SHEWMAC: an End-of-line SPC Scheme via Exponentially Weighted Moving Statistics*

Chih-Min Fan[1], Shi-Chung Chang[1], Ruey-Shan Guo[2]
Hui-Hung Kung[3], Jyh-Cheng You[3], Hsin-Pai Chen[3], Steven Lin[3], Chih-Shih Wei[4]
[1] Department of Electrical Engineering, National Taiwan University
[2] Department of Industrial and Business Administration, National Taiwan University
[3] Taiwan Semiconductor Manufacturing Corporation
[4] Vanguard International Semiconductor Manufacturing Corporation

Abstract

Due to the multiple-stream and sequence-disorder effects, a 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 has been proposed by Fan et al. (1999). The SHEWMAC scheme consists of a Shewhart, an exponentially weighted moving average (EWMA), and an exponentially weighted moving Cpk (EWMC) charts for jointly monitoring the mean and variance of WAT lot average sequence. This paper aims at the robust design of a SHEWMAC scheme and the analysis of its effectiveness. Simulation results show that the SHEWMAC scheme with robust parameters reduces about 15% of the time in detecting WAT trends as compared to either the Exponentially Weighted Mean Square (EWMS) or combined Shewhart-EWMA schemes generally used for trend detection. Field data validation also shows that the incorporation of SHEWMAC complements the existing end-of-line data monitoring system and in-line SPC schemes for process integration.

1. Introduction

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.

* This work was supported in part by the National Science Council of the Republic of China under Grant NSC 87-2416-H-002-018 and by Taiwan Semiconductor Manufacturing Corporation under Contract 85-S-047

0-7803-6374-4/00/$10.00©2000 IEEE

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. By exploiting the advantages of both SHEWMA and Cpk review schemes, the authors further developed an integrated WAT SPC scheme [5], SHEWMAC, for jointly monitoring mean and variance of wafer lot average sequence from WAT data. The 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.

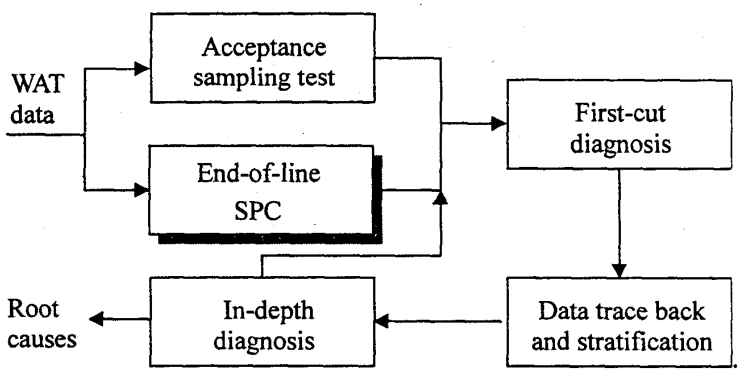


Fig. 1 Sequential detection and diagnosis approach for end-of-line quality control

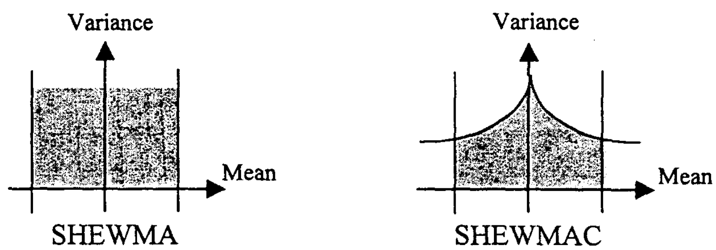


Fig. 2 In-control regions of SHEWMA and SHEWMAC

This paper aims at the robust design of a SHEWMAC scheme and the analysis of its effectiveness for a real fab. A robust set of SHEWMAC parameters are calculated by considering the requirement of false alarm rate and the wide range of process conditions in a fab. To highlight the advantages of robust SHEWMAC design, four SPC schemes are compared: CSE, EWMS, SHEWMA, and SHEWMAC. All of the four schemes use exponentially weighted moving statistics as monitoring statistics. The CSE scheme is the direct use of a combined Shewhart-EWMA scheme without considering the multiple-stream and sequence-disorder effects [7]. The EWMS (exponentially weighted mean square) scheme is usually adopted for detecting the combined mean change and variance increase [8]. The SHEWMA scheme is a special case of SHEWMAC scheme with no EWMC control limit.

The remainder of this paper is organized as follows. Section 2 first characterizes the SD and MS features of WAT data. The robust design of SHEWMAC scheme is described in Section 3. Section 4 then compares it with the other three schemes using exponentially weighted moving statistics as monitoring statistics. By using fab data, the effectiveness of SHEWMAC scheme is finally analyzed.

2. SEQUENCE-DISORDER & MULTIPLE-STREAM

Figure 3 demonstrates the generation process of a WAT data sequence. Let $\{\overline{X}_i\}$ 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.*

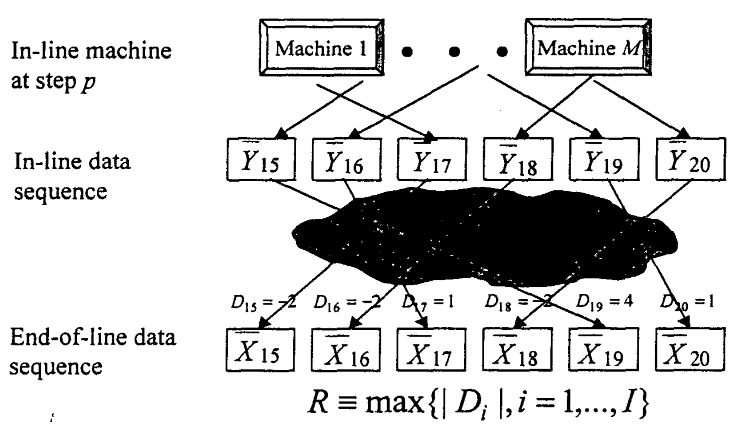


Fig. 3 Generation process of end-of-line WAT data

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 Fig. 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, $\{\overline{X}_i\}$, 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 $\{\overline{X}_i\}$.

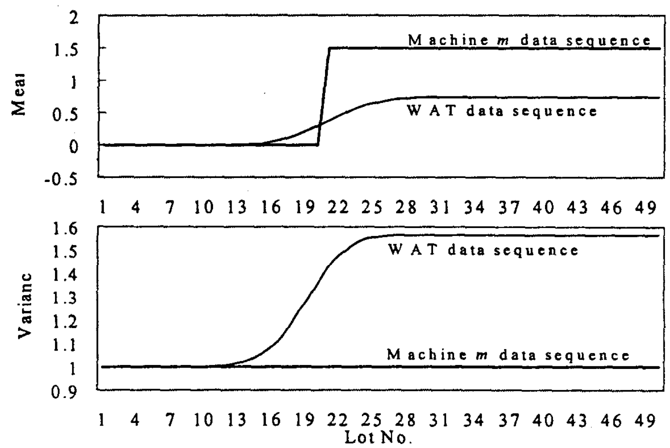


Fig. 4 The changes in mean and variance of $\{\overline{X}_i\}$

3. SHEWMAC SYSTEM

Figure 5 depicts the schematic diagram of SHEWMAC tool implementation. There are three function modules: Input Data Normalization, Control Charting, and Integrated Design. 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 Integrated Design module. The Integrated Design 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.

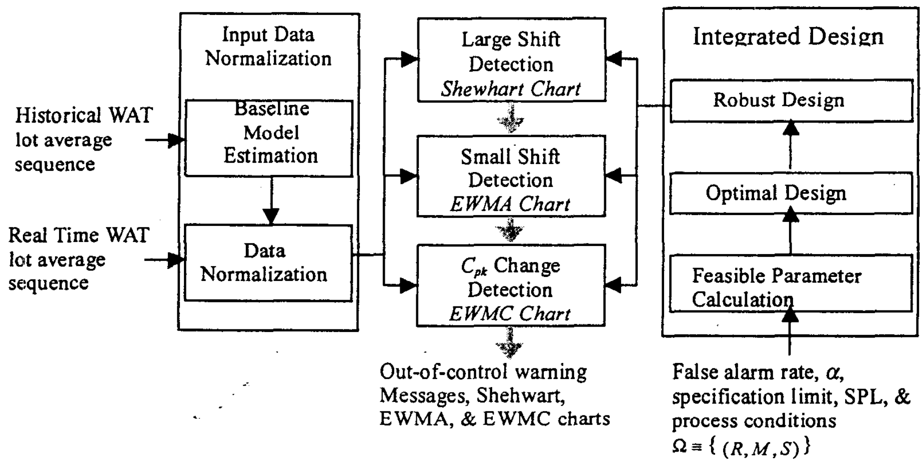


Fig. 5 Schematic diagram of SHEWMAC tool

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 ($\hat{\sigma}_{\overline{X}}^2$) of $\{\overline{X}_i\}$. This paper assumes that $\{\overline{X}_i\}$ follows a normal distribution. In specific, a moving range estimator [6] is adopted to estimate the variance $\hat{\sigma}_{\overline{X}} \approx 0.887\overline{MR}$, where $\overline{MR} = (\sum_{i=1}^{I_0} MR_i)/I_0$, $MR_i \equiv |\overline{X}_{i+1} - \overline{X}_i|, i = 1,2,\ldots,I_0$, 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 $\hat{\sigma}_{\overline{X}}^2$, the normalized metric $\overline{Z}_i \equiv (\overline{X}_i - \hat{\mu})/\hat{\sigma}_{\overline{X}}$ 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 $\{\overline{Z}_i\}$ in the Shewhart chart. The EWMA sequence is then generated by

$$\begin{aligned} A_i &= \lambda \overline{Z}_i + (1-\lambda)A_{i-1} \\ &= \sum_{q=0}^{i-1} W_{i\text{-}q} \overline{Z}_{i\text{-}q} + (1\text{-}\lambda)^i A_0, \quad i\text{=}1,2,\ldots, \end{aligned} \tag{1}$$

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

$$\begin{aligned} B_i &= \lambda \overline{Z}_i^2 + (1-\lambda)B_{i-1} \\ &= \sum_{q=0}^{i-1} W_{i\text{-}q} \overline{Z}_{i-q}^2 + (1\text{-}\lambda)^i B_0, \quad i\text{=}1,2,\ldots, \end{aligned} \tag{2}$$

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

$$C_i = (SPL - |\, A_i \,|)/(3\sqrt{V_i}), \tag{3}$$

where SPL is the normalized specification limit.

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

Integrated Design

The parametric design of SHEWMAC is 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 $\overline{Z}_i$, the average run length of an EWMA chart is usually characterized as a discrete state Markov chain [7]. Similar to this approach, the average run length of SHEWMAC is characterized as a two-variable, A_i and B_i, Markov chain.

Figure 6 depicts the design procedures in the Integrated Design module. Design inputs include a set of process conditions, $\Omega \equiv \{(R,M,S),\ R \in \mathbf{R}^+,\ M \in \mathbf{Z}^+,\ S \in \mathbf{R}^+\}$ and the requirement of false alarm rate, α. Design output is a robust selection of parameters, ($\bar{c}$, $\bar{\lambda}$, $\bar{h}$, $\bar{k}$). The philosophy of SHEWMAC design is to minimize the average run length $ARL1$ when process is out of control while maintaining the average run length $ARL0$ at a fixed level when process is in control. Thus the feasible parameter set satisfying the false alarm rate requiremnet is first calculated. Then the optimal parameters for individual process conditions in Ω are calculated. In practice, exact process conditions (R,M,S) cannot be known *a priori*. For the feasibility of implementation, a robust design of parameters is chosen by minimizing the detection delay in the wost case, so that the SHEWMAC scheme results in a satisfactory performance over possible conditions in Ω.

α $\qquad\qquad$ $\Omega = \{(R, M, S)\}$

Optimal Designs for Individual Process Conditions

- Feasible parameters calculation

$$\Phi = \{(c, \lambda, h, k) \mid ARL0(c, \lambda, h, k) = \frac{1}{\alpha}\}$$

- Optimal selection of parameters under R, M, S

$$(c^*, \lambda^*, h^*, k^*) \equiv Arg \underset{(c,\lambda,h,k)\in\Phi}{Min} ARL1(c, \lambda, h, k, R, M, S)$$

$$ARL1^*(R, M, S) \equiv ARL1(c^*, \lambda^*, h^*, k^*, R, M, S)$$

Robust Designs

- Robustness metric calculation

$$\Delta ARL1(c, \lambda, h, k, R, M, S) = ARL1(c, \lambda, h, k, R, M, S) - ARL1^*(R, M, S)$$

$$J(c, \lambda, h, k) = \underset{(R,M,S)\in\Omega}{Max} \Delta ARL1(c, \lambda, h, k, R, M, S)$$

- Robust selection of parameters

$$(\bar{c}, \bar{\lambda}, \bar{h}, \bar{k}) = Arg \underset{(c,\lambda,h,k)\in\Phi}{Min} J(c, \lambda, h, k)$$

$(\bar{c}, \bar{\lambda}, \bar{h}, \bar{k})$

Fig. 6 The design procedure of SHEWMAC scheme

4. Performance Evaluation

To highlight the advantages of robust SHEWMAC design, four SPC schemes are compared here: CSE, EWMS, SHEWMA, and SHEWMAC. All of the four schemes use exponentially weighted moving statistics as monitoring statistics. The CSE scheme is the direct use of a combined Shewhart-EWMA scheme without considering the multiple-stream and sequence-disorder effects [7]. The EWMS (exponentially weighted mean square) scheme is usually adopted for detecting the combined mean change and variance increase [8]. The SHEWMA scheme is a special case of SHEWMAC scheme with no EWMC control limit.

Table 1 lists the range of process conditions considered and the resultant design parameters for the CSE, EWMS, SHEWMA, and SHEWMAC schemes, each of which meets the false alarm rate of 0.27%. All the robust parameters in Table 1 are calculated by simulation.

Figure 7 demonstrates the *ARL*1 performance of the four schemes under different magnitudes of process shift. The smaller the *ARL*1 value, the faster the detection speed. It can be seen that the CSE scheme without considering MS & SD effects always results in the worst performance. When the shift size is small (S=1), the SHEWMA scheme is better than the EWMS scheme. However, when there is a large shift size (S=2), the EWMS scheme is superior to the SHEWMA scheme. No matter what the magnitudes of process shift, the SHEWMAC scheme always has the smallest ARL1 and is the best among the four schemes in detecting WAT trend.

Schemes	Robust Parameters	Range of Process Conditions
CSE	$(\bar{c}_C, \bar{\lambda}_C, \bar{h}_C)$ =(3.75, 0.19, 2.866)	$\{R=0,\ M=1,\ 1\le S\le 2\}$
EWMS	$(\bar{\lambda}_B, \bar{h}_B)$ =(0.03, 2.140)	$\{0\le R\le 50,\ 1\le M\le 5,\ 1\le S\le 2\}$
SHEWMA	$(\bar{c}_{SA}, \bar{\lambda}_{SA}, \bar{h}_{SA})$ =(3.25, 0.03, 2.523)	$\{0\le R\le 50,\ 1\le M\le 5,\ 1\le S\le 2\}$
SHEWMAC	$(\bar{c}_{SAC}, \bar{\lambda}_{SAC}, \bar{h}_{SAC}, \bar{k}, SPL)$ =(3.25, 0.03, 2.934, 0.801, 3)	$\{0\le R\le 50,\ 1\le M\le 5,\ 1\le S\le 2\}$

Table 1 Robust parameters for schemes comparison

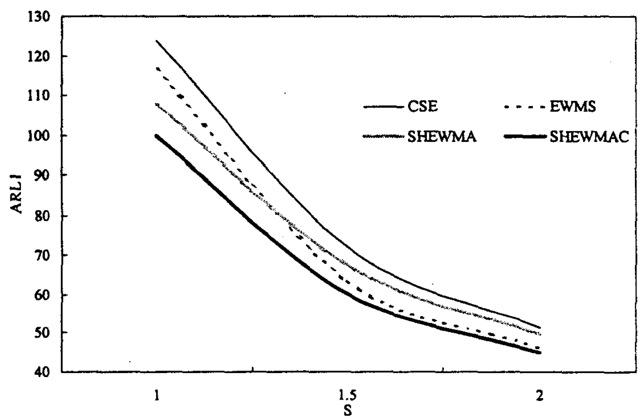
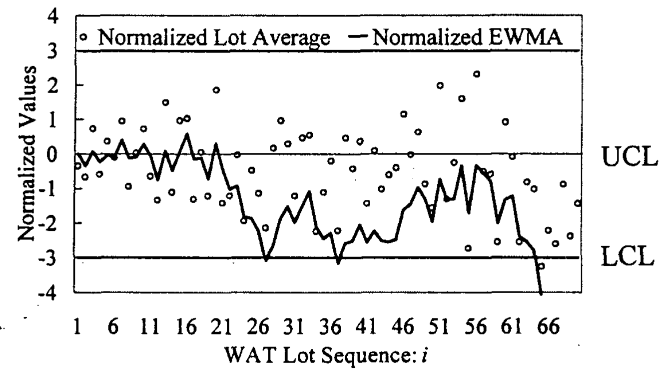


Fig. 7 Relation of ARL1 to S with R=25 and M=5

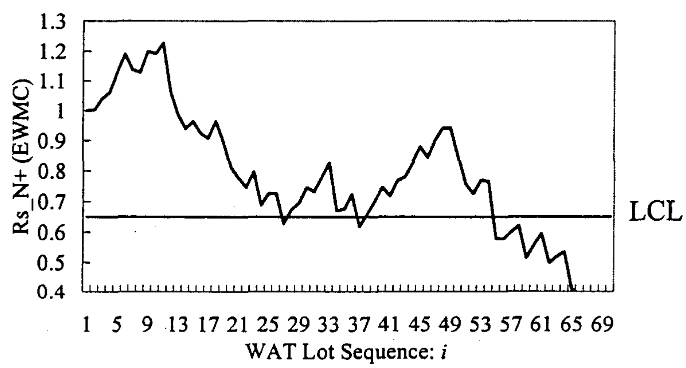
5. Field Data Application

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.

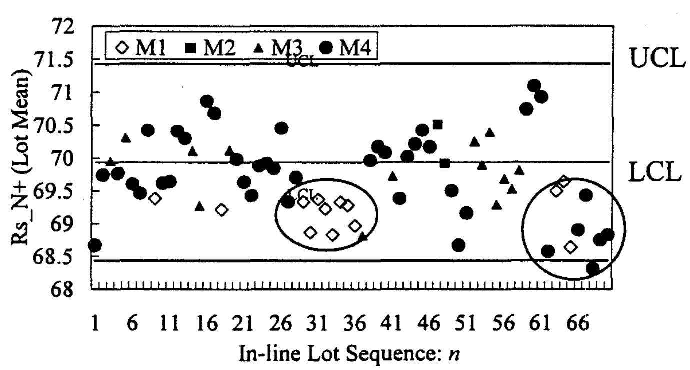
Through the data trace back and stratification functions of engineering data analysis (EDA) systm, it is found that N+ drain/source implant step is the root cause. Figure 8(c) demonstrates the Shewhart chart of Rs_N+ in the lot sequence and processing machines at the faulty step. It is obvious that M1 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.



(a) Combined Shewhart-EWMA chart



(b) EWMC chart; SPL is set as 3

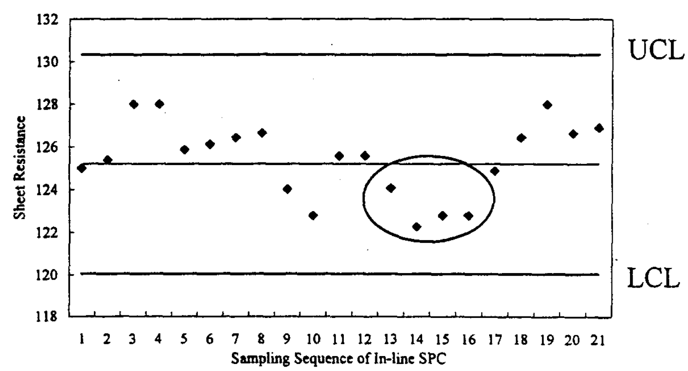


(c) Shewhart chart stratified by processing machines

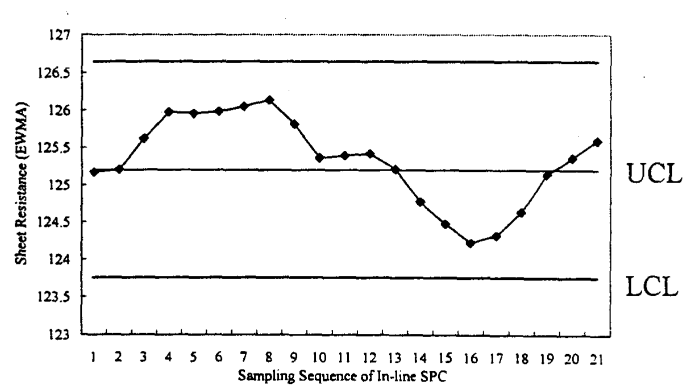
Fig. 8 Field data validation for SHEWMAC scheme

Can the fault be identified using in-line SPC for the N+ drain/source implant step? The in-line SPC at the N+ drain/source implant step monitors the sheet resistance, which is taken from the test wafer every 12 hours. Both Western Electric Rules (WER) and CSE schemes are adopted as the in-line SPC schemes.

The CSE charts for machine M4 at N+ drain/source implant step, during the tracking time of the 70 lots under investigation, are given in Figures 9(a) and 9(b). During the period of process shift, there are 23 lots, from the 48th to the 70th lot in Figure 8(c), processed by machine M4 for the N+ drain/source implant step. However, in the same period of time, only 4 data points, from the 13rd sampling point to the 16th sampling point in Figure 9, of sheet resistance are taken for in-line SPC. It can be seen that, using the in-line sheet resistance data, neither the CSE scheme nor the WER detects the large process shift in machine M4. As for the offset of machine M1, it is more difficult to detect by using in-line SPC because only two data points of sheet resistance are taken from M1 and the offset of M1 is much less than the magnitude of process shift in M4.



(a) Shewhart chart ($c = 3$)



(b) EWMA chart ($\lambda = 0.15$, $h = 2.96$)

Fig. 9 CSE charts for in-line SPC.

There are two reasons that the in-line SPC does not detect the process shift and machine offset in this case. 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. End-of-line SHEWMAC is thus complementary to the in-line SPC for process integration.

6. 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.

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, S.C. Chang, and C.S.Wei, "SHEWMA: an End-of-line SPC Scheme Using Wafer Acceptance Test Data," Accepted by *IEEE Transactions on Semiconductor Manufacturing*, 1999.

[5] C.M. Fan, S.C. Chang, R.S. Guo, H.H. Kung, J.C. You, H.P. Chen, S. Lin, and J. Wei, "SHEWMAC: an End-of-line SPC Scheme for Joint Monitoring of Process Mean and variance," *Proc. Int'l Symposium on Semiconductor Manufacturing*, pp. 233-236, 1999.

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

[7] 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.

[8] R. Domangue and S.C. Patch, "Some Omnibus Exponentially Weighted Moving Average Statistical Process Monitoring Schemes" *Technometrics*, vol. 33, pp. 299-314, Aug. 1991.