ISA Transactions 83 (2018) 107–125

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Research article

A survey of run-to-run control for batch processes

Kai Liu [a,b,c], YangQuan Chen [c,*], Tao Zhang [b], Siyuan Tian [b], Xi Zhang [a]

[a] School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing, 100083, China
[b] Lam Research Corporation, 4300 Cushing Pkwy, Fremont, CA 94538, USA
[c] School of Engineering, University of California, Merced, 5200 Lake Rd, Merced, CA 95340, USA

ARTICLE INFO

Article history:
Received 25 August 2017
Received in revised form 31 May 2018
Accepted 3 September 2018
Available online 21 September 2018

Keywords:
Run-to-run control
Survey
Batch process
Semiconductor manufacturing
Chemical mechanical planarization

ABSTRACT

Run-to-run (R2R) control is widely used in semiconductor manufacturing systems to minimize the process drift, shift and variability. The R2R controller adjusts control actions or recipes in a supervisory manner after each batch. This paper provides a comprehensive literature review of R2R control methods for the batch process. First, the principles of major R2R controllers are introduced and analyzed, such as exponentially weighted moving average (EWMA), double exponentially weighted moving average (d-EWMA), model predictive control (MPC), optimizing adaptive quality controller (OAQC), artificial neural network (ANN). Besides, simulation examples with different R2R controllers are made to compare the robustness and adaptability. Then, several case studies concerning a chemical mechanical planarization (CMP) process, a multi-input and multi-output (MIMO) system of the furnace process control and the management of blood glucose (BG) are presented. Finally, the paper concludes with some recommendations and directions for the future research.

1. Introduction

1.1. A background of R2R control

For many years, it has been suggested that automation with advanced process control (APC) will provide significant improvements in throughput, resulting in reductions in cost. As the size of wafer features become smaller and quality requirement becomes stricter year by year, the industry must innovate to maintain acceptable product yield, throughput, and overall equipment effectiveness (OEE). Statistical process control (SPC) is an industry-standard methodology for measuring and controlling quality during the manufacturing process. Although SPC is indeed a powerful technique for monitoring reducing variation in semiconductor manufacturing processes, it is limited by two major issues. First, the underlying assumption on which SPC is based is that the observations collected and plotted on control charts represent a random sample from a stable probability distribution. However, this assumption does not hold for many commonly encountered scenarios. Another limitation of SPC is that it is usually applied off-line. As a result, corrective actions suggested by SPC alarms typically occur too long after process shifts. One solution to this dilemma is run-to-run (R2R) control,[1] in which each "run" or batch could be a single wafer, several wafers, a lot, or any other grouping of semiconductor products undergoing the same set of process conditions. The R2R controller is adjusting control actions or recipes after each batch. The R2R batch process in semiconductor manufacturing is shown in Fig. 1.

In semiconductor manufacturing, each step is a complicated physiochemical batch process, and measurements are almost exclusively performed off-line, often slow, inconsistent, or skipped by operators. The basic philosophy of conventional semiconductor manufacturing process control is that if all settings that affect the process are set correctly, the machine will consistently produce a specified product. Using the conventional approach, manufacturing personnel act to relate machine settings to product characteristics. Therefore, in the past, semiconductor manufacturing process control was largely achieved by ensuring that process parameters were set on a machine controller according to machine-dependent recipes. The approach has yet to be fully realized, however, due to a number of factors, including variability in equipment performance, variability in incoming materials such as wafers and chemicals, increasingly complex processes, and a lack of adequate models relating process settings to product characteristics. To address significant unresolved problems, model-based R2R process control using sensor inputs, process models, and process control strategies to ensure that the process remains optimal for every batch, such as chemical mechanical planarization (CMP) process, chemical vapor deposition (CVD), lithography and polysilicon gate etching, rapid thermal processing (RTP), etc.

R2R control was first proposed by Sachs and his co-workers at in the beginning of 1990s [1,2]. Now, it is a critical component of the success of existing and next generation fabrication facilities [3].

* Corresponding author.
E-mail address: ychen53@ucmerced.edu (Y. Chen).

[1] Or run-by-run (RbR) control.

https://doi.org/10.1016/j.isatra.2018.09.005

List of abbreviations

ANN	Artificial Neural Networks
APC	Advanced Process Control
BG	Blood Glucose
CMP	Chemical Mechanical Planarization
CVD	Chemical Vapor Deposition
d-EWMA	Double Exponentially Weighted Moving Average
DOE	Design Of Experiment
EKF	Extended Kalman Filter
EWMA	Exponentially Weighted Moving Average
FDC	Fault Detection and Classification
ILC	Iterative Learning Control
LPCVD	Low Pressure Chemical Vapor Deposition
LRD	Long Range Dependence
LSE	Least Squared Error
MIMO	Multi-Input Multi-Output
MISO	Multi-Input Multi-Output
MPC	Model Predictive Control
MSE	Mean Squared Error
OAQC	Optimizing Adaptive Quality Controller
OEE	Overall Equipment Effectiveness
PCC	Predictor Corrector Control
RLS	Recursive Least Squares
RTP	Rapid Temperature Processing
SISO	Single Input Single Output
SPC	Statistical Process Control
ST	Self Tuning
T1DM	Type 1 Diabetes Mellitus
VM	Virtual Metrology

1.2. Basic structure of a R2R control system

APC has evolved rapidly in the semiconductor industry since 1990s, with R2R control emerging as the first technologically viable product of that evolution. One of the important elements of APC is R2R control, which is usually used to eliminate initial recipe bias, process shifts, and patterned disturbance [4,5]. Typically, R2R control techniques consist of two main steps: (i) Process modeling; (ii) Online model tuning and process control.

First, the process modeling usually assumes a linear regression model, which is built based on off-line experiments. It relates inputs variables u_k with output variables or response y_k. Second, the fitted model can be updated or refined with upcoming observed data and used to determine a control action, which is applied to maintain the process to the target value for the quality control.

It should be noted that R2R controllers do not directly regulate the process variables as used in semiconductor manufacturing. Instead, the control actions or recipes refer to setpoints of automatic PID controllers. The physical process variables are usually regulated by PID controllers within each run, such as temperature, flow, pressure, etc. However, after the measurement is taken, the specifications may change from batch to batch, due to the aging effects. In the wafer etching processes, for example, depletion of the etch solution or the degradation of thermocouples in high temperature furnaces can introduce trend or shift. Another example is the CVD process, deposition materials not only deposits on wafers but also on chambers, leading to strong autocorrelations of the next batches. In addition, the system may suffer from trend disturbances or a maintenance operation.

Therefore, the R2R controller, which involves both feedforward and feedback control actions, is acting as a supervisor over the automatic PID process in the inner loop [6]. The inner loop contains process variables in during each run and the sampling time in the inner loop is smaller than in the outer loop. New control actions or updated recipes are generated by the R2R controller in the outer loop after each run finished. Therefore, the control block diagram in Fig. 2 corresponds to internal model control (IMC) in control engineering [7–9].

Although there are some survey papers concerning on R2R batch process control, such as [10–13], some may be outdated and confined in the Fab[2] industry. In fact, R2R control techniques can be utilized in the biomedical field, which is a new research field in recent years. Therefore, after reviewing more than 200 relevant publications, a more comprehensive survey of R2R control algorithms for batch process control is evaluated in this paper. The purposes of this paper are: (i) To review literature on R2R control and encompass some novel methods in the past few years; (ii) To compare different R2R controllers and give some discussions; (iii) To include the case with the R2R batch process control in the glucose management; (iv) To propose some new aspects for the upcoming research directions on R2R batch process control.

This survey paper is organized as follows: Background and algorithms of R2R control are introduced in Sections 1 and 2. Section 3 evaluates different R2R controllers with simulation examples. Section 4 presents case studies on a CMP process, a MIMO system of the furnaces process and the blood glucose (BG) management. Topical publications on R2R control are categorized and several promising research directions are proposed in Section 5. Section 6 concludes the paper.

2. R2R control algorithms

The R2R control algorithm consists of 2 parts—modeling of the process (system) and control action calculation.

2.1. Process modeling

However, after the measurement is taken, the specifications may change from batch to batch, due to the aging effects. In the wafer etching processes, for example, depletion of the etch solution or the degradation of thermocouples in high temperature furnaces can introduce trend or shift. Another example is the CVD process, deposition materials not only deposits on wafers but also on chambers, leading to strong autocorrelations of the next batches. In addition, the system may suffer from trend disturbances or a maintenance operation. Therefore, the R2R batch process control is developed to serve as a "supervisor", indicating the proper adjustment of the automatic controllers for the process engineers.

Since both of the algorithms being described for the process are polynomial-based and linear, they can be represented using standard linear equation techniques:

$$y_k = \alpha + \beta u_k + \varepsilon_k \tag{1}$$

where $k = 1, 2, 3 \cdots$ denotes run or batch number, y_k is the system output, u_k is the control input (recipe). α, β denote intercept (shift) and slope (process gain) parameters of the regression process model respectively. $\{\varepsilon_t\}$ is the white noise sequence with variance σ^2, representing the measurement noise from the metrology devices. Note that u_k and y_k in Eq. (1) can be vectors if the system has multiple inputs and outputs, i.e. a MIMO system.

This dynamic model is then used to adjust the equipment settings to control the process outputs. In response to the need for

[2] Fab is the abbreviation for fabrication, typically refers to the semiconductor manufacturing.

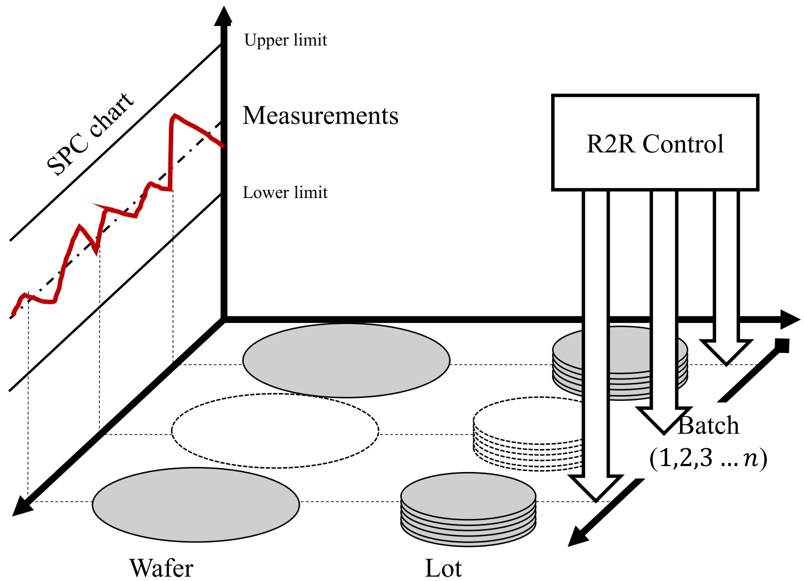

Fig. 1. R2R batch process in semiconductor manufacturing.

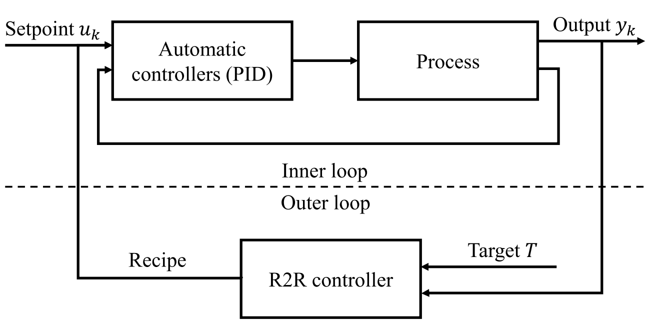

Fig. 2. Supervisory of R2R controller.

continuous process tuning solutions, R2R process control algorithms began to emerge in university and industrial research [14–19]. These methods were the first closed-loop feedback controllers to be used at the process level in the semiconductor industry.

2.2. Online model tuning and process control

2.2.1. Exponentially Weighted Moving Average (EWMA)

If α, β in model (1) are known, then the optimal control for system around a desired target T is:

$$u_k = \frac{T-\alpha}{\beta}. \tag{2}$$

Control actions are applied to the system prior the start of the run $k+1$. Mismatches between model and system are unavoidable in practical application, and system variations occur from R2R sometimes. To account for dynamic effects, R2R controller is using predicted response $\hat{y}_k = a_{k-1} + bu_{k-1}$ and $\hat{a}_{k+1} = a_k = \alpha, b = \beta$ at each run k. It regulates the process with time-varying intercept by adjusting α in Eq. (1) with the following equation:

$$u_k = \frac{T-a_k}{b}. \tag{3}$$

Obviously, the estimate of the slope is off-line while the estimate of the intercept is computed recursively updates a_k on a R2R basis:

$$a_k = \lambda(y_{k-1} - bu_{k-1}) + (1-\lambda)a_{k-1}, \tag{4}$$

where a_k is an exponentially weighted average of the historic deviations from the first run to the kth run and $0 < \lambda < 1$ is an adjustable tuning parameter (weight) for the EWMA filter [20]. The larger the forgetting factor λ, more recent measurements are weighted in the EWMA controller. Therefore, this EWMA "gradual model" R2R controller is able to track and compensate for gradual changes in the process as well as filter out random walk noise in the process [5].

The block diagram of EWMA controller is illustrated in Fig. 3. The prediction error $T - a_{k-1}$ is fed back to the controller after passing through the EWMA filter, which can be interpreted as a first-order filter for disturbance prediction. The discrete time unit delay z^{-1} is used in the figure.[3] Especially, if the noise ε_k follows an ARMA(1,1) model in Eq. (1), EWMA controller provides minimum variance control.

The R2R controller applies an EWMA filter to smooth the control action on a linear process and tries to keep the responses y_k at the optimal performance in the presence of process noise and drift. The algorithms are simple and straightforward therefore it is a natural starting point for an R2R control strategy [18]. It has been shown to provide good results in a number of applications [3,21–24].

From Eqs. (3) and (4), however, we could see that EWMA controller cannot compensate for a second order system. Besides, the type of disturbances should also be considered in the R2R batch process control. Therefore, Sachs et al. [2] recommend to use EWMA controller if the process does not suffer from a severe drift or a large shift.

2.2.2. Double exponentially weighted moving average (d-EWMA)

As discussed above, EWMA based R2R controller cannot regulate the process subject to a deterministic drift. Thus, a new R2R controller double EWMA filters are employed by Butler [14]. As shown in Fig. 4, the EWMA filter 1 is used to estimate the true output as the single-EWMA does and EWMA filter 2 is to estimate the trend d. This modification of the single EWMA controller makes it possible to compensate for the lag in target tracking when a process in undergoing a drift.

If the fitted process model is $y_k = a_{k-1} + bu_{k-1} + D_{k-1}$, the control action is calculated as follows:

$$u_k = \frac{T - a_k - D_k}{b}, \tag{5}$$

[3] Some papers refer it to the backshift operator B, i.e. $By_k = y_{k-1}$.

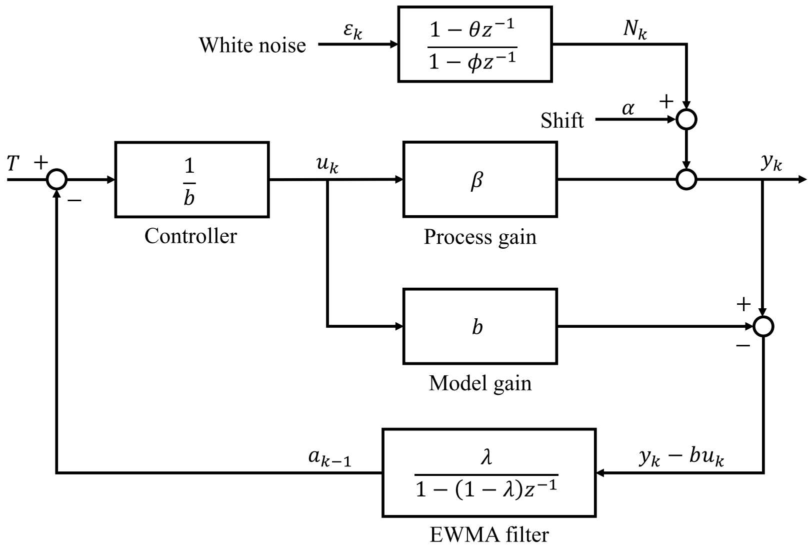

Fig. 3. Block diagram of EWMA based R2R control.

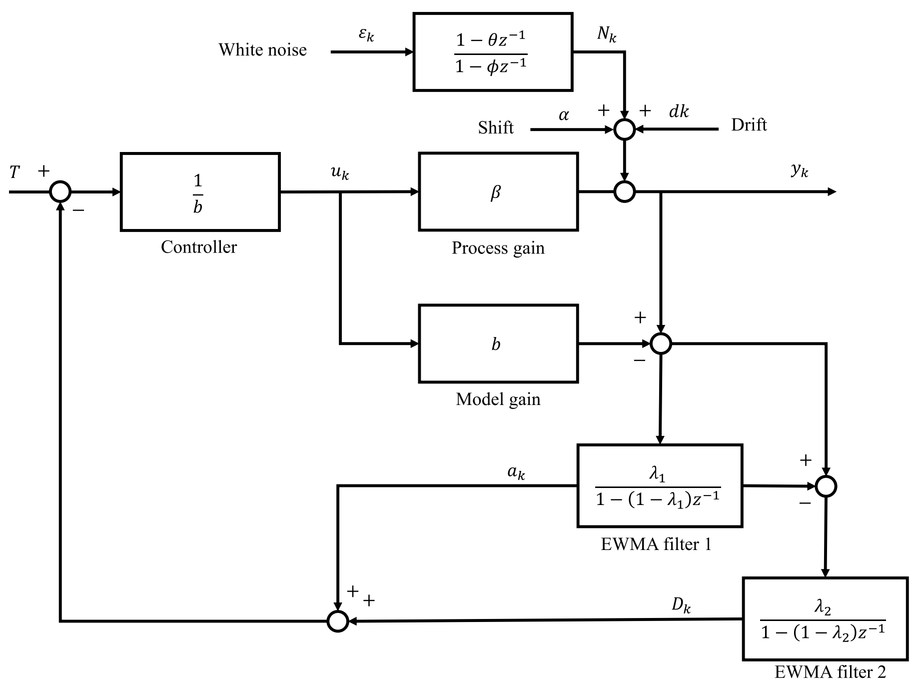

Fig. 4. Block diagram of d-EWMA based R2R control.

where the intercept a_k and trend D_k is given by two (double) EWMA filters:

$$a_k = \lambda_1(y_k - bu_{k-1}) + (1-\lambda_1)(a_{k-1} + D_{k-1}), \tag{6}$$

$$D_k = \lambda_2(y_k - bu_{k-1} - a_{k-1}) + (1-\lambda_2)D_{k-1}. \tag{7}$$

In the above d-EWMA scheme the sum of the updated intercept and trend components $(a_k + D_k)$ that is fed back to the controller. It should be also noted that Butler first proposed a PCC [21] scheme prior to d-EWMA [14] in the following algorithm:

$$a_k = \lambda_1(y_k - bu_{k-1}) + (1-\lambda_1)a_{k-1}, \tag{8}$$

$$D_k = \lambda_2(y_k - bu_{k-1} - a_{k-1}) + (1-\lambda_2)D_{k-1}. \tag{9}$$

From Eqs. (6) and (8), we could see the only difference between PCC and d-EWMA is the estimate of intercept a_k. Therefore, d-EWMA has a more straightforward asymptotic behavior compared with PCC [25]. With appropriate choice of λ_1, λ_2, d-EWMA controller can remove the drift completely from the process [14]. Due to the similarity between PCC and d-EWMA, only d-EWMA is discussed in the rest of paper.

2.2.3. *Model Predictive Control (MPC)*

Model Predictive Control (MPC) calculates a manipulated variable profile by using a model to optimize an open-loop performance objective over a prediction horizon. It has been successfully used in the industrial applications such as chemicals, refining, petrochemical, metallurgy, etc. [26].

In 1997, Mulins et al. first proposed MPC based R2R control for the CMP process [27]. The linear model predictive control (LMPC) algorithm employs a standard state space formulation for the process model.

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned} \tag{10}$$

where x_k is the process state variables at kth run, u_k is the process input (control signal) and y_k is the output (measurement). A, B, C and D are state matrix, input matrix, output matrix, feedforward matrix, respectively.

The control profile u_k is calculated based on the minimization of the objective function:

$$J = \min_{u^N} \sum_{j=0}^{\infty} (y_{k+j}^T Q y_{k+j} + u_{k+j}^T R u_{k+j} + \Delta u_{k+j}^T S \Delta u_{k+j}^T) \tag{11}$$

where N is the prediction horizon,[4] Q and S are symmetric, positive semi-definite weighting matrices, and R is a symmetric, positive definite weighting matrix. These four parameters are used to specify the response of the optimal controller [12].

The quadratic program of the Eq. (10) should be constrained by following inequality equations:

$$\begin{aligned} u_{\min} &\le u_{j+k} \le u_{\max} \\ y_{\min} &\le y_{j+k} \le y_{\max} \\ \Delta u_{\min} &\le \Delta u_{j+k} \le \Delta u_{\max} \end{aligned} \tag{12}$$

At the run k, the quadratic program of the solution u_k is achieved, thereafter, u_k is input into the process or system.

One key aspect of the use of MPC in a R2R control is application of dynamic systems, control of the MIMO system, ability to handle the rate of change of inputs. Another benefit is MPC can differentiate process and metrology noise, since process model is presented in the process state and output observation terms, while metrology noise can be characterized by the covariance matrix in the observer [26].

2.2.4. Optimizing Adaptive Quality Controller (OAQC)

In 1996, Castillo et al. proposed a multivariate self-tuning (ST) controller for run-to-run process using recursive least squares (RLS) algorithm to obtain on-line parameter estimations [28], and then they proposed an optimizing adaptive quality control (OAQC) to seek and maintain optimum operating conditions for a MIMO nonlinear quadratic process in 1998 [29]. The OAQC controller provides the desirable estimation of the nonlinear regression model such that the best control process is attained. Hence, the OAQC performs as an "optimizer" as well as a "controller". Moreover, because of recursive parameter estimation, the OAQC accounts for process nonlinearity, and the accuracy of the initial state of the control model need not be satisfied.

The OAQC consists of following features:

1. An on-line multivariate RLS model-fitting for identification of controller parameters is performed;
2. An experimental design is defined based on some optimality criteria;
3. A recipe-finding optimization step is performed.

Instead of solving these three problems in series, the OAQC attempts to solve them simultaneously in an incremental way at each run, speeding up the qualification phase of a process and reducing extreme disruptions to the process while the OAQC learning algorithms are running. An SPC chart is used either as a deadband to reduce noisy control moves, or simply as a process monitor to detect an out-of-control state while allowing the calculation of control moves at each run. Fig. 5 shows the control diagram of OAQC controller with the SPC chart.

The OAQC implements a second-order MIMO Hammerstein transfer model to approximate the process with nonlinearities as the following form:

$$y_k = y(0) + Nz_{k-1} + Mt + (I_p - C\mathcal{B})\varepsilon_k \tag{13}$$

where $y(0)$ is a vector $p \times 1$ intercepts or offsets; $z_k' = (u_k, u_k^2, u_k^{(i)} u_k^{(j)}$ $(i < j))$ is a $2n + n(n-1)/2$ vector, which contains quadratic expansion of input terms u_k; t denotes a vector containing run index k in p components; $\mathcal{B}$ is the backshift operator; ε_k is a sequence of white noise.

The one-step-ahead minimum mean square error (MSE) prediction of Eq. (13) yields to the expression:

$$\hat{y}_k = Ly_k + M(t+1) + Nz_k \tag{14}$$

where N is a $(p \times 2n + n(n-1)/2)$ vector contains parameters for both linear and quadratic terms.

Suppose the linear model to be estimated is written as follows:

$$\hat{y}_k = \hat{\boldsymbol{\theta}}' \phi_k \tag{15}$$

where $\hat{\boldsymbol{\theta}}' = [\hat{L}|\hat{N}|\hat{M}]$ is a $p \times (2n + 2p + n(n-1)/2)$ matrix of parameters, and $\boldsymbol{\phi}_k = (y_{k-1}|z_{k-1}|t)$ is a $1 \times (2n+2p)+(n(n-1)/2)$ vector of regressors. The estimation of $\hat{\boldsymbol{\theta}}$ is proceeded row by row.

Step 1: Compute the gain vector $(dim \times 1)$

$$\boldsymbol{K}_k = \frac{\boldsymbol{P}_{k-1}\phi_k}{\lambda + \phi_k' \boldsymbol{P}_{k-1}\phi_t}. \tag{16}$$

Step 2: Compute the one step ahead forecast errors (scalar)

$$e_k = y_k - \hat{\boldsymbol{\theta}}_k' \phi_k \tag{17}$$

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mathbf{K}_k e_k, \tag{18}$$

Step 3: Compute the precision matrix $(dim \times dim)$ proportional to the covariance matrix of the vectors of parameter estimates

$$\mathbf{P}_k = \frac{[\mathbf{I}_{dim} - \mathbf{K}_k \phi_k']\mathbf{P}_{k-1}}{\lambda} + \mathbf{R}_k, \tag{19}$$

$$\mathbf{R}_t = \frac{\mathbf{I}_{dim}\mathbf{K}_k'\mathbf{P}_{k-1}\phi_k}{dim}, \tag{20}$$

where $dim = 2(p+n) + n(n-1)/2$.

Note that Eq. (18) has an EWMA form with weights $\boldsymbol{K}$ depending on the variances of estimates $\boldsymbol{P}$. A discounting factor λ is useful for better transient performance. If $\lambda = 1$ it is the standard RLS algorithm [30].

It is desired to keep y as close as possible to the target T. For each run k, OAQC solves the following problem:

$$J = \min(\hat{y}_k - T)' W (\hat{y}_k - T) + (u_k - u_{k-1})' \Gamma (u_k - u_{k-1}) \tag{21}$$

where W is a diagonal weighting matrix with entries ω_i, denoting relative priority of each response, and Γ is a diagonal matrix of cost coefficient γ_i, denoting the relation to its previous setting. Optimization of J with respect to the u_k is carried out using a mixed penalty-barrier method, which inputs and outputs should stay within bounds. Castillo [28] provides the optimal control law that minimizes objective function in Eq. (21) as follows:

$$u_k^* = [(\hat{N}' W \hat{N} + \Gamma)^{-1} (\hat{N} W)(T - \hat{M}(t+1) - \hat{L} y_k) + \Gamma u_{k-1}]. \tag{22}$$

Therefore, u_k^* is fed to the next recipe if it satisfies the multivariate SPC chart.

[4] N usually equals to 1 in R2R batch control, i.e., one-step-ahead.

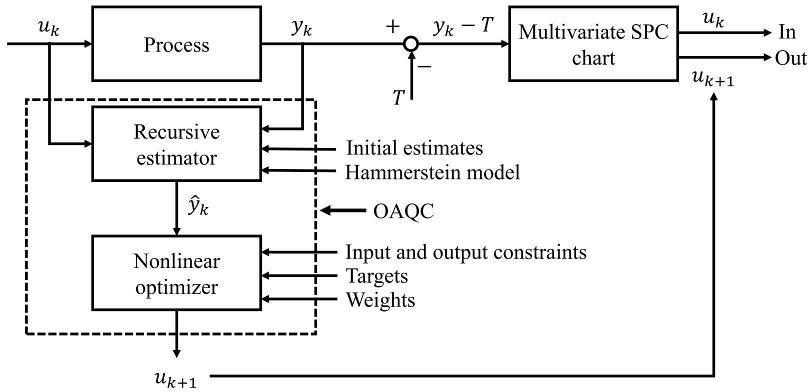

Fig. 5. Block diagram of OAQC-based R2R control.

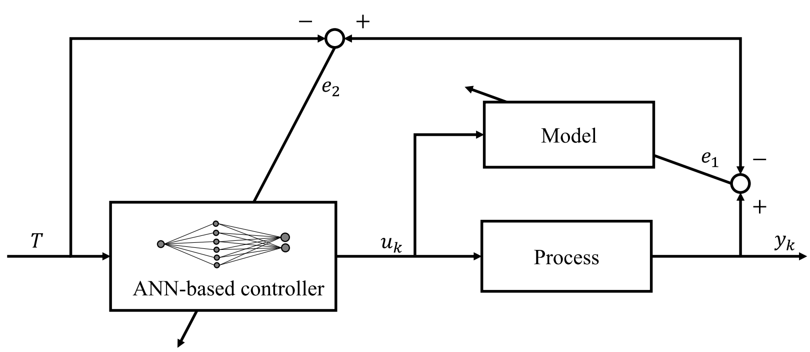

Fig. 6. Block diagram of ANN-based R2R control.

2.2.5. Artificial Neural Network (ANN)

In semiconductor manufacturing, virtual metrology (VM), which aims to predict properties of a wafer based on values using sensor data from production equipment and physical metrology values of preceding samples. It provides metrology data for every wafer, it can be leveraged to provide factory-wide R2R control at every process run. The data obtained for every process can then be used in a feed-forward and feedback control scheme to provide R2R control for every wafer. In 1996, the artificial neural network (ANN) based R2R controller was applied in CVD by Wang et al. [31]. The block diagram of ANN-based R2R control at the wafer level for an individual process is pictured schematically in Fig. 6.

The procedure of the ANN-based control scheme consists of 5 steps:

1. A set of input–output data is required to build the process.
2. An ANN-based controller is trained off-line after some initial settings of select network structure, search algorithms, constraints, etc.
3. The trained controller is used to tune the kth run of the process.
4. Update the process model by the upcoming data at $(k+1)$th run.
5. Calculate and updated new weights in the controller and go to Step 3.

ANN is one of the widely used non-parametric models for both classification and regression.

$$\hat{y}_k = \mathcal{F}(\mathbf{u}_k, \Theta_k), \tag{23}$$

where $\hat{y}_k$ is the estimate of the response; $\mathbf{u}_k = [u_1, u_2, u_3, \ldots, u_n]_k$ is the vector of process tuning variables; Θ_k denotes layer weights from input nodes to output nodes.

For each neuron, the activation function $\mathcal{F}$ can be a linear, hyperbolic tangent, or logistic sigmoid function. Suppose that there are M_1 input nodes, M_2 hidden nodes, the output of the network in Eq. (23) can be expressed with input variables and weights as follows:

$$\hat{y}_k = \sum_{j=1}^{M_2} W_j^{(2)} \mathcal{G}\Big(\sum_{i=1}^{M_1} W_{ji}^{(1)} \mathbf{u}_k\Big), \tag{24}$$

where $\mathcal{G}$ denotes an activation function from hidden layer to input node. $W_j^{(2)}$ and $W_{ji}^{(1)}$ represent the weight from the output node to jth hidden node and the weight from the jth hidden node to the ith input node, respectively.

Generally, before the start of the neural network training, the input and output data should be scaled in [-1, 1] to avoid the saturation functions before training of network.

Neural networks optimize the weights Θ_k by minimizing the objective loss function $E_k^{(1)}$ at kth run, for which the least squared error (LSE) between the targets and the network results is generally used:

$$E_k^{(1)}(\Theta_k) = \frac{1}{2}\sum_{k=1}^{M}(y_k - \hat{y}_k)^2, \tag{25}$$

where y_k and $\hat{y}_k$ are the target and prediction value of kth test sample, respectively.

$$\begin{aligned} E_k^{(1)}(\Theta_k) &= \frac{1}{2}\sum_{k=1}^{M}(y_k - \hat{y}_k)^2 \\ &= \frac{1}{2}\sum_{k=1}^{M}(y_k - \mathcal{F}(\mathbf{u}_k, \Theta_k))^2 \\ &= \frac{1}{2}\|\epsilon_k\|^2 \end{aligned} \tag{26}$$

where $\epsilon_k = y_k - \hat{y}_k$ is the estimate error evaluated at kth run.

$$E_k^{(1)}(\Theta_k + \Delta\Theta_k) \approx E_k^{(1)} + J^T\epsilon_k\Delta\Theta_k + \frac{1}{2}\Delta\Theta_k^T J^T J \Delta\Theta_k, \tag{27}$$

where $J = \partial E_k^{(1)}(\Theta_k + \Delta\Theta_k)/\partial\Delta\Theta_k$ is the Jacobian matrix for a small $\Delta\Theta_k$ if using the first order Taylor expansion.

Letting $J = 0$, we can obtain the adaptation of neural network weights $\Delta\Theta_k$ as

$$\Delta\Theta_k = -(J^T J)^{-1} J^T \epsilon^k. \tag{28}$$

The above iteration algorithm is called Gauss–Newton method. Yi et al. modified it by using Levenberg–Marquardt algorithm to avoid the large step size [32]. Similarly, we can obtain the updating laws of weights Θ_k for the neural network controller by minimizing $E_k^{(2)}$:

$$E_k^{(2)}(\Theta_k) = \frac{1}{2}\sum_{i=1}^{M}(T - \hat{y}_k)^2, \tag{29}$$

The weights Θ_k can affect the process output via the control signal $\mathbf{u}_k$. The error gradient can be computed using the chain rule as:

$$\frac{\partial E_k^{(2)}}{\partial\Theta_k} = \sum_{i=1}^{M}\frac{\partial E_k^{(2)}}{\partial\hat{y}_k}\frac{\partial\hat{y}_k}{\partial\mathbf{u}_k}\frac{\partial\mathbf{u}_k}{\partial\Theta_k}. \tag{30}$$

For a more detailed description, readers could refer to the literatures in [32–34].

In a nutshell, both OAQC and ANN can be considered as a self-tuning control scheme, which includes: building the process model, extracting the controller parameters, detecting the process shifts/drifts, determining the control action, and optimizing the controller.

3. Simulation examples

3.1. Example 1: R2R batch process control with/without EWMA controller

Consider a more general SISO model of Eq. (1) for the discrete part manufacturing of the batch process:

$$(1 - \omega z^{-1})y_k = \alpha + \beta u_{k-1} + N_k, \tag{31}$$

where the noise follows an ARMA(1,1) process is defined as:

$$N_k = \frac{1 - \theta z^{-1}}{1 - \phi z^{-1}}\varepsilon_k, \tag{32}$$

where ω represents the autocorrelation of the observed responses. The larger ω, the stronger the coupling effects. Clearly, if $d = \theta = \phi = \omega = 0$, the corresponding process in Eq. (31) will be reduced to Eq. (1).

In order to show the effectiveness of the EWMA controller, we run a simple simulation of Eq. (31) with $T = 10, \alpha = 2, \beta = b = 2, \theta = 0.7, \phi = 1.0, \omega = 0.1$. We set the tuning forgetting factor $\lambda = 0.1$ and $\lambda = 0.9$ to see EWMA controller how EWMA controller regulate the process. From first 20 runs in Fig. 7, it demonstrates that larger λ gives more aggressive response. Besides, the process is in completely off-target ($T = 10$) without the control action.

3.2. Example 2: A SISO process with different R2R controllers under a deterministic drift

Consider a process with a equipment aging drift by adding a trend d:

$$(1 - \omega z^{-1})y_k = \alpha + \beta u_{k-1} + dk + N_k \tag{33}$$

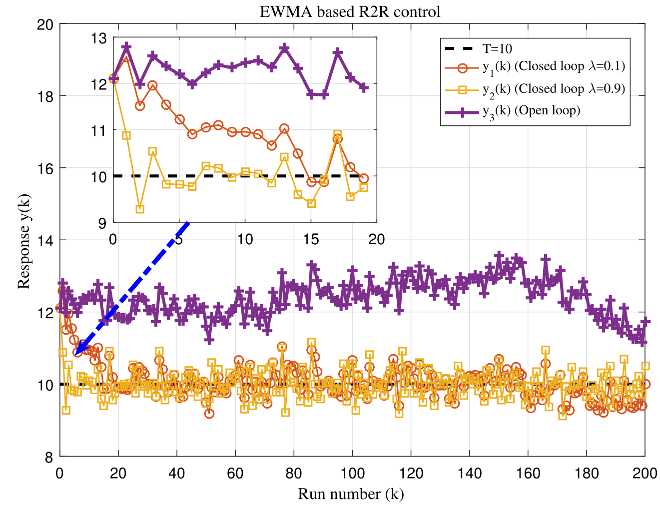

Fig. 7. EWMA controller with $\lambda = 0.1$ and $\lambda = 0.9$ compared with open loop response.

where $\omega = 0.1, \alpha = 2, \beta = b = 2, d = 1, \theta = 0.5, \phi = 1, T = 10$.

We set the same parameters in this example to compare the results in Eq. (31), using one forgetting factor $\lambda = 0.1$ for the single-EWMA controller, and using two forgetting coefficients $\lambda_1 = 0.1$ and $\lambda_2 = 0.5$ for the double-EWMA controller.

ANN-based controller with BP training algorithm learning rate $\alpha = 0.05$, one hidden layer with 5 nodes.

In addition, the trend d is changed from 0.5 to 0.1 at the run 100 to test the robustness of different R2R controllers. If a process is subject to a deterministic drift due to equipment aging, changes in the drift rate may lead to potential overshoot based on the time averaging of the d-EWMA drift estimator.

Control actions are applied to compensate the system's trend from run to run. From Fig. 8, the slopes of control actions of d-EWMA, ANN indicating the drift compensation, are correctly calculated at approximately 0.5 and 0.1. However, EWMA is somehow failed to track the target due to the inability of compensation for this drift.

Therefore, the d-EWMA R2R controller can compensate for the deterministic trend of the system with the best performance. Simulation conclusions are similar to literature [11] and [21].

3.3. Example 3: A MIMO process with a MPC controller under a deterministic drift

There are a large variety of multivariate process with nonlinearities. In this section, a quadratic process with a linear drift originated from the CMP process are used as an example.

The MIMO process is built as the state space model in Eq. (10):

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} B = \begin{bmatrix} 50.18 & 13.67 \\ -6.65 & 19.95 \\ 163.40 & 27.52 \\ 8.45 & 5.2500 \end{bmatrix}$$

$$C = \begin{bmatrix} -0.01 & 0.01 & 0 & 0 \\ 0 & 0 & 0 & -0.01 \end{bmatrix} D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We set the prediction horizon $p = 10$ and moving horizon $m = 3$. Targets are $T_1 = 1700$ for removal rate and $T_2 = 150$ for uniformity (see Fig. 9).

MPC-based controller keeps a good control of two output removal rate y_1 and uniformity y_2 within targets simultaneously. We have got similar evaluation results with [10].

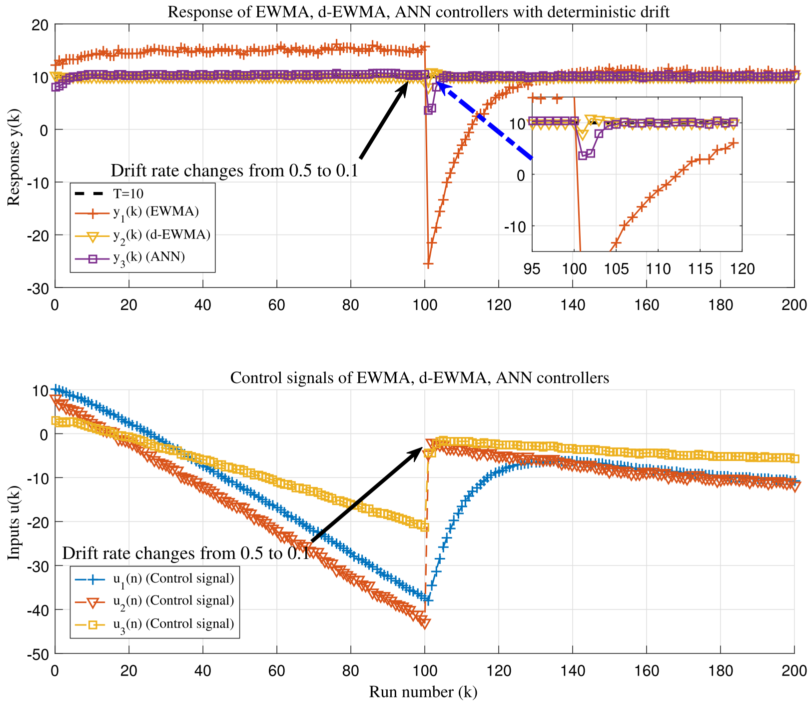

Fig. 8. Comparison of EWMA, d-EWMA, ANN controllers with deterministic drift.

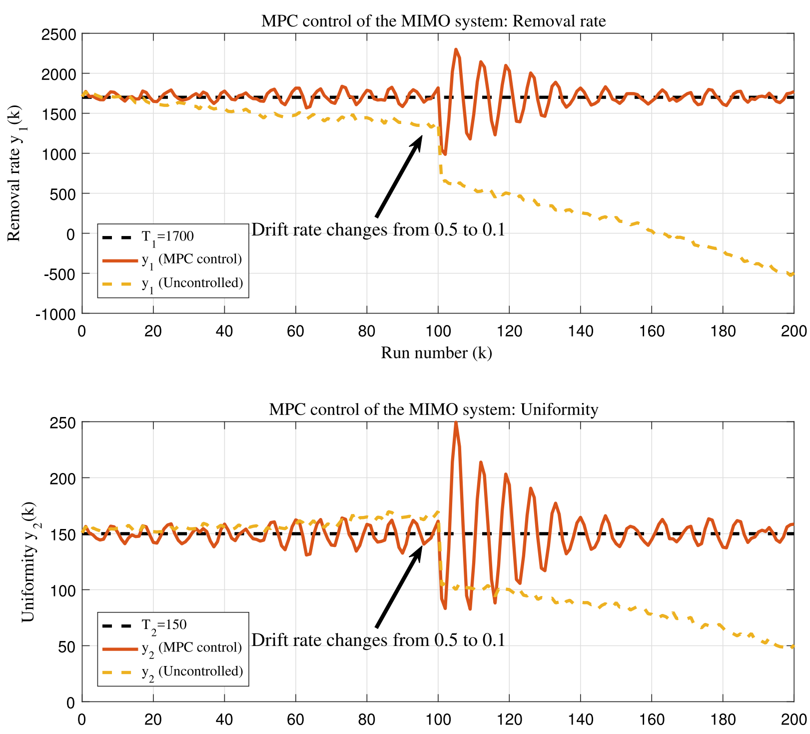

Fig. 9. MPC control simulation of 4×2 MIMO process.

4. R2R case studies

4.1. R2R process control for CMP in semiconductor manufacturing

CMP is the planarization method that has been selected by the semiconductor industry today [35]. In 1989, CMP method was developed by Davari et al. for global planarization of the interlayer dielectrics [36]. In the early 1990s, IBM successfully used the CMP process to deal with the slow dry etching problem of copper [37]. In addition, CMP has been integrated with the dual-damascene method to simplify the fabrication steps of the conducting wires in between the consecutive layers of a multilayer integrated circuit [37–39]. As the wafer size increases to 300 mm or more, equipment and process design improvements have been invoked to address these issues. Furthermore, since high-performance products are badly required in the fierce semiconductor market, with an

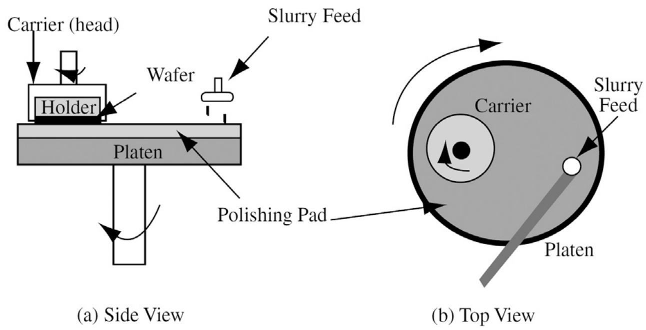

Fig. 10. CMP configurations.

increasing number of manufacturers, the quality of semiconductor products is gaining more significance.

The first R2R uniformity control of CMP was reported on a Strasbaugh CMP tool, which showed significant improvements in process capability [40]. It can be achieved through multivariate control of a CMP oxide process, such as thickness and uniformity etc. The observed drifts in CMP processes, and the availability of post-process measurements, motivate the use of a R2R strategy. A generic semiconductor control system framework has been under development, and is applied to the CMP control problem [41–46].

An illustrative of CMP process is shown in Fig. 10. The wafer is affixed to a wafer carrier via backside air and pressed face-down on a rotating platen holding a polishing pad. A slurry with abrasive material—alkaline slurry of colloidal silica for oxide or silicon polishing for example, is dripped onto the rotating platen during polishing. The slurry chemically attacks the wafer surface, converting the silicon top layer to a hydroxylated form, which is can be easily removed by the mechanical planarization process. The main difficulty exists in achieving a reliable film thickness since the removal rate is always changing over time in order to guarantee the within-wafer uniformity of the polish. Besides, due to non-constant relative pad velocity from the edge to the center, polish rates vary from the center to the edge of the wafer. Sporadic slurry and by-products are moving under the wafer, wafer bowing as a result of pressure, or machine drift in time of any of these parameters. Regarding the recipe parameters, typically, there are three primary parameters in a CMP recipe i.e. the down force, the platen rotation speed, and the carrier rotation speed while another variable in CMP recipe is the back pressure. Basically, if the non-uniformity problem is identified to be a center-slow-edge-fast process, back pressure can be used to push the back of a wafer thus accelerating the center polish rate thereafter the uniformity can be improved accordingly [47].

Performances of the EWMA, d-EWMA, OAQC R2R control algorithms are compared through simulation using the benchmark process model (plant) described in [10]

$$y[k] = C + f(u[k]) + \varepsilon_k + d_k, \tag{34}$$

where d_k is a linear drift with constant drift speed $d_k = \begin{bmatrix} 17 & 1.5 \end{bmatrix}'$ and ε_k is a normally distributed white noise with mean zero and covariance.

$$\Lambda = \begin{bmatrix} 665.64 & 0 \\ 0 & 5.29 \end{bmatrix}, \tag{35}$$

$f(u[k])$ is a full second-order polynomial function of the inputs with the following form:

$$f(u[k]) = \sum_{i=0}^{3} \sum_{j=0}^{3} \beta(i,j)\, u(i)\, u(j), \tag{36}$$

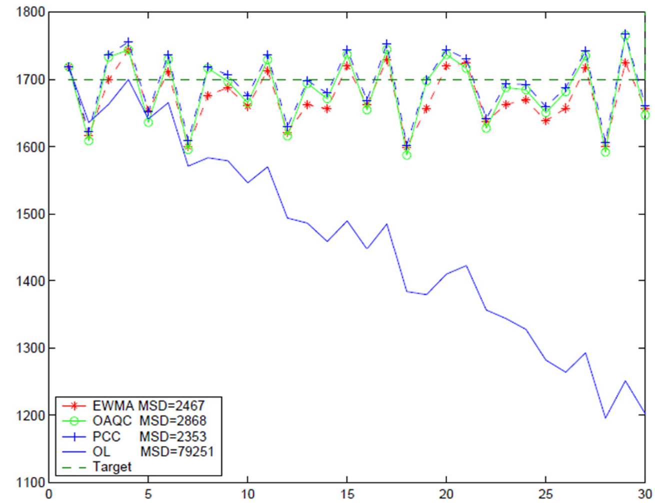

Fig. 11. Removal rate comparison of three R2R control algorithms [5,10].

Table 1
Comparison of results using PCC, EWMA and OAQC.

MSE	Removal rate	Non-uniformity
PCC	2353	16
EWMA	2467	18
OAQC	2868	19

where

$$\beta = \left[\begin{matrix} 1386.5 & 381.02 & 112.19 & 3778.8 & 21.301 \\ 1520.8 & 2365.6 & 2923.5 & 281.66 & 3.9419 \end{matrix}\right.$$

$$\begin{matrix} 8.7159 & 24.953 & 37.082 & 17.642 & 11.974 \\ 1.0754 & 1.406 & 0.33797 & 72.274 & 94.222 \end{matrix}$$

$$\left.\begin{matrix} 164.99 & 28.150 & 249.17 & 0.025067 \\ 26.175 & 13.505 & 36.691 & 32.929 \end{matrix}\right].$$

The simulation model for the EWMA and PCC controllers have the form given in Eq. (1) to Eq. (9), in which

$$A = \begin{bmatrix} 5.018 & -0.665 & 16.34 & 0.845 \\ 13.67 & 19.95 & 27.52 & 5.25 \end{bmatrix}, C = \begin{bmatrix} -138.21 \\ -627.32 \end{bmatrix},$$

and with the outputs: removal rate and within wafer non-uniformity. The target removal rate T_1 and the target non-uniformity T_2 are assumed to be 1800 and 300, respectively. Initial non-uniformity is set to 150. Simulation of the process is done using the three control schemes. The EWMA controller is simulated $\lambda = 0.6$ and the PCC with $\lambda_1 = 0.6$ and $\lambda_2 = 0.3$. Removal rate and non-uniformity are shown in Figs. 11 and 12, respectively.

The results obtained with the three control schemes are compared quantitatively by calculating the mean squared error (MSE) between the response and the target for each run. It is obvious from Fig. 11 that all three algorithms maintain good control of the simulated CMP process with linear drift and normally distributed white noise and process drift is well compensated for keeping the removal rate around the target value.

The mean square error (MSE) for both removal rate and non-uniformity is summarized in Table 1. The weighting factors used in the PCC algorithm are fixed parameters. Smith et al. used a self-tuning EWMA controller with the help of ANN function approximation, which dynamically updates the controller parameters to find the appropriate choice of these factors [48,49]. However such strategy needs significantly large amount of training data to find out the functional mapping between the disturbance state of the process and the corresponding weighting factors of EWMA.

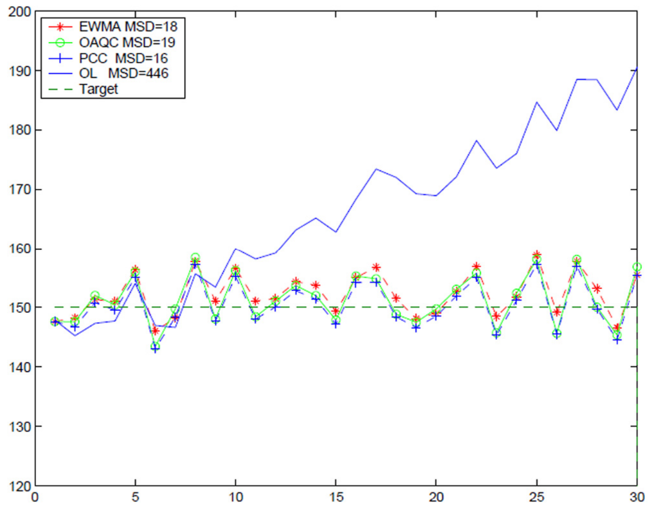

Fig. 12. Non-uniformity comparison of three R2R control algorithms [5,10].

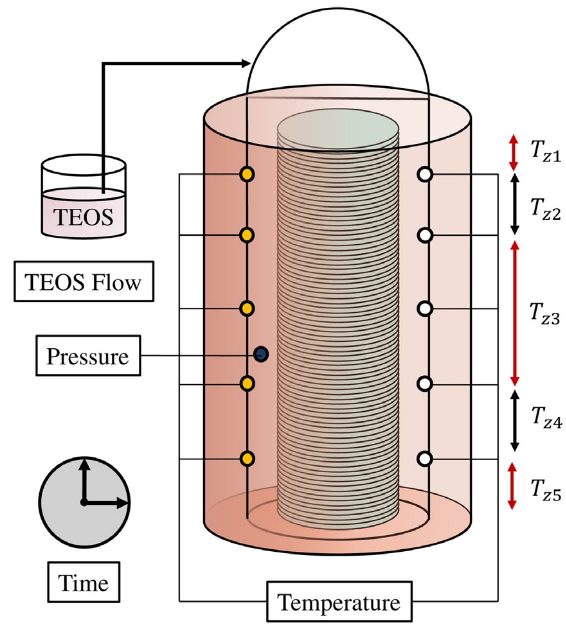

Fig. 13. Deposition parameters: Example of TEOS.

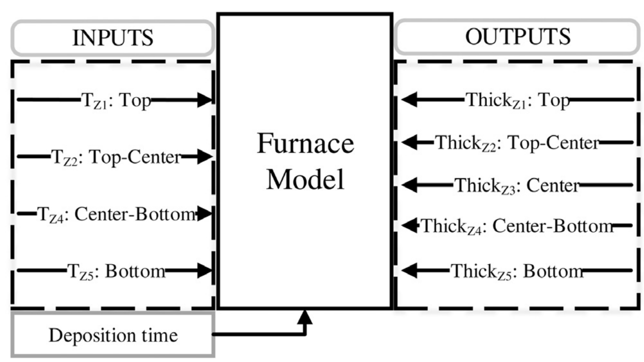

Fig. 14. MIMO model of furnaces process.

4.2. Implementation of MIMO R2R control algorithm on furnace process

As we know, tuning a process recipe is an arduous work for the MIMO system. Usually, the specifications of the outputs are interacted (coupled) with input parameters. In addition, the whole process system is subjected to the equipment drifts within preventive maintenance cycle and aging effects, which needs to be addressed on an ad-hoc basis by closely monitoring process and equipment SPC charts. This human dependency adds some potential errors thus resulting in serious loss of engineering productivity. Therefore, the R2R controller for MIMO furnace process can minimize the process deviation and simultaneously provide adequate level of automation to help human tasks.

As shown in Fig. 13, a vertical furnace has two families of vertical furnaces: low pressure chemical vapor deposition (LPCVD) and atmospheric furnaces [50,51]. The first study and work on LPCVD furnace processes R2R control have been done in ST-Microelectronics Crolles site [52]. A LPCVD recipe is defined by the pressure, the gas flows, the deposition time and the temperature. Once defined for a recipe those three parameters (pressure, gas flow and temperature) cannot be changed. The deposition time will adjust the total thickness deposited, and the 4 left temperatures zones will help to realize a good uniformity.

In Fig. 13, five heating zones are presented. The largest is the center zone which can never be changed. When the process gets out of control, the recipe parameters must be tuned on the LPCVD furnaces. Under this condition, the feedback R2R loop model will be a MIMO one: Multi-input (the deposition time and 4 temperature zones T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) which need to be tuned to reach multi-output (range and thickness specifications) in Fig. 14. T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5} should be adjusted by technicians to center the process. But with this adjustment method, the accuracy and uniformity of the process cannot be guaranteed necessarily for the increasing requirements. Automatic calculations should be the goal to control and adjust the temperatures. The previous run must help to center the next one. Therefore, a feedback R2R loop is required to reach this goal. This is a better way to regulate the process because such a loop is based on measurement from the process k in order to regulate the process $k+1$. Such studies can be find in literature [50,53].

Once the measurements were done, an analysis of variance was performed in order to extract the influent parameters for each zone. It should be noticed that the interactions are all first order. The expression of the mean thicknesses ($Thick_{Zn}$) for the 5 different zones of the furnace for a given deposition time can be written as:

$$Thick_{Z_{1,2,3,4,5}} = C_k + a_k \cdot T_{Z_1} + b_k \cdot T_{Z_2} + c_k \cdot T_{Z_4} + d_k \cdot T_{Z_5} \tag{37}$$

where C_k is a constant; T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}, the zone temperatures 1, 3, 4 and 5; a_k, b_k, c_k, d_k, the normalized coefficients of the model. A first description model has been built for TEOS and nitride processes. The constants and model coefficients can be found in [52]. Once the predictive model is extracted, tried and tested, R2R dashboard has been built.

Once the model is done, the key challenge of this project besides the production downtime on a furnace to perform DOE is the metrology tools automation development. The principle is as follows: the automation gets the setting from ProcessWorks, which means the model prediction (setting calculation and thickness prediction) from the previous run. These settings, time and temperatures, are send to the furnace via the automation. Then a batch is loaded and processed with these settings. At the end of the process, the 5 test wafers are shipped to the metrology tool (ellipsometer) and thicknesses $Thick_{Z_1 \sim Z_5}$ are measured. The values are send to the automation. The saved data are fed to the thickness feedback model. Then it is possible to have a model coefficient recalculation, the results will feed the model prediction for the next run. The principle of the optimizer is based on an iterative method. To go

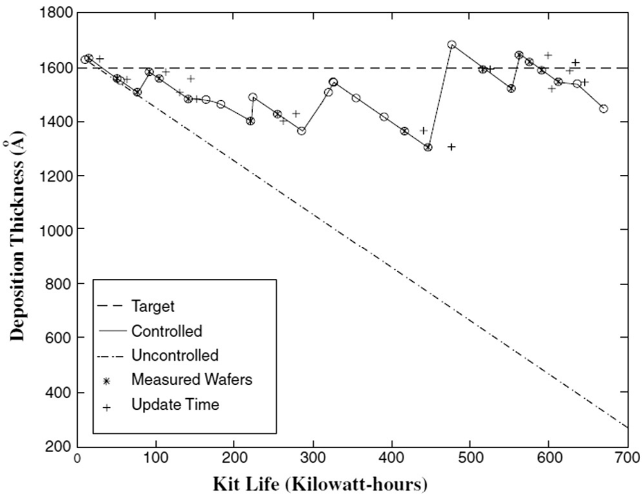

Fig. 15. MIMO model validation of deposition thickness [52].

further, the furnace aging effect has also been taking into account in the model through the integration of a corrective factor called K-factor. It is the ratio between actual thickness measurement with given parameters and the thickness predictions for the same input parameters. The equations given by $Thick_{Z_1 \sim Z_5}$ can be transformed into a matrix form, and the temperatures under a vector form.

$$\begin{pmatrix} Thick_{z_1} \\ Thick_{z_2} \\ Thick_{z_3} \\ Thick_{z_4} \\ Thick_{z_5} \end{pmatrix}_{(k+1)} = \begin{pmatrix} f_k(t, T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) \\ f_k(t, T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) \\ f_k(t, T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) \\ f_k(t, T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) \\ f_k(t, T_{Z_1}, T_{Z_2}, T_{Z_4}, T_{Z_5}) \end{pmatrix}_{\text{DOE}} \times \mathbf{K}_k \times \frac{t}{t_0} \tag{38}$$

Richard et al. point out that it should consider the center zone as the reference to predict the deposition time [52]. Therefore, the K-factor remains the same calculation, but now it is normalized to keep the center component at 1. Then, it iteratively updates with the following calculation:

$$\mathbf{K}_k = \mathbf{K}_{k-1} \times \left(\frac{Thick_{Measured}}{Thick_{Predicted}} \Big/ \frac{Thick_{Measured(Zone3)}}{Thick_{Predicted(Zone3)}} \right) \tag{39}$$

As illustrated in Fig. 15 from [5,52], using R2R control algorithms helps to attain a better uniformity and to better center the deposition thicknesses around the target.

4.3. Application of R2R control algorithm to BG management

Although R2R control algorithm has been widely applied to traditional batch processes such as semiconductor industry manufacturing, it can be easily applied to the biomedical engineering such as the research on the therapy of people with type 1 diabetes mellitus (T1DM). The protocol for T1DM is an autoimmune disease characterized by a lack of endogenous production of insulin, a hormone that increases glucose uptake within the body [54]. Diabetes is a disorder of the metabolism characterized by the inability of the pancreas to secrete sufficient insulin. When glucose levels remain high for extended periods of time (hyperglycemia) the patient is at risk for neuropathy, nephropathy, and other long-term vascular complications. Intensive insulin therapy requires three or more insulin injections per day or the use of an external insulin infusion pump in order to minimize the risk for complications.

The R2R algorithm exploits the repetitive nature of the insulin therapy regimen of the diabetic patient. The application of R2R control to help manage diabetes provides an opportunity to properly adjust the current insulin therapy of the patient. Therefore, the daily cycle of the patient with diabetes can be treated as a batch process. For a person with diabetes, each day serves as a single run.

An R2R algorithm was first used to suggest preprandial insulin doses to improve postprandial glucose concentration in [55]. Iterative learning control (ILC) algorithms were also proposed to optimize the operation of a batch process which can be utilized in the glucose management [56]. Tuo et al. proposed high order R2R control scheme to optimize insulin pump which has excellent robustness and could be a promising candidate therapy [57]. It is interesting to note that most of these research work are related to the Francis J. Doyles research group in Harvard John A. Paulson School of Engineering and Applied Sciences (https://thedoylegroup.org/).

In R2R batch process control, information about product quality from the previous run is used to determine the input for the next run. Based upon similarities between controlling semiconductor manufacturing and designing an effective long-term drug therapy, a R2R control algorithm is designed and validated with a case study on controlling blood coagulation in [58].

A block diagram of the R2R controller in BG management is illustrated in Fig. 16. The controller receives the measured glucose concentration, y_m, and compares it to the control objective. The controller then determines the desired insulin dose u and sends it to the insulin pump.

Depending on the insulin delivery route, the insulin is absorbed into the bloodstream with a specific pharmacokinetic (PK) profile. The insulin then acts to change the glucose concentration, which is affected by disturbances such as meals. The blood glucose concentration is measured by the sensor. Depending on the sensor type and placement, the sensor measurement may experience a dynamic lag [59].

From the BG measurements obtained at the end of the day, the insulin therapy can be adjusted for the next day. Each meal requires the appropriate insulin bolus to mediate rising BG concentrations. Therefore, the timing T and the amount Q of the insulin dose must be adjusted to yield the desired G_{Max} and G_{Min} value for each meal.

Generally, insulin dosing can be divided into two regimens: basal and bolus insulin. The basal insulin is required for fasting conditions [60,61]; while bolus insulin is calculated to correct for meals or hyperglycemia condition [62]. Both dosages should be adjusted over time. The development of external insulin infusion pumps as well as the introduction of rapid-acting insulin analogues has made automated intensive insulin therapy feasible. Algorithms are required to adjust the insulin therapy according to the measurements of glucose concentrations [63]. Use bolus for an example, for each meal, an update law is prescribed to correct the insulin bolus amount and timing for the next day. For the present application, $u_k(t)$ and $y_k(t)$ correspond to the insulin and glucose profiles, respectively, where t is the continuous time variable and k is the run number. The run index k represents the repetition of the 24-h daily routine of breakfast, lunch, and dinner meals.

The general R2R control algorithm is given below:

1. Parameterize the input $u_k(t)$ profile for run k. Also, consider a sampled version ψ_k, of the output $y_k(t)$, such that the input parameter vector ν_k, and the controlled variable vector ψ_k, have the same dimension. This gives

$$\psi_k = F(\nu_k). \tag{40}$$

2. Choose an initial guess for ν_k when $k = 1$.
3. Complete the run using the input $u_k(t)$ corresponding to $\nu_k(t)$. Determine ψ_k from the measurements $y_k(t)$.
4. Update the input parameters using

$$\nu_{k+1} = \nu_k + \mathbf{K}(\psi^r - \psi_k), \tag{41}$$

where $\mathbf{K}$ is an appropriate gain matrix and ψ^r represents the reference values to be attained.

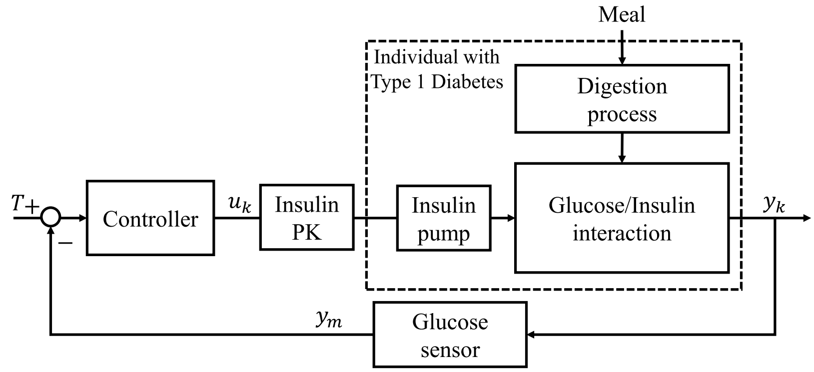

Fig. 16. A block diagram representation of the artificial pancreas [59].

5. Set $k = k + 1$ and repeat steps 3 and 4 until it converges.

The convergence of the R2R algorithm can be determined by analyzing the dynamics of the error for the closed-loop system, where the error is $e_k = \psi^r - \psi_k$. For this analysis, a linearized version of the system Eqs. (40) is used

$$\psi_k = \mathbf{S}\nu_k, \tag{42}$$

where $\mathbf{S} = \partial F/\partial \nu$ is the sensitivity matrix. Using (41) and (42), the linearized error dynamics are easily derived through the following manipulations:

$$\mathbf{S}\nu_{k+1} = \mathbf{S}\nu_k + \mathbf{SK}(\psi^r - \psi_k), \tag{43}$$

$$\psi^r - \mathbf{S}\nu_{k+1} = \psi^r - (\mathbf{S}\nu_k + \mathbf{SK}(\psi^r - \psi_k)),$$

$$\psi^r - \psi_{k+1} = \psi^r - \psi_k - \mathbf{SK}(\psi^r - \psi_k),$$

$$e_{k+1} = e_k\mathbf{SK}_{ek} = (\mathbf{I} - \mathbf{SK})e_k.$$

Thus, the stability and convergence of the algorithm depends on controller gain $\mathbf{K}$ and the sensitivity matrix $\mathbf{S}$. And there is no coupling between the meals. The initial guesses ν_1 and some other important variables settings can be found in [64,65]. For each meal, there are two outputs G_{Max} and G_{Min} and two inputs T and Q. To have the flexibility of taking BG measurements at different times breakfast (B), lunch (L), and dinner (D), a fixed glucose level can no longer be used. Then, the sampled output vector is

$$\psi_k = \begin{bmatrix} G(T_{B_1}) - G(T_{B_2}) \\ G(T_{L_1}) - G(T_{L_2}) \\ G(T_{D_1}) - G(T_{D_2}) \end{bmatrix}.$$

The manipulated variable ν_k is simply the dose of insulin (per gram of carbohydrate in the meal) corresponding to each meal of day k; $\nu_k = [Q_B\ Q_L\ Q_D]'$. Fig. 17 shows the R2R algorithms ability to control BG concentrations and converge to the desired preprandial goal of 90 mg/dl and postprandial goal of 140 mg/dl within 10 days. The dashed–dotted line on the top plot shows the open-loop BG concentration in the absence of insulin boluses. Glucose levels rise above 200 mg/dl and the patient remains in a hyperglycemic state for an extended period of time. On day 1, given by the solid line, the initial guesses for the insulin bolus amount and timing can be seen from the bottom plot. The corresponding BG concentration still remains outside of the desired boundaries for G_{Max} and G_{Min}. On day 2, the algorithm continues to compute the optimal insulin bolus amount and timing for each meal, as shown by the dashed line in the top and bottom plots. Eventually on day 10, represented by the thick solid line, G_{Max} and G_{Min} for each meal come within the desired bounds.

The similarities between the protocol for a person with type 1 diabetes and a batch recipe which motivate the application of the R2R technique include:

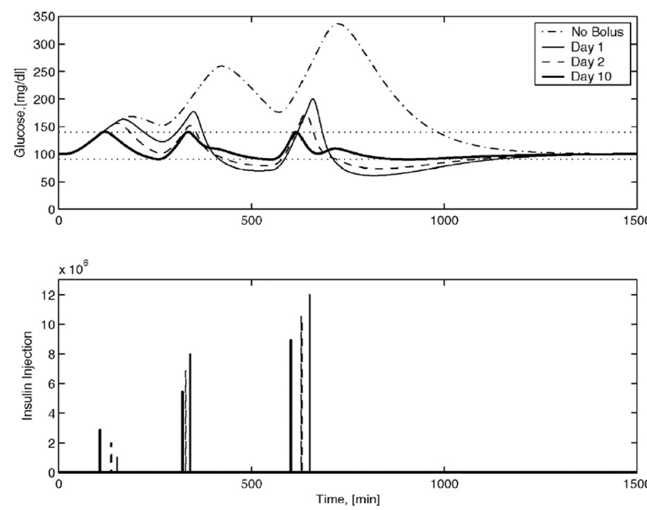

Fig. 17. Convergence of R2R control algorithm within 10 days [65,66].

1. The recipe (24-h cycle) for a patient consists of a repeated meal protocol (typically three meals) with some variance on meal type, timing, and duration;
2. There are key quality variables or time points of the BG profile that can be measured, reflecting the impact of the input insulin therapy regimen;
3. Since each patient is different, the insulin therapy needs to be adjusted individually.

Hence, the repetitive nature of the insulin therapy regimen of the diabetic patient adds availability to the R2R algorithm. Wang et al. proposed a novel adaptive R2R control strategy for insulin pump therapy optimization [67]. Good et al. applied R2R control to optimize drug dosage in order to regulate blood coagulation [58]. Lee et al. improved current continuous glucose monitors performance by a R2R strategy based on continuous wear that personalizes sensor calibration parameters using data from previous weeks use [68]. Palerm et al. presented a novel R2R control algorithm to adjust the meal-related insulin dose using only post-prandial BG measurements [69].

5. Literature overview and categorization

5.1. Literature overview

As what we have discussed before, we could summarize different R2R control methods in Table 2.

The searching phrases were “run to run” or “run by run” in the title and “control” in the abstract. Please note that the quotation

Table 2
Comparisons of R2R methods.

R2R controllers	Authors (year)	Benefits	Problems
EWMA	Sachs (1990)	Very simple structure	Cannot handle the drift
d-EWMA, PCC	Butler (1994), Butler (1993)	Handles the process with the linear drift	Used in almost linear systems
MPC, OAQC	Campbel (1997), Castillo (1998)	Able to control the MIMO system	Uses state space model, requires more senors
ANN	Wang (1996)	Useful in the complex or nonlinear process	Requires good training data sets
ILC	Xiong (2005)	Track the transient	Hard to incorporate, limited use

mark was used in order to make sure that unrelated papers were not involved. Because the exact name were used in the literature search, it is possible that many important R2R control publications may be missed.

From the selected paper, however, it can still provide some essential facts that R2R control will be a heated and promising area in the near future. As shown in Fig. 18, the science of R2R control will continue to expand as the academic and industry communities look to incorporating capabilities, which makes R2R control continue to be an integral part of the fabrication facility of the future [70–72].

From the perspective of the research field, 90% of the paper are discussed within Fab, while others are in the biomedical field, i.e. glucose management; From the perspective of the content, there are 94 paper talking about the EWMA controller from 1995 to 2018, or enhanced EWMA (d-EWMA, PCC, ANN, etc.) regarding to different conditions, such as when EWMA controllers subject to stochastic metrology delay. There are 30 paper directly talking about the CMP in the title. Based on the review of all the R2R paper, over 36% of articles are concerning CMP, especially from 1997 to 2008. Others are discussing about the etch process [73], lithography [74], CVD [75], crystallization [76], diabetes & BG [63] and statistical process adjustment methods for quality control. In fact, there are more CMP examples discussed in the paper without using the CMP key words in the title. In addition, there are about 39 papers talking about implementing VM into the R2R control algorithms, especially from 2005 to 2015. It can be foreseen that the above two aspects (CMP, VM) would be the promising fields in the future (see Fig. 19).

In addition, the textbook, *Run-to-Run Control in Semiconductor Manufacturing*, written by James Moyne et al., introduces what R2R control is and identifies key components of implementing R2R control for virtually any process [5]. Although it was published almost 20 years ago, it has been a classical book which should not be missed by readers interested in R2R process control. It helps the reader maximize the benefits obtainable through R2R control and illustrates the advantages of R2R control with manufacturing case studies.

It should be noted that the application of R2R control algorithms in the biomedical engineering is a relatively hot and promising topic. Due to the similarity between semiconductor batch process and insulin injection process, numerous novel R2R controllers have been successfully proposed to keep one subjects BG level within the normal range [65,69,60,77,62,78,79].

Besides, there are 18 patents concerning R2R control in the United States Patent and Trademark Office (http://www.uspto.gov). 12 patents are held by Advanced Micro Devices Inc. (Austin, TX); 2 patents by Applied Material Inc., Santa Clara, CA; 2 patents by Brooks Automation Inc., Chelmsford, MA; 1 patent by Texas Instruments Inc., Dallas, TX; 1 patent by Tokyo Electron Ltd., Tokyo, Japan; 1 patent by UC Santa Barbara, CA.

5.2. Literature categorization

The R2R control law is an efficient process control method that extracts advantages from both feedback and statistic control schemes. Note that an early R2R survey papers was published 20 years ago [11]. Wang et al. reviewed ILC, repetitive control, and R2R control in 2009 [71]. However, all of survey papers are only covered cases within Fab when talking about R2R control [80,13]. Therefore, from what we have discussed above, a more recent survey of R2R control algorithms for batch process control can be classified as follows:

5.2.1. EWMA

The EWMA feedback controller is one of the most popular R2R control schemes for adjusting certain semiconductor manufacturing process with a linear drift. Although the EWMA controller can guarantee a long-term stability, it usually requires a moderately large number of runs to bring the output of a process to its target. In order to reduce a possibly high rework rate, Tseng et al. proposed a variable discount factor to tackle the problem [81,82]. To achieve desired process output results, R2R control systems involve both feedforward and feedback control schemes in which R2R controllers are generally implemented through EWMA-based algorithms [83]. Control performance of both schemes was analytically derived and the determination of the optimal control parameters was provided and validation results using the field CMP data demonstrate the control effectiveness of both control schemes. Wang et al. analyzed the behavior of an EWMA controller with gain updating and compared it to that of an EWMA controller with intercept updating [84]. Zheng et al. proposed two intuitive ways, a “tool-based” approach and a “product-based” approach, to analyze the stability and performance analysis of mixed product R2R control [85]. Patel et al. presented a recursive scheme for optimizing the gain of an EWMA controller under stability constraints [86,8]. Firth et al. proposed an adaptive disturbance estimation method used to solve the inaccurate disturbance modeling when doing EWMA controller model assumptions [87]. Kazemzadeh et al. developed a quadratic R2R controller model using EWMA and d-EWMA scheme which improves capability of R2R process in the certain manufacturing process [88]. The influences of metrology delay on both the transient and asymptotic properties of the product quality with disturbance were analyzed under an EWMA R2R control [89]. Ko et al. proposed an intelligent adaptive process control method using dynamic deadband to minimize process error and to reduce process variance in semiconductor manufacturing [90].

5.2.2. d-EWMA

Butler and Stefani proposed the use of *in situ* ellipsometer to drive a new R2R supervisory controller, in which they first termed PCC, to alleviate the effect of machine and process drift in the polysilicon gate etch process [21,14]. Fan et al. proposed a triple-EWMA controller for the CMP process [91]. Hybrids of two R2R control schemes, PCC and double-EWMA schemes, for processes under deterministic drifts were presented in [92]. Chen et al. proposed an age-based d-EWMA controller applied to the CMP process, and it improved the control efficiency significantly [93,94]. In [95], an enhanced d-EWMA controller was proposed to further eliminate the off-target (non-random biases) of the process output. In [96], the d-EWMA controller, time-varying d-EWMA controller, age-based d-EWMA controller, and extending Kalman filter (EKF) controller have been applied to aluminum sputter deposition processes for predicting deposition rates and comparing their performances. A multi-objective monitoring approach was proposed by Wang et al. to monitor the semiconductor manufacturing process with d-EWMA R2R Controller, which can not only be used to monitor

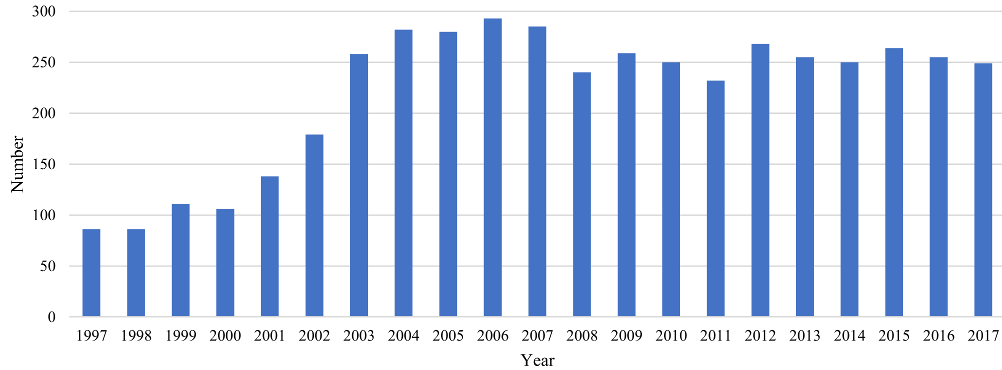

Fig. 18. Publications on R2R control.

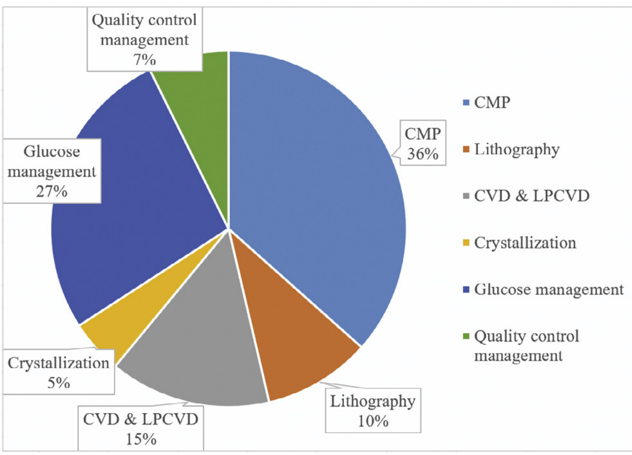

Fig. 19. R2R control applications in different categories.

the non-stationary drifted process, but also can reduce the missing rates [97]. Lee et al. proposed a robust R2R control algorithm, which uses the d-EWMA or optimal filter but solves a min–max problem to find the input adjustment that minimizes the worst-case predicted error on CMP process model [98,99]. The optimization model for finding the controller weights and the extension of this type of controllers to the multiple controllable factor case was described and illustrated in [100].

5.2.3. MPC

MPC is a model based control method, which requires a better identification of the process, such as constraints, nonlinearities, model uncertainties, etc. Therefore, a nonlinear model of MPC, an adaptive MPC and the robustness of MPC should be considered for the next-generation MPC technology as indicated in [26].

Lee et al. proposed a novel model predictive control technique called Batch-MPC (BMPC), which is based on a time-varying linear system model [101]. Cueli presented an iterative nonlinear model predictive control (INMPC), which incorporated ILC to an underlying nonlinear model [102].

Magni et al. modified the conventional MPC for Type 1 Diabetes [103]. Hovorka developed a nonlinear MPC controller by using Bayesian parameter estimation to maintain normoglycemia in subjects with type 1 diabetes [104].

More recently, Kwon developed a run-to-run-based model predictive controller (R2R-based MPC) for crystallization process [76, 105,106].

5.2.4. OAQC

The idea of the OAQC is to combine the model optimization step with the control step. It is done by using RLS to estimate the parameters of a Hammerstein model online and then optimizing it in order to obtain the best control action. Castillo [29] proposed and termed optimizing adaptive quality control (OAQC), using recursive least squares and a weight of the new parameter estimate with the previous values. It is interesting to note that the ST controller shares the similar algorithm with the OAQC. Self-tuning controllers are developed to provide on-line parameter estimation and control. A RLS algorithm is normally used to provide on-line parameter estimation to the controller, then the total cost function is minimized to obtain a recipe for the next run [107]. Jen et al. developed an ST controller offering a better performance compared with those of the control actions provided by OAQC [107].

Chen et al. applied a real-coded genetic algorithm to adaptively adjust the discount coefficients for d-EWMA controller with the help of the RLS identification method for model building [94].

Patel et al. presented a recursive scheme for optimizing the gain of an EWMA controller under stability constraints and model uncertainties [8].

5.2.5. ANN

ANN are quite useful for complex system modeling and classification of semiconductor manufacturing. It has been successfully used in the modeling of plasma etch process [108–111,73], CMP process [32,112], CVD [31], polymerization process [113].

Hankinson et al. developed a real-time ANN-based interactive R2R controller for the reactive ion etching (RIE) process [114]. Smith et al. proposed A self-tuning EWMA Controller utilizing ANN function approximation techniques to provide updates to the controller parameters [48].

Rietman et al. introduced an active neutral network control of wafer attributes in a plasma etch process [109] and claimed two patents [115,116] in 1995 and 1998.

Different types of ANN-based R2R process control system has been proposed for the CMP process [117–119]. The desired uniformity within wafer can be reached with the optimal polishing time while reducing the polishing time by 1/3.

Chang et al. proposed a VM structure which consists of a piecewise linear neural network to approximate the process drift and a fuzzy neural network is developed to control the fabrication outcome [120]. Kao et al. incorporated the so-called reliance index of VM of R2R control to gauge the reliability in the feedback loop [121]. Moreover, it provides an effective and economical solution for metrology prediction.

A well established VM system has following benefits in semiconductor manufacturing industry:

1. VM enhances the final yield by managing scrapped wafers appropriately [122,123].
2. VM enables a predictive maintenance based on real-time forecast of metrology data [122].

3. VM can detect process drifts timely and promptly [75].
4. VM enables R2R process control framework based on information obtained during or after the process in order to improve productivity [124,125].
5. VM reduces the cost and time needed for actual metrology [122,126].
6. VM can be included with online fault diagnosis and big data solutions [127,128].

Although VM has many advantages, there are some practical issues that need to be addressed. First, the number of input variables of a prediction model is huge as a large number of sensors are operating during the process. Second, due to preventive maintenance and daily wear and tear, the process is not stationary. Third, there are issues on how to incorporate the VM model into the R2R control system. In addition, poor training of the ANN may cause fluctuations in the weighting factors, resulting in additive noise and poor tracking of drift.

5.3. Promising fields and future work

As the quality specifications for semiconductor devices have become more and more stringent, R2R control has become an promising technology for semiconductor manufacturing.

In general, R2R control involves using previous measurement data to optimize a particular process step or set of steps. In fact, however, each identical machine has its own copy of the control algorithm. Algorithms may become out of sync with each other producing unanticipated results. It should be borne in mind that any changes may pose great risks to operations everytime recipe is updated for a highly automated system. Therefore, algorithms used to control the process must be separated from the machine control code from the aspect of engineers. A common way to do in the industry engineering practice is designating control algorithms into a dedicated R2R control system which frees the machine code from the burden of managing algorithms with cleaner code that is updated less often.

Stability analysis of EWMA algorithms has allowed for better determination of range of operation. A multi-algorithm approach to control has been developed to provide a wider range of controllability of systems. Phenomenological R2R control models and its libraries that combine stochastic information with physics and chemistry process control solutions are more directly optimized to financial parameters such as yield and throughput.

From the recent publications, the possible future work may be concerned in the following fields:

1. The R2R process control solution, especially for CMP process, will have to be automated to meet increasing demands on throughput and to reduce operator errors. The R2R controller will also control parameters that cannot (yet) be controlled at an *in situ* level, and will accommodate and utilize inter-process feedforward and feedback information to achieve industry-level control targets [129–131]. Therefore, a supervisory batch-based R2R control by taking both the batch information and the feedback quality information into account [80]. Meanwhile, some novel methods should be proposed to deal with the challenging tasks such as the system with metrology delay and uncertainties [9,25,132–134].
2. Hybrids of R2R, ILC and MPC control algorithm can be implemented, especially in high-mix environments where a large number of disturbances exist. ILC utilizes the system repetitions to improve the system control performance when the system executes a given task repeatedly [135]. Therefore, application of ILC technique can be naturally applied in Fab, such as the RTP system [136–140]. For example, higher-order modeling and control algorithms which incorporate a learning capability have been developed thus reducing the need for process identification ahead of control system deployment [141–143,13,144–147].
3. The MIMO control law is somewhat more complicated and the MIMO controller is thought to be more appropriate than SISO or single input multi-output (MISO) controllers especially in the heating process [52]. In order to achieve a robust and feasible control result in the MIMO control system, selection of optimum manipulated variables to controlled variables is specifically crucial [148–157].
4. The repetitive nature of an intensive glucose control therapy regimen is analogous to the semiconductor manufacturing batch process. The application of R2R batch control to the biomedical engineering can be viewed as a novel field with huge potentials [65,68,55,77,158].
5. The EWMA algorithm in R2R controller can provide minimum variance control for an ARMA noise series under the assumption that the stochastic noise belongs to the Gaussian distribution. However, the more challenging parts are the data with long range dependence (LRD), which belong to the non-Gaussian distribution. Therefore, the non-Gaussian cases can be handled by the generalized ARMA models with fractional order signal processing techniques [159–162].
6. VM and yield prediction, which consist of some predictive and corrective models for metrology outputs in function of the previous metrology outputs and the equipment parameters of current, will be incorporated into control solutions [163–166]. In addition, event-based control rule approaches, such as fault detection and classification (FDC) will be incorporated into the R2R control solutions, which involves monitoring the behavior of the manufacturing equipment during operation and detecting events that might affect the quality of the products [167,168].
7. The advent of big data solutions combined with cloud computing, machine learning and data mining has paved the way for significant opportunities in yield optimization and OEE improvement [128,169,170]. Thus, how to maximize the usage of these data and model the variability would be the priorities for all process engineers and product managers. A big-data platform and its applications for semiconductor manufacturing based on Hadoop framework are proposed in [171,172].

Each of these topics provides significant opportunities for research as well as benefits in application to industrial facilities.

6. Conclusion

The R2R batch process control has attracted much attention in research and has been widely used in practice. First, the principles of R2R control methods are introduced in this paper, such as EWMA, d-EWMA, MPC, OAQC, ANN algorithms. Second, simulation examples are conducted to compare the different R2R controllers. Then, three cases on CMP process, furnace uniformity control and BG management are studied with validation results to demonstrate the control effectiveness of R2R control scheme. Finally, we point out promising fields and possible future research directions of R2R batch process control.

Acknowledgments

The authors would like to thank three anonymous referees for their comments that led to a clear and better presentation of our work. The authors appreciate the feedback comments from Professor Youqing Wang of Beijing University of Chemical Technology (BUCT) for us to improve this survey.

References

[1] Sachs E, Guo RS, Ha S, Hu A. On-line process optimization and control using the sequential design of experiments. In: Proceedings of the VLSI technology, digest of technical papers. IEEE; 1990, p. 99–100.
[2] Sachs E, Guo RS, Ha S, Hu A. Process control system for VLSI fabrication. IEEE Trans Semicond Manuf 1991;4(2):134–44.
[3] Khan K, El Chemali C, Moyne J, Chapple Sokol J, Nadeau R, Smith P, et al. Yield improvement at the contact process through run-to-run control. In: Proceedings of the electronics manufacturing technology symposium. IEEE; 1999, p. 258–63.
[4] Del Castillo E. Statistical process adjustment for quality control, vol. 369. Wiley-Interscience; 2002.
[5] Moyne J, Del Castillo E, Hurwitz AM. Run-to-run control in semiconductor manufacturing. CRC press; 2000.
[6] May GS, Spanos CJ. Fundamentals of semiconductor manufacturing and process control. John Wiley & Sons; 2006.
[7] Adivikolanu S, Zafiriou E. Internal model control approach to run-to-run control for semiconductor manufacturing. In: Proceedings of the American control conference, vol. 1. IEEE; 1997, p. 145–9.
[8] Patel NS, Jenkins ST. Adaptive optimization of run-to-run controllers: The EWMA example. IEEE Trans Semicond Manuf 2000;13(1):97–107.
[9] Hwang S-H, Lin J-C, Wang H-C. Robustness diagrams based optimal design of run-to-run control subject to deterministic and stochastic disturbances. J Process Control 2018;63:47–64.
[10] Ning Z, Moyne JR, Smith T, Boning D, Castillo ED, Yeh JY, et al. A comparative analysis of run-to-run control algorithms in the semiconductor manufacturing industry. In: Proceedings of the advanced semiconductor manufacturing conference and workshop. IEEE; 1996, p. 375–81.
[11] Del Castillo E, Hurwitz AM. Run-to-run process control: literature review and extensions. J Qual Technol 1997;29(2):184.
[12] Campbell WJ, Firth SK, Toprac AJ, Edgar TF. A comparison of run-to-run control algorithms. In: Proceedings of the American Control Conference, vol. 3. 2002, p. 2150–5.
[13] Tan F, Pan T, Li Z, Chen S. Survey on run-to-run control algorithms in high-mix semiconductor manufacturing processes. IEEE Trans Ind Inf 2015;11(6):1435–44.
[14] Butler SW, Stefani JA. Supervisory run-to-run control of polysilicon gate etch using in situ ellipsometry. IEEE Trans Semicond Manuf 1994;7(2):193–201.
[15] Guldi RL, Jenkins CD, Oamminga G, Baum T, Foster T. Process optimization tweaking tool (POTT) and its application in controlling oxidation thickness. IEEE Trans Semicond Manuf 1989;2(2):54–9.
[16] Sachs E, Ingolfsson A, Ha S. Tuning a process while performing SPC: an approach based on the sequential design of experiments. In: Proceedings of the advanced semiconductor manufacturing conference and workshop. IEEE; 1990, p. 126–30.
[17] Leang S, Spanos CJ. Statistically based feedback control of photoresist application. In: Proceedings of the advanced semiconductor manufacturing conference and workshop. IEEE; 1991, p. 185–90.
[18] Ingolfsson A, Sachs E. Stability and sensitivity of an EWMA controller. Univ. Alberta School Bus Res Paper 1992;25(4):1–174.
[19] Moyne WP. Run by run control: interfaces, implementation, and integration (Ph.D. thesis), Massachusetts Institute of Technology; 1995.
[20] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.
[21] Butler SW, Stefani J. Application of predictor corrector control to polysilicon gate etching. In: Proceedings of the American control conference. IEEE; 1993, p. 3003–7.
[22] Sachs E, Hu A, Ingolfsson A. Run by run process control: combining SPC and feedback control. IEEE Trans Semicond Manuf 1995;8(1):26–43.
[23] Moyne J, Curry J, Solakhian V, Weaver T, Gwizdak R. Improving reliability, yield and throughput of chemical-mechanical planarization through process automation and control. In: Proceedings of the advanced semiconductor manufacturing conference: SEMICON Taiwan, vol. 98. 1998.
[24] El Chemali C, Moyne J, Khan K, Colt J, Chapple-Sokol J, Nadeau R, et al. Multizone Uniformity Control of a CMP process utilizing a pre and post-measurement strategy. In: Proceedings of the 46th international symposium on american vacuum society. 1998.
[25] Wang J, He QP, Qin SJ, Bode CA, Purdy MA. Recursive least squares estimation for run-to-run control with metrology delay and its application to STI etch process. IEEE Trans Semicond Manuf 2005;18(2):309–19.
[26] Qin SJ, Badgwell TA. An overview of industrial model predictive control technology. In: AIChE symposium series, vol. 93. New York, NY: American Institute of Chemical Engineers, 1971-c2002; 1997, p. 232–56.
[27] Mullins JA, Campbell WJ, Stock AD. Evaluation of model predictive control in run-to-run processing in semiconductor manufacturing. In: Process, equipment, and materials control in integrated circuit manufacturing III, vol. 3213. International Society for Optics and Photonics; 1997, p. 182–90.
[28] Castillo ED. A multivariate self-tuning controller for run-to-run process control under shift and trend disturbances. IIE Trans 1996;28(12):1011–21.
[29] Castillo ED, Yeh JY. An adaptive run-to-run optimizing controller for linear and nonlinear semiconductor processes. IEEE Trans Semicond Manuf 1998;11(2):285–95.
[30] Åström KJ, Wittenmark B. Adaptive control. Courier Corporation; 2013.
[31] Wang XA, Mahajan RL. Artificial neural network model-based run-to-run process controller. IEEE Trans Compon Packag Manuf Technol C 1996;19(1):19–26.
[32] Yi J, Sheng Y, Xu CS. Neural network based uniformity profile control of linear chemical-mechanical planarization. IEEE Trans Semicond Manuf 2003;16(4):609–20.
[33] Khan AA, Moyne JR, Tilbury DM. Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares. J Process Control 2008;18(10):961–74.
[34] Kang P, Lee HJ, Cho S, Kim D, Park J, Park CK, et al. A virtual metrology system for semiconductor manufacturing. Expert Syst Appl 2009;36(10):12554–61.
[35] Zantye PB, Kumar A, Sikder A. Chemical mechanical planarization for microelectronics applications. Mater Sci Eng R 2004;45(3):89–220.
[36] Davarik B, Koburger C, Schulz R, Warnock J, Furukawa T, Jost M, et al. A new planarization technique, using a combination of RIE and chemical mechanical polish (CMP). In: Proceedings of the international electron devices meeting technical digest. IEEE; 1989, p. 61–4.
[37] Ryan JG, Geffken RM, Poulin NR, Paraszczak JR. The evolution of interconnection technology at IBM. IBM J Res Dev 1995;39(4):371–81.
[38] Borst CL, Gill WN, Gutmann RJ. Chemical-mechanical polishing of low dielectric constant polymers and organosilicate glasses: fundamental mechanisms and application to IC interconnect technology. Springer Science & Business Media; 2013.
[39] Taur Y, Mii YJ, Nowak E, Frank D, Wong HS, Buchanan D, et al. CMOS scaling into the 21st century: 0.1 μm and beyond. IBM J Res Dev 1995;39(1.2):245–60.
[40] Boning DS, Moyne WP, Smith TH, Moyne J, Telfeyan R, Hurwitz A, et al. Run by run control of chemical-mechanical polishing. IEEE Trans Compon Packag Manuf Technol C 1996;19(4):307–14.
[41] Moyne JR, Telfeyan R, Hunvitz A, Taylor J. A process-independent run-to-run controller and its application to chemical-mechanical planarization. In: Proceedings of the advanced semiconductor manufacturing conference and workshop. IEEE; 1995, p. 194–200.
[42] Moyne JR, McAfee Jr LC. A generic cell controller for the automated VLSI manufacturing facility. IEEE Trans Semicond Manuf 1992;5(2):77–87.
[43] Telfeyan R, Moyne J, Hurwitz A, Taylor J. Demonstration of a process-independent run-to-run controller. In: Proceedings of the 187th meeting of the electrochemical society, vol. 20. 1995, p. 1–13.
[44] Telfeyan R, Moyne J, Chaudhry N, Pugmire J, Shellman S, Boning D, et al. A multilevel approach to the control of a chemical–mechanical planarization process. J Vac Sci Technol A 1996;14(3):1907–13.
[45] Mandur JS, Budman HM. Simultaneous model identification and optimization in presence of model-plant mismatch. Chem Eng Sci 2015;129:106–15.
[46] Duan Y, Liu M, Dong M, Wu C. A two-stage clustered multi-task learning method for operational optimization in chemical mechanical polishing. J Process Control 2015;35:169–77.
[47] Da L. Control of semiconductor manufacturing: CMP & thickness variations. (Ph.D. thesis), National Unverisy of Singapore; 2003.
[48] Smith TH, Boning DS. A self-tuning EWMA controller utilizing artificial neural network function approximation techniques. IEEE Trans Compon Packag Manuf Technol C 1997;20(2):121–32.
[49] Smith TH, Boning DS, Stefani J, Butler SW. Run by run advanced process control of metal sputter deposition. IEEE Trans Semicond Manuf 1998;11(2):276–84.
[50] Gumpher J, Bather W, Wedel D. LPCVD silicon nitride uniformity improvement using adaptive real-time temperature control. IEEE Trans Semicond Manuf 2003;16(1):26–35.
[51] Shinde S, Sonar A, Sun Y. Advanced process control for furnace systems in semiconductor manufacturing. In: Proceedings of the SEMI advanced semiconductor manufacturing conference. IEEE; 2013, p. 275–9.
[52] Richard B, Roussy A, Brun L, Pompier D, Alegret C, Pinaton J. Implementation of MIMO R2R control regulation on furnaces processes. In: Proceedings of the advanced semiconductor manufacturing conference. IEEE; 2010, p. 143–8.
[53] Cale TS, Crouch PE, Shen S, Tsakalis KS. Run-to-run adaptive optimization of a tungsten silicide LPCVD process. In: Proceedings of the 34th IEEE conference on decision and control, vol. 3. IEEE; 1995, p. 2474–5.
[54] Daneman D. Type 1 diabetes. Lancet 2006;367(9513):847–58.
[55] Zisser H, Jovanovič L, Doyle III F, Ospina P, Owens C. Run-to-run control of meal-related insulin dosing. Diabetes Technol Ther 2005;7(1):48–57.
[56] Wang Y, Zisser H, Dassau E, Jovanovič L, Doyle FJ. Model predictive control with learning-type set-point: Application to artificial pancreatic β-cell. AIChE J 2010;56(6):1510–8.
[57] Tuo J, Sun H, Shen D, Wang H, Wang Y. Optimization of insulin pump therapy based on high order run-to-run control scheme. Comput Methods Programs Biomed 2015;120(3):123–34.

[58] Good R, Hahn J, Edison T, Qin SJ. Drug dosage adjustment via run-to-run control. In: Proceedings of the American control conference, vol. 5. IEEE; 2002, p. 4044–9.
[59] Huyett LM, Dassau E, Zisser HC, Doyle III FJ. Glucose sensor dynamics and the artificial pancreas: the impact of lag on sensor measurement and controller performance. IEEE Control Syst 2018;38(1):30–46.
[60] Palerm CC, Zisser H, Jovanovič L, Doyle FJ. A run-to-run control strategy to adjust basal insulin infusion rates in type 1 diabetes. J Process Control 2008;18(3):258–65.
[61] Toffanin C, Sandri A, Messori M, Cobelli C, Magni L. Automatic adaptation of basal therapy for Type 1 diabetic patients: a Run-to-Run approach. Biomed Signal Process Control 2014;47(3):2070–5.
[62] Campos Delgado DU, Campos Cornejo F, Hernández Ordoñez M. Extension of the run-to-run control to multi-boluses schemes. In: Proceedings of the IEEE international conference on control applications. IEEE; 2008, p. 678–83.
[63] Wang Y, Dassau E, Doyle III FJ. Closed-loop control of artificial pancreatic-cell in type 1 diabetes mellitus using model predictive iterative learning control. IEEE Trans Biomed Eng 2010;57(2):211–9.
[64] Srinivasan B, Primus C, Bonvin D, Ricker N. Run-to-run optimization via control of generalized constraints. Control Eng Pract 2001;9(8):911–9.
[65] Owens C, Zisser H, Jovanovič L, Srinivasan B, Bonvin D, Doyle III FJ. Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus. IEEE Trans Biomed Eng 2006;53(6):996–1005.
[66] Doyle III FJ, Jovanovič L. Method and apparatus for glucose control and insulin dosing for diabetics, US Patent 7, 651, 845, 2010.
[67] Wang H, Wang Y. A novel adaptive run-to-run control strategy for insulin pump therapy optimization. In: Proceedings of the 32nd Chinese control conference. IEEE; 2013, p. 2939–44.
[68] Lee JB, Dassau E, Doyle FJ. A run-to-run approach to enhance continuous glucose monitor accuracy based on continuous wear. IFAC-PapersOnLine 2015;48(20):237–42.
[69] Palerm C, Zisser H, Jovanovič L, Doyle F. A run-to-run framework for prandial insulin dosing: handling real-life uncertainty. Internat J Robust Nonlinear Control 2007;17(13):1194–213.
[70] Boning D, Moyne W, Smith T, Moyne J, Hurwitz A. Practical issues in run by run process control. In: Proceedings of the advanced semiconductor manufacturing conference and workshop. IEEE; 1995, p. 201–8.
[71] Wang Y, Gao F, Doyle FJ. Survey on iterative learning control, repetitive control, and run-to-run control. J Process Control 2009;19(10):1589–600.
[72] Ringwood JV, Lynn S, Bacelli G, Ma B, Ragnoli E, McLoone S. Estimation and control in semiconductor etch: Practice and possibilities. IEEE Trans Semicond Manuf 2010;23(1):87–98.
[73] Card JP, Naimo M, Ziminsky W. Run-to-run process control of a plasma etch process with neural network modelling. Qual Reliab Eng Int 1998;14(4):247–60.
[74] Chemali C, Freudenberg J, Hankinson M, Bendik JJ. Run-to-run critical dimension and sidewall angle lithography control using the PROLITH simulator. IEEE Trans Semicond Manuf 2004;17(3):388–401.
[75] Lin TH, Hung MH, Lin RC, Cheng FT. A virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing. In: Proceedings of the IEEE international conference on robotics and automation. IEEE; 2006, p. 1054–9.
[76] Kwon JS-I, Nayhouse M, Orkoulas G, Ni D, Christofides PD. Run-to-run-based model predictive control of protein crystal shape in batch crystallization. Ind Eng Chem Res 2014;54(16):4293–302.
[77] Campos Cornejo F, Campos Delgado DU, Espinoza Trejo D, Zisser H, Jovanovič L, Doyle III FJ, et al. An advisory protocol for rapid-and slow-acting insulin therapy based on a run-to-run methodology. Diabetes Technol Ther 2010;12(7):555–65.
[78] Kim S, Lanz KJ, Evans CE, Gibson ER, Olesberg JT, Mallem M, et al. Real-time monitoring of glycerol and methanol to enhance antibody production in industrial pichia pastoris bioprocesses. Biochem Eng J 2015;94:115–24.
[79] Ma J, Zhao H, Schuck P. A histogram approach to the quality of fit in sedimentation velocity analyses. Anal Biochem 2015;483:1–3.
[80] Wang K, Han K. A batch-based run-to-run process control scheme for semiconductor manufacturing. IIE Trans 2013;45(6):658–69.
[81] Tseng ST, Chou RJ, Lee SP. A study on a multivariate EWMA controller. IIE Trans 2002;34(6):541–9.
[82] Tseng ST, Yeh AB, Tsung F, Chan YY. A study of variable EWMA controller. IEEE Trans Semicond Manuf 2003;16(4):633–43.
[83] Good R, Qin SJ. Stability analysis of double EWMA run-to-run control with metrology delay. In: Proceedings of the American control conference, vol. 3. IEEE; 2002, p. 2156–61.
[84] Wang J, He QP. EWMA run-to-run controllers with gain updating: Stability and sensitivity analysis. In: 2008 American control conference. IEEE; 2008, p. 2872–7.
[85] Zheng Y, Lin QH, Wang DSH, Jang SS, Hui K. Stability and performance analysis of mixed product run-to-run control. J Process Control 2006;16(5):431–43.
[86] Baras JS, Patel NS. Designing response surface model-based run-by-run controllers: A worst case approach. IEEE Trans Compon Packag Manuf Technol C 1996;19(2):98–104.
[87] Firth SK, Campbell WJ, Toprac A, Edgar TF. Just-in-time adaptive disturbance estimation for run-to-run control of semiconductor processes. IEEE Trans Semicond Manuf 2006;19(3):298–315.
[88] Kazemzadeh R, Karbasian M, Moghadam MB. Design and investigation of EWMA and double EWMA with quadratic process model in R2R controllers. Qual Quantity 2008;42(6):845–57.
[89] Wu MF, Lin CH, Wong DSH, Jang SS, Tseng ST. Performance analysis of EWMA controllers subject to metrology delay. IEEE Trans Semicond Manuf 2008;21(3):413–25.
[90] Ko HH, Kim JS, Kim J, Baek JG, Kim SS. Intelligent adaptive process control using dynamic deadband for semiconductor manufacturing. Expert Syst Appl 2011;38(6):6759–67.
[91] Fan SKS, Jiang BC, Jen CH, Wang CC. SISO run-to-run feedback controller using triple EWMA smoothing for semiconductor manufacturing processes. Int J Prod Res 2002;40(13):3093–120.
[92] Guo RS, Chen A, Chen JJ. Run-to-run control schemes for CMP process subject to deterministic drifts. In: Proceedings of the semiconductor manufacturing technology workshop. IEEE; 2000, p. 251–8.
[93] Chen A, Guo RS. Age-based double EWMA controller and its application to CMP processes. IEEE Trans Semicond Manuf 2001;14(1):11–9.
[94] Chen C-T, Chuang Y-C. An intelligent run-to-run control strategy for chemical–mechanical polishing processes. IEEE Trans Semicond Manuf 2010;23(1):109–20.
[95] Tseng ST, Song W, Chang YC. An initial intercept iteratively adjusted (IIIA) controller: an enhanced double EWMA feedback control scheme. IEEE Trans Semicond Manuf 2005;18(3):448–57.
[96] Chen JH, Kuo TW, Lee AC. Run-by-run process control of metal sputter deposition: combining time series and extended kalman filter. IEEE Trans Semicond Manuf 2007;20(3):278–85.
[97] Wang Y, Zheng Y, Gu XG, Huang L. Multi-objective Fault Monitoring for Semiconductor Manufacturing Process with dEWMA Run-to-Run Controller. In: Proceedings of the international conference on industrial informatics-computing technology, intelligent technology, industrial information integration. IEEE; 2015, p. 152–5.
[98] Ramaswamy S, Lee JH. Robust forecasts and run-to-run control for processes with linear drifts. In: Proceedings of the 2003 American control conference, 2003, vol. 5. IEEE; 2003, p. 3986–91.
[99] Lee JH, Kiew CM. Robust forecasts and run-to-run control for processes with linear drifts. J Process Control 2009;19(4):636–43.
[100] Castillo ED. Long run and transient analysis of a double EWMA feedback controller. IIE Trans 1999;31(12):1157–69.
[101] Lee KS, Chin IS, Lee HJ, Lee JH. Model predictive control technique combined with iterative learning for batch processes. AIChE J 1999;45(10):2175–87.
[102] Cueli JR, Bordons C. Iterative nonlinear model predictive control. Stability, robustness and applications. Control Eng Pract 2008;16(9):1023–34.
[103] Magni L, Forgione M, Toffanin C, Dalla Man C, Kovatchev B, De Nicolao G, et al. Run-to-run tuning of model predictive control for type 1 diabetes subjects: in silico trial. J Diabetes Sci Technol 2009;3(5):1091–198.
[104] Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, et al. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 2004;25(4):905.
[105] Kwon JS, Nayhouse M, Ni D, Christofides PD. Handling parametric drift in batch crystallization using predictive control with R2R model parameter estimation. IFAC-PapersOnLine 2015;48(8):912–7.
[106] Kwon JSI, Nayhouse M, Orkoulas G, Ni D, Christofides PD. A method for handling batch-to-batch parametric drift using moving horizon estimation: application to run-to-run MPC of batch crystallization. Chem Eng Sci 2015;127:210–9.
[107] Jen C-H, Jiang B, Fan S-K. General run-to-run (R2R) control framework using self-tuning control for multiple-input multiple-output (MIMO) processes. Int J Prod Res 2004;42(20):4249–70.
[108] Rietman E, Lory ER. Use of neural networks in modeling semiconductor manufacturing processes: An example for plasma etch modeling. IEEE Trans Semicond Manuf 1993;6(4):343–7.
[109] Rietman EA, Frye RC, Lory ER, Harry TR. Active neutral network control of wafer attributes in a plasma etch process. J Vac Sci Technol B 1993;11(4):1314–6.
[110] Kim B, May GS. An optimal neural network process model for plasma etching. IEEE Trans Semicond Manuf 1994;7(1):12–21.
[111] Rietman EA. A neural network model of a contact plasma etch process for VLSI production. IEEE Trans Semicond Manuf 1996;9(1):95–100.
[112] Jia X, Di Y, Feng J, Yang Q, Dai H, Lee J. Adaptive virtual metrology for semiconductor chemical mechanical planarization process using GMDH-type polynomial neural networks. J Process Control 2018;62:44–54.
[113] Zhang J. Batch-to-batch optimal control of a batch polymerisation process based on stacked neural network models. Chem Eng Sci 2008;63(5):1273–81.

[114] Hankinson M, Vincent T, Irani KB, Khargonekar PP. Integrated real-time and run-to-run control of etch depth in reactive ion etching. IEEE Trans Semicond Manuf 1997;10(1):121–30.
[115] Frye RC, Harry TR, Lory ER, Rietman EA. Active neural network control of wafer attributes in a plasma etch process, US Patent 5,467,883, 1995.
[116] Frye RC, Harry TR, Lory ER, Rietman EA. Active neural network control of wafer attributes in a plasma etch process, US Patent 5,737,496, 1998.
[117] Wang G-J, Chou M-H. A neural-Taguchi-based quasi time-optimization control strategy for chemical-mechanical polishing processes. Int J Adv Manuf Technol 2005;26(7–8):759–65.
[118] Wang GJ, Yu CH. Developing a neural network-based run-to-run process controller for chemical-mechanical planarization. Int J Adv Manuf Technol 2006;28(9–10):899–908.
[119] Wang GJ, Lin BS, Chang KJ. In-situ neural network process controller for copper chemical mechanical polishing. Int J Adv Manuf Technol 2007;32(1–2):42–54.
[120] Chang YJ, Kang Y, Hsu CL, Chang CT, Chan TY. Virtual metrology technique for semiconductor manufacturing. In: Proceedings of the international joint conference on neural networks. IEEE; 2006, p. 5289–93.
[121] Kao CA, Cheng FT, Wu WM. Preliminary study of run-to-run control utilizing virtual metrology with reliance index. In: Proceedings of the IEEE conference on automation science and engineering. IEEE; 2011, p. 256–61.
[122] Yung Cheng JC, Cheng FT. Application development of virtual metrology in semiconductor industry. In: Proceedings of the 31st annual conference of IEEE industrial electronics society. IEEE; 2005, p. 6–pp.
[123] Chen YT, Yang HC, Cheng FT. Multivariate simulation assessment for virtual metrology. In: Proceedings of the 2006 IEEE international conference on robotics and automation. IEEE; 2006, p. 1048–53.
[124] Khan AA, Moyne JR, Tilbury DM. An approach for factory-wide control utilizing virtual metrology. IEEE Trans Semicond Manuf 2007;20(4):364–75.
[125] Qin SJ, Cherry G, Good R, Wang J, Harrison CA. Semiconductor manufacturing process control and monitoring: A fab-wide framework. J Process Control 2006;16(3):179–91.
[126] Chen P, Wu S, Lin J, Ko F, Lo H, Wang J, et al. Virtual metrology: a solution for wafer to wafer advanced process control. In: Proceedings of the IEEE international symposium on semiconductor manufacturing. IEEE; 2005, p. 155–7.
[127] Srinivasan R, Qian MS. Online fault diagnosis and state identification during process transitions using dynamic locus analysis. Chem Eng Sci 2006;61(18):6109–32.
[128] Moyne J, Samantaray J, Armacost M. Big data capabilities applied to semiconductor manufacturing advanced process control. IEEE Trans Semicond Manuf 2016;29(4):283–91.
[129] Opila RL. Chemical mechanical planarization in IC device manufacturing III: Proceedings of the International Symposium. In: Proceedings of the international symposium. The Electrochemical Society; 2000.
[130] Solomon PR, Rosenthal P, Spartz M, Bosch-Charpenay S, Bosch O, Richter M. Advanced process control in semiconductor manufacturing. In: Proceedings of the ASQ world conference on quality and improvement. American Society for Quality; 2001, p. 185.
[131] Prabhu AV. Performance monitoring of run-to-run control systems used in semiconductor manufacturing. ProQuest; 2008.
[132] Liu T, Gao F, Wang Y. IMC-based iterative learning control for batch processes with uncertain time delay. J Process Control 2010;20(2):173–80.
[133] Ai B, Wong DSH, Jang SS, Zheng Y. Stability Analysis of EWMA Run-to-Run Controller Subjects to Stochastic Metrology Delay. IFAC Proc Vol 2011;44(1):12354–9.
[134] Gondhalekar R, Dassau E, Doyle FJ. Velocity-weighting & velocity-penalty MPC of an artificial pancreas: Improved safety & performance. Automatica 2018;91:105–17.
[135] Ahn H-S, Chen Y, Moore KL. Iterative learning control: Brief survey and categorization. IEEE Trans Syst Man Cybern C 2007;37(6):1099–121.
[136] Chen Y, Xu J-X, Lee TH, Yamamoto S. An iterative learning control in rapid thermal processing. In: Proceedings of the international conference on modeling, simulation and optimization. 1997, p. 189–92.
[137] Lee KS, Lee J, Chin I, Choi J, Lee JH. Control of wafer temperature uniformity in rapid thermal processing using an optimal iterative learning control technique. Ind Eng Chem Res 2001;40(7):1661–72.
[138] Lee KS, Lee JH. Iterative learning control-based batch process control technique for integrated control of end product properties and transient profiles of process variables. J Process Control 2003;13(7):607–21.
[139] Yang DR, Lee KS, Ahn HJ, Lee JH. Experimental application of a quadratic optimal iterative learning control method for control of wafer temperature uniformity in rapid thermal processing. IEEE Trans Semicond Manuf 2003;16(1):36–44.
[140] Cho W, Edgar TF, Lee J. Iterative learning dual-mode control of exothermic batch reactors. Control Eng Pract 2008;16(10):1244–9.
[141] Bode CA, Wang J, He QP, Edgar TF. Run-to-run control and state estimation in high-mix semiconductor manufacturing. Annu Rev Control 2007;31(2):241–53.
[142] Pan TH, Sheng BQ, Chang CC, Wong DSH, Jang SS. Run to run control based on adaptive cluster method for high-mix semiconductor manufacturing processes. In: Proceedings of the 29th Chinese control conference. IEEE; 2010, p. 3428–31.
[143] Wang J, He QP, Edgar TF. State estimation in high-mix semiconductor manufacturing. J Process Control 2009;19(3):443–56.
[144] Chang CC, Pan TH, Wong DSH, Jang SS. A G&P EWMA algorithm for high-mix semiconductor manufacturing processes. J Process Control 2011;21(1):28–35.
[145] Jiang X. Control performance assessment of run-to-run control system used in high-mix semiconductor manufacturing (Ph.D. thesis), University of Texas at Austin; 2012.
[146] Harirchi F, Vincent T, Subramanian A, Poolla K, Stirton B. Implementation of nonthreaded estimation for run-to-run control of high mix semiconductor manufacturing. IEEE Trans Semicond Manuf 2013;26(4):516–28.
[147] Prabhu AV, Edgar TF. A new state estimation method for high-mix semiconductor manufacturing processes. J Process Control 2009;19(7):1149–61.
[148] Baek KH, Han KP, Choi G, Kang HK, Jung ES, Song K, et al. Multiple input multiple output controller design to match chamber performance in plasma etching for semiconductor manufacturing. J Vac Sci Technol B 2013;31(6):062201.
[149] Good RP, Qin SJ. On the stability of MIMO EWMA run-to-run controllers with metrology delay. IEEE Trans Semicond Manuf 2006;19(1):78–86.
[150] Chen J, Wang F. PLS based dEWMA run-to-run controller for MIMO non-squared semiconductor processes. J Process Control 2007;17(4):309–19.
[151] Rajagopal R, Castillo ED. An analysis and mimo extension of a double EWMA run-to-run controller for non-squared systems. Int J Reliab Qual Saf Eng 2003;10(04):417–28.
[152] Lee HJ. Advanced process control and optimal sampling in semiconductor manufacturing. ProQuest; 2008.
[153] Rzepniewski AK, Hardt DE. Multiple input-multiple output cycle-to-cycle control of manufacturing processes (Ph.D. thesis), Massachusetts Institute of Technology; 2003.
[154] Liang C, Huang Y. Integrated product and process control for sustainable semiconductor manufacturing. Chin J Chem Eng 2011;19(2):192–8.
[155] Fan SKS, Chang YJ. Multiple-input multiple-output double exponentially weighted moving average controller using partial least squares. J Process Control 2010;20(6):734–42.
[156] Pan CC, Sun CI, Jang SS, Wong DSH, Yang GK. A MIMO R2R control using data-driven gain scheduling. Control Eng Pract 2011;19(11):1344–53.
[157] Benjamin R, Gerhard L, Stefan L. Multiple in/Multiple out, run to run controller for accurate and stable epitaxy processes: APC: Advanced process control. In: Proceedings of the 27th annual SEMI advanced semiconductor manufacturing conference. IEEE; 2016, p. 180–4.
[158] Herrero P, Pesl P, Reddy M, Oliver N, Georgiou P, Toumazou C. Advanced insulin bolus advisor based on run-to-run control and case-based reasoning. IEEE J Biomed Health Inf 2015;19(3):1087–96.
[159] Sheng H, Chen Y, Qiu T. Fractional processes and fractional-order signal processing: techniques and applications. Springer Science & Business Media; 2011.
[160] Liu K, Chen Y, Zhang X. An Application of the Seasonal Fractional ARIMA Model to the Semiconductor Manufacturing. IFAC-PapersOnLine 2017;50(1):8097–102.
[161] Liu K, Zhang X, Chen Y. Energy informatics and fractional calculus. In: ASME 2017 international design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers; 2017, V009T07A028–V009T07A028.
[162] Liu K, Chen Y, Zhang X. An Evaluation of ARFIMA (autoregressive fractional integral moving average) programs. Axioms 2017;6(2):16.
[163] Ferreira A, Roussy A, Condé L. Virtual metrology models for predicting physical measurement in semiconductor manufacturing. In: Proceedings of the advanced semiconductor manufacturing conference. IEEE; 2009, p. 149–54.
[164] Su YC, Lin TH, Cheng FT, Wu WM. Accuracy and real-time considerations for implementing various virtual metrology algorithms. IEEE Trans Semicond Manuf 2008;21(3):426–34.
[165] Dreyfus G. Neural networks: methodology and applications. Springer Science & Business Media; 2005.
[166] Fan SKS, Chang YJ. An integrated advanced process control framework using run-to-run control, virtual metrology and fault detection. J Process Control 2013;23(7):933–42.
[167] Xiong Z, Zhang J, Wang X, Xu Y. Run-to-run iterative optimization control of batch processes based on recurrent neural networks. Adv Neural Netw 2004;3174:97–103.
[168] Qin SJ, Cherry G, Good R, Wang J, Harrison CA. Control and monitoring of semiconductor manufacturing processes: challenges and opportunities. Dyn Control Proc Syst 2005;2004:125.
[169] Moyne J, Schulze B, Iskandar J, Armacost M. Next generation advanced process control: Leveraging big data and prediction. In: Proceedings of the 27th annual SEMI advanced semiconductor manufacturing conference. IEEE; 2016, p. 191–6.

[170] Moyne J, Samantaray J, Armacost M. Big data emergence in semiconductor manufacturing advanced process control. In: Proceedings of the 26th annual SEMI advanced semiconductor manufacturing conference. IEEE; 2015, p. 130–5.

[171] Watson S, Samantaray J, Scoville J, Moyne J. Big data analytics system, US Patent App. 13/929,615, 2014.

[172] Kang S, Chien WK, Yang JG. A study for big-data (Hadoop) application in semiconductor manufacturing. In: Proceedings of the IEEE international conference on industrial engineering and engineering management. IEEE; 2016, p. 1893–7.