Run-to-Run Control of CMP Process Considering Aging Effects of Pad and Disc

Argon Chen, Ruey-Shan. Guo, Y.L. Chou, C.L. Lin
Graduate Institute of Industrial Engineering
National Taiwan University
1, Roosevelt Rd. Sec. 1, Taipei, Taiwan 106
Fax: 886-2-23635856 Email: achen@ccms.ntu.edu.tw

Jowei Dun and S. A. Wu
Taiwan Semiconductor Manufacturing Co. (TSMC)
9, Creation Rd. I, Science-Based Industrial Park
Hsin-Chu, Taiwan, R.O.C.
Fax: 886-3-5790298 Email: sawu@tsmc.com.tw

Abstract - CMP processes are known to be erratic and unstable. A simple control strategy is to predict the run-to-run process removal rate and then adjust the processing time based on the prediction. EWMA and PCC techniques are two most often used prediction techniques. In this work, we revise the PCC design to take into account the ages of the abrasive pad and conditioning disc. It is shown that the proposed age-based technique can significantly improve the prediction capability and, thus, the control efficiency over conventional techniques.

INTRODUCTION

CMP Processes

Chemical-mechanical polishing (CMP), as a newly developed planarization technique, is demonstrated to be the only global planarization technique available and therefore is considered to be a strategically important technology for next-generation 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 (Figure 1). 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 smoothes down from repeated polishing. The carrier and platen rotate at variable speeds, typically on the order of 30 rpm. Tools differ in a number of wafers that may be simultaneously polished; single-wafer, dual-wafer, and five-headed tool exist.

The control of CMP processes, however, is known to be quite 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 usually employed to restore the abrasive surface of the pad, but the lifetimes of the pad and the conditioning disc remain quite limited.

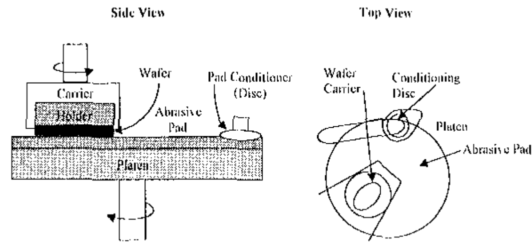

Fig. 1 CMP Process

As a result of these problems, a widely used approach is to use a number of send-ahead or dummy wafers to condition and/or calibrate the tool before or after each lot of wafers. It is then the goal of this research to employ a process control strategy to improve the performance of CMP processes. Currently, CMP process control is almost exclusively based on post-process measurements on test wafers. The availability of these post-process measurements and the typical process drifts due to components' wear-out motivate the use of Run-to-Run (RtR) control strategy.

Run-to-run (RtR) Controllers

The use of Exponentially Weighted Moving Average (EWMA) statistic for estimating process deviations has been widely studied and adopted in practice. The EWMA statistic has been shown to be quite effective even for processes subject to various kinds of process disturbances. In particular, applications in the semiconductor process industry have shown that the EWMA statistic is also capable of bringing processes with linear drift under control [1]. Suppose that the process output Y_t can be controlled linearly by an input variable X_t and is subject to a natural process disturbance ε_t ($\sim N(0, \sigma^2)$) and a systematic disturbance δ_t. The process model can be written as:

$$\begin{aligned} Y_t &= \alpha + \beta X_t + \varepsilon_t + \delta_t \\ &= (\alpha + \delta_t) + \beta X_t + \varepsilon_t \end{aligned}$$

where α is the linear process model's intercept term and β is the system gain that translates the input variable's size to

0-7803-5403-6/99/$10.00 ©1999 IEEE.

the process output's responded size. The EWMA statistic is then used to estimate the size of the process's intercept (α) plus its systematic deviation at time $t+1$ (δ_{t+1}):

$$a_t = w(Y_t - \beta X_t) + (1-w)a_{t-1}$$
$$= \sum_{j=1}^{t}[w(1-w)^{t-j}(Y_j - \beta X_j)]$$

where the weight w is usually set between 0 and 1 and a_t is an exponentially weighted average of the historic deviations $\alpha+\delta_j$'s. Therefore, the process output at time $t+1$ is estimated to be:

$$\hat{Y}_{t+1} = a_t + \beta X_{t+1}.$$

In order to keep the process output at a pre-determined target level (T), we obtain the process recipe at time $t+1$:

$$X_{t+1} = \frac{T - a_t}{\beta}.$$

Such a controller is insufficient for processes subject to systematic wear-out. Several authors [2][3] have addressed this problem and propose using two EWMA formulas: one for estimating "step-change" deviation and the other for estimating the process "drift" speed.

$$\begin{aligned} a_t &= w_1(Y_t - \beta X_t) + (1-w_1)a_{t-1} \\ p_t &= w_2(Y_t - \beta X_t - a_{t-1}) + (1-w_2)p_{t-1} \end{aligned} \quad (1)$$

where w_1 and w_2 are the weights for the first and second EWMA equations respectively and p_t is to estimate the size of the process drift from t to $t+1$. Thus, the process recipe at $t+1$ should be:

$$X_{t+1} = \frac{T - (a_t + p_t)}{\beta}.$$

Such a control scheme is referred to as Predictor Corrector Control (PCC) scheme, termed by Butler and Stefani [2].

In this research, we first propose an adjustment to the PCC control scheme. The revised PCC control formula, referred to as Double-EWMA (D-EWMA) scheme, is then further refined to take the process age into consideration. The methodology is finally applied to the CMP process.

AGE-BASED RtR CONTROLLER

Revised PCC Controller – D-EWMA Controller

We first propose an adjustment of the PCC formula such that the process age can be taken into consideration. We refer to this adjusted PCC formula as Double EWMA (D-EWMA) formula as distinct from the original PCC formula. This D-EWMA formula is:

$$\begin{aligned} a_t &= w_1(Y_t - \beta X_t) + (1-w_1)(a_{t-1} + p_{t-1}) \\ p_t &= w_2(Y_t - \beta X_t - a_{t-1}) + (1-w_2)p_{t-1} \end{aligned} \quad (2)$$

As can be observed in (2), the only difference from the original PCC formula is that in the first EWMA formula for a_t, we add p_{t-1} into the formula. As a result, we can obtain a better asymptotic behavior of the controller:

$$\lim_{t\to\infty} E(a_t) = \alpha + c\sigma t$$
$$\lim_{t\to\infty} E(p_t) = c\sigma$$

The two estimates a_t and p_t here have very clearly defined meanings. a_t in this D-EWMA formula represents an estimate for $\alpha+\delta_t$ and p_t is an estimate for the size of the process drift from t to $t+1$. Together a_t+p_t is the estimate for $\alpha+\delta_{t+1}$. It can be shown that the process output with D-EWMA control will converge to the desired target T.

Age-based D-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 the process's age at the time of sampling. In this section, we will develop an age-based D-EWMA formula in that the unequally spaced data can be accommodated along with the age data.

First, let t_i denote the process age at the ith sampling. In Equation (2), the two EWMA equations need to be modified to take the age into consideration. a_i now becomes the estimate for the intercept (α) plus the systematic deviation at the ith sample (δ_i). The first EWMA equation in (2) should be modified to:

$$a_i = w_1(Y_i - \beta X_i) + (1-w_1)[a_{i-1} + (t_i - t_{i-1})p_{i-1}] \quad (3)$$

where p_{i-1} is an estimate for the drifting size per unit time at the (i-1)th sample. From sample i-1 to sample 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 samples, by the ith sample 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 (3). We also need to modify the second EWMA formula for estimating the drifting speed at the ith sample (p_i):

$$p_i = w_2\left(\frac{Y_i - \beta X_i - a_{i-1}}{t_i - t_{i-1}}\right) + (1-w_2)p_{i-1} \quad (4)$$

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

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

to keep the process output on target.

It should be noted that such a modification to accommodate the process age is only possible for the D-EWMA formula. For the original PCC formula, the modification becomes almost impossible because of its unclear definitions of a_t and p_t.

AGE-BASED CMP D-EWMA CONTROLLER

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 [4][5]. 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. D-EWMA is therefore needed to capture the changes of the removal rate. The D-EWMA prediction of the removal rate at ith observation ($\hat{R}_i$) can be expressed as:

$$a_i = w_1 R_{i-1} + (1 - w_1)\hat{R}_{i-1} \tag{5}$$

$$p_i = w_2(R_{i-1} - a_{i-1}) + (1 - w_2)p_{i-1} \tag{6}$$

$$\hat{R}_i = a_i + p_i \tag{7}$$

These are D-EWMA prediction equations corresponding to Equation (2) without considering the process age. Similar to the D-EWMA controller, the first EWMA equation (5) is to estimate the level of the removal rate (a_i) and the second EWMA equation (6) is to capture its changing speed (p_i). In equation (7), 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 conditioning disc can be acquired along with the removal rate data. Figure 2 shows the trend of the removal rate over an entire lifetime of one disc. During the lifetime of a disc, preventive maintenance (PM) was performed seven times. Each time the abrasive pad was replaced with a new one.

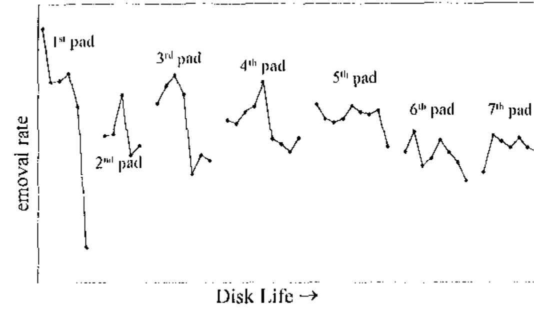

Fig. 2 Trend of removal rate over a disc's lifetime

It can be 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 in the disc's lifetime. For example, the 1st pad in the beginning of the disc's life has the removal rate decreasing drastically but this decreasing trend becomes almost indiscernible for the pads near the end of the disc's 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's age upon which the pad is operated. 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 Equations (3)-(4), which takes into consideration the age of the pad and the age of the disc.

First, the initial changing speed of the removal rate (p_0) for each abrasive pad is estimated based on the age (τ) of the disc:

$$p_0(\tau) = c + b\tau\,, \tag{8}$$

where c is the changing speed of the removal rate for a brand-new disc (τ=0) and b represents how the changing speed changes as the disc ages.

Second, given the initial changing speed of the removal rate for each new pad, the D-EWMA formula in equation (5)-(7) is then revised to accommodate the age (t) of the pad. Let t_i denote the age of the pad at the ith observation. We propose the following age-based D-EWMA prediction scheme:

$$a_i = w_1 R_{i-1} + (1 - w_1)\hat{R}_{i-1} \tag{9}$$

$$p_i = w_2\left(\frac{R_{i-1} - a_{i-1}}{t_{i-1} - t_{i-2}}\right) + (1 - w_2)p_{i-1} \tag{10}$$

$$\hat{R}_i = a_i + (t_i - t_{i-1})p_i \tag{11}$$

We can then model the data shown in Figure 2 to estimate the values of all the parameters (c, b, w_1, and w_2) in equations (8)-(11). Figure 3 shows this estimated model.

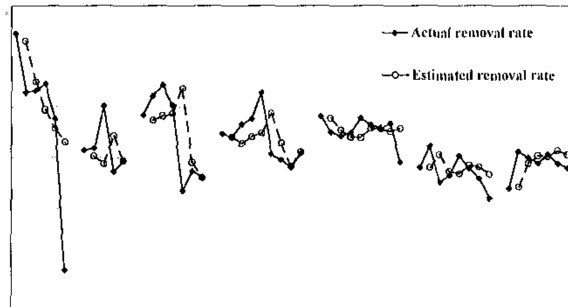

Fig. 3 Estimated removal rate vs. actual removal rate

The model estimated using data in Figure 2 can be now used to predict the removal rate for other discs.

Figure 4 shows the prediction for another disc using the model estimated above.

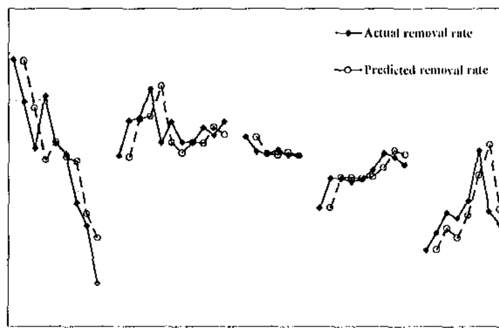

Fig. 4 Removal rate prediction

CONCLUSIONS

We summarize the performance of EWMA, D-EWMA and age-based D-EWMA schemes by comparing their mean squared error (MSE) of prediction in Table 1.

Table 1 Comparison among EWMA, D-EWMA and age-based D-EWMA

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

It is recommended that when the age data for the abrasive pad and the brushing 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.

REFERENCES

1. A. Ingolfsson and E. Sachs, "Stability and Sensitivity of an EWMA Controller," *Journal of Quality Technology*, vol. 25, no. 4, 1993.

2. S. Bulter and J. Stefani, "Supervisory Run-to-Run Control of Polysilicon Gate Etch Using *In Situ* Ellipsometry," IEEE Transactions Semiconductor Manufacturing, vol. 7, no. 2, May 1994.

3. T. Smith, D. Boning, J. Stefani, and S. Bulter, "Run by Run Advanced Control of Metal Sputter Deposition," *IEEE Trans. Semiconductor Manufacturing*, 11, pp. 276-284, 1998.

4. D. Boning, A. Hurwitz, J. Moyne, W. Moyne, S. Shellman, T. Smith, J. Taylor, and R. Telfeyan, "Run by Run Control of Chemical Mechanical Polishing," *IEEE Trans. Components, Packaging and Manufacturing Technology – Part C*, vol. 19, no. 1, 1996.

5. H. W. Chiou and L. J. Chen, "PID Run to Run Control of CMP Removal Rate," CMP-MIC Conference, pp. 375-382, 1997.