

Age-Based Double EWMA Controller and Its Application to CMP Processes

Argon Chen and Ruey-Shan Guo, *Member, IEEE*

Abstract—In the originally proposed run-by-run control scheme, the EWMA statistic is used as an estimate of the process deviation from its target. However, the controller based on the EWMA statistic is not sufficient for controlling a wearing out process. The PCC controller has been thus proposed to enhance the run-by-run controller capability. In this paper, we first reexamine the fundamentals of the PCC formulations and propose an adjustment that is advantageous in controlling processes subject to both random shifts and drifts. The adjusted PCC controller is then further refined to take into account the process age. This age-based double EWMA scheme is then applied to the CMP process, which is known in the semiconductor industry to be rather unstable.

Index Terms—CMP process, EWMA, run-by-run control.

I. INTRODUCTION

THE USE of the exponentially weighted moving average (EWMA) statistic for estimating process deviations has been widely studied and adopted in practice. Box and Jenkins [1] show that a controller based on the EWMA statistic is a minimum mean square error (MMSE) controller when the underlying process disturbance follows the IMA(1, 1) (first-order integrated moving average) process. An EWMA control scheme can be also shown equivalent to an optimal PID controller for a second-order dynamic process under the IMA(1, 1) disturbance [2]. In practice, the EWMA statistic has been shown to be quite effective even for processes subject to disturbances other than the IMA process. In particular, applications in the semiconductor process industry, known as run-by-run process control, have shown that the EWMA statistic is also capable of bringing processes with linear drift under control [5], [8]–[10]. Suppose at the i th run of the process the output Y_i can be controlled linearly by an input variable X_i and is subject to a natural process disturbance ε_i ($\sim N(0, \sigma^2)$) and a systematic disturbance δ_i . The process model can be written as

$$Y_i = \alpha + \beta X_i + \varepsilon_i + \delta_i = (\alpha + \delta_i) + \beta X_i + \varepsilon_i \quad (1)$$

where α is the linear process model's intercept term and β is the system gain that translates the input variable size to the size of process response. In this study, β is assumed known or closely estimated. It should be noted that in the control

engineering literature an important subject is to identify the system dynamics [3], [4], whereas the focus of run-by-run control schemes is mostly on the systematic disturbance since the run-to-run post-process measurements are rather independent. The EWMA statistic, here, is used to estimate the size of the process intercept (α) plus its systematic deviation at the $(i + 1)$ th run (δ_{i+1}):

$$a_i = w(Y_i - \beta X_i) + (1 - w)a_{i-1} \\ = \sum_{j=1}^i [w(1 - w)^{i-j}(Y_j - \beta X_j)] \quad (2)$$

where the weight w is usually set between 0 and 1; and a_i is an exponentially weighted average of the historic deviations $\alpha + \delta_j$ from the first run to the i th run. Thus, the process output at the $(i + 1)$ th run is estimated to be

$$\hat{Y}_{i+1} = a_i + \beta X_{i+1}. \quad (3)$$

In order to keep the process output at a predetermined target level (T), we obtain the process recipe for the $(i + 1)$ th run

$$X_{i+1} = \frac{T - a_i}{\beta}. \quad (4)$$

Such a controller is insufficient for processes subject to systematic wear-out. Several authors [6], [7] have addressed this problem and propose using two EWMA formulas

$$a_i = w_1(Y_i - \beta X_i) + (1 - w_1)a_{i-1} \\ p_i = w_2(Y_i - \beta X_i - a_{i-1}) + (1 - w_2)p_{i-1} \quad (5)$$

where w_1 and w_2 are the weights for the first and second EWMA equations, respectively, and p_i is used to compensate for the error incurred by a_i . $a_i + p_i$ is then used to estimate $\alpha + \delta_{i+1}$ and the process recipe at $i + 1$ becomes

$$X_{i+1} = \frac{T - (a_i + p_i)}{\beta}. \quad (6)$$

Such a control scheme is referred to as predictor corrector controller (PCC) scheme, termed by Butler and Stefani [6].

In this paper, we first examine the time domain behavior of the PCC scheme. Based on the observation, we propose an adjustment that has a more straightforward asymptotic behavior. The adjusted PCC formula is then further refined to take the process age into consideration. The CMP process will be finally used as an example to illustrate the methodology.

Manuscript received July 2, 2000. This research was supported in part by NSC89-2218-E-002-009, Taiwan, R.O.C.

A. Chen is with the Graduate Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan 106 (e-mail: achen@ccms.ntu.edu.tw).

R.-S. Guo is with the Department of Industrial Management and Business Administration, National Taiwan University, Taipei, Taiwan (e-mail: rsguo@ccms.ntu.edu.tw).

Publisher Item Identifier S 0894-6507(01)01039-9.

II. PCC AND DOUBLE EWMA FORMULA

Suppose that the process is subject to a linear drift. That is,

$$Y_i = \alpha + \beta X_i + \varepsilon_i + ci\sigma = (\alpha + ci\sigma) + \beta X_i + \varepsilon_i \quad (7)$$

where the process output is systematically drifting away by a size of $c\sigma$ per process run. To see how the estimates a_i and p_i work, we can examine the process asymptotic behavior. As time approaches infinity, we may write (see Appendix A)

$$E(a_i) = \alpha - \frac{c\sigma}{w_I} + c\sigma(i+1) + o(i)$$

where

$$\lim_{i \rightarrow \infty} o(i) = 0 \quad (8)$$

and

$$\lim_{i \rightarrow \infty} E(p_i) = \frac{c\sigma}{w_I}. \quad (9)$$

From (8) and (9), the asymptotic expected values of a_i and p_i do not have clearly readable meanings. However, after substituting (8) and (9) into (6) and (7), we can obtain the steady-state process output

$$\lim_{i \rightarrow \infty} E(Y_i) = T. \quad (10)$$

Equation (10) shows that, under a linear drift disturbance, a process controlled by the PCC scheme is a stable system and its expected output converges to the desired target T .

The estimate for $\alpha + \delta_{i+1}$ in the PCC formula can be further rewritten as (see Appendix B)

$$\begin{aligned} a_i + p_i &= (w_I + w_2 - w_I w_2) \sum_{j=1}^i e_j + w_I w_2 \sum_{j=1}^i \sum_{k=1}^j e_k \\ &\quad + a_0 + (w_I i + 1)p_0 \\ &= w_I \sum_{j=1}^i e_j + w_{II} \sum_{j=1}^i \sum_{k=1}^j e_k + a_0 + (w_I i + 1)p_0 \end{aligned} \quad (11)$$

where $e_k = Y_k - T$. This is in effect an Integral-double-Integral (I-II) controller, in contrast to the proportional-integral-derivative (PID) controller. In this I-II controller, the control action is proportional to the summation of the output errors and to the summation of summations of the output errors. The I-II controller can be shown to be an MMSE controller for processes subject to IMA(2, 2) disturbances [1]. In (11), we can observe that the controller's integral constant is $w_I = w_I + w_2 - w_I w_2$ and the double-integral constant is $w_{II} = w_I w_2$. We define a *control space* of the I-II controller as the space filled up by all possible settings of control constants (w_I, w_{II}), as shown in Fig. 1. In Fig. 1, the controller allows both w_I and w_{II} to be set between 0 and 1. For the PCC controller, we usually set w_1 and w_2 between 0 and 1 as well. This, however, leads to a limited control region in the (w_I, w_{II}) control space, as the shaded area shown in Fig. 1.

Now, we would like to propose an adjustment of the PCC formula such that the control region can be extended to the entire

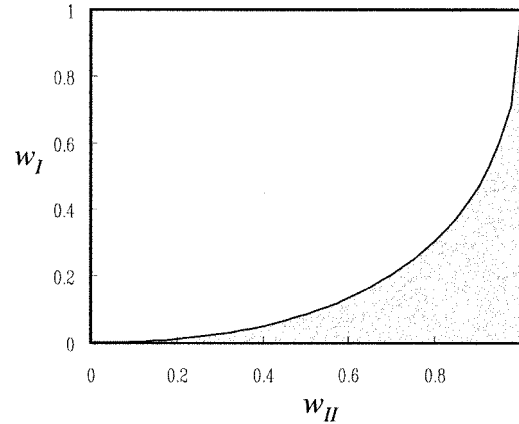
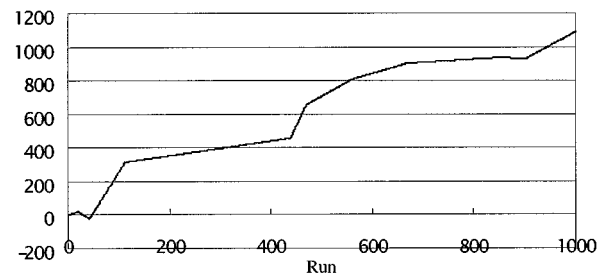
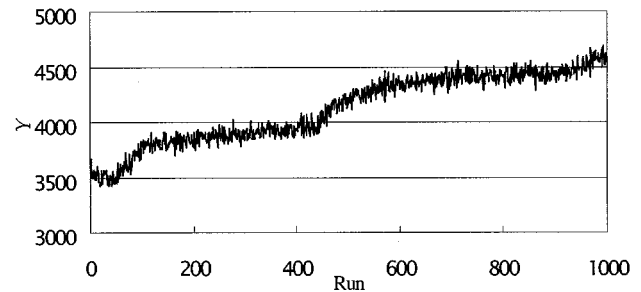


Fig. 1. Control space of I-II controller.



(a) Systematic deviation - random drifts



(b) Process output

Fig. 2. A typical random drift process.

space and the asymptotic estimates have more straightforward meanings. This proposed double EWMA (d-EWMA) formula is distinct from the original PCC formula:

$$\begin{aligned} a_i &= w_I(Y_i - \beta X_i) + (1 - w_I)(a_{i-1} + p_{i-1}) \\ p_i &= w_2(Y_i - \beta X_i - a_{i-1}) + (1 - w_2)p_{i-1}. \end{aligned} \quad (12)$$

As can be observed in (12), the only difference from the original PCC formula is that in the first EWMA formula, we add p_{i-1} into the formula. As a result, the I-II controller becomes (see Appendix C)

$$a_i + p_i = w_I \sum_{j=1}^i e_j + w_2 \sum_{j=1}^i \sum_{k=1}^j e_k + (i+1)p_0 + a_0. \quad (13)$$

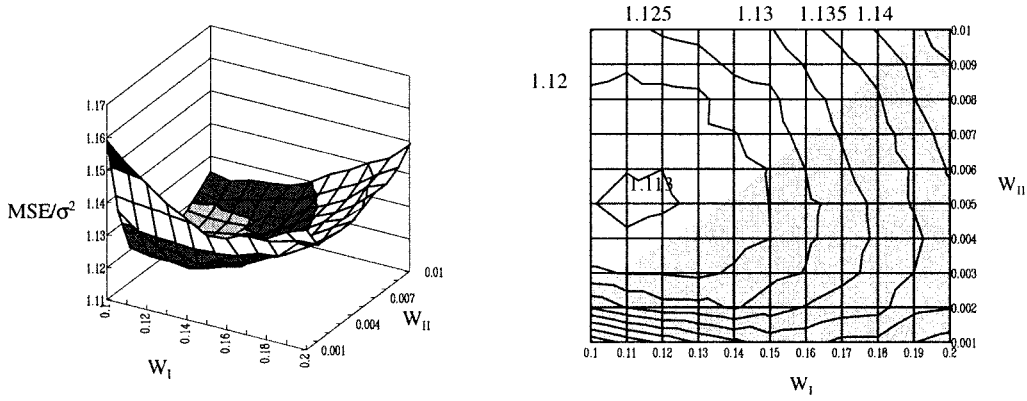


Fig. 3. Contour plots for the I-II control efficiency over the control space.

That is, $w_I = w_1$ and $w_{II} = w_2$ and with $0 < w_1 < 1$ and $0 < w_2 < 1$ we will have an I-II controller that fills up the entire control space.

With this d-EWMA formula, we can also obtain a more straightforward asymptotic behavior of the controller

$$E(a_i) = \alpha + ci\sigma + o(i)$$

where

$$\lim_{i \rightarrow \infty} o(i) = 0 \quad (14)$$

and

$$\lim_{i \rightarrow \infty} E(p_i) = c\sigma. \quad (15)$$

Equations (14) and (15) are inductively derived in Appendix D. This result is more straightforward than the original PCC formula. Moreover, the two estimates a_i and p_i here have clearly defined meanings. a_i in this d-EWMA formula represents an estimate for $\alpha + \delta_i$ and p_i is an estimate for the size of the process drift from the i th run to the $(i + 1)$ th run. Together, $a_i + p_i$ is the estimate for $\alpha + \delta_{i+1}$. Substituting (14) and (15) into (6) and (7), the process output can be found also converging to the desired target T .

Though both controllers (PCC and d-EWMA) are “unbiased” controllers, the advantages of d-EWMA controller over the PCC controller are twofold.

- 1) d-EWMA controller is a direct form of I-II controller.
- 2) The asymptotic behavior of a_i and p_i in the d-EWMA formula reveals more straightforward meanings.

To illustrate these two advantages, we simulate a random drift process where $T = 3500$, $a_0 = \alpha = 3000$, $\beta = 10$, and $p_0 = 0$. The process drift speed c is a random number over process runs, i.e.,

$$c_i = \begin{cases} c_{i-1}, & \text{with probability of 0.99} \\ \sim N(0.02, 0.05), & \text{with probability of 0.01.} \end{cases} \quad (16)$$

A typical process with such random drift disturbances is shown in Fig. 2.

To evaluate the control efficiency, we use a normalized mean squared error (MSE/σ^2) as the performance measure. MSE is defined as

$$MSE = \frac{\sum_{i=1}^n (Y_i - T)^2}{n} \quad (17)$$

where n is the total number of runs. MSE/σ^2 is then a measure normalized against the variance (σ^2) of the natural disturbance (ε_i). Fig. 3 shows the contour plot for the control efficiency over the control space.

As can be seen, the optimal I-II control setting ($w_I \cong 0.115$ and $w_{II} \cong 0.005$) falls outside the control region (shaded area) of the original PCC controller. That is, if we use the PCC controller and restrict the values of w_1 and w_2 in the interval of $(0, 1)$, the PCC controller would never be as effective as an optimal I-II controller (or d-EWMA controller).

III. AGE-BASED DOUBLE EWMA CONTROLLER

In many applications, the data sampling time is not equally spaced. This leads to an invalid result in the d-EWMA formula. However, the data collected often comes with an indication of the process age at the time of sampling. In practice, the process age is usually defined as the actual usage time of the process since its most recent preventive maintenance (PM). In this section, we will develop an age-based d-EWMA formula where the unequally spaced data can be accommodated along with the age data.

First, let t_i denote the process age at the i th run. In (12), the two EWMA equations need to be modified to take age into consideration. The first EWMA equation in (12) should be modified to

$$a_i = w_1(Y_i - \beta X_i) + (1 - w_1)[a_{i-1} + (t_i - t_{i-1})p_{i-1}] \quad (18)$$

where p_{i-1} is an estimate for the drift size per unit time at the $(i - 1)$ th run. From run $i - 1$ to run i , the process has grown $(t_i - t_{i-1})$ older. Since the process continues to drift away at a rate of p_{i-1} between the two runs, by the i th run the process has further deviated for an amount of $(t_i - t_{i-1})p_{i-1}$. This explains the second term of the EWMA formula in (18). We also need to

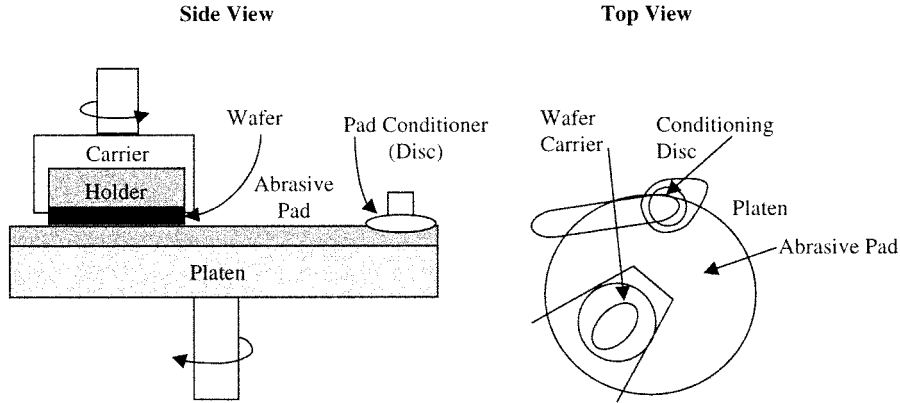


Fig. 4. A typical CMP process.

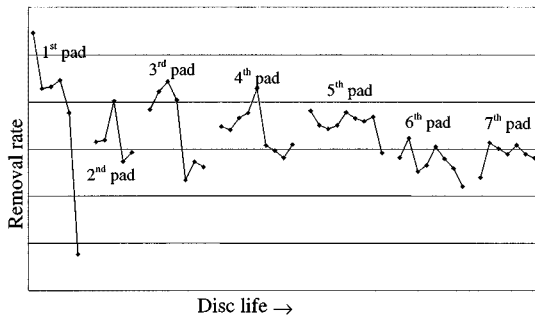


Fig. 5. Trend of removal rate over a disc lifetime.

modify the second EWMA formula for estimating the drifting speed at the i th run (p_i)

$$p_i = w_2 \left(\frac{Y_i - \beta X_i - a_{i-1}}{t_{i-1} - t_i} \right) + (1 - w_2)p_{i-1} \quad (19)$$

where $Y_i - \beta X_i - a_{i-1}$ represents the amount the process drifts between run $i - 1$ and run i . The recipe at the $(i + 1)$ th run should be thus set at

$$X_{i+1} = \frac{T - [a_i + (t_{i+1} - t_i)p_i]}{\beta} \quad (20)$$

to keep the process output on target.

It should be noted that such a modification to accommodate the process age is only possible with the d-EWMA formula. For the original PCC formula, the modification would require more care because of its different definitions of a_i and p_i .

IV. APPLICATION TO CMP PROCESS WITH AGING PAD AND DISC

Chemical-mechanical polishing (CMP), as a newly developed planarization technique, is demonstrated to be the most effective global planarization technique and therefore is considered to be a strategically important technology for multi-level device. A typical CMP process consists of a wafer carrier, a rotating platen with a replaceable abrasive pad mounted on the surface, and a pad conditioner (Fig. 4). The wafer is held face down by the carrier, which presses the wafer against the polishing pad. The pad conditioner (or conditioning disc) rotates and sweeps across the pad to roughen the pad surface as it

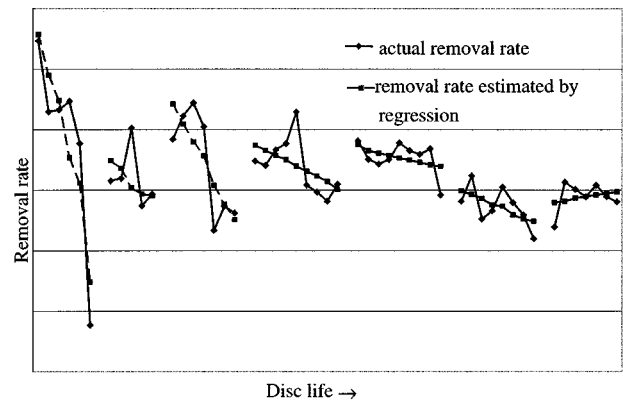


Fig. 6. Removal rate estimated by linear regression.

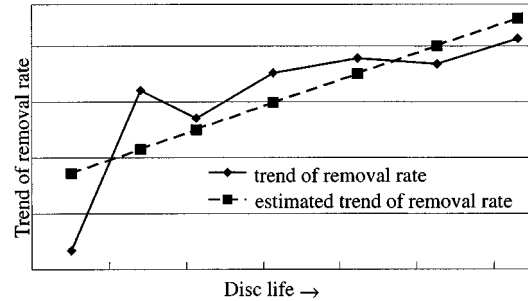


Fig. 7. Estimating decreasing rate of removal rate by linear regression.

smoothes down from repeated polishing. The carrier and platen rotate at variable speeds, typically on the order of 30 rpm. Tools differ in the number of wafers that may be simultaneously polished; single-wafer, dual-wafer, and five-headed tools exist.

The control of CMP processes, however, is known to be difficult because of poor understanding of the process, degradation (wear-out) of polishing pads, inconsistency of the conditioner, and the lack of *in situ* sensors. Because the process includes mechanical abrasion of the surface, the polishing pad wears rapidly. Concurrent or sequential "conditioning" is sometimes employed to restore the abrasive surface of the pad, but the lifetimes of the pad and the conditioning disc remain quite limited. To tackle some of these problems, a widely used approach is to use a number of send-ahead or dummy wafers to recalibrate the tool before or after each lot of wafers. Currently, most CMP

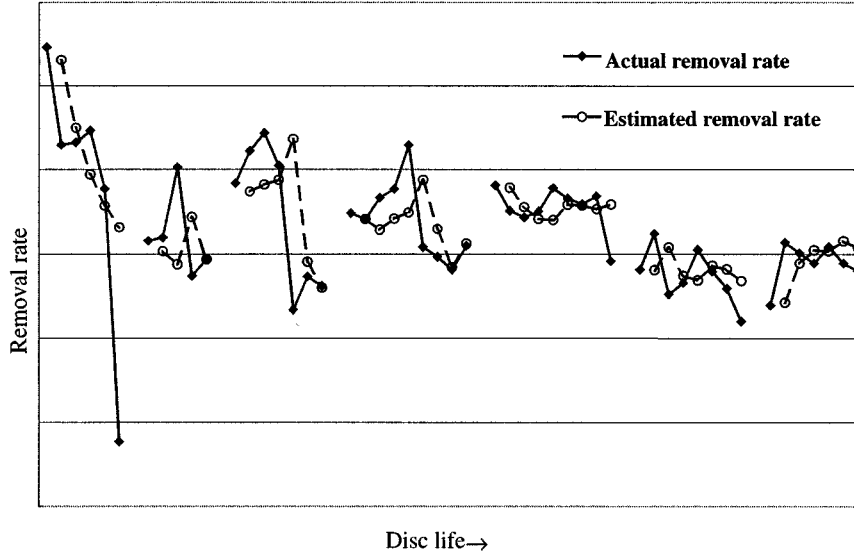


Fig. 8. Estimated removal rate model versus actual removal rate.

process control is based on postprocess measurements on test wafers. It is the goal of this research to employ a process control strategy to improve the performance of CMP processes.

A simple strategy for controlling the CMP process is to predict the run-to-run process removal rate and then adjust the polishing time based on the prediction [8], [11]. EWMA and PCC (or double EWMA) techniques are the two most often used prediction techniques. Based on the age-based d-EWMA controller presented earlier, we will design an age-based run-to-run prediction technique that takes into account the ages of the abrasive pad and conditioning disc. We will show how this proposed technique could improve the prediction capability and, thus, the control efficiency through actual CMP production data.

In a CMP process, both the abrasive pad and the conditioning disc are wearing out quickly. Because of the combination of chemical and mechanical processes during polishing, the wear-out process becomes quite irregular. Simple EWMA prediction of the removal rate is not sufficient. The more sophisticated d-EWMA may be needed to capture the changes of the removal rate. The d-EWMA prediction of removal rate at $(i + 1)$ th observation can be expressed as

$$a_i = w_1 R_i + (1 - w_1) \hat{R}_i \quad (21)$$

$$p_i = w_2 (R_i - a_{i-1}) + (1 - w_2) p_{i-1} \quad (22)$$

$$\hat{R}_{i+1} = a_i + p_i \quad (23)$$

where R_i is the actual removal rate observed at the i th sample and \hat{R}_i is the predicted removal rate of the i th sample. These are d-EWMA prediction equations corresponding to (12) without considering the process age. Similar to the d-EWMA controller, the first EWMA equation (21) is to estimate the level of the removal rate (a_i) and the second EWMA equation (22) is to capture its changing speed (p_i). In (23), the removal rate at observation $i + 1$ is then predicted by adding together the level estimate a_i and the anticipated change (p_i) from observation i to $i + 1$.

In practice, the age of the abrasive pad and the brushing disc can be acquired along with the removal rate data. Fig. 5 shows the trend of the removal rate over an entire lifetime of one disc. During the lifetime of a disc, PM was performed seven times. Each time the abrasive pad was replaced with a new one.

It is observed that the removal rate has the tendency to decrease as the pad becomes older and wears out over time. However, the removal rate's decreasing trends seem to be different for pads operated at different points of the disc lifetime. For example, the first pad in the beginning of the disc life has the removal rate decreasing drastically but this decreasing trend becomes almost indiscernible for the pads near the end of the disc life. This observation helps us understand how the ages of pad and disc affect the removal rate. The removal rate is basically decreasing as the pad gets older. But the removal rate's decreasing trend for each pad is again affected by the disc age. The older the disc, the less the removal rate's decreasing trend. Based on this observation, we can design an age-based d-EWMA prediction scheme, similar to (18)–(20), which takes into consideration the age of the pad and the age of the disc.

First, the removal rate for each abrasive pad is estimated by simple linear regression as shown in Fig. 6.

It can be observed from Fig. 6 that the decreasing trend of removal rate indeed becomes less significant as the disc grows older. This characteristic of removal rate trend is modeled in a linear regression model as shown in Fig. 7.

The model for the changing rate of removal rate, p , is expressed as

$$p(\tau) = d + b\tau \quad (24)$$

where

τ is the age of disc,

d is the removal rate trend for a pad operated under a brand-new disc ($\tau = 0$), and

b represents how the removal rate trend changes as the disc ages.

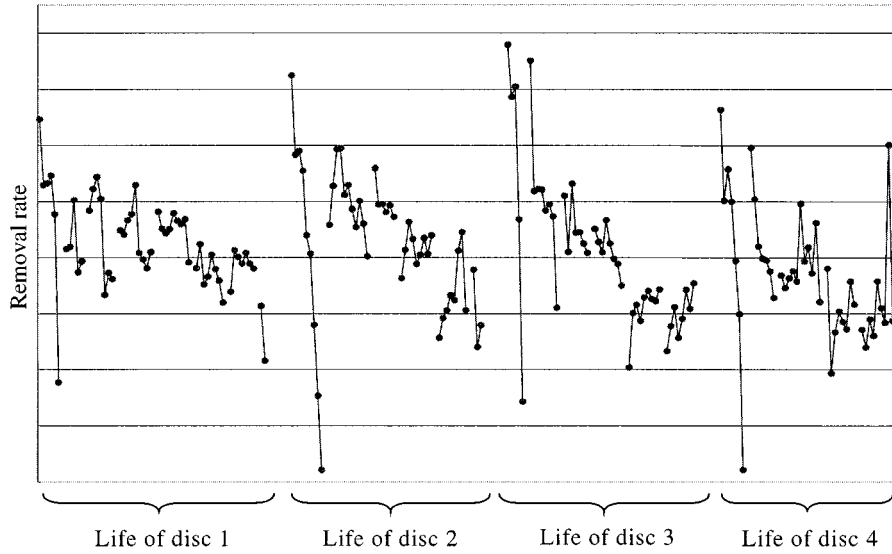


Fig. 9. Trend of removal rate over four discs.

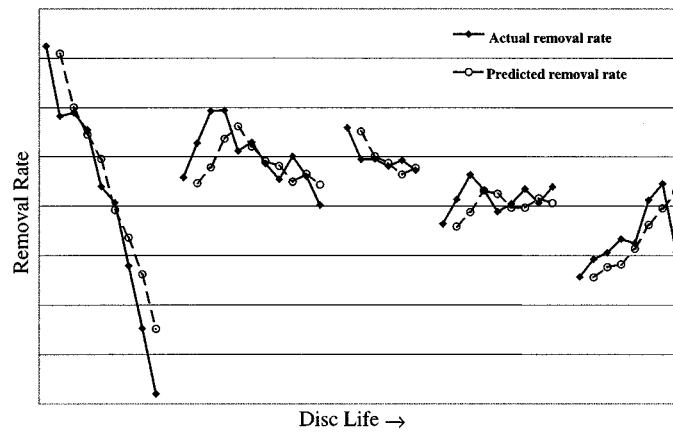


Fig. 10. Removal rate prediction for disc 2.

d and b are both estimated with statistical significance (can be also seen from Figs. 6 and 7). Based on (24), each new pad's initial trend of the removal rate, p_0 , is then estimated by

$$p_0(\tau_j) = d + b\tau_j$$

where τ_j is the disc age at which the j th pad is first used.

Given the initial changing trend of the removal rate for each new pad, the formula in (21)–(23) is then revised to accommodate the age (t) of the pad. Let t_i denote the age of the pad at the i th observation. We propose the following age-based d-EWMA prediction scheme

$$a_i = w_1 R_i + (1 - w_1) \hat{R}_i \quad (25)$$

$$p_i = w_2 \left(\frac{R_i - a_{i-1}}{t_i - t_{i-1}} \right) + (1 - w_2) p_{i-1} \quad (26)$$

$$\hat{R}_{i+1} = a_i + (t_{i+1} - t_i) p_i. \quad (27)$$

The data shown in Fig. 5 is then used to estimate the parameters, w_1 and w_2 . Fig. 8 shows this estimated model.

TABLE I
COMPARISON AMONG EWMA, PCC AND TIME-BASED PCC

	Prediction MSE	Improvement (%) over EWMA
EWMA	191.683	-
d-EWMA	178.484	6.88%
Age-based d-EWMA	158.163	17.48%

The model estimated using data in Fig. 5 can be now used to predict the removal rate for other discs. Fig. 9 shows the raw removal rate data for four discs and the discontinuities within each disc represent pad changes.

Fig. 10 shows the prediction for disc 2 using the model estimated from disc 1.

We summarize the performance of EWMA, d-EWMA and age-based d-EWMA prediction schemes by comparing their predicted mean squared error (MSE) in Table I.

It is recommended that when the age data for the abrasive pad and the conditioning disc is available, the proposed age-based d-EWMA scheme should be used to predict the removal rate. The prediction improvement is about 17% better than the EWMA scheme and 11% better than the d-EWMA scheme.

V. CONCLUSION

In this paper, we have presented an adjustment of the original PCC controller. This adjustment has a cleaner form and is shown equivalent to the form of I-II controller. The adjusted PCC controller enables us to accommodate the process age into the formula. An age-based d-EWMA controller is then developed. The performance of this age-based d-EWMA controller is also illustrated through the example of CMP removal rate prediction. The results show that the proposed controller improves the control efficiency significantly.

APPENDIX A

PCC controller:

$$\begin{aligned} a_i &= w_1(Y_i - \beta X_i) + (1 - w_1)a_{i-1} \\ &= w_1(\alpha + c_i\sigma + \varepsilon_i) + (1 - w_1)a_{i-1} \\ p_i &= w_2(Y_i - \beta X_i - a_{i-1}) + (1 - w_2)p_{i-1} \\ &= w_2(\alpha + c_i\sigma + \varepsilon_i - a_{i-1}) + (1 - w_2)p_{i-1}. \end{aligned}$$

Using a matrix expression, we obtain

$$\begin{aligned} \begin{bmatrix} a_i \\ p_i \end{bmatrix} &= \begin{bmatrix} 1 - w_1 & 0 \\ -w_2 & 1 - w_2 \end{bmatrix} \begin{bmatrix} a_{i-1} \\ p_{i-1} \end{bmatrix} \\ &+ \begin{bmatrix} w_1\alpha + w_1\varepsilon_i + w_1c_i\sigma \\ w_2\alpha + w_2\varepsilon_i + w_2c_i\sigma \end{bmatrix}. \end{aligned}$$

Let

$$\begin{aligned} U(i) &= \begin{bmatrix} a_i \\ p_i \end{bmatrix}, \quad A = \begin{bmatrix} 1 - w_1 & 0 \\ -w_2 & 1 - w_2 \end{bmatrix} \\ B(i) &= \begin{bmatrix} w_1\alpha + w_1\varepsilon_i + w_1c_i\sigma \\ w_2\alpha + w_2\varepsilon_i + w_2c_i\sigma \end{bmatrix}. \end{aligned}$$

Then,

$$\begin{aligned} U(0) &= \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \\ A &= \begin{bmatrix} 1 & 0 \\ \frac{-w_2}{w_2 - w_1} & 1 \end{bmatrix} \begin{bmatrix} 1 - w_1 & 0 \\ 0 & 1 - w_2 \end{bmatrix} \begin{bmatrix} \frac{1}{w_2 - w_1} & 0 \\ \frac{w_2}{w_2 - w_1} & 1 \end{bmatrix}. \end{aligned}$$

We obtain

$$\begin{aligned} A^i &= \begin{bmatrix} 1 & 0 \\ \frac{-w_2}{w_2 - w_1} & 1 \end{bmatrix} \begin{bmatrix} (1 - w_1)^i & 0 \\ 0 & (1 - w_2)^i \end{bmatrix} \\ &\cdot \begin{bmatrix} 1 & 0 \\ \frac{w_2}{w_2 - w_1} & 1 \end{bmatrix} \\ &= \begin{bmatrix} (1 - w_1)^i & 0 \\ \frac{w_2}{w_2 - w_1} ((1 - w_2)^i - (1 - w_1)^i) & (1 - w_2)^i \end{bmatrix} \end{aligned}$$

and, thus as shown at the bottom of the page. Finally,

$$\begin{aligned} a_i &= \alpha(1 - w_1)^i + \sum_{j=1}^i (1 - w_1)^{i-j} (w_1\alpha + w_1\varepsilon_j + w_1c_j\sigma) \\ p_i &= \frac{\alpha w_2}{w_2 - w_1} ((1 - w_2)^i - (1 - w_1)^i) + \sum_{j=1}^i \\ &\cdot \left\{ \frac{w_2}{w_2 - w_1} ((1 - w_2)^{i-j} - (1 - w_1)^{i-j}) \right. \\ &\cdot (w_1\alpha + w_1\varepsilon_j + w_1c_j\sigma) \\ &\left. + (1 - w_2)^{i-j} (w_2\alpha + w_2\varepsilon_j + w_2c_j\sigma) \right\}. \end{aligned}$$

As i approaches infinity:

$$E(a_i) = \alpha - \frac{c\sigma}{w_1} + c\sigma(i + 1) + o(i) \quad \text{where} \quad \lim_{i \rightarrow \infty} o(i) = 0$$

$$\left(\begin{aligned} &\because E(\varepsilon_i) = 0 \text{ and as } i \rightarrow \infty, \\ &\sum_{j=1}^i j(1 - w_1)^{i-j} = \frac{i+1}{w_1} - \frac{1}{w_1^2} + o(i) \end{aligned} \right).$$

$$\begin{aligned} U(i) &= AU(i-1) + B(i) = A[AU(i-2) + B(i-1)] + B(i) = A^i U(0) + \sum_{j=1}^i A^{i-j} B(j) \\ &= \begin{bmatrix} (1 - w_1)^i & 0 \\ \frac{w_2}{w_2 - w_1} ((1 - w_2)^i - (1 - w_1)^i) & (1 - w_2)^i \end{bmatrix} \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \\ &+ \sum_{j=1}^i \begin{bmatrix} (1 - w_1)^{i-j} & 0 \\ \frac{w_2}{w_2 - w_1} ((1 - w_2)^{i-j} - (1 - w_1)^{i-j}) & (1 - w_2)^{i-j} \end{bmatrix} \begin{bmatrix} w_1\alpha + w_1\varepsilon_j + w_1c_j\sigma \\ w_2\alpha + w_2\varepsilon_j + w_2c_j\sigma \end{bmatrix} \end{aligned}$$

Similarly,

$$\lim_{i \rightarrow \infty} E(p_i) = \frac{c\sigma}{w_1}$$

APPENDIX B

$$\begin{aligned} a_i &= w_1(e_i + p_{i-1}) + a_{i-1} \\ &= w_1 \sum_{j=1}^i e_j + w_1 \sum_{j=1}^{i-1} p_j + a_0 + w_1 p_0 \\ p_i &= w_2 e_i + p_{i-1} \\ &= w_2 \sum_{j=1}^i e_j + p_0 \\ a_i + p_i &= (w_1 + w_2) \sum_{j=1}^i e_j + w_1 \sum_{j=1}^{i-1} p_j \\ &\quad + (w_1 + 1)p_0 + a_0 \\ &= (w_1 + w_2) \sum_{j=1}^i e_j + w_1 \sum_{j=1}^{i-1} \left(w_2 \sum_{k=1}^j e_k + p_0 \right) \\ &\quad + a_0 + (w_1 + 1)p_0 \\ &= (w_1 + w_2) \sum_{j=1}^i e_j + w_1 w_2 \sum_{j=1}^{i-1} \sum_{k=1}^j e_k \\ &\quad + (w_1 i + 1)p_0 + a_0 \\ &= (w_1 + w_2 - w_1 w_2) \sum_{j=1}^i e_j + w_1 w_2 \sum_{j=1}^i \sum_{k=1}^j e_k \\ &\quad + (w_1 i + 1)p_0 + a_0. \end{aligned}$$

APPENDIX C

$$\begin{aligned} a_i &= w_1 e_i + p_{i-1} + a_{i-1} \\ &= w_1 \sum_{j=1}^i e_j + \sum_{j=1}^{i-1} p_j + a_0 + p_0 \\ p_i &= w_2 e_i + p_{i-1} \\ &= w_2 \sum_{j=1}^i e_j + p_0 \\ a_i + p_i &= (w_1 + w_2) \sum_{j=1}^i e_j + \sum_{j=1}^{i-1} p_j + 2p_0 + a_0 \\ &= (w_1 + w_2) \sum_{j=1}^i e_j + \sum_{j=1}^{i-1} \left(w_2 \sum_{k=1}^j e_k + p_0 \right) \\ &\quad + a_0 + 2p_0 \\ &= (w_1 + w_2) \sum_{j=1}^i e_j + w_2 \sum_{j=1}^{i-1} \sum_{k=1}^j e_k \\ &\quad + (i + 1)p_0 + a_0 \end{aligned}$$

$$\begin{aligned} &= w_1 \sum_{j=1}^i e_j + w_2 \sum_{j=1}^i \sum_{k=1}^j e_k \\ &\quad + (i + 1)p_0 + a_0. \end{aligned}$$

APPENDIX D

Double EWMA controller can be rewritten as

$$\begin{aligned} a_i + p_i &= (1 - w_1 - w_2)(a_{i-1} + p_{i-1}) \\ &\quad + p_{i-1} + (w_1 + w_2)(\alpha + c\sigma i + \varepsilon_i) \\ p_i &= w_2(Y_i - bX_i - a_{i-1}) + (1 - w_2)p_{i-1} \\ &= (-w_2)(a_{i-1} + p_{i-1}) + p_{i-1} + w_2(\alpha + c\sigma i + \varepsilon_i). \end{aligned}$$

Let $Z_i = a_i + p_i$ and $w = w_1 + w_2$. Then,

$$Z_i = (1 - w)Z_{i-1} + p_{i-1} + w(\alpha + c\sigma i + \varepsilon_i) \quad (D1)$$

$$p_i = (-w_2)Z_{i-1} + p_{i-1} + w_2(\alpha + c\sigma i + \varepsilon_i). \quad (D2)$$

Take the expected values of (D2)

$$E(p_i) = (-w_2)E(Z_{i-1}) + E(p_{i-1}) + w_2(\alpha + c\sigma i).$$

As $i \rightarrow \infty$, $E(p_{i-1}) \rightarrow E(p_i)$ and thus $E(Z_{i-1}) \rightarrow \alpha + c\sigma i$. Let $E(Z_i) = \alpha + c\sigma(i + 1) + o(i)$ where $\lim_{i \rightarrow \infty} o(i) = 0$. Now, take the expected values of (D1)

$$E(Z_i) = (1 - w)E(Z_{i-1}) + E(p_{i-1}) + w(\alpha + c\sigma i). \quad (D3)$$

As $i \rightarrow \infty$, substituting $\alpha + c\sigma(i + 1) + o(i)$ and $\alpha + c\sigma i + o(i)$ into $E(Z_i)$ and $E(Z_{i-1})$, we write

$$\begin{aligned} \alpha + c\sigma(i + 1) + o(i) \\ = (1 - w)[\alpha + c\sigma i + o(i)] + E(p_{i-1}) + w(\alpha + c\sigma i) \end{aligned}$$

and therefore $\lim_{i \rightarrow \infty} E(p_{i-1}) = c\sigma$ and $E(a_i) = \alpha + c\sigma i + o(i)$ where $\lim_{i \rightarrow \infty} o(i) = 0$.

ACKNOWLEDGMENT

The authors would like to thank J.-J. Chen, C.-P. Tung, Y.-L. Chou, and C.-L. Lin for their help in preparing the figures. They would also like to thank Dr. J. Dun and S.-A. Wu of TSMC who provided the valuable CMP data for this study.

REFERENCES

- [1] G. Box and M. Jenkins, *Time Series Analysis—Forecasting and Control*. Oakland, CA: Holden-Day, 1974.
- [2] G. Box and A. Luceno, *Statistical Control—By Monitoring and Feedback Adjustment*. New York: Wiley, 1997.
- [3] K. Ogata, *Modern Control Engineering*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1990.
- [4] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [5] E. Sachs, A. Hu, and A. Ingolfsson, "Run by run process control; Combining SPC and feedback control," *IEEE Trans. Semiconductor Manufact.*, vol. 8, pp. 26–43, Feb. 1995.
- [6] S. Bulter and J. Stefani, "Supervisory run-to-run control of polysilicon gate etch using *in situ* ellipsometry," *IEEE Trans. Semiconductor Manufact.*, vol. 7, pp. 193–201, May 1994.
- [7] T. Smith, D. Boning, J. Stefani, and S. Butler, "Run by run advanced process control of metal sputter deposition," *IEEE Trans. Semiconductor Manufact.*, vol. 11, pp. 276–284, May 1998.

- [8] D. Boning, W. Moyné, T. Smith, J. Moyné, R. Telfeyan, A. Hurwitz, S. Shellman, and J. Taylor, "Run by run control of chemical-mechanical polishing," *IEEE Trans. Components, Packaging, Manufact. Technol. C*, vol. 19, pp. 307–314, Oct. 1996.
- [9] T. Smith and D. Boning, "A self-tuning EWMA controller utilizing artificial neural network function approximation techniques," *IEEE Trans. Components, Packaging, and Manufact. Technol. C*, vol. 20, pp. 121–132, Mar. 1997.
- [10] E. Del Castillo and A. Hurwitz, "Run-to-run process control: Literature review and extensions," *J. Quality Technol.*, vol. 29, pp. 184–196, Apr. 1997.
- [11] H. W. Chiou and L. J. Chen, "PID run to run control of CMP removal rate," in *CMP-MIC Conf.*, 1997, pp. 375–382.



Argon Chen received his B.S. degree from National Taiwan University in 1986. In 1992 and 1994, he earned M.S. degrees in industrial engineering and in statistics, respectively, from Rutgers University. He received his Ph.D. degree in industrial engineering in 1995 also from Rutgers University.

From 1995 to 1996, Dr. Chen worked for AT&T Bell Labs, Middletown, New Jersey. As a Consultant for the division of globalization, provisioning, technology planning and systems, he worked in the areas of system analysis, system modeling and design, and

system integration. In 1996, Dr. Chen joined National Taiwan University (NTU) as a Faculty Member of the Graduate Institute of Industrial Engineering. He is now an Associate Professor at NTU, where he is also a Principle Investigator of several research projects funded by the government or industries in Taiwan. His research interests include industrial statistical inference, quality engineering, computer integrated manufacturing, supply chain management, and information technology applications.

Dr. Chen currently serves on the Editorial Board of IIE Transactions on Quality and Reliability Engineering.



Ruey-Shan Guo (S'90–M'91) received the B.S. degree from National Taiwan University in 1983 and M.S. degree from MIT in 1987. He received his Ph.D. degree in mechanical engineering with a major in manufacturing and a minor in solid state physics in 1991, also from MIT. He obtained his MBA degree from San Jose State University in 1994.

From 1991 to 1995, he was a Senior Process Engineer at National Semiconductor Fairchild Research Center, Santa Clara, CA, where he worked in the area of factory layout design, contamination control, statistical quality control and computer-aided manufacturing.

From 1995 till now, he has been an Associate Professor in the Graduate Institute of Business Administration and Graduate Institute of Industrial Engineering of National Taiwan University. During his current position, he has many projects with Taiwan IC manufacturers, mostly in the areas of CIM, quality control, CMP run-to-run process control, supply chain management, and foundry business model. He teaches undergraduate and graduate courses in operations management, advanced quality control, enterprise resource planning, and supply chain management. He won the best teaching award in National Taiwan University in May 2000.

Dr. Guo has served in the program committee and as session chair of ISSM from 1997–2000.