Dynamic Support Vector Regression Control System for Overlay Error Compensation With Stochastic Metrology Delay

Article *in* IEEE Transactions on Automation Science and Engineering · September 2019

DOI: 10.1109/TASE.2019.2935179

Marzieh Khakifirooz
Tecnológico de Monterrey

Dynamic Support Vector Regression Control System for Overlay Error Compensation With Stochastic Metrology Delay

Marzieh Khakifirooz, *Member, IEEE*, Chen-Fu Chien, *Member, IEEE*, and Ying-Jen Chen

***Abstract*—This study aims to develop a robust monitoring system for advanced control and compensation of the overlay errors based on ϵ-insensitive support vector regression (SVR), considering metrology delay. The proposed ϵ-insensitive SVR control system has the ability to solve quadratic optimization problems in real settings. To investigate the consistency and reliability of the proposed algorithm, a simulation study based on empirical data was conducted to validate the solution quality enhancement by the proposed approach. The stability of the system under metrology delay was investigated when Lyapunov stability function takes place as the kernel function of the ϵ-insensitive SVR optimization system. For sensitivity analysis, we compared and analyzed the effect of noise and time-varying metrology delay, within an online process with a simulation study based on empirical data. This approach can effectively reduce the misalignment of the overlay errors through the self-tuning process of ϵ-insensitive SVR and provide real-time decision aid for process engineers.**

***Note to Practitioners*—In practice, there are dynamic metrology delays that have not been adequately addressed. This study developed a robust monitoring system that can consider metrology delay for advanced control and effective compensation of the overlay errors. A study based on empirical data has validated the practical viability of the proposed approach. Indeed, the proposed algorithm can obtain a high degree of reliability for the measurement data in the complicated semiconductor fabrication process. Indeed, the developed solution is implemented in real practice.**

***Index Terms*—ϵ-insensitive support vector regression (SVR), intelligent manufacturing, Lyapunov mapping function, metrology delay, overlay error, stochastic time-delay system (TDS).**

Manuscript received January 13, 2019; revised June 23, 2019; accepted August 11, 2019. This article was recommended for publication by Associate Editor S. Mishra and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2634-F-007-001 and Grant MOST 108-2634-F-007-008 and in part by Micron Foundation. *(Corresponding author: Chen-Fu Chien.)*

M. Khakifirooz is with the Department of Industrial Engineering, Tecnológico de Monterrey, Monterrey 64849, Mexico.

C.-F. Chien is with the Department of Industrial Engineering and Engineering Management, National Tsing Hua University (NTHU), Hsinchu 30013, Taiwan, and also with the Artificial Intelligence for Intelligent Manufacturing Systems (AIMS) Research Center, Ministry of Science and Technology, Hsinchu 30013, Taiwan (e-mail: cfchien@mx.nthu.edu.tw).

Y.-J. Chen is with DALab Solutions x Associates (DALabx) Ltd., Hsinchu 30013, Taiwan.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org

Digital Object Identifier 10.1109/TASE.2019.2935179

I. Introduction

SEMICONDUCTOR manufacturing consists of a lengthy process for printing of multiple integrated circuit patterns in successive layers on the wafer through photolithography tools such as steppers or scanners that are capital-intensive and also the bottleneck for wafer fabrication [1]. The scanner is used to superimpose a masking pattern on the top of a wafer. Overlay errors denote the displacements between the actual position and the placement of the mask image over the wafer [2], [3]. The overlay errors require to be compensated rather than migrated. Migration leads to desired results when a potential risk is either properly segregated or controlled, while compensation minimizes the risks of accumulated error. Indeed, compensation is a preventive control. Therefore, as the tolerance of critical dimensions of integrated circuits is tightly and slightly increasing, effective compensation of the overlay errors is needed for advanced technology transformation and maintaining competitive advantages of leading semiconductor companies.

A variety of advanced process control (APC) and advanced equipment control (AEC) approaches have been developed to control and reduce process overlay errors for yield enhancement [4], [5]. The crafted APC approaches incorporated collections of methodologies, including statistical inferences [6], [7] machine learning [8], run-to-run (R2R) control [9], stochastic methods [10], and virtual metrology (VM) [11].

In addition to various feedback controllers based on exponentially weighted moving averages (EWMA) estimations, a number of alternatives techniques including matrix completion [12], artificial neural networks [13], image processing [14], and state-space control design [15] have been proposed to identify influential process variables for the overlay error. Furthermore, overlay error compensation has been used as feedforward control signals to support process control of other manufacturing steps including dry-etching or chemical-mechanical polishing [16].

The main concern in the design of control systems is linked to the flow of signals in the closed system. A number of models have been developed in control design based on the offline data obtained from experimentation. Only a few studies have dealt with real-time data-based control designs, while most of them have assumed a boundary in a compact set of state variables [17]. However, for the stability of the system in an online setting, system errors would be involved in the information, and hence a constant range for the data cannot be assumed. Focusing on real settings, this study aims to develop an approach using self-determined self-learning models from empirical data.

Due to the need for providing rapid feedback to the process control, the lack of real-time metrology data has caused extensive limitations in the R2R control. Generally, there is a gap to transfer measurement data between the metrology tools and the production line, which is called metrology delay. Most semiconductor manufacturing processes are suffering from issues caused by the metrology delays due to the time needed for measurements, metrology capacity, and the queueing time spending between the process tool and the metrology station [18]. The stability and performance of the control process are affected by the metrology delay. Moreover, in the real-time data-based control design, the delay is not fixed but flows stochastically.

On the other hand, tuning of control parameters rapidly and optimally is required for achieving an acceptable control performance in modern wafer fabrication facilities (fabs). However, most of the control models cannot update themselves due to the dynamic nature of the controlling process, and thus the system may suffer from modeling inaccuracies.

The interior design of a control system should be equipped with a function approximator that can minimize the total risk. However, most approximators, such as neural networks and polynomial estimators, are only minimizing empirical risks. The limited training set, compared with the number of free parameters, can cause a high generalization risk of overfitting. By minimizing the empirical risk, in combination with generalization risk, a better approximation technique for reducing the total risk called structural risk minimization (SRM) [19] can be used. Support vector regression (SVR) uses the principle of SRM [20].

This study used the modified SVR formulation as an optimization function to support the online control process. The modified version of SVR, instead of inequality constraints, takes equality constraints and a quadratic cost function into account, called ϵ-SVR [19]. The ϵ-SVR has a strong mathematical foundation with a high generalization ability and can find the global minimum for regression estimation problems, while avoiding the trap of finding the local minima. Several studies have used SVR to perform approximations of unknown nonlinear functions in adaptive control systems [21], [22]. However, only a few studies have extended the application of this method into the semiconductor industry. Song *et al.* [23] used the least square SVR to establish a quantitative analysis model for gas sensor components in a metal–oxide–semiconductor. Guo *et al.* [24] developed online sequential learning optimization based on ϵ-SVR optimization, incremental extreme learning machine, and proximal SVR for process monitoring of critical dimension in the dry-etching process.

This study proposes an approach for controlling overlay factors with dynamic metrology delay that is equipped with ϵ-SVR and a quadratic cost function for the feedback R2R controller to increase the capability for dealing with the needs for process control. The main contributions of this study are as follows:

1) Introducing ϵ-SVR optimization technique as a learning-based control system with disturbance rejection capability for compensating misalignment during the photolithography process of wafer fabrication.
2) Considering the real-time control process into a learning-based control system.
3) Improving the stability of the control system in the presence of unmeasurable disturbances (i.e., metrology delay, process delay, process variation, and measurement noise) by adjusting the kernel function of ϵ-SVR as stability rules.

The remainder of this article is organized as follows. Section II describes the fundamentals for this study. Section III describes the proposed on-line control approach based on ϵ-SVR. Section IV estimates the validity of the proposed approach with simulation and empirical data. Section V concludes this research with a discussion on contributions and future research directions.

II. Fundamentals

A. Overlay Error

The overlay errors are measured from the displacements between the present and previous exposure layers, through the box-in-box design [6]. The overlay errors can be attributed to intrafield and interfield errors [6]. The interfield overlay errors are the result of mismatch problems between the mask and the wafer. The intrafield overlay errors are due to fitment problems between the light source filter lens and the mask. The interfield overlay errors are measured at the center of the wafer, whereas the intrafield errors are measured with respect to the center of the exposure field.

Due to increases in dimensions of wafers, the variables affecting the overlay errors have become progressively complicated. In this study, the proposed overlay model by Chang *et al.* [2] has been adopted to estimate the overlay error of the scanner. The proposed overlay model in [2] considered 10 variables including mask-related intrafield overlay errors such as translation, rotation, and magnification, and wafer-related interfield overlay errors involving expansion and rotation, in the x-axis (d_{x+X}) and y-axis (d_{y+Y}), respectively, as follows:

$$d_{x+X} = T_{x+X} + S_X X - (\theta_\omega + \phi)Y + (M_i + M_a)x - (\theta_r + \theta_a)y + \varepsilon_{x+X}, \quad (1)$$

$$d_{y+Y} = T_{y+Y} + S_Y Y + \theta_\omega X + (M_i - M_a)y + (\theta_r - \theta_a)x + \varepsilon_{y+Y} \quad (2)$$

where (X, Y) and (x, y) denote the interfield and the intrafield coordinate systems, respectively.

The corresponding conventional models based on empirical data for models in (1) and (2) are presented by models (3) and (4), respectively,

$$\hat{d}_{x+X} = t_x + s_X X - r_X Y + m_x x - r_x y \quad (3)$$

$$\hat{d}_{y+Y} = t_y + s_Y Y + r_Y X + m_y y + r_y x. \quad (4)$$

The parameters of the overlay variables in (1) and (2) are derived from the coefficients of the models in (3) and (4), as summarized in Table I.

TABLE I
ESTIMATED COEFFICIENT PARAMETERS OF OVERLAY ERROR MODELS

Overlay error model parameters	Estimated coefficient
Translation in x-axis	$E(t_x) = T_{x+X}$
Translation in y-axis	$E(t_y) = T_{y+Y}$
Scale in x-axis	$E(s_X) = S_X$
Scale in y-axis	$E(s_Y) = S_Y$
Wafer rotation	$E(r_Y) = \theta_\omega$
Orthogonally	$E(r_X - r_Y) = \phi$
Isotropic Magnification	$E\left(\frac{m_x + m_y}{2}\right) = M_i$
Reticle rotation	$E\left(\frac{r_x + r_y}{2}\right) = \theta_r$
Asymmetric magnification	$E\left(\frac{m_x - m_y}{2}\right) = M_a$
Asymmetric rotation	$E\left(\frac{r_x - r_y}{2}\right) = \theta_a$

B. R2R Control With Metrology Delay

The implementation of APC in semiconductor manufacturing affected with an inherent problem known as metrology delay. Thus far, numerous APC techniques in the semiconductor industry have failed because of infrequent measurement data and extensive metrology delays. VM is deemed as the most popular technique and is a potential solution for overcoming these difficulties [25]–[27]. However, most of the existing studies focused on the improvements of the control performance of controllers rather than the stability properties of the control system when using VM methodology. Only few research has been done to investigate the ways of enhancing the stability of R2R-type controllers under measurement delay [28], [29], and only few research has addressed the dynamic nature of time delay in semiconductor manufacturing [23], [30].

To address the stochastic time delay in monitoring the overlay variables, this study designed two scenarios. The first scenario considered that the delay happened due to the time needed for measurements, metrology capacity, and the waiting time in the wafer queue between the production station and the metrology station along with the wafer-to-wafer (W2W) frequency. The W2W measurement scheme overlooks each wafer independently within a lot and needs sampling information from across the wafer to give rapid feedback for potential adjustments of process parameters for the subsequent wafers. Although integrated metrology [31] can help achieve a quick feedback, still there could be a lag of one to six wafer delay in receiving the feedback signals, using integrated metrology.

The second scenario considered real-time processing delay, due to the communications between the sensors and the controller, inside the controller, and between the controller and the actuator. In these cases, the process delays are assessed from when a signal is observed, to when the transformation takes place in the actuator. In addition, sometimes process delay happens because of bottleneck tools/processes in the system.

C. Dynamic Control System

The notations and terminologies for implementing the ϵ-SVR control system are listed as follows.

μ, λ	Parameters of zero-inflated poison (ZIP) distribution.
$\bar{x}, s^2$	Sample mean and sample variance of ZIP distribution.
$R_{\text{emp}}(.)$	The empirical risk function of the control system.
C(.)	Cost function for optimization objective.
u_t	Process input for run t, $t \geq 1$.
Q_t	Process output for run t, $t \geq 1$.
d_t	Process disturbance for run t, $t \geq 1$.
E_t	Deviation from the target for run t, $t \geq 1$.
T	Target of overlay variables.
ε_t	White noise for run $t, t \geq 1$.
$\boldsymbol{\omega}$	Vector of control system parameters.
$\boldsymbol{h}$	Mapping function.
$\mathbb{C}$	ϵ-SVR regularization parameter.
ϵ	ϵ-SVR epsilon parameter.
b	ϵ-SVR bias term.
$\kappa(.,.)$	ϵ-SVR kernel function.
$(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)$	ϵ-SVR support vector.
$\mathbf{x}_t$	State vector in the state-space model for run t.
$\boldsymbol{A}, \boldsymbol{B}$	Coefficient matrices in the state-space model.
M	Upper bound of the output variable.
σ	Admissibility parameter.
η	Learning rate parameter of online ϵ-SVR.
N	Number of samples per each run.
l, j	The length of process delay and measurement delay, respectively.
V_{Lyap}	Lyapunov stability function.
$\boldsymbol{P}$	Lyapunov positive definite symmetric matrix.
K	Number of folds for cross-validation.
i, t	Index for process run, $1 \leq i \leq t$.
k	Index for cross-validation folds, $1 \leq k \leq K$.
δ	Upper bound for Lyapunov–Razumikhin function.
m	Total number of runs.

The semiconductor manufacturing system is controlled mainly through the feedback loops. When a change in one process variable causes a dynamic variation in other variables, the system provides feedback loop to the origin step and makes a functional relationship or autocorrelation dynamics among process variables. This situation occurs quite often in semiconductor manufacturing. Therefore, dynamic approaches are required to be admitted for efficiently improving operational and performance monitoring of wafer fabrication.

The proposed dynamic model can compensate most of the process dynamics and noise disturbances in the semiconductor manufacturing process, such as multivari-

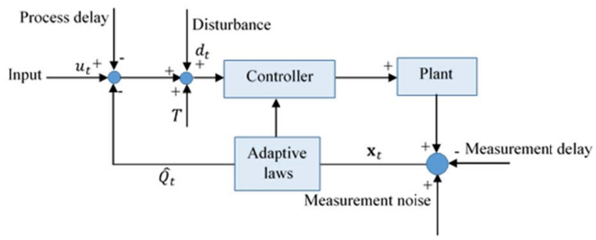

Fig. 1. Block diagram of the proposed control system.

ate process offset and process gain, the quadratic effect of process variables, autocorrelation and deterministic drifting effects, stochastic metrology delay, and nonstationary disturbances.

The proposed dynamic model that used to carry out the optimization and identification of empirical risk of approximation is described as follows:

$$\min\ R_{\text{emp}}(\boldsymbol{\omega}) = \frac{1}{t}\sum_{i=1}^{t} \mathrm{C}(T+\varepsilon_i, \hat{Q}_i)$$
$$\text{s.t. } \hat{Q}_t = \boldsymbol{h}(\hat{Q}_{t-1}, \ldots, \hat{Q}_1, u_{t-1}, \ldots, u_1, d_t)' \quad (5)$$

where both C(.) as the cost function and $\boldsymbol{h}(.,.)$ as the mapping function are assumed to be twice continuously differentiable. This model structure is intended for modeling a discrete-time dynamic system, with $\forall t : u_t, Q_t \in R$.

The block diagram of the proposed control system is illustrated in Fig. 1. According to Fig. 1, the input is the variation in each overlay variable from the target, which is affected by system delay, noise, and disturbance. The sensors continually and effectively measure the value of each overlay variable. The size of the actual variation which is an output of the manufacturing system compared with the input is augmented by the control plant in the fab.

The difference between the input and the output from the control device efficiently generated an impulse that activated the adaptive laws for implementing appropriate actions. In addition, uncertainties, including process time variance, measurement delays, measurement noise, and unknown disturbances, are variables which significantly affect the performance of the control system.

III. Proposed Approach

To address the overlay error problem in real settings, a dynamic ϵ-SVR controller is designed to implement corrective actions with links to feedback signals. Fig. 2 presents the general operational framework of the proposed control system. The proposed control system consists of the following criteria: 1) optimization analysis and providing the cost and constraints of the optimization problem for overlay error compensation; 2) an online ϵ-SVR algorithm for a real-time monitoring system; 3) stability analysis for learning algorithm; 4) stability analysis of the control system with delay; and 5) adjustment of the kernel mapping function for encompassing stochastic time delay.

A. Optimization Analysis

Consider a class of a single-in, single-out (SISO) system in the following form:

$$\hat{\mathbf{x}}_t = f(\mathbf{x}_1, \cdots, \mathbf{x}_{t-1}) + g(\mathbf{x}_1, \cdots, \mathbf{x}_{t-1})u_{t-1} + d_t$$
$$\hat{Q}_t = \hat{\mathbf{x}}_t \quad (6)$$

where unknown f and g functions are bounded and no prior knowledge is required for bounding. The state vector of the system assumes to be estimated through the optimization process of the control loop. To have a controllable system for the model in (6), the following assumptions are required:

Assumption 1: $\lim_{t\to\infty} E(\hat{Q}_t) = T + \varepsilon_t$

Assumption 2: $\lim_{t\to\infty} Var(\hat{Q}_t) < \infty$.

The control objective is designed to provide the control signal based on the system and an adaptation law for adjusting control parameters as illustrated in Fig. 1. Therefore, the state vector of the approximator function in (6) follows the desired trajectory state (target) even in the presence of disturbance. Consequently, the tracking error in (7) is aimed to converge to zero

$$E_t = \hat{Q}_t - (T + \varepsilon_t) \quad (7)$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon)$ and $E(\varepsilon_t)$ is a near-zero variance predictor. Indeed, ε_t is joined to the constant value of the target T, to model a stochastic target variable.

The cost function determines how adequately the given model is working with the noisy actual data. In this study, the cost function is selected as the quadratic ϵ-insensitive cost function as follows:

$$C_t = \begin{cases} 0, & \text{if } |\hat{Q}_t - (T+\varepsilon_t)| < \epsilon \\ (\hat{Q}_t - (T+\varepsilon_t))^2 - \epsilon, & o.w. \end{cases} \quad (8)$$

where $\epsilon \geq 0$.

A quadratic cost function is designed to make the convex quadratic cost function, and thus build a smooth strategy to determine the hyperparameters for SVR algorithm.

Considering a set of training points $\{(u_1 + d_1, T + \varepsilon_1), \ldots, (u_{t-1}+d_{t-1}, T+\varepsilon_{t-1})\}$, the ϵ-SVR is trained to map from the input space to the feature space in the presence of the disturbance, d_t. The kernel function $\kappa(u_t, \boldsymbol{u})$ handles the mapping role in the feature space.

Therefore, the linear equation in (9) is called the dual form of the optimization problem in (5) with regard to (8)

$$\min_{\alpha_i,\alpha_i^*} \frac{1}{2}\sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right)^T \left(\kappa(u_i, \boldsymbol{u}) + \frac{1}{\mathbb{C}}\right)\left(\alpha_i - \alpha_i^*\right)$$
$$+\epsilon \sum_{i=1}^{t}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{t}(T+\varepsilon_i)\left(\alpha_i - \alpha_i^*\right)$$
$$\text{s.t. } \sum_{i=1}^{t}\left(\alpha_i - \alpha_i^*\right) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq \mathbb{C} \quad (9)$$

where nonzero values of $(\alpha_t - \alpha_t^*)$ are called support vectors and are bounded by positive unknown number C, as the regularization parameter. The dual problem in (9) in terms of the kernel matrix is converted by introducing the Lagrangian

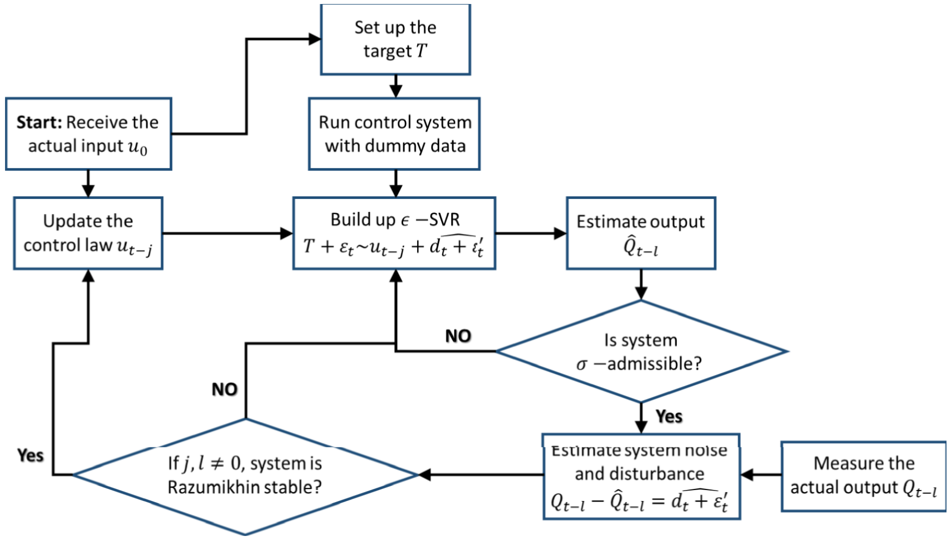

Fig. 2. Dynamic LS-SVR control system framework.

multiplier and applying the Karush–Kuhn–Tucker (KKT) conditions on a convex quadratic form of SVR model with affine constraints [20].

The main objective of the control system is to find the optimal control laws by the optimization model in (9), while the optimal control laws are used for minimization of the cost function. Therefore, regarding the kernel optimization method in (9), the corresponding approximation function in (6) should be estimated using the kernel function. Hence, the estimated output in (6) is

$$\hat{Q}_t = \sum_{i=1}^{t} \left(-\alpha_i + \alpha_i^*\right)\kappa(u_i, \boldsymbol{u}) + b \tag{10}$$

where $b = -\sum_{i=1}^{t}(-\alpha_i + \alpha_i^*)\kappa(u_i, \boldsymbol{u}) + T + \varepsilon_i - \epsilon$ is the bias term.

In linear system dynamic models in (5) and (6), the actual value of the input variable is unknown at the current run, before the metrology station forwards the value of process variables to the production station. Therefore, the input value of the current run, u_t, can be updated by the parameters of the approximator function at the last run as follows:

$$\hat{u}_t = u_{t-1} + \frac{E_{t-1}}{\frac{\partial \kappa(u_{t-1}, \boldsymbol{u})}{\partial u_{t-1}} \sum_{i=1}^{t-1}\left(-\alpha_i + \alpha_i^*\right)\kappa(u_i, \boldsymbol{u})}. \tag{11}$$

For details, refer to the Appendix.

B. Online ϵ-SVR Algorithm

With regard to the procedure of all the learning-based algorithms, including ϵ-SVR, the model first learns from training points and then evaluates by testing points. In addition, the key assumption is that the distribution for both training and test points is fixed over time, and that points are selected based on identically and independently distributed (*i.i.d.*) rule. However, due to the nature and design of a control system, an online algorithm can receive only one sample within each run. Therefore, with online learning, no distribution assumption is required.

On-Line ϵ −SVR $(\epsilon, \mathbb{C}, \eta)$

1 $\alpha \leftarrow 0$
2 $\alpha^* \leftarrow 0$
3 **For** $t \leftarrow 1$ **to** m **do**
4 Receive $\hat{u}_t$
5 $\hat{Q}_t = \sum_{i=1}^{t}(-\alpha_i + \alpha_i^*)\,\kappa(u_i, \boldsymbol{u}) + b$
6 Receive $(T + \varepsilon_t)$ and estimate $\mathbb{C}_t$
7 $\alpha_{t+1}^* \leftarrow \alpha_t^* + min\big(max\big(\eta\big[\big(T + \varepsilon_t - \hat{Q}_t\big) - \epsilon\big], -\alpha_t^*\big), \mathbb{C} - \alpha_t^*\big)$
8 $\alpha_{t+1} \leftarrow \alpha_t + min\big(max\big(\eta\big[\big(\hat{Q}_t - T - \varepsilon_t\big) - \epsilon\big], -\alpha_t\big), \mathbb{C} - \alpha_t\big)$
9 Randomly split the data into K disjoint sets
10 **For** $k \leftarrow 1$ **to** K **do**
11 **Return** $\hat{Q}_{t,k}$ **and** $\mathbb{C}_k$
12 **If** $\forall k \leftarrow 1$ **to** K: $|\mathbb{C}_k - \mathbb{C}_{k'}| \leq 2\sqrt{M}\big|\hat{Q}_{t,k} - \hat{Q}_{t,k'}\big|$ **then**
13 **Return** $\hat{Q}_{t+1}$ **and** $\hat{u}_{t+1}$
14 **End For**
15 **End For**

Fig. 3. Online algorithm of ϵ-SVR.

Fig. 3 gives a pseudocode of the online algorithm with arbitrary positive semidefinite (PSD) kernel function, while being in the presence of bias and disturbance. The general online setting for a control system involves t runs. At the tth run (step 3 in Fig. 3), the algorithm will receive the estimated input, $\hat{u}_t$, which is equivalent to $(u_t + \hat{d}_t)$ (step 4 in Fig. 3) and make the prediction of $\hat{Q}_t$ (step 5 in Fig. 3). Then, it can receive the true label $(T + \varepsilon_t)$ (step 6 in Fig. 3) and estimate the cost $_t$. The objective of online setting is to minimize the cumulative cost $\sum_{i=1}^{t}(T + \varepsilon_i, \hat{Q}_i)$ over all t runs [32].

The online version of the ϵ-SVR algorithm (steps 7 and 8 in Fig. 3) can be obtained by the application of a stochastic gradient descent with the dual-objective function of SVR (more details can be found in [33]).

To avoid inefficiencies due to sequential computations and speed up of the processing time, a parallel execution algorithm is developed to check the stability of optimal cumulative cost function. Steps 9–14 in Fig. 3 are related to stability checking of online learning algorithm which is discussed in Section III-C.

C. Local Stability for Learning Algorithm

The local stability of a function denotes the control limits that allow function changes on a time horizon. In particular, additional condition on the ϵ-SVM (support vector machine), controller is considered for imposing local stability of the closed-loop control system. Lipschitz continuity or $\sigma-$ admissibility condition [34] is used for local stability checking of online $\epsilon-$SVM.

The σ-admissibility condition is integrated through the cross-validation scenario to enhance the stability of the proposed online learning algorithm. In particular, the training data split in K folds, in which for each pair of folds, the admissibility condition is checked based on the following definition:

Definition 1: A cost function C is $\sigma-$ admissible with respect to the output class $\boldsymbol{Q}$ if there exists $\sigma \in R^+$ such that for any two outputs $Q_t{}'$, $Q_t'' \in \boldsymbol{Q}$ and for all label information $(T + \varepsilon_t)$

$$|\mathrm{C}(Q_t', T + \varepsilon_t) - \mathrm{C}(Q_t'', T + \varepsilon_t)| \leq \sigma |Q_t' - Q_t''|. \quad (12)$$

This assumption holds for the quadratic cost functions where the output set and the set of target values are bounded by some $M \in R^+$ such that

$\forall Q_t \in \boldsymbol{Q}, |Q_t| < M$ and $|T + \varepsilon_t| < M$.

Stability-based learning bound for quadratic ϵ-SVR is $\sigma-$ admissible with $\sigma = 2\sqrt{M}$ [34]. Through this condition, when an optimization method is applied, the constraints in (9) will particularly hold within a limited tolerance.

D. Global Stability for Control System

For stability analysis of a controller under time-delay systems (TDSs), Lyapunov method is applied in this study. Considering the stability analysis of linear TDS, the proposed approach is validated for all similar systems intending to minimize the quadratic cost function.

Considering the system in (6), the TDS model of (6) can be derived as follows [35]:

$$\mathbf{x}_t = \boldsymbol{A}\mathbf{x}_{t-1} + \boldsymbol{B}(f(\mathbf{x}_1, \cdots, \mathbf{x}_{t-j}) + g(\mathbf{x}_1, \cdots, \mathbf{x}_{t-j})u_{t-l} + d_t),$$
$$\hat{Q}_t = \mathbf{x}_t \quad (13)$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are known as constant state and input matrices, respectively. In addition, j and l denote the length of delay caused by the metrology tool and queuing system from the past process, respectively. The general theory of Lyapunov's method that can guarantee the stability of linear TDS [36] is described as follows.

Proposition 1: Consider the following Lyapunov function of the system in (13):

$$V_{\mathrm{Lyap}}(t) = \mathbf{x}_t^T \boldsymbol{P}\mathbf{x}_t. \quad (14)$$

The stability of the dynamic system in (13) can be granted if and only if $\boldsymbol{P}$ will be a PSD matrix and $(d)/(dt)V_{\mathrm{Lyap}}(t) < 0$.

The stability analysis of the model in (13) could be reduced to the analysis of the higher order system without delay if Lyapunov–Razumikhin theory [37] is applied to the system. The underlying condition of Razumikhin method for the ideal estimation of Lyapunov $\boldsymbol{P}$ matrix in (14) can be explained by the following proposition.

Proposition 2: Consider that the solution of the model in (13) begins inside the ellipsoid $V_{\mathrm{Lyap}}(t) = \mathbf{x}_t^T P\mathbf{x}_t < \delta$. At each time when the system receives an information by delay if the solution is going to leave this ellipsoid at some time $t' \geq t$, then Lyapunov–Razumikhin theory will check the condition and will not let the solution leave the ellipsoid region as long as it satisfies the following Razumikhin condition:

$$\mathbf{x}_{t'-j}^T \boldsymbol{P}\mathbf{x}_{t'-j} \leq \mathbf{x}_t^T \boldsymbol{P}\mathbf{x}_t. \quad (15)$$

or

$$V_{\mathrm{Lyap}}(t' - j) \leq V_{\mathrm{Lyap}}(t) < \delta, \quad 0 \leq j < t'. \quad (16)$$

E. Kernel Function of ϵ-SVR

To choose the kernel function for ϵ-SVR, Lyapunov condition of global stability as the kernel function is considered, while the mapping function is considered as linear.

Indeed, Lyapunov–Razumikhin condition can be used for stability checking, due to the linear mapping performance of Lyapunov function and its PSD nature. Bhatia and Elsner [38] have shown that Lyapunov function in (14) can be rephrased to say that its map is invertible. The following proposition states the conditions under which Lyapunov stability function characterizes the innerproduct particularity and can be used as the kernel function:

Proposition 3: Lyapunov inner product: suppose the stability property for the linear time-invariant (LTI) system at (13), under Lyapunov function. Therefore, for $\boldsymbol{P}$ as a PSD matrix

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{\boldsymbol{P}} = \mathbf{x}_t^T \boldsymbol{P}\mathbf{x}_t \quad (17)$$

defines an inner product. Therefore, the above inner product, induced by Lyapunov function, can be considered as the inner product of the system in (13), and subsequently for the optimization problem in (9) (see [39]).

This study used Lyapunov inner product as the kernel function for ϵ-SVR optimization. Without any loss of generality, the KTT conditions are still valid for this modification. Through this amendment, the convergence to the stability status under delay can be speeded up, while still keeping Razumikhin condition in the closed loop of the control system, which can facilitate the convergence of online learning of the ϵ-SVR controller.

IV. Performance Analysis

To estimate the validity of the proposed control system, the online ϵ-SVR controller proposed in Fig. 3 is used to estimate the input–output and the parameters of a control system, while the admissibility checking is conducted through a cross-validation scenario, and stability is carried out based on Lyapunov mapping operator as a kernel function of ϵ-SVR.

To run the simulation for the proposed control system, first, the input value is extracted from the empirical data with the target value equal to zero. However, for changing the constant target value to a variable, a random variable ε_t from $N(0, 10^{-15})$ is added to the constant target value

Fig. 4. Density probability plot of total overlay error on x- and y-axes (vertical blue dashed line is $T+\varepsilon_t = 0$).

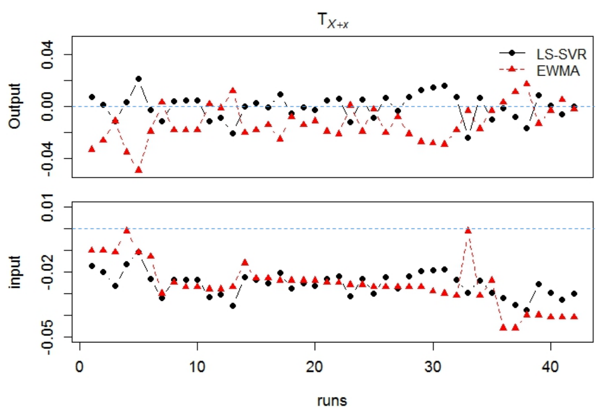

Fig. 5. Trend chart for T_{x+X} (horizontal blue dashed line is $T + \varepsilon_t = 0$).

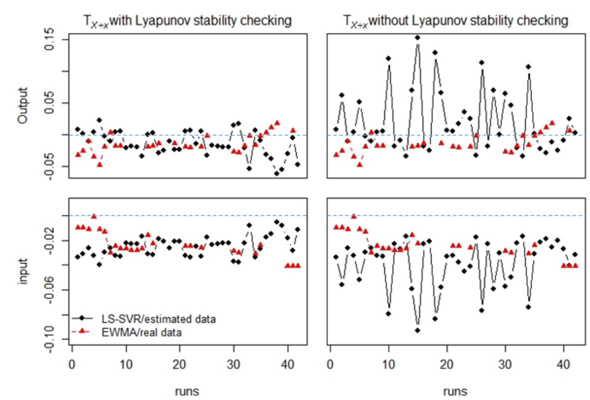

Fig. 6. Trend chart for T_{x+X} for the model with stochastic (normal) delay (horizontal blue dashed line is $T + \varepsilon_t = 0$).

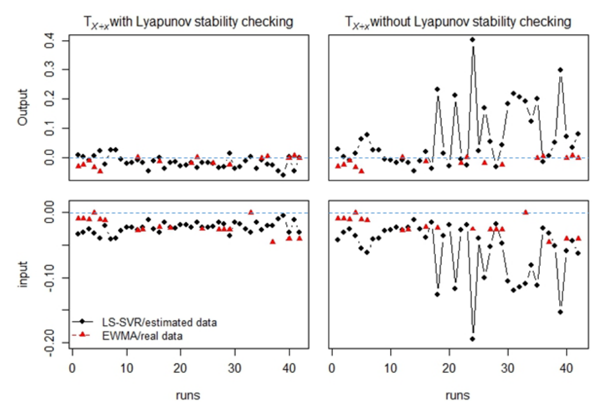

Fig. 7. Trend chart for T_{x+X} for the model with stochastic (fast and deep) delay (horizontal blue dashed line is $T + \varepsilon_t = 0$).

at each run. To prepare the minimum requirements of training data for learning algorithm of ϵ-SVR, a sequence of dummy data is generated based on the linear relationship of $T + \varepsilon_{t,1} \sim T + \varepsilon_{t,2}$.

The EWMA controller that is a common APC/R2R solution in practice is selected as a baseline for performance comparison.

To validate the performance of the proposed approach, the overlay error model in [2] and the sampling strategy suggested by Chien *et al.* [32] are integrated to estimate the corresponding overlay error on the x- and y-axes.

At each run, the cumulative overlay error measurements for the x- and y-axes are calculated for the performance comparison between the ϵ-SVR and EWMA controllers. The density probability plot is used to visualize the control performance for the cumulative overlay error compensation on the x- and y-axes. Fig. 4 illustrates this comparison for 42 runs. In summary, according to the illustration in Fig. 4, the proposed ϵ-SVR controller has archived an improvement of at least 24% and 8% of the variation and 96% and 97% of overlay error reduction for x- and y-axes, respectively.

This study used two metrics: range (18) and root-mean-squared error (RMSE) (19) to evaluate the variation in individual overlay variables for t runs

$$\text{Range} = \max_t \hat{Q}_t - \max_t \hat{Q}_t \tag{18}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{t} [\hat{Q}_i - (T + \varepsilon_i)]^2}{t}}. \tag{19}$$

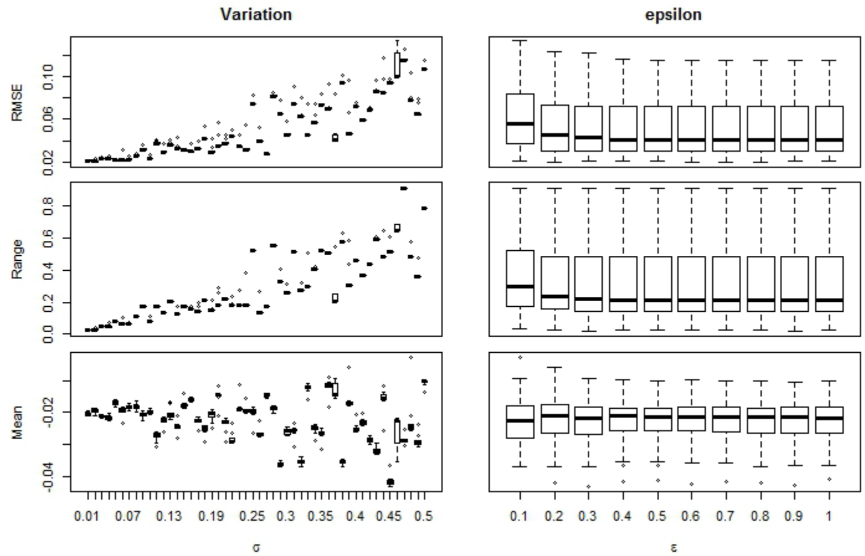

Fig. 8. Box plot for T_{x+X} for the model with stochastic (fast and deep) delay, random variation, and random ϵ.

TABLE II

VARIATION AND RANGE IMPROVEMENT FOR OVERLAY VARIABLES

Initial variation boundary	Variables	Estimated Parameters	Evaluation Tools	
			RMSE	Range
$(10^{-5}, 10^{-4})$	T_{y+Y}	Q_t	−2%	−5%
		u_t	5%	228%
$(10^{-4}, 10^{-3})$	T_{x+X}	Q_t	94%	46%
		u_t	6%	67%
	θ_ω	Q_t	427%	3%
		u_t	5%	91%
	θ_a	Q_t	−10%	−11%
		u_t	5%	172%
$(10^{-3}, 10^{-2})$	ϕ	Q_t	400%	1%
		u_t	5%	176%
	θ_r	Q_t	19%	5%
		u_t	0%	97%
$(10^{-2}, 10^{-1})$	S_Y	Q_t	37%	23%
		u_t	53%	235%
$(10^{-1}, 1)$	S_X	Q_t	67%	52%
		u_t	2%	132%
	M_i	Q_t	9%	16%
		u_t	1%	331%
	M_a	Q_t	14%	38%
		u_t	14%	97%

Fig. 5 represents the simulation result for the comparison between EWMA and ϵ-SVR for the input and output of the T_{x+X} variable. Consequently, ϵ-SVR has a smoother variation and a better compensation performance (e.g., lower variance and closer to target) than EWMA, given the input and output values of T_{x+X}, respectively.

To investigate the effect of disturbance on the system, plus having an overall overview of the performance of the proposed ϵ-SVR controller, the overlay variables are classified into four groups based on the variation in individual overlay variables. The results of the percentage improvement of the total variation and the range for all the input and output variables of the proposed ϵ-SVR compared with the EWMA controller are summarized in Table II. Variables in Table II are classified based on the variation in E_t, where E_t is calculated from empirical data and the result of the EWMA controller. The results have shown that the proposed ϵ-SVR controller tightens up the excellent performance bound and eventually achieves a lower cost in comparison to the EWMA control system, even in the presence of an extensive disturbance. The RMSE and range percentages in Table II indicate how much the result of ϵ-SVR is better than EWMA control. That is, when the variation increases, the advantages of ϵ-SVR are more tangible.

For investigating the performance of the proposed control system in the presence of a delay, this study designs two simulation scenarios. First, when delay happens during the W2W process control (process delay) for each of 25 runs with a maximum length of 6, and second, when delay occurs during the real-time processing (measurement delay). The duration of the delay is set to nine lags during the W2W process control for measurement delay. To illustrate the delay situations, random numbers are generated from the zero-inflated Poisson (ZIP) distribution, with parameters $\lambda = (s^2 + \bar{x}^2 - \bar{x})/(\bar{x})$, $\mu = (s^2 - \bar{x})/(s^2 + \bar{x}^2 - \bar{x})$, where s^2 denotes the sample variance, and $\bar{x}$ is the sample mean.

For the delay with a normal speed, the stochastic delay is generated randomly through ZIP distribution with the parameter setting at $\lambda = 2$ and $\mu = 0.6$. Similarly, for the fast and deep delay, the scenario is designed when the stochastic delay occurred at random from ZIP ($\lambda = 3$, $\mu = 0.3$).

To investigate the stability of the system under uncertainty, the kernel function of the ϵ-SVR controller is set in the form of Lyapunov function in (17). The adaptive control laws are added to the system as Razumikhin condition in (16). Figs. 6 and 7 illustrate the performance of the proposed model for normal and fast–deep delays for the overly variable T_{x+X}, respectively. The results have shown that under a fast and deep delay, while the stability of the system is more susceptible, Lyapunov mapping function and Razumikhin condition can perfectly grant the stability of the system.

In the above analysis, the parameter ϵ remains as a parameter that cannot freely be controlled by the process control designer, while due to the roles and designations of this parameter, it should remain fixed and controllable. Providentially, the positive definite Lyapunov kernel function can keep the ϵ parameter fixed and solve the optimization problem in a 2-D space when only parameters $\boldsymbol{P}$ and $\mathbb{C}$ are required to be estimated. To verify this fact, a scenario is designed for one simulated overlay variable from $N(0, \sigma)$ when σ is changed from 0.01 to 0.5 with a lag of 0.01, and the setting for the parameter ϵ is adjusted to the interval [0.1,1] with a lag of 0.1. This simulation runs under the fast and deep delay situation. Three statistics including range (18), RMSE (19), and the average of variation from zero are used to show the performance of parameter ϵ under the fast and deep delay. The results are illustrated in Fig. 8. The boxplots of ϵ variation in Fig. 8 can be used as a guideline to select the best setting for ϵ, considering the level of variation in the system. Although with a higher variation the system performance may lose its resistance, the variation in ϵ almost remains the same.

V. Conclusion

A. Discussion

This study has developed a novel, online ϵ-SVR controller, synthesized with the predictive feedback scheme and a self-tuning algorithm, to dynamically adjust the control parameters for compensating the overlay errors while effectively considering stochastic metrology delay.

In comparison to other regression-based methods such as Kernel ridge regression (KRR) or partial least square (PLS), Lagrange multipliers in ϵ-SVR use the concept of support vectors. The advantage of support vectors for a new test observation can eliminate the use of whole training data (such as the KRR), and only support vectors could be used for prediction. Therefore, the speed of prediction in ϵ-SVR is faster in comparison to other regression-based techniques.

The benefits of ϵ-SVR are also associated with its capability to support a variety of cost functions and the bias term. Indeed, ϵ-SVR is equivalent to KRR if ϵ in (8) is set to be zero. In addition, the quadratic cost function of ϵ-SVR brings an additional advantage in comparison to other SVR models such as least squares-support vector regression (LS-SVR), showing its robustness to non-Gaussian noise and disturbances. Furthermore, the proposed approach can also support the decisions for determining equipment backups [40] for maintaining productivity and yields for semiconductor manufacturing.

Another advantage of ϵ-SVR is the potential to describe nonlinear input–output relationships that some methods such as PLS may not be able to deal with it. However, the performance of ϵ-SVR for small data with multi-collinearity is not as promising as the PLS technique [41].

B. Concluding Remarks

As uncertain variables in the online control system such as delays may diminish the performance and stability of the control system, yet little research has been done to address this issue for controlling overlay errors. The proposed approach is equipped with a time-delay compensation method, based on Lyapunov kernel function, which can improve the practical viability of the ϵ-SVR controller. In particular, this study designed two types of experiments with regard to the length and speed of the delay, to estimate the validity of the model and examine the stability characteristics of the ϵ-SVR controller in practical situation. The experimental results have shown that the proposed approach has a robust and qualified performance than conventional EWMA approaches. Furthermore, according to the feedback of domain experts, this approach is viable for utilization in real setting in the presence of unmeasurable uncertainty for controlling the overlay errors during the photolithography process.

C. Future Research Directions

Future studies should be done to examine the practical performance of the proposed approach under various settings of uncertainties. While this study has considered stochastic time delay for compensating overlay errors, further research is needed to design the manufacturing system and the information infrastructure for effectively estimating potential time delay in the network control system to empower real-time decision for smart production. Further studies can be done to modify the configuration of the proposed approach and extend it to other semiconductor manufacturing processes such as dry-etching and chemical-mechanical polishing in different contexts.

The limitations of the existing approaches can be traced in part to the lack of a framework within which different uncertain factors can be considered in light of the dynamic nature of process changes for APC. Due to the complexity and emergence in process control of semiconductor industry, other learning-based optimization models can be investigated in future studies. Furthermore, due to more adoption of APC/AEC, R2R, and VM for smart production, the processing time in advanced wafer fabs has become uncertain. Since semiconductor manufacturing is a complicated flexible job-shop scheduling problem (FJSP) [43], future research should be done to minimize production cycle time while considering uncertain processing time and satisfying the precedence relationships of the jobs and other constraints. More studies should be done to empower the proposed hybrid strategy Industry 3.5 [44] that is based on the existing Industry 3.0 platform to address the needs of flexible decision and smart production for the visions of Industry 4.0.

Appendix

For the discrete-time state model in (5) without considering the disturbance effect, a Taylor expansion can be expressed by

$$Q_{t+1} = \boldsymbol{h}(Q_{t-1}, u_{t-1}) + \sum \frac{1}{r!} \frac{\partial^r [\boldsymbol{h}(Q_{t-1}, u_{t-1})]}{\partial u_{t-1}^r} (u_t - u_{t-1})^r + \sum \frac{1}{r!} \frac{\partial^r [\boldsymbol{h}(Q_{t-1}, u_{t-1})]}{\partial q_{t-1}^r} (\boldsymbol{q}_t - \boldsymbol{q}_{t-1})^r \tag{A.1}$$

where $\boldsymbol{h}(Q_{t-1}, u_{t-1})$ is the output at time $t-1$, and $\boldsymbol{q}_{t-1} = (Q_{t-1}, \ldots, Q_1)$. Here, $\boldsymbol{h}(Q_{t-1}, u_{t-1})$ is equivalent to Q_t.

The output Q_t is highly sensitive to the input u_t [42], that is,

$$\left|\frac{\partial[\boldsymbol{h}(Q_{t-1},u_{t-1})]}{\partial u_{t-1}}\right| \gg \left|\frac{\sum \frac{1}{r!}\frac{\partial^r[\boldsymbol{h}(Q_{t-1},u_{t-1})]}{\partial q_{t-1}^r}(\Delta q_t)^r}{\Delta u_t}\right| \quad \text{(A.2)}$$

where Δ is the increment operator. From this assumption, we can drop the third term on the right-hand side of (A.1) to represent the model (6) by

$$Q_{t+1} = Q_t + \frac{\partial[\boldsymbol{h}(Q_{t-1},u_{t-1})]}{\partial u_{t-1}}\Delta u_t + R(q_{t-1},u_{t-1})\Delta u_t. \quad \text{(A.3)}$$

For simplification, $(\partial[\boldsymbol{h}(Q_{t-1},u_{t-1})])/(\partial u_{t-1})$ is denoted by $h'(Q_{t-1})$. The remainder term $R(q_{t-1},u_{t-1})$ approaches zero at a faster rate than Δu_t approaches zero. Thus, the model can be derived by neglecting the remainder $R(q_{t-1},u_{t-1})\Delta u_t$ in (A.3) as follows:

$$Q_{t+1} \approx Q_t + h'(Q_{t-1})\Delta u_t. \quad \text{(A.4)}$$

We consider that $h'(Q_{t-1})$ exists and it is estimated using ϵ-SVR approach. From (A.4), the control law can be determined directly if the increment Δu_t behaves linearly, such that

$$\begin{aligned} u_t &= u_{t-1} + \Delta u_t \\ \Delta u_t &= \frac{Q_t^* - Q_t}{h'(Q_{t-1})} \end{aligned} \quad \text{(A.5)}$$

where Q_t^* is the desired trajectory (the target value).

The feedback control law in (A.5) will be practical if and only if $h'(Q_{t-1}) \neq 0$.

Now, considering ϵ-SVR estimates a new point as (10), the estimated function $h'(Q_{t-1})$ can be computed by the feedback control law as follows:

$$\begin{aligned} h'(Q_{t-1}) &= \frac{\partial\left[\sum_{i=1}^{t}\left(-\alpha_i+\alpha_i^*\right)\kappa(u_i,\boldsymbol{u})+b\right]}{\partial u_{t-1}} \\ &= \frac{\partial\kappa(u_i,\boldsymbol{u})}{\partial u_{t-1}}\sum_{i=1}^{t}\left(-\alpha_i+\alpha_i^*\right)\kappa(u_i,\boldsymbol{u}). \end{aligned} \quad \text{(A.6)}$$

With regard to (7), the control law can be determined as follows:

$$u_t = u_{t-1} + \frac{E_{t-1}}{\frac{\partial\kappa(u_{t-1},\boldsymbol{u})}{\partial u_{t-1}}\sum_{i=1}^{t-1}\left(-\alpha_i+\alpha_i^*\right)\kappa(u_i,\boldsymbol{u})}. \quad \text{(A.7)}$$

References

[1] C.-J. Kuo, C.-F. Chien, and C.-D. Chen, "Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 1, pp. 103–111, Jan. 2011.

[2] W. Chang *et al.*, "Statistical overlay error prediction for feed forward and feedback correction of overlay errors, root cause analysis and process control," U.S. Patent 15 123 980, Jan. 19, 2017.

[3] C.-F. Chien and C.-Y. Hsu, "UNISON analysis to model and reduce step-and-scan overlay errors for semiconductor manufacturing," *J. Intell. Manuf.*, vol. 22, no. 3, pp. 399–412, Jun. 2011.

[4] C.-F. Chien, Y.-J. Chen, and C.-Y. Hsu, "A novel approach to hedge and compensate the critical dimension variation of the developed-and-etched circuit patterns for yield enhancement in semiconductor manufacturing," *Comput. Oper. Res.*, vol. 53, pp. 309–318, Jan. 2015.

[5] M. Khakifirooz, C. F. Chien, and Y.-J. Chen, "Bayesian inference for mining semiconductor manufacturing big data for yield enhancement and smart production to empower industry 4.0," *Appl. Soft Comput.*, vol. 68, pp. 990–999, Jul. 2018.

[6] S. Lee and S. G. Kang, "A study of improving overlay accuracy by applying a new outlier handling method in FPD manufacturing," *Int. J. Prod. Res.*, vol. 53, no. 14, pp. 4249–4265, 2015.

[7] S. C. Horng, "Compensating modeling overlay errors using the weighted least-squares estimation," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 1, pp. 60–70, Feb. 2014.

[8] M. Khakifirooz, C.-F. Chien, and M. Fathi, "Compensating misalignment using dynamic random-effect control system: A case of high-mixed wafer fabrication," *IEEE Trans. Autom. Sci. Eng.*, to be published. doi: 10.1109/TASE.2019.2894668.

[9] C.-F. Chien, Y.-J. Chen, C.-Y. Hsu, and H.-K. Wang, "Overlay error compensation using advanced process control with dynamically adjusted proportional-integral R2R controller," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 473–484, Apr. 2014.

[10] Y. Jiao and D. Djurdjanovic, "Stochastic control of multilayer overlay in lithography processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 404–407, Aug. 2011.

[11] C.-F. Chien and Y.-J. Chen, "Manufacturing intelligence and smart production for industry 3.5 and empirical study of decision-based virtual metrology for controlling overlay errors," in *Proc. VLSI-DAT*, Apr. 2016, pp. 1–4.

[12] Z. Wang, M. Liu, and M. Dong, "Low-rank manifold optimization for overlay variations in lithography process," *J. Process Control*, vol. 62, pp. 11–23, Feb. 2018.

[13] H.-F. Kuo and A. Faricha, "Artificial neural network for diffraction based overlay measurement," *IEEE Access*, vol. 4, pp. 7479–7486, 2016.

[14] J. Park *et al.*, "Exact and reliable overlay metrology in nanoscale semiconductor devices using an image processing method," *Proc. SPIE*, vol. 13, no. 4, Oct. 2014, Art. no. 041409.

[15] F. He and Z. Zhang, "An empirical study-based state space model for multilayer overlay errors in the step-scan lithography process," *RSC Adv.*, vol. 5, no. 126, pp. 901–906, Oct. 2015.

[16] M. Khakifirooz, M. Fathi, and C.-F. Chien, "Modelling and decision support system for intelligent manufacturing: An empirical study for feedforward-feedback learning-based run-to-run controller for semiconductor dry-etching process," *Int. J. Ind. Eng., Theory, Appl., Pract.*, vol. 25, no. 6, pp. 828–842, Nov. 2018.

[17] D. Anberg, D. M. Owen, B.-H. Lee, S. Shetty, and E. Bouche, "A study of feed-forward strategies for overlay control in lithography processes using CGS technology," in *Proc. ASMC*, May 2015, pp. 395–400.

[18] R. P. Good and S. J. Qin, "On the stability of MIMO EWMA run-to-run controllers with metrology delay," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 1, pp. 78–86, Feb. 2006.

[19] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.

[20] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Inf. Process. Lett. Rev.*, vol. 11, no. 10, pp. 203–224, Oct. 2007.

[21] C. P. Chen, S. K. Tiong, J. D. Tan, S. P. Koh, and Y. C. Fong, "Online support vector based gas emission prediction system for generation power plant," *J. Fundam. Appl. Sci.*, vol. 10, no. 5, pp. 472–485, 2018.

[22] O. Naghash-Almasi and M. H. Khooban, "PI adaptive LS-SVR control scheme with disturbance rejection for a class of uncertain nonlinear systems," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 135–144, Jun. 2016.

[23] K. Song, Q. Wang, J. Li, and H. Zhang, "Quantitative measurement of gas component using multisensor array and NPSO-based LS-SVR," in *Proc. I2MTC*, May 2013, pp. 1740–1743.

[24] L. Guo, J. H. Hao, and M. Liu, "An incremental extreme learning machine for online sequential learning problems," *Neurocomputing*, vol. 128, pp. 50–58, Mar. 2014.

[25] W.-M. Wu, F.-T. Cheng, T.-H. Lin, D.-L. Zeng, and J.-F. Chen, "Selection schemes of dual virtual-metrology outputs for enhancing prediction accuracy," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 2, pp. 311–318, Apr. 2011.

[26] S. A. Lynna, N. MacGearailt, and J. V. Ringwood, "Real-time virtual metrology and control for plasma etch," *J. Process Control*, vol. 22, pp. 666–676, Apr. 2012.

[27] M.-F. Wu, C.-H. Lin, D. S.-H. Wong, S.-S. Jang, and S.-T. Tseng, "Performance analysis of EWMA controllers subject to metrology delay," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 3, pp. 413–425, Aug. 2008.

[28] R. Good and S.-J. Qin, “Performance synthesis of multiple input-multiple output (MIMO) exponentially weighted moving average (EWMA) run-to-run controllers with metrology delay,” *Ind. Eng. Chem. Res.*, vol. 50, no. 3, pp. 1400–1409, Oct. 2011.

[29] M. Jin and F.-G. Tsung, “Smith–EWMA run-to-run control schemes for a process with measurement delay,” *IIE Trans.*, vol. 41, no. 4, pp. 346–358, 2009.

[30] E. K. Boukas, and Z.-K. Liu, *Deterministic and Stochastic Time-Delay Systems*. Basel, Switzerland: Birkhauser, 2002.

[31] A. P. Shanmugasundram, H. Armer, and A. T. Schwarm, “Method, system and medium for process control for the matching of tools, chambers and/or other semiconductor-related entities,” U.S. Patent 7 082 345. Jul. 25, 2006.

[32] C.-F. Chien, K.-H. Chang, and C.-P. Chen, “ Design of a sampling strategy for measuring and compensating for overlay errors in semiconductor manufacturing,” *Int. J. Prod. Res.*, vol. 41, no. 11, pp. 2547–2561, 2003.

[33] S. Vijayakumar and S. Wu, “Sequential support vector classifiers and regression,” in *Proc. SOSO*, 1999, pp. 610–619.

[34] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundation of Machine Learning*. Cambridge, MA, USA: MIT Press, 2012, pp. 267–277.

[35] R. Shahbazi and M.-R. Akbarzadeh, “PI adaptive fuzzy control with large and fast disturbance rejection for a class of uncertain nonlinear systems,” *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 187–197, Feb. 2008.

[36] E. Fridman, *Introduction to Time-Delay Systems: Analysis and Control*. Basel, Switzerland: Birkhauser, 2014, pp. 243–272.

[37] S. Zhang and M.-P. Chen, “A new Razumikhin theorem for delay difference equations,” *Comput. Math. Appl.*, vol. 36, nos. 10–12, pp. 405–412, Nov. 1998.

[38] R. Bhatia and L. Elsner, “Positive linear maps and the Lyapunov equation,” in *Linear Operators and Matrices* (Operator Theory: Advances and Applications), vol. 130, I. Gohberg and H. Langer, Eds. Basel, Switzerland: Birkhauser, 2001, pp. 107–120.

[39] I. Kalashnikova, B. van Bloemen Waanders, S. Arunajatesan, and M. Barone, “Stabilization of projection-based reduced order models for linear time-invariant systems via optimization-based eigenvalue reassignment,” *Comput. Methods Appl. Mech. Eng.*, vol. 272, pp. 251–270, Apr. 2014.

[40] C.-F. Chien and C.-Y. Hsu, “A novel method for determining machine subgroups and backups with an empirical study for semiconductor manufacturing,” *J. Intell. Manuf.*, vol. 17, no. 4, pp. 429–439, Aug. 2006.

[41] R. Grbić, D. Sliškovic, and E. K. Nyarko, “Application of PLS and LS-SVM in difficult-to-measure process variable estimation,” in *Proc. IEEE 8th Int. Symp. Intell. Syst. Inform.*, Sep. 2010, pp. 313–318.

[42] X. Yuan, Y. Wang, and L. Wu, “Composite feedforward-feedback controller for generator excitation system,” *Nonlinear Dyn.*, vol. 54, no. 4, pp. 355–364, 2008.

[43] T. Jamrus, C.-F. Chien, M. Gen, and K. Sethanan, “Hybrid particle swarm optimization combined with genetic operators for flexible job-shop scheduling under uncertain processing time for semiconductor manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 32–41, Feb. 2018.

[44] C.-F. Chien, C.-W. Chou, and H.-C. Yu, “A novel route selection and resource allocation approach to improve the efficiency of manual material handling system in 200-mm wafer fabs for industry 3.5,” *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 4, pp. 1567–1580, Oct. 2016.

Marzieh Khakifirooz (M’19) received the M.S. degree in industrial statistics and the Ph.D. degree in industrial engineering and engineering management from National Tsing Hua University (NTHU), Hsinchu, Taiwan.

She is currently an Assistant Professor with the Department of Industrial Engineering, Tecnológico de Monterrey, Monterrey, Mexico. She has outstanding practical experience from various global consultancies for high-tech industries. Her current research interests include the application of optimization in smart manufacturing, Industry 4.0, decision-making, and machine learning.

Dr. Khakifiroozis is an active member of the System Dynamic Society and the Institute of Industrial and Systems Engineers (IISE).

Chen-Fu Chien (M’03) received the B.S. degree (Hons.) with double majors in industrial engineering and electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 1990, and the M.S. degree in industrial engineering and the Ph.D. degree in decision sciences and operations research from the University of Wisconsin–Madison, Madison, WI, USA, in 1994 and 1996, respectively. He attended the PCMPCL Executive Training from Harvard Business School, Boston, MA, USA, in 2007.

From 2002 to 2003, he was a Fulbright Scholar with the University of California at Berkeley, Berkeley, CA, USA. From 2005 to 2008, he was on-leave as the Deputy Director of Industrial Engineering Division, Taiwan Semiconductor Manufacturing Company (TSMC). He is currently a Tsinghua Chair Professor and a Micron Chair Professor with NTHU. He is also the Convener of the Industrial Engineering and Management Program, Ministry of Science and Technology (MOST), the Director of the Artificial Intelligence for Intelligent Manufacturing Systems (AIMS) Research Center of MOST, the NTHU–TSMC Center for Manufacturing Excellence, and the Principal Investigator for the MOST Semiconductor Technologies Empowerment Partners (STEP) Consortium. His current research interests include decision analysis, big data analytics, modeling and analysis for semiconductor manufacturing, and manufacturing intelligence. He has authored or coauthored more than five books, 170 journal articles, and 11 case studies in Harvard Business School. His book on Industry 3.5 (ISBN 978-986-398-380-4) that proposes Industry 3.5 as a hybrid strategy for emerging countries to migrate for intelligent manufacturing is one of the bestselling books in Taiwan. He holds ten U.S. invention patents on semiconductor manufacturing.

Dr. Chien was a recipient of the National Quality Award, the Executive Yuan Award for Outstanding Science and Technology Contribution, the Distinguished Research Awards, the Tier 1 Principal Investigator (top 3%) from MOST, the Distinguished University-Industry Collaborative Research Award from the Ministry of Education, the University Industrial Contribution Awards from the Ministry of Economic Affairs, the Distinguished University-Industry Collaborative Research Award, the Distinguished Young Faculty Research Award from NTHU, the Distinguished Young Industrial Engineer Award, the Best IE Paper Award, the IE Award from Chinese Institute of Industrial Engineering, the Best Engineering Paper Award and the Distinguished Engineering Professor by Chinese Institute of Engineers in Taiwan, the 2011 Best Paper Award for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and the 2015 Best Paper Award for the IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. He has been invited to give keynote lectures at international conferences including APIEMS, C&IE, FAIM, IEEM, IEOM, IML, ISMI, and leading universities worldwide.

Ying-Jen Chen received the Ph.D. degree in industrial engineering and engineering management (IEEM) from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2013.

He was a Post-Doctoral Researcher with the Artificial Intelligence for Intelligent Manufacturing Systems (AIMS) Research Center, Ministry of Science and Technology, Hsinchu, and the NTHU-TSMC Center for Manufacturing Excellence, Hsinchu. He is currently a Manager of DALab Solutions x Associates (DALabx) Ltd., Hsinchu. His current research interests include quality engineering, advanced process control (APC), data mining and big data analytics, and yield enhancement.

Dr. Chen was a recipient of the Best Paper Award at the CIIE Annual Meeting in 2010, 2012, and 2014 and the Best Presentation Award at the 2015 International Symposium on Semiconductor Manufacturing Intelligence (ISMI2105). His works have been published in the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, *Computers & Operations Research*, *Computers & Industrial Engineering*, and the *International Journal of Production Research*.