq 0 \quad \forall r \in O^t.
 \end{aligned} \tag{2}$$

Model (2) is equivalent to the following problem:

$$\begin{aligned}
 & \min_{u_i, v_r} \frac{\sum_{i \in I^t} u_i x_{ik}}{\sum_{r \in O^t} v_r y_{rk}} \\
 & \text{s.t. } \frac{\sum_{i \in I^t} u_i x_{ik}}{\sum_{r \in O^t} v_r y_{rj}} \leq 1 \quad \forall j \in S^t, \\
 & \quad u_i \geq 0 \quad \forall i \in I^t, \\
 & \quad v_r \geq 0 \quad \forall r \in O^t.
 \end{aligned} \tag{3}$$

For Stage 1, Model (3) can be rewritten as follows:

$$\begin{aligned}
 & \min \frac{u_{HC} HC_k + u_{MC} MC_k + u_S S_k + u_T T_k}{v_L L_k} \\
 & \text{s.t. } \frac{u_{HC} HC_j + u_{MC} MC_j + u_S S_j + u_T T_j}{v_L L_j} \leq 1 \quad \forall j \in S^1, \\
 & \quad u_{HC} \geq 0, u_{MC} \geq 0, u_S \geq 0, u_T \geq 0, v_L \geq 0.
 \end{aligned} \tag{4}$$

Let $\frac{u_{HC}}{v_L} = w_{HC}$, $\frac{u_{MC}}{v_L} = w_{MC}$, $\frac{u_S}{v_L} = w_S$, and $\frac{u_T}{v_L} = w_T$,

Model (4) is rewritten as follows:

$$\begin{aligned}
 & \min w_{HC} \frac{HC_k}{L_k} + w_{MC} \frac{MC_k}{L_k} + w_S \frac{S_k}{L_k} + w_T \frac{T_k}{L_k} \\
 & \text{s.t. } w_{HC} \frac{HC_j}{L_j} + w_{MC} \frac{MC_j}{L_j} + \\
 & \quad w_S \frac{S_j}{L_j} + w_T \frac{T_j}{L_j} \leq 1 \quad \forall j \in S^1, \\
 & \quad w_{HC} \geq 0, w_{MC} \geq 0, w_S \geq 0, w_T \geq 0.
 \end{aligned} \tag{5}$$

Model (5) has a nice interpretation. Indeed, HC/L is the number of headcount needed per layer, which is a commonly used labor productivity in the industry. More precisely, the reciprocal of the labor productivity. MC/L measures the equipment productivity; S/L represents the return rate on space, and T/L is the layer cycle time. Therefore, Model (5) evaluates record k by examining the weighted productivity on four aspects. Without assigning subjective predetermined weights of four aspects, weights are selected in favor (minimize the weighted index) of record k , while normalizing other records' score being no larger than one using the same weights.

In contrast to Stage 1, Stage 2 is more straightforward since there is only one resource and one output. The conventional single ratio approach can be easily adopted. The corresponding Model (1) $t=2$ will give the same results as using wafer per layer, "how many layers needed to become a complete wafer?" as the comparison metric.

4 PERFORMANCE ANALYSIS

This section presents the initial economic efficiency analysis results for a real case in Taiwan. The study is based on the comparison of four different production sites (fabs) within the same corporate. Thus this study has a more similar comparing basis than the investigation conducted by Leachman, Ding, and Chien (2007) since their study compares records from different companies around the world.

In the study, records for the last three years are collected for each fab, which represent monthly resources consumed and outputs produced. There are total 156 records and the data are pooled together for comparison based on Model (1). Figures 2 and 3 are box plots for the CCR-efficiency scores on production efficiency and manufacturability, respectively. In fact, the y -axis represents the reciprocal of ϕ_k^{t*} in Model (1); a larger value is better. The x -axis associates with four fabs. There are some differences among fabs on production efficiency (Figure 2); Fab C has the worst average production efficiency and Fabs A and D perform better than the other two. Fab C has poor manufacturability CCR-efficiency score (Figure 3), which indicates Fab C produces more complex products than the others. Fab C is worse than other three fabs within 10% in production efficiency, however, it has significant gap (disadvantage) to the others in terms of manufacturability.

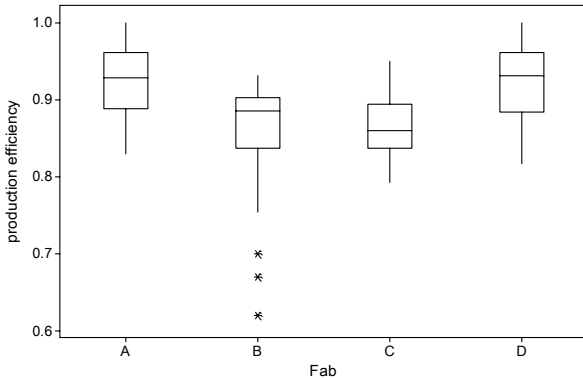


Figure 2: Box plot of production efficiency

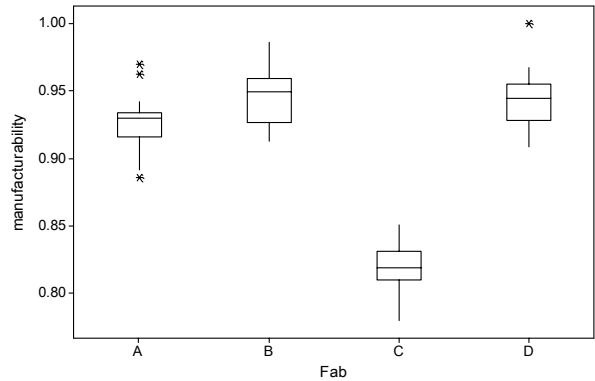


Figure 3: Box plot of manufacturability

Another popular DEA model proposed by Banker et al. (1984) is as follows:

$$\begin{aligned}
 & \max_{\theta_k^t, \lambda_j} \theta_k^t \\
 & s.t. \sum_{j \in S^t} x_{ij} \lambda_j \leq x_{ik} \quad \forall i \in I^t, \\
 & \quad \sum_{j \in S^t} y_{rj} \lambda_j \geq \theta_k^t y_{rk} \quad \forall r \in O^t, \\
 & \quad \sum_{j \in S^t} \lambda_j = 1, \\
 & \quad \lambda_j \geq 0 \quad \forall j \in S^t;
 \end{aligned} \tag{6}$$

where $t \in \{1,2\}$ representing Stages 1 and 2. Model (6) has additional convexity constraint, $\sum_{j \in S^t} \lambda_j = 1$, comparing to (1). The difference of optimal values for a particular record k between models (1) and (2), $\frac{\phi_k^{t*}}{\theta_k^{t*}}$, measures the scale effect, named scale inefficiency (Banker, Charnes, and Cooper 1984). Clearly, $\frac{\phi_k^{t*}}{\theta_k^{t*}} \leq 1$; $\frac{\phi_k^{t*}}{\theta_k^{t*}} = 1$ indicates that records k is at the proper production scale, the most productive production scale (MPSS). $\frac{\phi_k^{t*}}{\theta_k^{t*}} < 1$ reveals that k is scale inefficient, namely k is not at the MPSS (Banker Charnes and Cooper 1984), too large or too small. Smaller $\frac{\phi_k^{t*}}{\theta_k^{t*}}$ shows more difference to the proper size it should be. Moreover, the optimal solution of (1) related to k , $\sigma_k = \sum_{j \in S^t} \lambda_j^*$, provides information on k regarding to the relative scale position to the corresponding MPSS. $\sigma_k = 2$ suggests that k is double to the size of MPSS, and scaling down in production scale may increase the productivity.

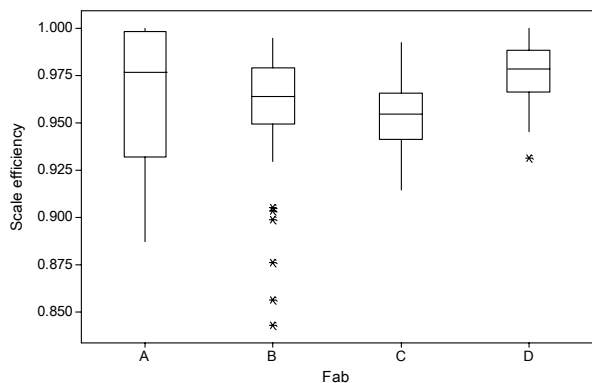


Figure 4: Box plot of scale efficiency in production

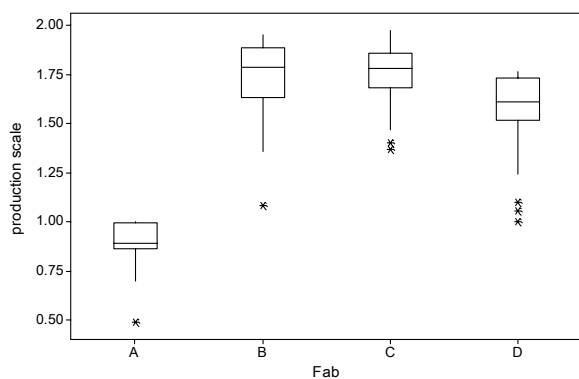


Figure 5: Box plot of scale efficiency in manufacturability

Table 1: Correlation coefficients between factors and production efficiency

	tool utilization	cycle time	yield
production efficiency	-0.19	-0.66	0.54
$1/\phi_k^*$			

$\sigma_k = 0.8$ says size of k is only 80% to the right size it should be.

Figures 4 and 5 are box plots for scale efficiency and the relative scale to the MPSS, respectively. Figure 4 shows that all records of Fabs B and C are scale inefficient, they are never in MPSS. Fab A is smaller than other three fabs. Fab A is ever only half of the MPSS while Fabs B and C can be as large as twice of the MPSS (Figure 5).

We also study the factors that determine the performance, particularly the production efficiency. Table 1 summarizes the correlation coefficients between production efficiency and three factors: tool utilization, cycle time and

yield. The correlation is measured using records of four fabs. It is found that utilization and cycle time are negatively associated with production efficiency ($1/\phi_k^*$). It means that longer cycle time associates with using more resources or producing less outputs. On the other hand, yield is positively correlated with production efficiency. Higher yield rate comes along with better performance in production efficiency.

The initial results demonstrate that the proposed method is able to provide useful information for fab operation improvements. More detailed study is on-going and complete results will be available.

5 CONCLUSION

Semiconductor industry is very competitive, in which efficiently utilizing resources to provide products and services is essential for maintaining competitive advantages. Knowing whether the resource is properly utilized is the foundation for future improvements and/or decision making. This study aims to investigate economic efficiency on fabrication operations. We propose a two-stage fabrication production model, which clearly defines and explains the “real” performance of fab production operations and the non-production issues manufacturability. A DEA model is built to provide an overall performance index considering different aspects. A single performance index can used to evaluate and rank the performance for period review. Furthermore, according to a real case, an ex post relative efficiency analysis is conducted and the initial results are reported. The initial results show that our model gives a clear picture on fabrication facility performance and better explanations on performance difference. It also hints that a complete case study can help providing diagnosis for inefficient decision making units and identifying factors affecting performance.

REFERENCES

Banker, R.D., A. Charnes, and W. W. Cooper. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30:1078–1092.

Camp, R. C. 1989. *Benchmarking: The search for industrial best practices that lead to superior performance*. New York, NY and Milwaukee, WI, Quality Resources and ASQC Quality Press.

Carbone, T. A. 2000. Measuring efficiency of semiconductor manufacturing operations using data envelopment analysis (DEA). In *Proceeding 2000 IEEE/SEMI Adv. Semiconductor Manufacturing Conference*, 56–62.

Charnes, A., W.W. Cooper, and E. Rhodes. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6):429-444.

- Chien, C.-F., A. Hsiao, and I. Wang. 2004. Constructing semiconductor manufacturing performance indexes and applying data mining for manufacturing data analysis. *Journal of the Chinese Institute of Industrial Engineer* 21(4):313-327.
- Leachman, R.C., and D. A. Hodges. 1996. Benchmarking semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 9(2):158-169.
- Leachman, R.C., S. Ding, and C.-F. Chien. 2007. Economic efficiency analysis of wafer fabrication. *IEEE Transactions on Automation Science and Engineering* 4(4):501-512
- Nickerson, J. A., and T. W. Sloan. 1997. Data reduction techniques and hypothesis testing for analysis of benchmarking data. *International Journal of Production Research* 37(8):1717-1741.

AUTHOR BIOGRAPHIES

WEN-CHIH CHEN received the B.S. degree in Industrial Engineering from the National Tsing Hua University, Taiwan in 1995. He received his Ph.D. from School of Industrial and Systems Engineering at Georgia Institute of Technology, USA, in 2003. Dr. Chen is an Assistant Professor at Department of Industrial Engineering and Management in National Chiao Tung University, Taiwan. His current research interests focus on performance evaluation and decision making analysis in semiconductor manufacturing. Email: wenchih@faculty.nctu.edu.tw.

CHEN-FU CHIEN was born in 1966 in Taiwan, Republic of China. He received the B.S. degree with double majors (with Phi Tao Phi Honor) in Industrial Engineering and Electrical Engineering from the National Tsing Hua University, Taiwan in 1990. He received the M.S. and Ph.D. degrees in Industrial Engineering, with two minors in Statistics and Business, from the University of Wisconsin-Madison, U.S.A. in 1994 and 1996, respectively. Dr. Chien is a Professor in the Department of Industrial Engineering and Engineering Management, National Tsing Hua University (NTHU). Since 2005, he is on-leave from NTHU to serve as Deputy Director of Industrial Engineering Division in Taiwan Semiconductor Manufacturing Company that is the world largest semiconductor foundry. He is a Fulbright Scholar in the Department of Industrial Engineering and Operations Research, University of California-Berkeley (2002-2003) and a Visiting Professor in Cambridge University (2004-2005, sponsored by Royal Society). He is a Steering Committee of Industrial Engineering Division in National Science Council, Taiwan. He is a member of IEEE, IIE, and INFORMS. He is now Board Member of the Chinese Institute of Decision Sciences and Chinese Institute of Industrial Engineers (CIIE). He received the Distinguished Young Industrial Engineer Award, the Distinguished Young Faculty Research Award

from NTHU, Best Paper Award from CIIE, Best Research Awards from the National Science Council, Distinguished Industrial Collaboration Award from the Ministry of Education, Best Engineering Paper Award by Chinese Institute of Engineers, and Tier 1 Principal Investigator (Top 3%) of National Science Council (2005-2008), Taiwan. His research and development efforts center on decision analysis, data mining, modeling and analysis for semiconductor manufacturing, and production strategy.

MING-HSUAN CHOU is a M.S. student in the Department of Industrial Engineering and Engineering Management National Tsing Hua University, Taiwan. Her research interests include data envelopment analysis, performance evaluation and operations research.