SCIENCE CHINA
Information Sciences

• LETTER •

August 2019, Vol. 62 089201:1–089201:3
https://doi.org/10.1007/s11432-018-9593-1

Kalman filtering-based supervisory run-to-run control method for semiconductor diffusion processes

Zhao XU[1], Changhua HU[2*] & Jinwen HU[3]

[1]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;*
[2]*Xi'an Research Institute of High-Technology, Xi'an 710025, China;*
[3]*School of Automation, Northwestern Polytechnical University, Xi'an 710072, China*

Received 16 May 2018/Accepted 16 July 2018/Published online 15 January 2019

Citation Xu Z, Hu C H, Hu J W. Kalman filtering-based supervisory run-to-run control method for semiconductor diffusion processes. Sci China Inf Sci, 2019, 62(8): 089201, https://doi.org/10.1007/s11432-018-9593-1

Dear editor,

In response to the need for continuous tuning solutions in run-to-run (R2R) manufacturing processes, R2R control began to emerge in academic and industrial research two decades ago. It is aimed at minimizing process drift, shift and variability between machine runs and decreasing cost in semiconductor companies. However, gaps still exist between the current development of R2R technologies and their industry-wide commercialization. The most widely used R2R control approaches are the exponentially-weighted moving average (EWMA) approach [1, 2], the double-EWMA controller [3], and the variable EWMA controller [4]. However, they suffer from system limitations such as the limited dimension of system models and limited types of disturbances [5]. In addition, current R2R technologies deal with process shifts or drifts by assuming the process gain to be constant and its estimate to be accurate by appropriately designing the experiments. In fact, the gain may also vary during the process, thus affecting the final product quality. Ignoring the effect of the time-varying process gain or its inaccurate estimate will lead to unexpected output errors when using the current R2R technologies if the gain varies significantly or the estimation error is large. The semiconductor diffusion process is still commonly carried out by using the traditional R2R control system. Diffusion refers to the entire process of adding a dopant to the surface of the wafer at a high temperature. The process is likely to be unstable when it is affected by production noises owing to equipment degradation and environment changes. Furthermore, current R2R technologies can guarantee the stability of single-input-single-output (SISO) or multiple-input-single-output (MISO) systems with unknown process shift and drift and constant process gain. However, the technologies that can guarantee the stability of multiple-input-multiple-out (MIMO) systems with unknown process shift and drift and time-varying process gain are still open for research.

A Kalman filtering-based supervisory R2R control (KFSR) method is presented for overcoming the limitations of traditional R2R control approaches for semiconductor diffusion processes. The Kalman filter is applied to adaptively estimate the system model parameters by incorporating the observed output measurements, following which an optimization-based control scheme is designed to select the optimal setting points. The developed KFSR method is adapted to systems of arbitrary dimensions and with more types of disturbances, and the system stability is proven to be guaranteed by the KFSR method.

System modeling. Consider the following linear R2R manufacturing system:

$$y_k = Hx_k + B_k u_k + v_k, \tag{1}$$

---

[*] Corresponding author (email: hch_reu@sina.com)

where $k > 0$ is an integer denoting the run index or batch number in R2R manufacturing, $y_k \in \mathbb{R}^p$ is the output measurement for the concerned product quality, $x_k \in \mathbb{R}^q$ is a state related to the process modeling error, $u_k \in \mathbb{R}^r$ is the control input, v_k is the random measurement noise, H and B_k are matrices of appropriate dimensions. p, q and r are positive integers. The desired output is denoted by y^* and the set of all output measurements up to time k is defined as $Y_k = \{y_1, \ldots, y_k\}$. In this study, we assume that $r \geqslant p$ because control variables should be in general no less than the independent output variables. It is clear to see that Hx_k is actually the intercept of the model that includes the modeling error and model drift. The following dynamic model of x_k is assumed:

$$x_k = Ax_{k-1} + w_k, \tag{2}$$

where A is the dynamic matrix and w_k is the random noise that drives x_k. Note that the model is general enough to include the most common intercept and process drift dynamics as shown in [6]. The matrices A and H are known as a priori and assumed to be constant. x_k is the state we need to estimate at each run for control. We use $\hat{x}_{j|i}$ to denote the estimate of x_j based on the measurements up to the i-th run and $\tilde{x}_{j|i} = x_j - \hat{x}_{j|i}$ to denote the corresponding estimation error. v_k and w_k are assumed to be zero-mean white Gaussian noises subject to

$$\mathrm{E}\begin{bmatrix} v_k \\ w_k \end{bmatrix}\begin{bmatrix} v_l \\ w_l \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} R_k\delta_{kl} & 0 \\ 0 & Q_k\delta_{kl} \end{bmatrix},$$

where δ_{kl} is the Kronecker delta; R_k and Q_k are assumed to be positive definite and bounded by $R_k \leqslant R$ (i.e., $R - R_k$ is semi-positive definite) and $Q_k \leqslant Q$, respectively. Although R_k and Q_k may not be known for each run, assuming that their bounds R and Q are known is natural. Throughout this article, $\mathrm{E}[\cdot]$ denotes the mathematical expectation, $\|\cdot\|$ is the 2 norm of vectors or matrices, and I is the identity matrix of appropriate dimension and T the transpose operation.

We use B_e to denote the estimate of B_k from the historical data that is assumed to be of a full row rank so that all the output parameters can be controlled. To guarantee the estimate is unbiased, $B_k \equiv B_e$ is indispensable to be unchanged as required by many other literatures. However, this condition may never hold in real systems as they always undergo dynamic changes owing to environmental changes or disturbances. In the following part, we further investigate the control performance in the general case by designing a suitable control strategy where B_k is given by

$$B_k = B_e + \Delta_k, \tag{3}$$

where Δ_k is a white Gaussian noise independent of v_k and w_k and subject to $\mathrm{E}[\Delta_k\Delta_k^{\mathrm{T}}] = G_k \leqslant G$ for all $k > 0$, and we obtain the following conclusions.

State estimation. The framework of Kalman filtering is applied for state estimation.

(1) Prediction step:

$$\begin{aligned} &\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1}, \\ &P_{k|k-1} = AP_{k-1|k-1}A^{\mathrm{T}} + Q. \end{aligned} \tag{4}$$

(2) Correction step:

$$\begin{aligned} &\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - B_eu_k - H\hat{x}_{k|k-1}), \\ &P_{k|k} = \left(P_{k|k-1}^{-1} + H^{\mathrm{T}}W^{-1}H\right)^{-1}, \end{aligned} \tag{5}$$

where $W = R + (U + \|u^*\|)^2G$, $K_k = P_{k|k}H^{\mathrm{T}}W^{-1}$, U is a given threshold related to the system control capability and u^* is a given reference control input.

R2R control. In R2R manufacturing, different control schemes are employed based on different optimization criteria according to different task requirements [7]. Nonetheless, the basic ideas are the same, i.e., to formulate the control design problem as a quadratic optimization problem. In this study, we adopt the same idea and determine the control input of each run by using

$$\begin{aligned} &u_k = \mathop{\arg\min}_{u \in \mathbb{R}^r} \|u - u^*\|^2 \\ &\text{s.t. } y^* = H\hat{x}_{k|k-1} + B_eu_k, \end{aligned} \tag{6}$$

where u^* is the user given reference control input. We can obtain the solution of the optimization problem (6) in a straightforward manner as

$$\begin{aligned} u_k = &\left(I - B_e^{\mathrm{T}}(B_eB_e^{\mathrm{T}})^{-1}B_e\right)u^* \\ &+ B_e^{\mathrm{T}}(B_eB_e^{\mathrm{T}})^{-1}(y^* - H\hat{x}_{k|k-1}). \end{aligned} \tag{7}$$

Remark 1. In real implementations, u_k may deviate largely from the reference input u^* owing to the effect of random shift or drift on $\hat{x}_{k|k-1}$ and exceeds the system capability, actually meaning that the manufacturing process has been out of control. Therefore, in real implementations, an upper bound U can be set such that an alarm is triggered for machine maintainable to prevent the process to be out of control when $\|u - u^*\| > U$.

Theorem 1. If $\mathrm{E}[\Delta_k] \equiv 0$, $\mathrm{E}[\tilde{x}_{1|0}] = 0$, $P_{1|0}$ is a positive definite matrix subject to $\mathrm{E}[\tilde{x}_{1|0}\tilde{x}_{1|0}^{\mathrm{T}}] \leqslant P_{1|0}$, the pair $\{A, H\}$ is observable, and the control input u_k is given by (7) subject to $\|u - u^*\| \leqslant U$, then it holds that $\mathrm{E}[y_k] = y^*$ and $\mathrm{E}[\|y_k - y^*\|^2] \leqslant \sigma^2$ for all $k > 0$ and some positive number σ^2.

Proof. First, defining $\tilde{x}_{k+1|k} = x_{k+1|k} - \hat{x}_{k+1|k}$, we have

$$\begin{aligned}\tilde{x}_{k+1|k} = {} & A\left(I - K_kH\right)\tilde{x}_{k|k-1} - AK_kv_k \\ & - AK_k\Delta_ku_k + w_k, \end{aligned} \tag{8}$$

which implies

$$\begin{aligned}\mathrm{E}\left[\tilde{x}_{k+1|k}\right] &= A\left(I - K_kH\right)\mathrm{E}\left[\tilde{x}_{k|k-1}\right] \\ &\quad - AK_k\mathrm{E}\left[\Delta_k\right]\mathrm{E}\left[u_k\right] \\ &= A\left(I - K_kH\right)\mathrm{E}\left[\tilde{x}_{k|k-1}\right].\end{aligned}$$

Hence, $\mathrm{E}[\tilde{x}_{1|0}] = 0$ implies that $\mathrm{E}[\tilde{x}_{k|k-1}] = 0$ for all $k > 0$. Thus,

$$\mathrm{E}\left[y_k\right] - y^* = H\mathrm{E}\left[\tilde{x}_{k|k-1}\right] + \mathrm{E}\left[\Delta_k\right]\mathrm{E}\left[u_k\right] = 0.$$

On the other hand, it holds that

$$\begin{aligned}&\mathrm{E}\left[\tilde{x}_{k+1|k}\tilde{x}_{k+1|k}^{\mathrm{T}}\right] \\ &= A\left(I - K_kH\right)\mathrm{E}\left[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}\right]\left(I - K_kH\right)^{\mathrm{T}}A^{\mathrm{T}} \\ &\quad + AK_k\mathrm{E}\left[\Delta_ku_ku_k^{\mathrm{T}}\Delta_k^{\mathrm{T}}\right]K_k^{\mathrm{T}}A^{\mathrm{T}} \\ &\quad + AK_kR_kK_k^{\mathrm{T}}A^{\mathrm{T}} + Q_k.\end{aligned} \tag{9}$$

$\mathrm{E}\left[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}\right] \leqslant P_{k|k-1}$ implies that

$$\begin{aligned}\mathrm{E}\left[\tilde{x}_{k+1|k}\tilde{x}_{k+1|k}^{\mathrm{T}}\right] &\leqslant \left(P_{k|k-1}^{-1} + H^{\mathrm{T}}W^{-1}H\right)^{-1} + Q \\ &= P_{k+1|k}.\end{aligned}$$

Thus, $\mathrm{E}[\tilde{x}_{1|0}\tilde{x}_{1|0}^{\mathrm{T}}] \leqslant P_{1|0}$ implies that $\mathrm{E}[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}] \leqslant P_{k|k-1}$ for all $k > 0$. According to the conclusions in [8], there exists a positive definite matrix Σ such that $P_{k|k-1} \leqslant \Sigma$ for all $k > 0$. Therefore, it holds that

$$\begin{aligned}\mathrm{E}\left[\|y_k - y^*\|^2\right] \leqslant & \|H\|^2\mathrm{trace}(\Sigma) + (U + \|u^*\|)^2 \\ & \times\mathrm{trace}(G) + \mathrm{trace}(R) \\ & \triangleq \sigma^2.\end{aligned} \tag{10}$$

Simulation. We compare our proposed R2R controller with the conventional EWMA controller. The model used in the simulation is as follows:

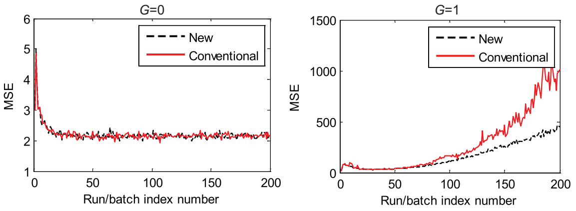
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B_e = \begin{bmatrix} 1 & 0.5 & 1.2 \\ 2 & 0.3 & 0.8 \end{bmatrix}.$$

B_k is a Gaussian process variable given bellow:

$$B_k \sim N(B_e, G \times I_3),$$

where I_3 is a three-dimensional identity matrix.

The model is divergent and Figure 1 shows the MSE of the output errors by new and conventional controller with and without Δ_k in (3). It shows that if there is no disturbance in real system ($B_k = B_e$), the two methods achieve similar and acceptable performance. However, this condition may never hold in real systems as they always undergo dynamic changes owing to environmental changes or disturbances. In such cases, if $\Delta_k \neq 0$ in (3) and $G = 1$, our proposed method achieves a much better performance compared to the conventional EWMA controller.



(a) (b)

Figure 1 (Color online) MSE of the output errors by new and conventional controller with (a) $\Delta_k = 0$, $G = 0$ and (b) $\Delta_k \neq 0$, $G = 1$.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61603303, 61803309, 61573365, 61833016), Natural Science Basic Research Program of Shaanxi (Grant No. 2017JQ6005), Fundamental Research Funds for the Central Universities (Grant No. G2017KY0008), and Innovation Development Foundation of Loving Students (Grant No. ASNIF2015-1502).

References

1 Sachs E, Hu A, Ingolfsson A. Run by run process control: combining SPC and feedback control. IEEE Trans Semicond Manufact, 1995, 8: 26–43
2 Tan F, Pan T H, Li Z M, et al. Survey on run-to-run control algorithms in high-mix semiconductor manufacturing processes. IEEE Trans Ind Inf, 2015, 11: 1435–1444
3 Butler S W, Stefani J A. Supervisory run-to-run control of polysilicon gate etch using in situ ellipsometry. IEEE Trans Semicond Manufact, 1994, 7: 193–201
4 Tseng S T, Tsung F, Liu P Y. Variable EWMA run-to-run controller for drifted processes. IIE Trans, 2007, 39: 291–301
5 Kim H, Park J H, Lee K S. A Kalman filter-based R2R control system with parallel stochastic disturbance models for semiconductor manufacturing processes. J Process Control, 2014, 24: 119–124
6 Wang Y, Hussein I I. Awareness coverage control over large-scale domains with intermittent communications. IEEE Trans Autom Control, 2010, 55: 1850–1859
7 Qin S J, Cherry G, Good R, et al. Semiconductor manufacturing process control and monitoring: a fab-wide framework. J Process Control, 2006, 16: 179–191
8 Anderson B D O, Moore J B. Detectability and stabilizability of time-varying discrete-time linear systems. SIAM J Control Optim, 1981, 19: 20–32