Continuously Improving Methods for Increasing the Running Efficiency of Equipment in 300-mm Semiconductor Fabrication

Yu-Chih Wang and Lee-Ing Tong

***Abstract*—Overall equipment efficiency (OEE) is widely adopted in semiconductor manufacturing to assess and enhance equipment productivity. The rate efficiency, which is a major component of OEE, is utilized to evaluate the achievement of actual production rate to the equipment's theoretical production rate. However, the rate efficiency is significantly influenced by factors beyond the equipment and, therefore, is inadequate for deciding whether a low production rate is due to the equipment itself. Such an inaccurate evaluation of production rate of equipment usually results in productivity loss. Hence, this study develops a novel metric, called running efficiency (RUNE), to compare the actual and theoretical production rates of the equipment. The RUNE is not affected by the production environment and can be employed to determine sources of equipment's variation. Additionally, a RUNE management procedure is also proposed. The management procedure incorporates an automatic target-setting scheme to set the theoretical production time of every motion in equipment to obtain the RUNE value of equipment. An exponentially weighted moving average control chart is then utilized in the management procedure to monitor the RUNE value of equipment. A real case from a 300-mm fabrication in Taiwan is employed to demonstrate the effectiveness of the proposed method.**

***Index Terms*—Overall equipment efficiency (OEE), productivity, rate efficiency, running efficiency (RUNE).**

I. Introduction

MANY semiconductor manufacturers have switched from 200-mm fabrications (fabs) to 300-mm fabs to lower the unit cost of chips in recent years. However, the investment of a 300-mm fab is approximately US$3–4 billion, and the equipment (also called processing tools) costs contribute approximately two thirds to three quarters of the total investments. Therefore, the manufacturers attempt to fully utilize the processing tools to maximize the productivities. The overall equipment efficiency (OEE), which is a composite index of the availability efficiency, operational efficiency, rate efficiency, and quality efficiency, is widely adopted to evaluate and improve the productivities of processing tools in the semiconductor industry. The OEE is derived from the total productive maintenance (TPM), which was promoted by the Japan Institute of Plan Maintenance (JIPM) in 1977 [1]. The OEE originally stood for overall equipment effectiveness. Semiconductor Equipment and Materials International (SEMI) has renamed it overall equipment efficiency, since the metric is expressed entirely in terms of time. The standard and definition of OEE are described in detail in other studies [2], [3].

Numerous OEE-based studies can be found in the field of productivity improvement of processing tools. For instance, Konopka [4] proposed a capacity utilization bottleneck efficiency system (CUBES) to analyze bottlenecks in production lines and to enhance the total manufacturing productivity. Pomorski [5] focused on the application of OEE and the loss analysis of major processing tools to optimize the performances of the constraint processing tools. Freck [6] used OEE to improve targeted constraint processing tool groups in order to elevate the fab throughput. Gouvêa da Costa and Pinheiro de Lima [7] carried out an exploratory field research on the misunderstandings and misuses of OEE in the automotive industry in Brazil. Huang *et al.* [8] considered OEE as a quantitative metric for measuring the productivities of the processing tools and concluded that OEE is better than traditional metrics for identifying the problems and underlying improvements that are needed to increase productivity.

Although OEE has been extensively adopted in the semiconductor industry, its rate efficiency has limitations. For instance, Jansson and Lesshammar [9] studied three OEE case applications in different industries, indicating that the rate efficiency is measured based on the maximum production speed and actual production speed. However, the maximum production speed may vary during different periods. Moreover, defining the maximum production speed is occasionally difficult. Leachman [10] concluded that, while acknowledging the potential value of OEE, many companies are daunted by the rigorous effort to collect and maintain the required data. Even when the same product is manufactured on the same processing tool, the rate efficiency may vary due to different operators or incorrect records of tool status. Ljungberg [11] attributed most losses of tool capacities and improvement opportunities to the underestimated theoretical production time. Therefore, an overestimated rate efficiency erroneously leads fab managers to believe that the processing tools with declining production rate still remain normal in terms of production rates. Ron and Rooda [12] noted that, in addition to excluding environment influential factors, a metric should be directed to the processing tools. Such a metric can be utilized to improve the performance of processing tools. That study also eliminated equipment-independent losses and replaced the rate efficiency with a rate factor, which is the ratio of actual manufacturing number of items to maximum number of items during production time.

Manuscript received January 08, 2008; revised November 20, 2009. First published April 22, 2010; current version published May 05, 2010.

The authors are with the Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: ycwangk@anet.net.tw; litong@cc.nctu.edu.tw).

Digital Object Identifier 10.1109/TSM.2010.2045679

0894-6507/$26.00 © 2010 IEEE

According to the definition of OEE given by SEMI [2]

$$\begin{aligned}\mathrm{OEE} = &\text{ availability efficiency}\\ &\times \text{ operational efficiency}\\ &\times \text{ rate efficiency} \times \text{ quality efficiency}\end{aligned} \quad (1)$$

with

$$\text{Rate efficiency} = \frac{\text{Theoretical production time for units}}{\text{Actual production time for units}}. \quad (2)$$

The true rate efficiency should be determined based on a fully loaded bottleneck machine because the production rate of the bottleneck machine determines the production rate of a processing tool. This study proposes the following running efficiency (RUNE) metric to exclude all factors apart from the processing tools and increase the measuring accuracy of the rate efficiency.

The RUNE is given by $\mathrm{RUNE} = T/A$, where T and A denote the theoretical and actual production times of a motion or an activity in a processing tool for each wafer, respectively. Both T and A are determined when a processing tool continuously manufactures wafers with the same recipe. The RUNE value of a processing tool can be evaluated from the ratio of the theoretical duration of wafer to wafer (W2W) to the steady-state actual duration of W2W on the bottleneck motion in a processing tool. The rate efficiency can be calculated using RUNE and bottleneck efficiency as follows:

$$\begin{aligned}\text{Rate efficiency} &= \frac{\text{Theoretical production time for units}}{\text{W2W time of bottleneck}}\\ &\quad \times \frac{\text{W2W time of bottleneck}}{\text{Production time for units}}\\ &= \text{RUNE} \times \text{bottleneck efficiency}\end{aligned} \quad (3)$$

where $\text{W2W time of bottleneck}/\text{Production time for units}$ denotes the bottleneck efficiency.

Because recipe changes and noncontinuously processing of wafers influence the bottleneck efficiency, the bottleneck efficiency can be utilized to evaluate how factors beyond the equipment affect the rate efficiency. Additionally, the RUNE can be utilized to evaluate the true production rate of a tool because its maximum production rate is determined by the production rate of the bottleneck motion inside the tool. Moreover, the RUNE values of motions in a processing tool can be evaluated from the ratio of the theoretical motion times to the actual motion times. Therefore, declining RUNE value of a processing tool facilitates the use of RUNE values of motions to identify the root causes of a declining production rate.

However, a 300-mm fab contains thousands of processing tools. Each processing tool includes numerous motions inside, possibly having various theoretical times under different recipes. Therefore, monitoring the RUNE values of a processing tool by thoroughly monitoring the ratio of the target production time to each wafer time spent on all critical motions in a processing tool is extremely difficult. This study proposes an automatic target-setting scheme to obtain the theoretical motion times and W2W durations of all critical motions for calculating the RUNE values under various recipes. Since the exponentially weighted moving average (EWMA) control chart is very effective against small process shifts and is a nonparametric procedure, it is then employed in the RUNE management procedure to monitor the RUNE values. Consequently, fab managers can effectively evaluate the RUNE value of each processing tool and detect a decline in production rate of a processing tool quickly. In the meantime, the maximum productivities of processing tools and the reduction of wafer production time can be achieved by maximizing the RUNE values of processing tools.

The remainder of this study is organized as follows. Section II reviews the conventional assessment method regarding the rate efficiency of a processing tool. Section III then describes the construction of the proposed RUNE management procedure. In Section IV, a real case is applied to demonstrate the effectiveness of the proposed method. This section also compares the results of the conventional PWPH-based management procedure and the proposed RUNE management procedure. Concluding remarks are finally made in Section V.

II. Review of the Conventional Accessing Method of the Rate Efficiency

According to (2), the rate efficiency can be evaluated by the ratio of the cumulative theoretical wafer production time to the cumulative actual wafer production time. However, due to the diversity of IC processes, a processing tool generally uses numerous recipes to produce wafers, with each recipe generally having a different theoretical production time. Therefore, evaluating the rate efficiency of a processing tool, depending on an individual lot, is meaningless. The ratio of actual wafers produced per hour (AWPH) to peak wafers produced per hour (PWPH) is then applied to evaluate the rate efficiency of a processing tool. The AWPH value is the wafer number produced by a tool in an hour. The PWPH value is a theoretical wafer number produced by a tool in an hour. The PWPH value of each recipe on a processing tool can be obtained by reviewing the production log of a processing tool under normal operating conditions and continuous production of wafers with the same recipe. Once wafers with different recipes are manufactured on a processing tool, the theoretical PWPH value can be obtained by the weighted average of PWPH values of all produced recipes by using the wafer quantities as the weights of the PWPH values.

However, the low loading of a processing tool may cause low work in progress (WIP) for the same recipe to be processed continuously. Additionally, an unbalanced WIP profile may cause the production recipes to change very frequently. Such noncontinuous processing of WIP with the same recipe leads to a low achievement rate of PWPH, since the bottleneck motion in a processing tool is not fully loaded. Hence, fab managers cannot easily determine whether the low rate efficiency of a processing tool originates from either the processing tool or from factors beyond the processing tool.

Identifying the root cause of poor rate efficiency of a processing tool in daily management requires that engineers measure the AWPH of a processing tool, which continuously produces WIP with the same recipe, for each recipe. A low ratio of AWPH to PWPH of a recipe on a processing tool implies a recipe with low rate efficiency on the tool. This finding also reveals that the production rate of the recipe on the processing

tool has decreased. However, frequently updating the theoretical PWPH values for all recipes and thoroughly monitoring all PWPH achievement rates of the processing tools for all recipes manually based on the PWPH-based measurement procedure is extremely difficult. Therefore, engineers often monitor the gap between AWPH and PWPH inaccurately and attempt to rationalize with reasons such as noncontinuous production and frequent recipe changeovers. Corrective action is taken to increase the production rate of a processing tool when the rate efficiency declines significantly.

Consider the enormous investment of 300-mm processing tools. Evaluating rate efficiency based on the achievement rate of PWPH cannot detect the declining production rate of a processing tool effectively. Such a method of assessing and enhancing the rate efficiency may incur lost productivity, resulting in a loss in throughput and a prolonged cycle time.

III. Continuous Improving Method for Enhancing the RUNE

Taylor and Heragu [13] remarked that reducing the variability, although not as effective as reducing the theoretical production time, is often a less expensive option for lowering the wafer cycle time. Delp *et al.* [14] also combined the utilization of equipment groups with the variability of the processing time, wafer arrival rate, and the availability of equipment in order to identify the capacity constraining equipment groups, and thus lower the wafer cycle time. Note that no opportunity of improvement should be missed in the highly competitive semiconductor manufacturing market. Accordingly, this study proposes RUNE, which is a novel metric for accurately accessing and enhancing the production rate of a processing tool, to replace the rate efficiency. The RUNE management procedure is also proposed. The management procedure incorporates an automatic target-setting scheme to set the theoretical values for the motion time and the W2W duration of each critical motion to easily obtain the RUNE value of a processing tool. The RUNE management procedure also implements EWMA control charts to monitor the RUNE values of processing tools. The entire RUNE management procedure can be divided into four steps, namely identifying critical motions, removing the nonsteady state data, obtaining theoretical values for the critical motions and employing the EWMA control chart to monitor the RUNE values. The four steps are described in detail as follows.

A. Step 1) Identify the Critical Motions in Equipment

Dettmer [15] restated that there is only one bottleneck in a system at any given time. That one bottleneck limits the output of the entire system. Although a processing tool has many motions, each with different theoretical time as determined by the recipe, the intensive monitoring resources can be reduced by concentrating on the bottleneck motions. The bottleneck motions can significantly influence the RUNE of a processing tool for all recipes. Fig. 1 illustrates the sequence of motions and required motion times of recipes A and B, respectively, in a processing tool. The rectangles in Fig. 1 represent the process motions, and the arrows represent the transfer motions, performed by robots, between process motions. The transfer motions only occur when upstream process motions are finished and the following process motions are available for the next wafer. Obviously, the bottleneck motions of recipes A and B are Process 2 (150 s) and Process 3 (150 s), respectively. Both processes are then identified as the critical motions of the processing tool. Throughput of the processing tool for recipes A and B can be maximized by fully utilizing the critical motions. Therefore, the W2W durations of Processes 2 and 3 are applied to access the RUNE values of this tool for recipes A and B, respectively. The production rate for recipe A (or B) declines once the actual W2W duration of the critical motion exceeds its theoretical time. Therefore, fab managers can detect the decline in production rate and determine whether this is a result of the critical motion or of noncritical motions, by simultaneously monitoring the W2W duration and motion time of the critical motion.

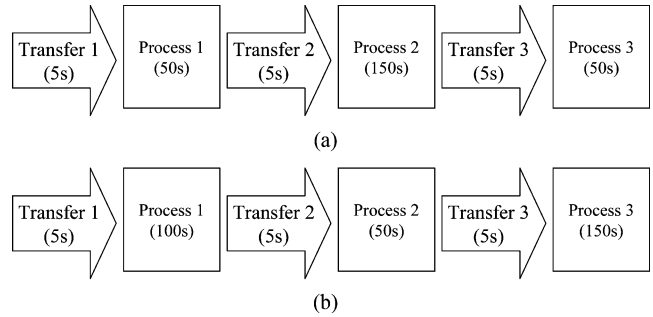

Fig. 1. Sequence of motions and required motion times in a processing tool for recipes A and B. (a) Required motion times for recipe A. (b) Required motion times for recipe B.

B. Step 2) Remove the Nonsteady-State Data

The next step after identifying the critical motions is to determine the theoretical values of the motion times and the W2W durations of each critical motion. The theoretical value of PWPH for each recipe is evaluated under the normal operating condition, which is processed continuously under WIP with the same recipe. Therefore, the requirement of continuous processing of WIP with the same recipe can be converted into an algorithm that removes nonsteady state data. Different eliminating algorithms for different tool types in a real fab can be performed automatically using computer integrated manufacturing (CIM) technology. The remaining data are the data showing the steady-state production rates of the processing tools. For instance, in Fig. 1, the W2W duration between the first wafer of recipe B and the last wafer of recipe A after the completion of Process 3 is 205 s. However, the steady-state output duration of recipe B on Process 3 is 155 s (Transfer 3 takes 5 s and Process 3 takes 150 s, respectively). Therefore, the data of the first wafer of recipe B are removed. Consequently, the influential factors apart from the processing tools, such as recipe changeovers, insufficient amount of WIP, and other environment factors, can be precisely excluded.

C. Step 3) Obtain the Theoretical Values for the Critical Motions

Although the monitoring resources can be effectively reduced by just monitoring the critical motions, the required resources are still enormous since the critical motions are performed interactively and have different theoretical values according to different recipes. Therefore, the theoretical values and specifi-

cations of the motion time and of the W2W duration for each critical motion are very difficult to set directly from the recipe for the subsequent management procedure. This study proposes an automatic target-setting scheme to save engineers' resources for setting the theoretical values and specifications, and to construct a continuously improving process for the RUNE. This automatic target-setting scheme obtains the theoretical values of the motion time and the W2W duration according to the steady-state data obtained in Step 2). Analysis of variance (ANOVA) [16] is applied to test whether the mean values of the motion time and W2W duration of the critical motions in different processing tools are significantly different. Since Duncan's multiple range test is very effective at detecting differences between means when real differences exist, it is a widely used procedure for comparing all pairs of means [16], [17]. Duncan's multiple range test is then applied to discover which tool groups have the smallest mean values if the ANOVA F-test is significant.

To apply Duncan's multiple range test for equal sample sizes n, the standard error of a treatment averages is given as

$$S_{\bar{y}_{i.}} = \sqrt{\frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{N-a}\cdot\frac{1}{n}}. \tag{4}$$

Duncan's significant ranges can then be converted into a set of $a-1$ least significant ranges R_p for $p = 2, 3, \ldots, a$, by computing

$$R_p = r_\alpha(p, f)S_{\bar{y}_{i.}} \quad \text{for} \quad p = 2, 3, \ldots, a \tag{5}$$

where α denotes the significance level and f denotes the number of degrees of freedom for error.

The differences between $a(a-1)/2$ pairs of treatment means are compared with the least significant range R_p, for $p = 2, 3, \ldots, a$. It begins with largest versus smallest, which would be compared with R_a. Next, the difference of the largest and the second smallest is compared with R_{a-1}. This process is continued until the differences of possible $a(a-1)/2$ pairs of means have been tested. Once the difference is greater than the corresponding R_p, the pair of means is said to be significantly different.

Based on Duncan's multiple range test, the smallest mean values of the motion time and of the W2W duration on each critical motion are then set as the theoretical values of the motion time and the W2W duration, respectively, for the recipe on all processing tools of the same type. The theoretical values of the motion time and the W2W duration are then periodically and automatically updated through the same procedure using CIM technology. Therefore, the process not only save resources for setting theoretical values, but also ensures that the new theoretical values increase equipment productivity.

D. Step 4) Monitor the RUNE Values by Using the EWMA Control Chart

The RUNE value for each wafer on the critical motion can be obtained by dividing the theoretical value obtained in Step 3) by the actual motion time. The RUNE value for each wafer on a processing tool can also be obtained by dividing the theoretical duration of W2W by the actual duration of W2W. Therefore, all actual motion times for different recipes on the same critical motion can be monitored with the same control charts to minimize the numbers of control charts, as well as the actual W2W duration of the critical motion. The productivity loss of equipment usually comes from the slight decline in production rate. This is because fab managers usually consider this slight decline as the result of noncontinuously processing of WIP with the same recipe. The EWMA control chart [18] is very effective against small process shifts, and is almost a perfectly nonparametric (distribution-free) procedure. Hence, the EWMA control chart is employed to monitor the RUNE value for each wafer on the critical motions and processing tools. Since larger RUNE value implies a higher production rate, the EWMA control chart only has a lower control limit (LCL). The parameter z_i and the LCL are given by

$$z_i = \lambda x_i + (1-\lambda)\, Z_{i-1}, \qquad \text{where } 0 < \lambda \leq 1 \text{ is a constant} \tag{6}$$

$$\mathrm{LCL} = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}\left[1-(1-\lambda)^{2i}\right]} \tag{7}$$

where

μ_0 target value of process;

L width of the control limits;

σ standard deviation of observations x_i.

Note that, as i gets larger, the LCL will approach steady-state value and is given by

$$\mathrm{LCL} = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}}. \tag{8}$$

Montgomery [19] noted that $0.05 \leq \lambda \leq 0.25$ works well in practice and $\lambda = 0.1$ is a popular choice. Based on the results of Lucas and Saccucci [19], the EWMA control chart in this study uses $\lambda = 0.1$ and $L = 2.814$, which will result in an in-control average run length $(\mathrm{ARL}_0) \cong 500$ and an ARL for detecting a shift of one standard deviation in the mean $(\mathrm{ARL}_1) \cong 10.3$.

Fig. 2 displays the entire framework of the RUNE management procedure. The RUNE management procedure not only automatically and periodically updates the small theoretical values for the motion times and the W2W durations of all critical motions, but also quickly detects the declining RUNE values by employing EWMA control charts. Thus, fab managers can easily determine whether the production rate of a processing tool is decreasing and can quickly infer the root causes in order to take corrective actions.

IV. Case Study

A real case from a Taiwanese 300-mm semiconductor fabrication was applied to demonstrate the effectiveness of the proposed RUNE management procedure. (To ensure confidentiality, the equipment ID and recipe ID were renamed, and the data were transformed.) Four processing tools of the same type, namely CVDA1, CVDA2, CVDA3, and CVDA4, provide one deposition recipe for the entire IC process. Each processing tool mainly has two process chambers (Ch-A and Ch-B), and each chamber has two sites for processing wafer simultaneously.

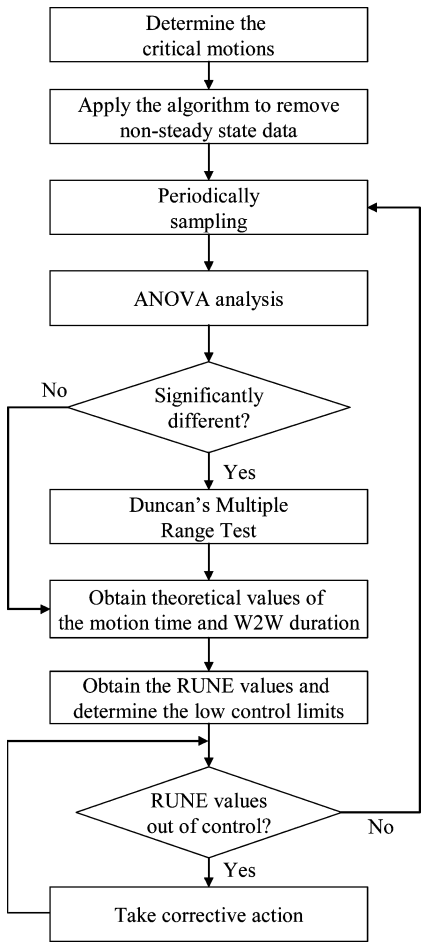

Fig. 2. Framework of RUNE management procedure.

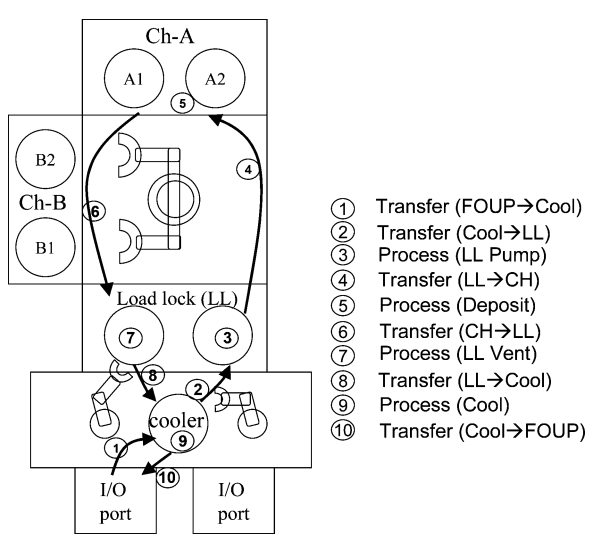

Fig. 3. Sequence motions in CVDA1-CVDA4.

Fig. 3 depicts the sequence of motions in the tools. All motions can be categorized into two groups, namely the transfer group (Motion 1, 2, 4, 6, 8, and 10) and the process group (Motion 3, 5, 7, and 9).

TABLE I
RANDOMLY SAMPLED MOTION TIMES IN THE PROCESSING TOOL (UNIT: SECONDS)

Group	Process Group				Transfer Group					
Motion	3	5	7	9	1	2	4	6	8	10
Time	20	130	19	50	5	4	12	12	4	5

A. Step 1) Identify the Critical Motions

To identify the critical motions, each motion time was randomly selected from the production log of the processing tool. Table I summarizes the motion times, revealing that Motion 5 had the longest motion time (130 s). Therefore, Motion 5 was identified as the critical motion, and its duration of W2W, denoted by 5_W2W, was employed to evaluate the RUNEs of the four processing tools. Because past experience indicated that an unsmooth or abnormal setting in a robot usually causes the process chambers to idle, the monitored list also included the motion times of Motion 4 and Motion 6 between the load lock (LL) and process chambers. Since the only difference between Motion 4 and Motion 6 is the transfer direction, only the motion time of Motion 4 was monitored to avoid redundant monitoring. Finally, the motion times of Motion 4 and Motion 5, and the duration of 5_W2W were selected for monitoring in the RUNE management procedure.

B. Step 2) Remove the Nonsteady-State Data

Because Ch-A and Ch-B are both process chambers and the wafer could be processed at either Ch-A or Ch-B, the motion time of Motion 5 and the duration of 5_W2W on Ch-A and Ch-B were monitored separately. Since the robot transfer distances were different from LL to Ch-A and Ch-B, the motion times of Motion 4 in the four processing tools were compared based on the average motion time from LL to Ch-A and to Ch-B. Ten observations of continuously processing of wafers with the same recipe on each monitored motion in each processing tool were sampled to ensure that the sampled data could represent the steady-state production rates of the processing tools. Table II summarizes the observations.

C. Step 3) Obtain the Theoretical Values

1) ANOVA for the Performances of Critical Motions From Various Processing Tools: ANOVA was applied to test whether the mean values of the motion times of Motion 4 and Motion 5, and of the duration of 5_W2W from various tools are significantly different. Table III summarizes the ANOVA results from the data in Table II. These results indicate that there are significant differences between the mean values of the four tools on the motion time of Motion 4 and the duration of 5_W2W, respectively. Duncan's multiple range test was then applied to determine which tool groups had the significantly smallest mean values of motion time of Motion 4 and duration of 5_W2W, respectively. These mean values can then be set as the theoretical values of motion time of Motion 4 and the duration of 5_W2W, respectively.

2) Duncan's Multiple Range Test for the Motion Times of Various Processing Tools: The least significant ranges R_p of (5) with $\alpha = 0.05$ were obtained and are summarized in Table IV. According to the result of Duncan's multiple

TABLE II
OBSERVATIONS FROM CVDA1–CVDA4 (UNIT: SECONDS). (a) MOTION TIMES OF MOTION 4 FROM VARIOUS TOOLS. (b) MOTION TIMES OF MOTION 5 FROM PROCESS CHAMBERS IN VARIOUS TOOLS. (c) DURATIONS OF 5_W2W FROM PROCESS CHAMBERS IN VARIOUS TOOLS

no	CVDA1	CVDA2	CVDA3	CVDA4
1	11	11.5	11	13.5
2	11	12	11.5	13.5
3	11.5	12	11.5	14.5
4	12	12.5	11	14
5	12	11.5	11.5	13.5
6	11.5	11.5	11	13.5
7	11.5	11.5	12	13.5
8	12	11.5	11.5	13.5
9	12	11.5	11.5	13.5
10	11.5	11.5	12	14
mean	11.6	11.7	11.45	13.7

(a)

no	CVDA1-A	CVDA1-B	CVDA2-A	CVDA2-B	CVDA3-A	CVDA3-B	CVDA4-A	CVDA4-B
1	131	130	131	131	131	130	131	131
2	130	130	131	130	130	130	131	130
3	131	131	130	130	130	131	130	130
4	130	130	131	129	131	131	130	131
5	131	131	130	130	130	130	129	130
6	130	130	130	129	131	131	131	130
7	130	131	131	130	129	130	131	131
8	131	130	129	130	131	131	132	130
9	130	130	130	130	131	130	129	131
10	129	130	131	131	131	130	130	131
mean	130.3	130.3	130.4	130	130.5	130.4	130.4	130.5

(b)

no	CVDA1-A	CVDA1-B	CVDA2-A	CVDA2-B	CVDA3-A	CVDA3-B	CVDA4-A	CVDA4-B
1	174	173	175	170	174	171	178	179
2	174	175	174	171	174	172	180	179
3	173	174	176	171	174	171	179	214
4	174	173	175	176	174	172	179	212
5	174	177	175	170	174	174	179	214
6	174	174	174	173	173	174	179	202
7	173	174	175	175	174	174	178	182
8	174	174	175	175	174	171	179	179
9	174	173	175	175	175	179	178	178
10	174	175	174	171	174	175	183	174
mean	173.8	174.2	174.8	172.7	174	173.3	179.2	191.3

(c)

TABLE III
ANOVA FOR THE PERFORMANCES OF CRITICAL MOTIONS FROM VARIOUS TOOLS. (a) ANOVA FOR THE MOTION TIMES OF MOTION 4. (b) ANOVA FOR THE MOTION TIMES OF MOTION 5. (c) ANOVA FOR THE DURATIONS OF 5_W2W

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-level
Motion 4	293.2938	7	41.8991	74.97	<0.001
Error	84.9500	152	0.5589		

(a)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-level
Motion 5	164.4750	7	23.4964	1.62	0.135
Error	2210.3000	152	14.5414		

(b)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-level
5_W2W	2482.1938	7	354.5991	6.04	<0.001
Error	8917.5498	152	58.6681		

(c)

range test, CVDA1, CVDA2 and CVDA3 were in the same tool group, with the smallest mean motion time of Motion 4. Additionally, Ch-A and Ch-B in CVDA1, CVDA2 and CVDA3 were in the same tool group with the smallest mean value of 5_W2W duration. Fig. 4 displays the results of Duncan's multiple range test for the motion time of Motion 4 and the duration of 5_W2W from various processing tools.

D. Step 4) Monitor the RUNE Values by Employing the EWMA Control Chart

Because the motion times of Motion 5 in the four processing tools showed no significant difference in the ANOVA F-test, the mean value of the motion times of Motion 5 in Ch-A and Ch-B of the four processing tools could then be set as the theoretical value of the motion time of Motion 5. Based on Duncan's multiple range test, the theoretical value of Motion 4 was set to the mean value of its motion times in CVDA1, CVDA2 and CVDA3. The theoretical value of the 5_W2W was set to the mean 5_W2W duration of Ch-A and Ch-B in CVDA1, CVDA2 and CVDA3. Table V presents the theoretical values of motion times of Motion 4 and Motion 5, and of the duration of 5_W2W. Accordingly, the RUNE values of critical motions for each wafer can be calculated by dividing the theoretical values by the observations from Table II. For example, the RUNE value of Motion 4 for each wafer can be calculated by dividing the theoretical value (11.583) by the observations in Table II. The mean value 1.000 and the standard deviation 0.032 of the RUNE values of CVDA1, CVDA2 and CVDA3 on Motion 4 were utilized to obtain $\mathrm{LCL} = 0.979$ using (6)–(8) with $\lambda = 0.1$ and $L = 2.814$. To verify the effectiveness of EWMA control chart, Table VI lists the RUNE and EWMA values of Motion 4, Motion 5, and 5_W2W duration for each wafer on CVDA1 and CVDA4 for benchmarking. The data denoted by $*$ in Table VI are denoted as the out-of-control sample points since these data fall outside the LCLs.

Since the 5_W2W duration represents the output duration of the processing tool, the low mean RUNEs of 5_W2W durations of Ch-A and Ch-B in CVDA4 indicate that CVDA4 has low RUNE and this reveals a low production rate. The Motion 4 was then strongly suspected to be the root cause of the low RUNE of CVDA4, since the mean RUNE value of Motion 4 in CVDA4 was significantly lower than those of other processing tools. Engineers examining CVDA4 found that the decline in RUNE values of Motion 4 resulted from an abnormal setting of the lift pins in Ch-A and Ch-B. The normal setting of the lift pins in both sites of Ch-A or Ch-B was to lift both wafers from the

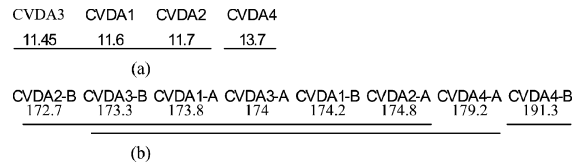

Fig. 4. Results of Duncan's multiple range test. (a) Pairwise comparison of mean motion times of Motion 4 from various tools. (b) Pairwise comparison of mean durations of 5_W2W from process chambers in various tools.

TABLE IV
LEAST SIGNIFICANT RANGES

	Motion time of Motion 4	Duration of 5_W2W
R2	0.331	5.517
R3	0.348	5.810
R4	0.359	6.005
R5	--	6.122
R6	--	6.239
R7	--	6.317
R8	--	6.395

TABLE V
THEORETICAL VALUES OF CRITICAL MOTIONS (UNIT: SECONDS)

	Motion time of Motion 4	Motion time of Motion 5	Duration of 5_W2W
Theoretical value	11.583	130.350	173.800

TABLE VI
RUNE AND EWMA VALUES FOR EACH OBSERVATION OF THE CRITICAL MOTIONS IN CVDA1 AND CVDA4. (a) RUNE AND EWMA VALUES OF MOTION 4. (b) RUNE AND EWMA VALUES OF MOTION 5. (c) RUNE AND EWMA VALUES OF 5_W2W

Machine	CVDA1		CVDA4	
no	RUNE	EWMA	RUNE	EWMA
1	1.053	1.005	0.858	0.986
2	1.053	1.010	0.858	0.973*
3	1.007	1.010	0.799	0.956*
4	0.965	1.005	0.827	0.943*
5	0.965	1.001	0.858	0.934*
6	1.007	1.002	0.858	0.927*
7	1.007	1.002	0.858	0.920*
8	0.965	0.999	0.858	0.914*
9	0.965	0.995	0.858	0.908*
10	1.007	0.997	0.827	0.900*
mean	0.999	(LCL= 0.979)	0.848	(LCL= 0.979)

(a)

Machine	CVDA1-A		CVDA1-B		CVDA4-A		CVDA4-B	
no	RUNE	EWMA	RUNE	EWMA	RUNE	EWMA	RUNE	EWMA
1	0.995	1.000	1.003	1.000	0.995	1.000	0.995	1.000
2	1.003	1.000	1.003	1.001	0.995	0.999	1.003	1.000
3	0.995	0.999	0.995	1.000	1.003	0.999	1.003	1.000
4	1.003	1.000	1.003	1.000	1.003	1.000	0.995	1.000
5	0.995	0.999	0.995	1.000	1.010	1.001	1.003	1.000
6	1.003	1.000	1.003	1.000	0.995	1.000	1.003	1.000
7	1.003	1.000	0.995	1.000	0.995	1.000	0.995	1.000
8	0.995	0.999	1.003	1.000	0.988	0.998	1.003	1.000
9	1.003	1.000	1.003	1.000	1.010	1.000	0.995	0.999
10	1.010	1.001	1.003	1.000	1.003	1.000	0.995	0.999
mean	1.000	(LCL = 0.997)	1.000	(LCL = 0.997)	1.000	(LCL = 0.997)	0.999	(LCL = 0.997)

(b)

Machine	CVDA1-A		CVDA1-B		CVDA4-A		CVDA4-B	
no	RUNE	EWMA	RUNE	EWMA	RUNE	EWMA	RUNE	EWMA
1	0.999	1.000	1.005	1.000	0.976	0.998	0.971	0.997
2	0.999	1.000	0.993	1.000	0.966	0.994	0.971	0.994
3	1.005	1.000	0.999	1.000	0.971	0.992*	0.812	0.976*
4	0.999	1.000	1.005	1.000	0.971	0.990*	0.820	0.961*
5	0.999	1.000	0.982	0.998	0.971	0.988*	0.812	0.946*
6	0.999	1.000	0.999	0.998	0.971	0.986*	0.860	0.937*
7	1.005	1.000	0.999	0.998	0.976	0.985*	0.955	0.939*
8	0.999	1.000	0.999	0.998	0.971	0.984*	0.971	0.942*
9	0.999	1.000	1.005	0.999	0.976	0.983*	0.976	0.946*
10	0.999	1.000	0.993	0.998	0.950	0.980*	0.999	0.951*
mean	1.000	(LCL = 0.994)	0.998	(LCL = 0.994)	0.970	(LCL = 0.994)	0.915	(LCL = 0.994)

(c)

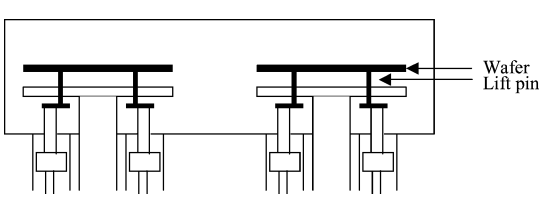

Fig. 5. Cross-sectional figure of the process chamber.

TABLE VII
MEAN RUNE VALUES OF ALL CRITICAL MOTIONS IN CVDA1–CVDA4

Tool_ID	Mean RUNE of 5_W2W	Mean RUNE of Motion 4	Mean RUNE of Motion 5
CVDA1	0.999	1.000	1.000
CVDA2	1.000	0.991	1.001
CVDA3	1.001	1.013	0.999
CVDA4	0.942	0.846	0.999
mean	0.986	0.962	1.000

TABLE VIII
PWPH-BASED RATE EFFICIENCIES OF CVDA1–CVDA4

Tool_ID	Production hours (a)	Up hours (b)	Operational efficiency (b/a)	AWPH (c)	PWPH (d)	Rate Efficiency (c/d)
CVDA1	13.70	23.76	0.58	47.50	57.00	0.83
CVDA2	12.36	23.66	0.52	49.10	57.00	0.86
CVDA3	7.61	18.50	0.41	56.40	57.00	0.99
CVDA4	8.18	23.69	0.35	48.70	57.00	0.85
mean	10.46	22.40	0.46	50.43	57.00	0.88

robot simultaneously. However, the lift pins of CVDA4 lifted the wafer first in one site, then in the other site. Fig. 5 illustrates the cross-sectional figure of the process chamber. The abnormal setting of the lift pins lengthened the staying time of the wafer in Ch-A and Ch-B, and on the robot. Therefore, the mean RUNE value of Motion 4 in CVDA4 declined to around 84.6% (mean RUNE value of $\mathrm{Motion}\ 4 = 0.846$), and the mean RUNE value of CVDA4 declined to around 94.2% (mean RUNE value of $5_\mathrm{W2W} = 0.942$). Table VII summaries the mean RUNE values of the critical motions in the four processing tools.

Table VIII compares the conventional PWPH-based management procedure with the RUNE management procedure, in which the PWPH-based rate efficiencies are obtained directly from the semiconductor fabrication. A lower operational efficiency in Table VIII implies a low ability of WIP to load bottleneck motions inside the tools. Due to non-continuous processing of WIP, the rate efficiencies of the processing tools are spread widely, thus providing meaningless information of production rates of the processing tools to fab managers. Since the rate efficiency of CVDA4, i.e., 0.85, was extremely close to

the rate efficiencies of CVDA1 (0.83) and CVDA2 (0.86), the accessing results could not be used by the engineers to detect the abnormal pin setting of Ch-A and Ch-B in CVDA4. Comparing the rate efficiencies in Table VIII with RUNE values in Table VII reveals that the RUNE value is clearly more accurate and effective than the PWPH-based rate efficiency since the RUNE value excludes factors beyond the equipment and adopt CIM technology and statistical process control (SPC) to detect the significant decline of production rate of a processing tool efficiently. Also, the RUNE value helps fab managers to prioritize the corrective actions for tools with a declining production rate. Hence, limited engineer manpower, working time, and capital can be utilized to maximize the fab throughput and to reduce the wafer manufacturing cycle time by using RUNE management procedure.

V. Conclusion

The rate efficiency of OEE has been extensively employed in semiconductor manufacturing to raise the production rate of a processing tool. However, whether the low production rate of a processing tool is caused by the processing tool is difficult to determine, since the rate efficiency is significantly influenced by factors apart from the processing tool. The RUNE is a metric that includes only the factors from the processing tool when accessing the production rate of a processing tool. The RUNE management procedure is proposed using an automatic target-setting scheme and EWMA control charts. The automatic target-setting scheme periodically and automatically decreases the theoretical values, thus easily obtaining the RUNE values of a processing tool. The EWMA control chart is utilized as an easy method to detect whether the RUNE value is declining. The mean RUNE value obtained from different processing tools of the same type can be applied to represent the overall RUNE of the tool group of the same type. Therefore, fab managers can then easily and accurately monitor the production rates of all processing tools by monitoring the overall RUNE values of different processing tool groups. The proposed RUNE management procedure thus allows fab managers to maximize the productivities of processing tools in a 300-mm fabrication and to reduce the wafer cycle time.

References

[1] SEMATECH, Overall Equipment Efficiency (OEE) Guidebook Revision 1.0 SEMATECH, 1995.

[2] *Standard for Definition and Measurement of Equipment Productivity, Semiconductor Equipment and Material International*, SEMI E79-0200, 2000.

[3] *Standard for Definition and Measurement of Equipment Reliability, Availability, and Maintainability*, SEMI E10-0701, 2001.

[4] J. M. Konopka, "Capacity utilization bottleneck efficiency system (CUBES)," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 18, no. 3, pp. 484–491, Sep. 1995.

[5] T. Pomorski, "Managing overall equipment effectiveness [OEE] to optimize factory performance," in *Proc. IEEE Int. Symp. Semicond. Manuf. (ISSM)*, 1997, pp. A33–A36.

[6] R. W. Freck, "Using overall equipment effectiveness (OEE) and the equipment improvement process (EI) to improve fab throughput," in *Proc. IEEE/SEMI Adv. Semicond. Manuf. Conf. (ASMC)*, 2000, pp. 469–471.

[7] S. E. Gouvêa da Costa and E. Pinheiro de Lima, "Uses and misuses of the overall equipment effectiveness for production management," in *Proc. IEEE SEMI Adv. Int. Semicond. Manuf. Sci. Symp.*, 2002, pp. 816–820.

[8] S. H. Huang, J. P. Dismukes, J. Sui, Q. Su, M. A. Razzak, R. Bodhale, and D. E. Robinson, "Manufacturing productivity improvement using effectiveness metrics and simulation analysis," *Int. J. Prod. Res.*, vol. 41, no. 3, pp. 513–527, Feb. 2003.

[9] P. Jansson and M. Lesshammar, "Evaluation and improvement of manufacturing performance measurement systems—The role of OEE," *Int. J. Oper. Prod. Manag.*, vol. 19, no. 1, pp. 55–78, 1999.

[10] R. C. Leachman, "Closed-loop measurement of equipment efficiency and equipment capacity," *IEEE Trans. Semicond. Manufact.*, vol. 10, no. 1, pp. 84–97, Feb. 2000.

[11] Õ Ljungberg, "Measurement of overall equipment effectiveness as a basis for TPM activities," *Int. J. Oper. Prod. Manag.*, vol. 18, no. 5, pp. 495–507, 1998.

[12] A. J. de Ron and J. E. Rooda, "Equipment effectiveness: OEE revisted," *IEEE Trans. Semicond. Manuf.*, vol. 18, no. 1, pp. 190–196, Feb. 2005.

[13] G. D. Taylor and S. S. Heragu, "A comparison of mean reduction versus variance reduction in processing times in flow shops," *Int. J. Prod. Res.*, vol. 37, no. 9, pp. 1919–1934, Jun. 1999.

[14] D. Delp, J. Si, and J. W. Fowler, "The development of the complete X-factor contribution measurement for improving cycle time and cycle time variability," *IEEE Trans. Semiconduct. Manufact.*, vol. 19, pp. 352–362, Aug. 2006.

[15] H. W. Dettmer, *Goldratt's Theory of Constraints: A System Approach to Continuous Improvement*. Milwaukee, WI: ASQC Quality, 1997, pp. 2–28.

[16] D. C. Montgomery, *Design and Analysis of Experiments*, 4th ed. New York: Wiley, 1996, pp. 67–77.

[17] D. B. Duncan, "Multiple range and multiple F tests," *Biometrics*, vol. 11, no. 1, pp. 1–42, Mar. 1955.

[18] D. C. Montgomery, *Introduction to Statistical Quality Control*, 4th ed. New York: Wiley, 2001, pp. 239–249.

[19] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: Properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–16, Feb. 1990.

Yu-Chih Wang received the B.S. and M.S. degrees in industrial engineering and management from National Chiao Tung University, Hsinchu, Taiwan, in 1997 and 1999, respectively, where he is currently working toward the Ph.D. degree.

His research interests include productivity improvement of wafer fabrication, cycle time reduction and capacity analysis for a semiconductor firm.

Lee-Ing Tong received the B.S. degree from National Cheng Kung University, Tainan, Taiwan, and the M.S. and Ph.D. degree from the University of Kentucky, Lexington, all in statistics.

She is a Professor with the Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan. She was an Associate Professor with the Department of Mathematics, University of Massachusetts-Dartmouth. Her current research activities include quality engineering and management and statistics in industrial applications.