A time-varying weights tuning method of the double EWMA controller

Chao-Ton Su*, Chun-Chin Hsu

Department of Industrial Engineering and Management, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan

Received 2 July 2003; accepted 15 March 2004

Abstract

In recent years, the exponential-weighted-moving-average (EWMA) statistic based controllers are popular in semiconductor manufacturing. However, the single EWMA controller is not sufficient for compensating for the wear-out process. Thus, a double EWMA controller was proposed to enhance the capability for controlling the drifting process. In the literatures, in the solution of the double EWMA controller, only the “trade-off” solution weights are used to tune the controller. However, it is a fixed weight tuning method, and it is known that a time-varying weight will produce a superior performance over that of a fixed one (J. Quality Technol. 29 (1997) 184). Therefore, this study aims to develop a heuristic time-varying weights tuning method for the double EWMA controller. The numerical results showed that the proposed time-varying tuning method possesses an improvement of at least 10% over that of the fixed weight scheme.

Keywords: Double EWMA controller; Trade-off; All-bias; All-variance

1. Introduction

Variation reduction to enhance product quality in a process is a crucial determinant of competitiveness for the manufacturing industries. Statistical process control (SPC) and engineering process control (EPC) are two important techniques for process variation reduction. The main objective of SPC is the on-line “monitoring” of the process, and it has been successfully used in discrete parts manufacturing industries. The objective of the EPC technique is process “adjustment”, which means the use of feedback control algorithms to maintain the process as close as possible to the desired target value.

Recently, a control algorithm called the exponential-weighted-moving-average (EWMA) controller is in vogue in the semiconductor manufacturing industry, and particularly in the chemical mechanical polishing (CMP) process. In the literature, the use of the single EWMA controller had been shown to be effective for reducing the process variability. However, a new problem then began to emerge. Many processes were starting to exhibit drifts in the performance of the equipment. Such drifts were often caused by worn-out tools or other systematic causes of deterioration. When using the single EWMA controller to compensate for the drifting process, the controlled process output will exhibit considerable offsets. Butler and Stefani [1] proposed using two EWMA equations to control the drifting process: one equation for estimating the drift, the other for estimating the step change deviation. It can be shown that the double EWMA controller is a minimum mean square error (MMSE) controller when the disturbance model IMA(2, 2) exists in the process [2].

Until recently, the only solution of the double EWMA controller was to use the fixed weights solution to tune the controller, until Del Castillo [2] proposed an optimization form for solving the problem. The objective of the optimization form was to compromise on the transient effect and the long run variance. He called the solution the “trade-off” solution weights. Although the trade-off solution is shown to be effective, it still is a fixed weight control scheme. In
the literature, many researches have shown that the EWMA controller with time-varying parameters possesses better performance than the fixed weight tuning method. However, available researches all focus on the single EWMA controller. Therefore, there is a great need for developing a time-varying weights tuning method to enhance the performance of the double EWMA controller.

In this study, an effective time-varying weights tuning approach is developed by adding the discount factor into the double EWMA parameters. The advantage of introducing the discount factor is to quickly compensate for the initial transient effect. A simulation study demonstrates that the proposed tuning approach significantly outperforms the fixed “trade-off” control scheme.

This study is organized as follows: Section 2 reviews the double EWMA controller. Section 3 presents the optimization form for solving the double EWMA controller problem, as proposed by Del Castillo [2]. In Section 4, the proposed time-varying tuning method is specified. Section 5 implements the proposed tuning method by using the software of Matlab/Simulink. In addition a comparison between the fixed trade-off weights control scheme and the proposed time-varying weights tuning method is made by running Monte Carlo simulations.

2. The double EWMA controller

The EWMA statistic, sometimes called a geometric moving average (GMA), was suggested by Roberts [3]. The use of the EWMA statistic has two distinct purposes [4]: as control charts [5–9] and as time series forecasts [10,11]. Recently, this statistic has been widely used for process adjustment purposes [12–20].

In semiconductor manufacturing, EWMA controllers are sometimes called bias tuning controller. The EWMA-based controllers are used for compensating against disturbances which affect the run-by-run variability in quality characteristics [17]. The use of the single EWMA controller is related to the pure-integral (I) control action, which is a part theme of the well-known PID controller. Box and Jenkins [10] showed that the single EWMA statistic is a minimum mean square error (MMSE) controller when the IMA(1,1) disturbance model exists in the process. The IMA(1,1) is an important nonstationary time series model in the study of process regulation and adjustment [7]. Other than the IMA(1,1), the single EWMA controller has been shown to perform effectively in some disturbance models such as the step and ramp with slow drift rate disturbance model. However, the single EWMA controller cannot compensate for a ramp disturbance with severe drift speed. That is, a considerable offset will be exhibited in the controlled process output.

Butler and Stefani [1] extended the single EWMA controller with another EWMA equation in order to compensate for the ramp disturbance model. In this section, we will briefly introduce the double EWMA controller that was proposed by Butler and Stefani [1]. Note that they do not call it a double EWMA controller, but refer to it as a predictor corrector control (PCC) scheme.

Consider a drifting process model as follows:

\[ y_t = \alpha + \beta x_{t-1} + \delta t + \epsilon_t, \]  

where \( y_t \) is the process output, \( \alpha \) is the intercept term, \( \beta \) is the system gain, \( x_t \) is the manipulate variable, \( \delta \) denotes the drifting speed, and \( \epsilon_t \) is the white noise term. A PCC control equation for \( x_t \) can be expressed as

\[ x_t = \frac{T - (a_t + D_t)}{b}, \]  

where \( T \) is the target value, \( b \) is the estimate of \( \beta \) which can be obtained off-line by using designed experiments in a pre-control phase. \( a_t \) and \( D_t \) can be expressed as follows:

\[ a_t = A_t(y_t - b x_{t-1}) + (1 - A_t)a_{t-1}, \quad 0 < A_t \leq 1, \]  

\[ D_t = A_t D_{t-1} + (1 - A_t)D_{t-1}, \quad 0 < A_t \leq 1, \]  

where \( A_t \) and \( A_t \) are the weights for the first and second EWMA equations. Note that if we set \( A_t = 0 \) or \( A_t = 0 \), then the double EWMA controller will reduce to a single-EWMA controller. From Eqs. (3) and (4), it is clear that the performance of a double EWMA controller depends on selecting both parameters of \( A_t \) and \( A_t \). To appropriately select both parameters, the stability conditions of controller parameters should be held to as follows [2]:

\[ 1 - \frac{1}{2} \xi(A_t + A_t) - \frac{1}{2} \sqrt{\xi^2 ((A_t + A_t)^2 - 4 A_t A_t)} \leq 1, \]  

\[ 1 - \frac{1}{2} \xi(A_t + A_t) + \frac{1}{2} \sqrt{\xi^2 ((A_t + A_t)^2 - 4 A_t A_t)} \leq 1, \]  

where \( \xi = \beta/b \) is the bias in the gain estimate.

As time approaches infinity, each \( a_t \) and \( D_t \) works as follows:

\[ \lim_{t \to \infty} E[a_t] \to \alpha + \delta(t + 1) - \delta/A_t, \]  

\[ \lim_{t \to \infty} E[D_t] \to \delta/A_t. \]  

We can see that \( a_t \) is an asymptotical estimate of the ramp disturbance with a bias term \( \delta/A_t \), and \( D_t \) is the asymptotical estimate of the bias term. Thus, \( a_t + D_t \) becomes an asymptotically unbiased one-step-ahead estimate of the ramp disturbance model. Another form of Eqs. (3) and (4) can be expressed as follows:

\[ a_t = A_t \left[ \frac{1 - (1 - A_t)B}{1 - 2B + B^2} \right] y_t, \]  

\[ D_t = A_t \left( \frac{1}{1 - B} \right) y_t. \]
where $B$ is the backshift operator ($By_t = y_{t-1}$). We can see that the $a_t$ term is a second order filter ($\lambda_2 \neq 0$) and can filter out the trend. The term, $D_t$, is the first order filter that can filter out the process offsets. Therefore, the PCC control equations can be expressed as follows:

$$a_t + D_t = \hat{\lambda}_1 \left[\frac{1 - (1 - \hat{\lambda}_2)B}{1 - 2B + B^2}\right] y_t + \hat{\lambda}_2 \left(\frac{1}{1 - B}\right)y_t$$

$$= w_1 \sum_{i=1}^{t} y_i + w_2 \sum_{j=1}^{t} y_j,$$  \hspace{1cm} (10)

where, $w_1 = \hat{\lambda}_1 + \hat{\lambda}_2 - \hat{\lambda}_1\hat{\lambda}_2$ and $w_2 = \hat{\lambda}_1\hat{\lambda}_2$.

From Eq. (10), we can see that the double EWMA controller is not a discrete PID form, but an integral-double-integral (I–II) form [21]. It can be shown that the I–II controller is an MMSE controller when the IMA(2,2) disturbance model affects the process. Compared to the non-stationary IMA(1,1) process (i.e. drift in any direction with equal probability), the IMA(2,2) noise model can be interpreted as a process that experiences random changes in the drifting speed in Eq. (1) [2]. Fig. 1 shows a block diagram of the double EWMA controller model when a ramp disturbance model exists in the process.

3. Trade-off tuning method

The control strategies of EWMA controllers can simply be divided into time-invariant and time-varying weights control schemes. The time-invariant control scheme means that the EWMA weights do not change with time, but that the weights are fixed to control the process. Del Castillo [15] presented a solution of balancing the adjustment and output variances to control the single-EWMA controller. He also designed a trade-off solution of transient and steady-state performance to control the double-EWMA controller (1999). The time-varying control scheme is sometimes called a self-tuning or adaptive control, because

![Double EWMA controller diagram](image)
Therefore, the optimization form can be modeled as follows:

\[
\min_{k_1,k_2} \ k_1 \text{VAR}(y_t) + k_2 \text{MSD}
\]

S.T. \quad 0 < k_1 \leq 1, \quad 0 < k_2 \leq 1, \quad (13)

where the parameters \((k_1,k_2)\) are determined by the engineers. For \((k_1,k_2) = (0,1)\), it is an all-bias solution. For \((k_1,k_2) = (1,0)\), it becomes an all-variance solution. If we set \(k_1 = k_2 = 1\), then a trade-off solution will be obtained. Del Castillo [2] suggested that keeping the trade-off solution weights to control the process would provide adequate performance than a variety case. For example: if we first use the all-bias solution to cancel out the transient effect, and then abruptly change it to the all-variance solution at a specific time, then the controlled process will incur a new transient at that specific time. Fig. 2 shows the above condition in which, we simulate the process with a drift rate of 0.5. By solving Eq. (13), we obtain the all-bias solution (0.1811, 0.7917) and the all-variance solution (0.0173, 0.1067). Suppose we change the control scheme at run 100. Clearly, a new transient occurs at run 100.

4. Proposed time-varying weights tuning method

In order to enhance the performance of the double EWMA controller, a simple but effective time-varying weights tuning algorithm will be proposed in this section. We will first present a preliminary model of the time-varying control scheme and then modify it to be our proposed tuning method.

It was intuition to initially use the all-bias solution to bring the process on target, and then use the all-variance solution to reduce the process oscillations around the target. However, it has shown to be an inefficient tuning method from the viewpoint of “abrupt change”. Therefore, we are attempting to use the viewpoint of “gradual change”, which means using higher weights first and then “gradually” reducing them to the all-variance solution weights. We call this control strategy a GC (gradual change) control scheme, and it can be expressed as follows:

\[
\lambda_1(t) = \lambda_{1,v} + (f_1)^t,
\]

\[
\lambda_2(t) = \lambda_{2,v} + (f_2)^t,
\]

where \(f_1 (0 \leq f_1 < 1)\) and \(f_2 (0 \leq f_2 < 1)\) denote the discount factor. \(\lambda_{1,v}\) and \(\lambda_{2,v}\) individually represent the \(a_i\)'s and \(D_i\)'s all-variance solution weight. From Eqs. (14) and (15), we can see that \(\lambda_1(t)\) and \(\lambda_2(t)\) approach the all-variance solution as time approaches to infinity.

However, there is a problem for adding a discount factor in Eq. (14). Consider a drifting process in Eq. (1). Fig. 3 shows the results of the controlled process (without noise term) when the discount value \(f_1\) is large, moderate and 0. We can see that a large discount value makes it difficult for the process to converge to the target value. On the contrary, when the discount value is 0, the process converges quickly to the target. Therefore, we are tempering it by setting the discount value in Eq. (14) to be 0. That is

\[
\lambda_1(t) = \lambda_{1,v},
\]

\[
\lambda_2(t) = \lambda_{2,v} + (f)^t,
\]

where \(f (0 \leq f < 1)\) expresses the discount factor. We call Eqs. (16) and (17) the MGC (modified gradual change) control scheme.

From Eq. (10), it can be shown that if the double EWMA controller with the fixed weights control scheme, then the following equation holds as:

\[
(\lambda_1, \lambda_2) = (\max\{\lambda_1, \lambda_2\}, \min\{\lambda_1, \lambda_2\})
\]

\[
= (\min\{\lambda_1, \lambda_2\}, \max\{\lambda_1, \lambda_2\}).
\]
Fig. 3. Controlled output with $f_1$ being large, moderate and 0.

Fig. 4. MGC-1 versus MGC-2 tuning method.

But, for the time-varying weights, the above equation does not hold. Therefore, there are two cases in the MGC control scheme, they are MGC-1:

$$\hat{\lambda}_1(t) = \min\{\lambda_{1,v}, \lambda_{2,v}\},$$  \hspace{1cm} (18)

$$\hat{\lambda}_2(t) = \max\{\lambda_{1,v}, \lambda_{2,v}\} + (f)^\gamma$$  \hspace{1cm} (19)

and MGC-2:

$$\hat{\lambda}_1(t) = \max\{\lambda_{1,v}, \lambda_{2,v}\},$$  \hspace{1cm} (20)

$$\hat{\lambda}_2(t) = \min\{\lambda_{1,v}, \lambda_{2,v}\} + (f)^\gamma.$$  \hspace{1cm} (21)

In order to compare the tuning method between MGC-1 and MGC-2, we assume the discount factor $f$ to be fixed as a constant. Fig. 4 shows the controlled process under MGC-1 versus MGC-2 control scheme. It shows that both control schemes converge to the target at almost the same time, but the controlled output under the MGC-2 tuning method shows much smaller offsets from the target. Thus, we will adopt the MGC-2 to be our proposed time-varying weights tuning method for the EWMA controller.

The advantage of adding a discount factor in our proposed tuning method is the quick response to the initial transient effect. In Fig. 5, the dash and solid line individually represent the controlled output ‘with’ and ‘without’ adding the discount value. For the case of $f 
eq 0$, the double EWMA controller compensates for the initial transient effect more quickly than in the case of $f = 0$. From the proposed tuning equations, we know that the performance of the controlled process output depends on setting the discount factor ($f$). Even though a larger discount value can quickly compensate for the initial transient effect, it may cause oscillations. Therefore, we will present how to determine the discount parameter in the following section.

5. Simulation results

In this section, we first present how to determine the discount factor, and then we implement the proposed tuning method by using the software of Matlab/Simulink version 4.1. Finally, we will make a comparison between the fixed weight trade-off solution and the proposed time-varying tuning method.
5.1. Implementation

For the control performance characterization, the normalized mean square error (MSE/$\sigma^2$) is used as the performance measure. The prediction MSE is defined as follows:

$$\text{MSE} = \frac{\sum_{i=1}^{m}(y_t - T)^2}{m}.$$ 

Therefore, the normalized mean square error is the measure of inflation for the controlled process produced against the natural disturbance ($\varepsilon$). To implement the proposed tuning method, the first task is to determine the discount factor. The objective of the chosen discount factor is to minimize the normalized mean square error. Consider a ramp process model in Eq. (1) with $\alpha = 0$, $T = 0$, $\zeta = \beta/b = 1$, $\delta = 1$, $\sigma^2 = 1$ and $m = 200$. By solving Eq. (13), we obtain the all-variance solution weight ($\lambda_{1r}, \lambda_{2r}$) = (0.0247, 0.1486). Therefore, our proposed tuning method can be expressed as follows:

$$\lambda_r^1 = \max\{0.0247, 0.1486\},$$

$$\lambda_r^2(t) = \min\{0.0247, 0.1486\} + (f)^t.$$ 

A plot of the normalized mean square errors of the controlled process output, obtained from a series of trial values of the discount factor, is shown in Fig. 6. The minimum value of MSE/$\sigma^2$ appears at $f^* = 0.92$. Table 1 shows a similar process for determining the discount factor under other drifting speeds. We can see that whatever the drifting speed is, the optimal discount factor always appears at $f^* = 0.92$. Thus, it shows the robustness for determining the discount factor under our proposed tuning method. Substituting $f^* = 0.92$ into Eq. (23), Fig. 7 shows the controlled process output, and it shows that observations wander around the target value ($T = 0$) with MSE/$\sigma^2$ = 1.2481. In addition, if we use the fixed trade-off solution weights to control the process, then MSE/$\sigma^2$ = 1.6183. Therefore, our proposed time-varying
The tuning method is 22.88% better than the fixed weights control scheme. The following section will make a more detailed comparison between the trade-off solution, and the proposed tuning method, by using the Monte Carlo simulations.

5.2. Comparisons

In order to validate the effectiveness of the proposed time-varying weight tuning method, Monte Carlo simulations are performed under various random seeds. In these Monte Carlo simulations, we assume the target value $T = 0$, $\alpha = 0$, $\xi = \beta/h = 1$ and $e_i \sim N(0, 1)$. At each simulation, the performance index, $\text{MSE}/\sigma^2$, is calculated based on the simulation results of 200 runs ($m$) and 200 simulation cycles (initial seed from 0 to 199).

Table 2 shows the comparison results between the fixed trade-off solution weights, and the proposed time-varying tuning methods under various drifting speeds. The estimated standard deviation errors are shown in the parentheses. We can see that the performance between the two control schemes is not significantly different with the slow drifting speed (say $\delta = 0.1$). But, when the drifting speed is moderate to large (say $\delta \geq 0.5$), then the proposed time-varying tuning method is much better than the fixed trade-off control scheme. The last column of Table 2 shows the percent improvement of the proposed time-varying tuning method over the fixed trade-off control scheme. It shows that the larger the drifting speed, the more improved the performance. Thus, it is recommended that when a drifting disturbance model exists in the process, using the proposed method to tune the double EWMA controller will produce a satisfactory performance.

6. Conclusion

In order to enhance the performance of the double EWMA controller, a simple but effective time-varying weights tuning method was developed. From this study, we know that adding a discount factor into the first EWMA controller equation will decrease the converging speed. But, by adding the discount factor into the second EWMA controller equation, one can quickly compensate for the initial transient effect. We also have shown the robustness of determining the discount factor under the proposed tuning method. Finally, through Monte Carlo simulations, we have shown that the proposed time-varying weights tuning method possesses a significant improvement over the fixed trade-off solution.
weights control scheme, especially for the process with a moderate to large drifting rate.

Although the proposed tuning method was implemented via simulation, it is nevertheless anticipated to improve the performance of the double EWMA controller in an actual process. Further research could extend the proposed time-varying tuning technique to the multiple inputs multiple outputs (MIMO) system. In addition, the determination of the stability conditions and proving the robustness of the time-varying weights tuning strategy are also important issues.

References