SHEWMAC: an End-of-Line SPC Scheme for Joint Monitoring of Process Mean and Variance

Chih-Min Fan[+], Shi-Chung Chang[+],
Ruey-Shan Guo[*]
[+]Department of Electrical Engineering
[*]Dept. of Industrial Mgmt. and Business Admin.
National Taiwan University
scchang@cc.ee.ntu.edu.tw

Hui-Hung Kung, Jyh-Cheng You, Hsin-Pai Chen,
Steven Lin, John Wei
Taiwan Semiconductor Manufacturing Co.
hhkung@tsmc.com.tw

Abstract – Under the effects of multiple-stream and sequence-disorder, process change caused by one machine at an in-line step may result in changes in both the mean and variance of end-of-line wafer acceptance test (WAT) data sequence. To speed up trend detection of WAT data without resorting to an intensive computing power, an end-of-line SHEWMAC scheme is proposed, which combines a Shewhart, an exponentially weighted moving average (EWMA), and an exponentially weighted moving Cpk (EWMC) charts for jointly monitoring the mean and variance of wafer lot average sequence from WAT data. In view of the wide ranges of process conditions and low volume of each product in a foundry fab, a data normalization technique is adopted to aggregate data of similar products and a new design method is developed to generate a robust set of scheme parameters. Simulation and field data validation show that SHEWMAC is superior to the combined Shewhart-EWMA scheme in shift detection speed and is complementary to the in-line SPC.

I. INTRODUCTION

Statistical Process Control (SPC) techniques have been widely adopted in semiconductor fabrication for the purpose of in-line process monitoring and control. Nevertheless, from the viewpoint of process integration, statistical stability at the in-line level does not guarantee the stability of the whole IC fabrication process. Quality control techniques such as acceptance sampling test, trend/control chart, and variance decomposition have therefore been applied and/or extended by fabs [1]-[3] to the monitoring and control of end-of-line wafer acceptance (WAT) data.

WAT data provides the integral statistics about process stability and product performance. It has the salient features of *sequence disorder* (SD) and *multiple streams* (MS) due to operation dispatching as compared with in-line data of individual machines and/or fabrication steps. In presence of the two features, a change caused by one machine at an in-line step may result in changes in both the mean and variance of a WAT data sequence. A current industrial practice groups WAT data over a period of time (window) and monitors mean, variance or process capability index (Cpk) of data groups respectively. In specific, a control chart of Cpk may serve to detect combined changes in mean and variance, and a window size of one week is taken for grouping so that trend patterns can be extracted under the salient features of WAT data sequence.

In many of the aforementioned WAT monitoring schemes, the control limits and window size are determined empirically because in-line SPC techniques do not apply directly. As a result, window size and control limits thus selected have led to slow process fault detection or frequent false alarms. There have been a lack of solid foundation for the design and analysis of WAT SPC schemes, especially for a fab where product types, process characteristics, and the intensities of SD and MS effects vary widely and frequently.

In [4], the authors proposed a framework of end-of-line quality control (Figure 1) and focused on the end-of-line SPC module. A SHEWMA scheme was developed and implemented in a foundry fab. It is a methodology for generating robust design parameters for the simultaneous application of Shewhart and EWMA control charts to WAT data. Filed data validation shows that the incorporation of SHEWMA control charts complements the existing end-of-line data monitoring/analysis system and in-line SPC schemes for process integration. It indeed improves the false alarm rate, detection speed and diagnosis efficiency from the current practice without resorting to an intensive computing power as the approach taken by [3].

By exploiting the advantages of both SHEWMA and Cpk review schemes, a new and integrated WAT SPC scheme, SHEWMAC, is developed in this paper for jointly monitoring mean and variance of wafer lot average sequence from WAT data. The proposed SHEWMAC scheme consists of a Shewhart, an EWMA, and an exponentially weighted moving Cpk (EWMC) control charts. Figure 2 illustrates the potential advantage of SHEWMAC over SHEWMA. The shaded areas in the mean-versus-variance plots are the respective in-control regions of SHEWMA and SHEWMAC derived by our analysis under approximately the same false alarm rate. It is obvious that when both process mean and variance change together, the monitored statistics are more likely to fall outside the in-control region of SHEWMAC. Namely, SHEWMAC is more sensitive in detecting a combined mean and variance change at a given false alarm rate. Compared with the currently used simple Cpk scheme for batch review, the SHEWMAC scheme has the advantage of easier scheme parameter design and rolling review.

The remainders of this paper will first characterize the SD and MS features of WAT data. A SHEWMAC system is then designed for industrial applications. Finally, by using simulation and fab data, the effectiveness of SHEWMAC scheme is validated.

0-7803-5403-6/99/$10.00 ©1999 IEEE.

II. SEQUENCE-DISORDER & MULTIPLE-STREAM

Figure 3 demonstrates the generation process of a WAT data sequence. Let $\{\bar{X}_i\}$ be a random sequence representing wafer lot averages of a WAT measurement item, where i is the lot output sequence index at the WAT step. In general, affected by different product flows and dispatching polices, the cycle time from a process step p to the end-of-line WAT step varies among lots. As a result, the lot with a sequence label n at step p very likely has a different lot sequence label i at the WAT step. This is defined as the *sequence-disorder effect*. Note that the processing of a lot may require more than 300 steps and each step may be processed by any one of a machine group. Define a *stream* as a sequence of machines that a lot goes through during its fabrication process. There are many possible streams in a fab and the resultant WAT measurements among different streams vary due to machine-to-machine variation. This is defined as the *multiple-stream effect*.

A triplet of process conditions (R, M, S) are defined to characterize these two salient features of WAT data, where

- R is the SD range from the monitored step p to WAT step (defined in Figure 3),
- M is the total number of machines in the monitored step p, and
- S is the potential magnitude of a shift (in standard deviation unit).

For example, when (R, M, S)=(*15*, *2*, *1.5*), the changes in both mean and variance of WAT data in end-of-line lot sequence, $\{\bar{X}_i\}$, in contrast with those in in-line lot sequence from the abnormal machine m is demonstrated in Figure 4. It can be seen that an in-line shift on machine m ramps and then levels off in the WAT data sequence, where the magnitude of leveling off part is reduced and the variance increases as compared with the original in-line shift. It is clear that to enhance the WAT shift detection speed, the end-of-line SPC scheme should have the capability to simultaneously detect changes in both mean and variance of $\{\bar{X}_i\}$.

III. SHEWMAC SYSTEM

Figure 5 depicts the schematic diagram of SHEWMAC tool implementation. There are three function modules: Input Data Normalization, Control Charting, and Robust Parameter Generation. In a foundry fab, daily generation of WAT data of each product type may be statistically "rare". To increase the sample size, WAT data inputs are first normalized so that data of different products belonging to the same processing technology can be aggregated to reach a scale of statistical significance. A normalized data sequence can then be monitored lot-by-lot by the Control Charting module based on the scheme parameters from the Robust Parameter Generation module. The Robust Parameter Generation module takes the requirement of false alarm rate and the possible range of process conditions $\Omega \equiv \{(R, M, S)\}$ as inputs. It evaluates the scheme performance and generates a robust set of SHEWMAC parameters over a wide range of process conditions. The outputs of the SHEWMAC scheme include a Shewhart, an EWMA, and an EWMC control charts of the normalized WAT lot average sequence, and a warning signal when a data point is out of control.

Data Normalization

The objective here is to use the historical WAT lot average sequence to establish the baseline behavior, and later normalize the real time WAT lot average sequence based on this baseline. The baseline behavior consists of the long-term mean ($\hat{\mu}$) and variance ($\hat{\sigma}^2_{\bar{X}}$) of $\{\bar{X}_i\}$. This paper assumes that $\{\bar{X}_i\}$ follows a normal distribution. In specific, a moving range estimator [5] is adopted to estimate the variance $\hat{\sigma}_{\bar{X}} \approx 0.887\overline{MR}$, where $\overline{MR} = (\sum_{i=1}^{I_0} MR_i)/I_0$, $MR_i \equiv |\bar{X}_{i+1} - \bar{X}_i|$, $i = 1,2,...,I_0$, and I_0 is the number of samples. This estimator is unbiased, is robust with respect to shifts in the process mean, and can model the machine-to-machine variation among lots well. Given $\hat{\mu}$ and $\hat{\sigma}^2_{\bar{X}}$, the normalized metric $\bar{Z}_i \equiv (\bar{X}_i - \hat{\mu})/\hat{\sigma}_{\bar{X}}$ will be approximately normally distributed and can be used as the common metric for all products.

Control Charting

In the Control Charting module, the Shewhart chart tests if the average of a lot is normal; the EWMA chart tests if there is any small WAT shift; and the EWMC chart tests if the slight changes in mean and variance result in a significant changes in Cpk. Warning messages from these three charts provide information about the occurrence and the extent of a process shift. If only the EWMA or EWMC chart detects an abnormal trend, there could be a small process shift. When there is a large trend in the EWMA and EWMC charts and a data point out of Shehwart control limits at the same time, a large process shift may have occurred.

Let the monitored statistics be $\{\bar{Z}_i\}$ in the Shewhart chart. The EWMA sequence is then generated by

$$A_i = \lambda\bar{Z}_i + (1-\lambda)A_{i-1}$$
$$= \sum_{q=0}^{i-1} W_{i-q}\bar{Z}_{i-q} + (1-\lambda)^i A_0, \quad i=1,2,..., \qquad (1)$$

where $W_{i-q} = \lambda(1-\lambda)^q$, $0 < \lambda \le 1$, and the initial value A_0 is usually set as zero. To get the Cpk values in EWMC chart, the variance is first estimated by $V_i = B_i - A_i^2$, where

$$B_i = \lambda\bar{Z}_i^2 + (1-\lambda)B_{i-1}$$
$$= \sum_{q=0}^{i-1} W_{i-q}\bar{Z}_{i-q}^2 + (1-\lambda)^i B_0, \quad i=1,2,..., \qquad (2)$$

is an exponentially weighted moving estimator of mean square and B_0 is usually set as 1. Given A_i and B_i, the EWMC sequence is then generated by

$$C_i = Min(USL - A_i, A_i - LSL)/(3\sqrt{V_i}), \qquad (3)$$

where USL and LSL are the upper and lower specification limits respectively.

In summary, SHEWMAC scheme parameters consists of quadruplet (c, λ, h, k), where c is the Shewhart control limit gain, λ is the EWMA weighting factor, h is the

EWMA control limit gain, and k is the EWMC control limit gain. Once the SHEWMAC parameters (c, λ, h, k) are available, control limits of Shewhart chart, EWMA chart, and EWMC chart are then set as $\pm c$, $\pm h\sqrt{\lambda/(2-\lambda)}$, and k respectively.

It is clear that A_i and B_i is a moving average and a moving mean square of $\{\bar{Z}_i, \bar{Z}_{i-1}, ..., \bar{Z}_1\}$ respectively with exponentially decreasing weighting coefficients, i.e., they tend to emphasize on utilizing the most recently collected data. To pop out the underlying trend in SD and MS data, a large window size (a small weighting factor λ) is needed. However, if the weighting factor λ is too small, the EWMA and EWMC will not be sensitive to process change and the detection speed will be slow. The other three parameters c, h, and k should also be designed in accordance with the choice of λ to maximize the detection speed and maintain a desirable false alarm rate.

Robust Parameter Generation

Figure 6 depicts the design procedures in the Robust Parameter Generation, which are based on the concept of run length. The run length is a random variable characterizing the number of observations that an SPC scheme takes to generate an out-of-control signal after the occurrence of a process change. In view of the fact that in Eq. (1), each EWMA value A_i is an interpolation of its former value A_{i-1} and the present normalized lot average $\bar{Z}_i$, the average run length of an EWMA chart is usually characterized as a discrete state Markov chain [6]. Similar to this approach, the Robust Parameter Generation module models the SHEWMAC as a two-variable, A_i and B_i, Markov chain. The robust design of SHEWMAC has two folds: to maximize average run length $ARL0$ for a normal process and to minimize average run length $ARL1$ after a process becomes abnormal. In practice, exact process conditions (R, M, S) cannot be known *a priori*. For the feasibility of implementation, the optimal parameters for each process condition in Ω is first calculated. Then a robust design of parameters is chosen so that the SHEWMAC scheme results in a satisfactory performance over possible conditions in Ω.

IV. VALIDATION

Simulation

As the proposed SHEWMAC is a simultaneous application of Shewhart, EWMA, and EWMC schemes, it therefore combines all the advantageous features of the three. Figure 7 demonstrates the simulation result that EWMC is good at median shift (1.5~2.5 sigma) detection, EWMA is superior in small shift (<1.5 sigma) detection and Shewhart is suitable for large shift (over 3 sigma) detection. Whichever the shift condition is, the detection speed of SHEWMAC equals the fastest of the three.

Field Data Application

A 0.26 μm logic device is selected with a focus on monitoring WAT item of Rs_N+, which represents the sheet resistance of N+ structure. In this case, the SHEWMAC parameters are chosen as (c, λ, h, k)=(3.25, 0.11, 2.90, 0.65) and the corresponding SHEWMAC control charts are demonstrated in Figures 8(a) and 8(b). The SHEWMAC generates seven warning messages, one from the Shewhart chart at the 65th lot, three from the EWMA chart at the 27th, 37th, and 64th lots, and the other three from the EWMC chart at the 27th, 37th, and 55th lots respectively.

Through the data trace back and stratification functions of engineering data analysis (EDA) systm, it is found that N+ drain/source implant step is the root cause. Figure 8(c) demonstrates the Shewhart chart of Rs_N+ in the lot sequence and processing machines at the faulty step. It is obvious that M1 had a significant machine offset from the 29th to 36th lots in its in-line lot sequence as compared to the other machines. Also, a process shift occurred at M4 starting from the 62th lot in its in-line lot sequence. In this case, it is validated that EWMA and EWMC charts are supperior to the Shewhart chart in detecting the samll machine offset of M1. Also, since the EWMC chart reflects the changes in both mean and variance, it enhances the shift detection of M4 by 10 lots as compared to the EWMA chart.

The in-line SPC at the N+ drain/source implant step monitors the sheet resistance taken from the test wafer every 12 hours. It did not detect the two shifts in this case. There may be two reasons. First, the in-line measurements may be less sensitive to the process change as compared to the WAT measurements taken from product wafer. Second, the sampling rate in in-line level is much less than that of WAT. SHEWMAC is thus complementary to the in-line SPC for process integration.

V. CONCLUSIONS

In this paper, an end-of-line SPC scheme, SHEWMAC, is proposed to monitor the simultaneous changes in mean and variance of WAT lot average sequence. Simulation and field data validation show that SHEWMAC is superior to the combined Shewhart-EWMA scheme in shift detection speed and is complementary to the in-line SPC. Full field implementation of the scheme will be reported in the near feature.

REFERENCES

[1] D. K. Michelson, "Statistically Calculating Reject Limits at Parametric Test," *Proc. IEEE/CPMT Int'l Electronics Manufacturing Technology Symposium*, pp. 172-177, 1997.

[2] F. N.H. Montijn-Dorgelo and H. J. ter Host, "Expert System for Test Structure Data Interpretation," *IEEE Proc. on Microelectronic Test Structures*, pp. 172-177, 1997.

[3] J. Pak, R. Kittler and P. Wen, "Advanced Methods for Analysis of Lot-to-Lot Yield Variation," *Proc. Int'l Symposium on Semiconductor Manufacturing*, pp. E17-E20, 1997.

[4] C.M. Fan, R.S. Guo and S.C. Chang, "An Integrated Fault Detection Scheme for Wafer Acceptance Test Data," *Proc. Int'l Symposium on Semiconductor Manufacturing*, pp. 440-443, 1998.

[5] E. Yashchin, "Monitoring variance components," *Technometrics*, vol. 36, no. 4, pp. 379-393, 1994.

[6] M.S. Saccucci and J.M. Lucas, "Average run lengths for exponentially weighted moving average control schemes using the Markov chain approach," *Journal of Quality Technology*, vol. 22, no. 2, pp. 154-162, 1990.

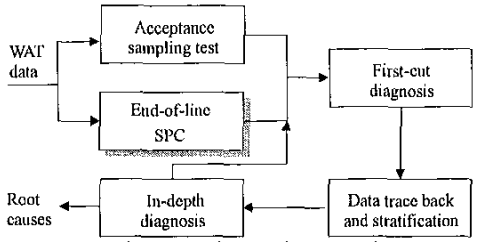


Figure 1: Sequential detection and diagnosis approach for end-of-line quality control

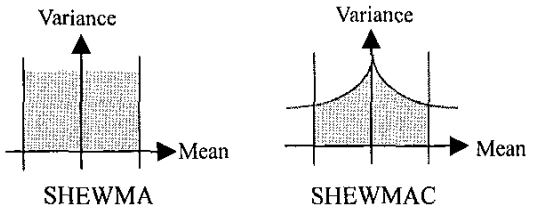


Figure 2: In-control regions of SHEWMA and SHEWMAC

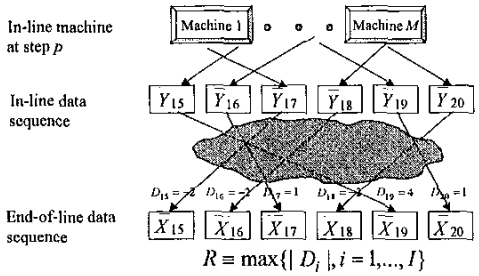


Figure 3: Generation process of end-of-line WAT data

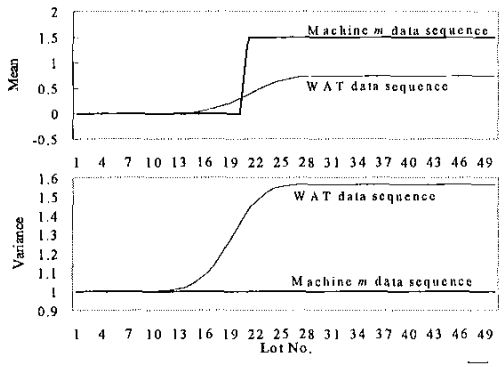


Figure 4: The changes in mean and variance of $\{\bar{X}_i\}$

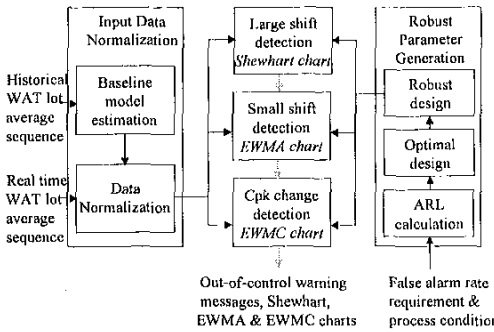


Figure 5: Schematic diagram of SHEWMAC tool

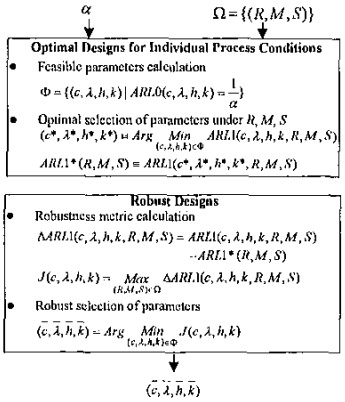


Figure 6: The design procedure of SHEWMAC scheme

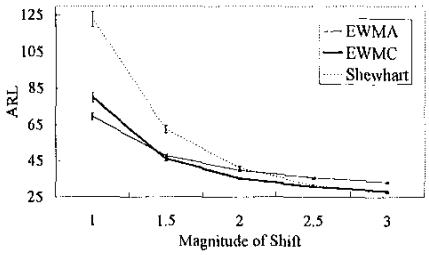
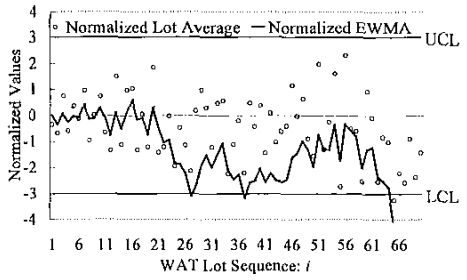
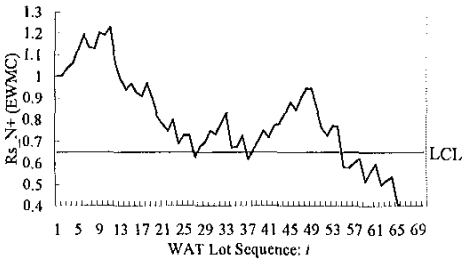


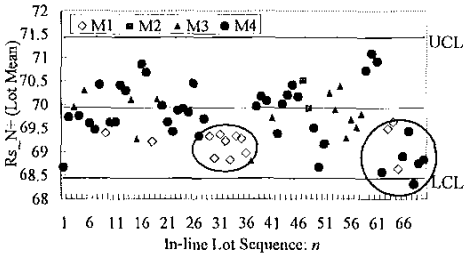
Figure 7: ARL versus magnitude of shift; R=25 and M=3



(a) Combined Shewhart-EWMA chart



(b) EWMC chart; USL and LSL are set as ±3



(c) Shewhart chart stratified by processing machines

Figure 8: Field data validation for SHEWMAC scheme