Controlling the Misalignment of Wafer Fabrication in Dynamic High-Mixed Manufacturing Environment

Article · April 2018

CITATIONS
0

READS
280

4 authors, including:

Marzieh Khakifirooz
Tecnológico de Monterrey
59 PUBLICATIONS **555** CITATIONS
SEE PROFILE

Panos Pardalos
University of Florida
1,728 PUBLICATIONS **48,323** CITATIONS
SEE PROFILE

Michel Fathi
University of North Texas
113 PUBLICATIONS **762** CITATIONS
SEE PROFILE

Compensating Misalignment Using Dynamic Random-Effect Control System: A Case of High-Mixed Wafer Fabrication

Marzieh Khakifirooz, *Student Member, IEEE*, Chen-Fu Chien, *Member, IEEE*, and Mahdi Fathi, *Member, IEEE*

***Abstract*— It is vital to have an exclusive modification in semiconductor production process because of meeting differentiated customer demands in dynamic and competitive global minuscule semiconductor technology market and the highly complex fabrication process. In this paper, we propose a control system based on the dynamic mixed-effect least-square support vector regression (LS-SVR) control system for overlay error compensation with stochastic metrology delay to minimize the misalignment of the patterning process. Moreover, for the stability of the control system in the presence of metrology delay and to deal with nonlinearity among the overlay factors, the novel Lyapunov-based kernel function is merged with the LS-SVR controller. The proposed controller's operation has been validated and implemented by a major semiconductor manufacturer in Taiwan. The experiments are verified that mixed-effect LS-SVR controller has the higher validity and higher efficiency in comparison with the exponentially weighted moving average (EWMA) and threaded EWMA controllers which had been previously implemented at the company or applied in similar studies.**

***Note to Practitioners*—Due to high production complexity in semiconductor manufacturing process, a meticulous and intelligent process control is needed to achieve higher throughput and customer satisfaction. Monitoring a complex system is challenging because the process components and variables operate autonomously and interoperate with other manufacturing segments. This paper proposes a novel run-to-run (R2R) control system to compensate the overlay error during the photolithography process that efficiently deals with the high-mixed manufacturing environment and metrology delay.**

***Index Terms*— High-mixed process, intelligent manufacturing, Lyapunov stability, metrology delay, overlay error, photolithography process, recipe-based system, support vector regression (SVR).**

Manuscript received revised May 6, 2018; accepted January 13, 2019. This paper was recommended for publication by Associate Editor Z. Yin and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 107-2634-F-007-002 and Grant MOST 107-2634-F-007-009. *(Corresponding author: Chen-Fu Chien.)*

M. Khakifirooz and C.-F. Chien are with the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan, and also with the Artificial Intelligence for Intelligent Manufacturing Systems Research Center, Ministry of Science and Technology, Taipei 10622, Taiwan (e-mail: khakifirooz.marzieh@gapp.nthu.edu.tw; cfchien@mx.nthu.edu.tw).

M. Fathi is with the Department of Industrial and Systems Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: fathi@ise.msstate.edu).

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

Digital Object Identifier 10.1109/TASE.2019.2894668

I. Introduction

SEMICONDUCTOR devices are continuing to shrink in size so that process engineers and researchers are facing these issues daily that how they can adopt a more authentic monitoring system to get rid of receiving intensive out of control errors and enhance the yield [1], [2]. More critically, the lithography process deals with entire new difficulties attached to development in post-lithographic technologies. Consequently, the factory integration team is required to investigate challenges to ensure infrastructure readiness for the lithography process and improve the advanced process control (APC) with the tighten control limits. Therefore, the overlay error of the lithography process is selected for further investigation in this study.

Run-to-run (R2R) control has been extensively adapted to analyze a variety of challenges in the process monitoring of complex semiconductor manufacturing (see [2]–[4]). For more details, one can refer to comprehensive reviews of studies on APC methods (the Kalman filtering technique [5], stochastic sequential optimization [6], artificial neural network [7], and feedforward-feedback learning-based controller [8]).

Misalignment in photolithography process is studied in several papers [9]–[11]. Also, several papers considered mathematical modeling for overlay alignment [12]–[14]. Additional prior reviews of overlay error compensation in photolithography process appeared in [15]–[18]. A research of literature revealed several studies that design a proper APC system for high-mixed semiconductor plant such as threaded exponentially weighted moving average (tEWMA) [19], a combined product and tool disturbance estimator [20], and a cycle forecasting EWMA controller [21].

Due to the need for the provision of rapid feedback to the process control, the lack of real-time metrology data causes extensive limitations in the R2R control. Most semiconductor manufacturing processes suffer from issues caused by metrology delays due to the time needed for measurements, metrology capacity, and the waiting time in the wafer queue between the processing tool and the metrology station [22]. The stability and performance of the process will be affected by the metrology delay. Moreover, since quality measurements perform online, the delay would not be fixed but flows stochastically.

On the other hand, tuning of control parameters quickly and optimally is extensively required to achieve an acceptable control performance for intelligent manufacturing of modern fabs. However, most of the control models cannot update autonomously, as the dynamics of the system vary during online control, and thus it may endure from modeling inaccuracies. A function approximator should minimize the total risk, yet most approximators such as neural networks and polynomial estimators minimize empirical risks. The limited training set compared to the number of free parameters, can cause a high generalized risk of overfitting. By minimizing the empirical risk, in combination with generalized risk, a more efficient approximation technique for reducing the total risk called structural risk minimization (SRM) [23] can be obtained. The SRM technique implements the support vector regression (SVR). Both the optimal control problems and SVR methods are a type of optimization models. Hence, one could try to merge these two formulations.

In this study, we aim to develop the multiple-input single-output (MISO) controller based on least-square SVR (LS-SVR) to minimizing the unmeasurable disturbance affected by the stochastic delay from metrology tools to fabrication tools and process noise. The LS-SVR controller [24] has gained popularity due to its promising performance in minimizing the regret function. Several studies have adopted the SVR method for monitoring dynamic multiple nodes process [25].

The contributions of proposed LS-SVR method in this study are: 1) to set up the tuned parameters of the LS-SVR controller for the high-mixed recipe system; 2) to deal with the unmeasurable delay and disturbance during the lithography process; 3) to compensate the misalignment of overlay factors in the high-mixed environment; and 4) to investigate the stability of the system in the presence of stochastic delay and deal with nonlinearity among the overlay factors.

The remainder of this paper is organized as follows: Section II introduces the MISO system framework for overlay factors. Section III presents the recipe-based control system in semiconductor manufacturing with a discussion of properties of random effect LS-SVR controller including the Lyapunov-based polynomial-kernel function. Section IV demonstrates the simulation experiments for a high-mixed plant and analyzes the sensitivity of the proposed LS-SVR controller. Section V includes analysis of manufacturing data. Finally, we conclude this paper in Section VI.

II. Fundamentals

A. Notation and Terminologies

The notation and terminologies used in this paper are listed as follows.

i	Recipe index.
j	Overlay factor index.
t	Process run index.
k	Fold index for cross validation.
N	Number of overlay factors in the system.
K	Number of fold for cross validation.
c_j	Indicator for controller of overlay factor j.
p_j	Indicator for plant of overlay factor j.
R_i	Indicator for recipe i.
u_{tj}	Input variable for overlay factor j at run t.
Q_{tj}	Process output for overlay factor j at run t.
d_{tj}	Process disturbance for overlay factor j at run t.
E_{tj}	Deviation from the target for overlay factor j at run t.
z_{tj}	Random effect for overlay factor j at run t.
T	Target of overlay factors.
ε_t	White noise in a process at run t.
τ	Process gain (parameter of EWMA/tEWMA controller).
a	Parameter of EWMA/tEWMA controller.
θ	Fixed discount factor in EWMA/tEWMA controller.
$\mathbf{x}_{tj}$	State vector of overlay factor in state-space model for overlay factor j at run t.
$C(.)$	LS-SVR cost function from optimization.
$\mathbb{C}$	LS-SVR regularization parameter.
b	LS-SVR bias term.
$\kappa(.,.)$	LS-SVR kernel function.
$(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)$	LS-SVR support vector.
$u_{t(j)}$	Input variable for overlay factor j with highest E_t at run t.
$Q_{t(j)}$	Process output for overlay factor j with highest E_t at run t.
$C_{t(j)}(.)$	Cost function for for overlay factor j with highest E_t at run t.
R_{emp}	Empirical risk function of control system.
$f, g, \mathbf{h}$	Mapping functions in stat-space model.
δ_{t_i}	Stochastic drift for recipe i at run t.
M	Upper bound for total overlay error.
σ	Admissibility parameter.
η	LS-SVR online learning parameter.
V_{Lyap}	Lyapunov stability function.
$\mathbf{P}, \mathbf{P}^*$	Lyapunov positive definite symmetric matrices.
λ, γ	Parameters of Zero-Inflated Poison (ZIP) distribution.
$\mathbf{I}$	Identity matrix/vector.
m, n	Intercept and power parameters for polynomial kernel function.
$\boldsymbol{\mu}$	Mean vector.
$\boldsymbol{\Sigma}$	Variance covariance matrix.

B. Multiple-Input Single-Output Control System

A system in which multiple inputs are used to govern a single output is called a MISO system. Regards to the communication status among input variables, the MISO system can carry noncollaborative and collaborative control strategies. The noncollaborative control plan is equivalent to a single-input-single-output (SISO) system with a single feedback loop for the most critical input variable [26]. On the other hand, the collaborative system, either in serial or parallel structure [27], benefits from the dynamic status of each plant to improve the performance of monitoring system considering the restrictions of process inputs and individual outputs. This paper designs a customized serial collaborative MISO controller considering overlay factors' relations and attitudes.

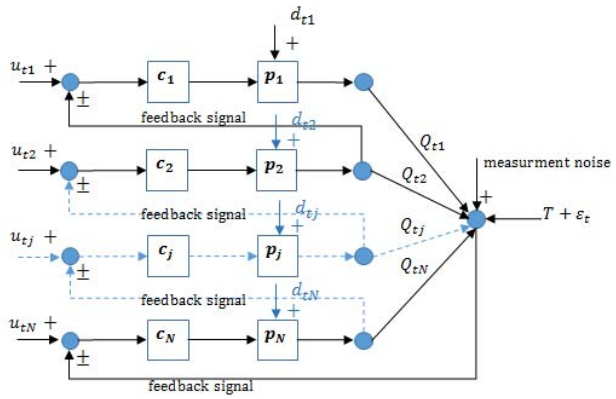

Fig. 1. MISO control system for overlay error compensation.

A regular linear ten-factor overlay model used by [17] is contemplated in this paper. Regarding the model in [17], although ten factors were investigated to compensate for the overlay error, the target point for all factors remains the same. Therefore, the MISO model is well suited for our study considering the collaboration and dependency among all overlay variables.

C. MISO Controller for Overlay Factors

Assume an R2R-MISO system where plants are arranged decreasingly based on the deviation of control outputs and the target point of the last run. The procedure of a cascade control system [28] is well used for controlling such dynamic control system where the system can be dynamic as a mix of linear or nonlinear plants. However, the linearity or nonlinearity remains fix at the entire process. Generally, a cascade control system consists of at least two control loops, at least one inner loop and one outer loop for closed-loop system operation. Load disturbance that enforces into the inner loop(s) can prevent before it extends to the whole system (outer loop). Therefore, if the inner loop contains a significant disturbance, the system can react and compensate the disturbance faster before the outer loop run or the whole system will be affected.

Consider Fig. 1, including N input variables, each serial plant p_j, $j = 1, \ldots, N$ at tth run, defines the effect of each manipulated variable u_{tj} over the corresponding estimated output $(\widehat{Q}_{tj})$. The nonmeasurable disturbance d_{tj} deviates each output variables from their corresponding control plant p_j. A set of serial controllers c_j are designed to minimize such deviation. Each c_j is updated based on a lower level measurement output Q_{tj+1}, which is corrupted by the noise signal. The level or order of each serial plant and controller is dynamic at $(t+1)$th run according to the error function in (1) at tth run, in a way that overlay factors with higher deviations have priority to monitor at the next run

$$\mathbf{E}_t = \mathbf{Q}_t - \mathbf{I}(T + \varepsilon_t) \tag{1}$$

where $\mathbf{I}$ is a identity vector.

D. Research Framework for the Proposed R2R-MISO Control System

Algorithm 1 presents the general operational structure of the proposed dynamic monitoring system in this paper. The initial analysis criteria of the presented control system are: 1) optimization analysis; 2) providing the cost and constraints of the optimization problem; 3) the multiple input nodes estimation; and 4) stabilization of the estimated results.

Algorithm 1: R2R-MISO

for $l : 1 \rightarrow t$ **do**
  Receive $(\boldsymbol{u}_l + d_l)$ for N overlay factors;
  for $j : 1 \rightarrow N$ **do**
    Create $(u_{l(j)} + d_l)$;
    Estimate $\widehat{Q}_{l(j)}$;
    Receive $(T + \varepsilon_l)$;
    Update $c_{(j)}$ controller by $\widehat{Q}_{l(j+1)})$;
    if $\left|C_{(j)} - C_{(j+1)}\right| \leq \varepsilon_l$ **then**
      $\widehat{Q}_{l(j+1)}, \ldots, \widehat{Q}_{l(N)} = \widehat{Q}_{l-1(j+1)}, \ldots, \widehat{Q}_{l-1(N)}$;
      Break loop;
    end
  end
end
Randomly split the data into K disjoint set;
if $\forall k : 1 \rightarrow K : \Pr\left(\sup\left|C_{(j)k} - C_{(j+1)k'}\right| \leq \sigma\right) \geq 1 - \eta$ **then**
  Return $\widehat{u}_{l(j)}$, $\widehat{Q}_{l(j)}$ and $C_{(j)}$
end

Several other phenomena including the delay of information flowing from the metrology tool to the control plant, and being a bottleneck make severe disturbances. Furthermore, at the bottleneck, multiple recipes are associated with the single machine. This situation can be more complicated when multiple products are manufactured during a specific time, and the production line schedule is a mixed schedule. To arrange this complication, a straightforward mixed system as a guideline for advanced high-mixed fab has investigated.

III. Recipe-Based Control System for Semiconductor Manufacturing

This section introduces the mixed-recipe control system based on the LS-SVR controller. The nonlinearity among the overlay factors is also considered as another challenging issue. Assume the recipe-based controller based on a mixed-recipe system, with different levels of complexity as shown in Figs. 2–4. Fig. 2 shows a particular case of a mixed-recipe schedule where each recipe is periodically applied to the process after a certain number of runs, δ (fixed drift). Within a fixed drift recipe schedule system, the subsequent run's control action is based on the previous run when the same recipe was on that tool. If the recipe schedule is random (stochastic drift, δ_t) (Fig. 3), the subsequent run's control action is controlled based on the output of the previous run, when the same recipe and drift were on that tool. The system becomes even more complicated with a dynamic schedule and in the presence of other recipes in the system, as shown in Fig. 4.

In this paper, we consider a system with high-mixed recipe schedule with a random drift similar to Fig. 4. The random drift recipe schedule shown in Fig. 4 considers as the random

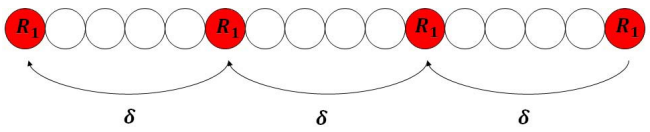

Fig. 2. Mixed-recipe schedule system with fixed drift.

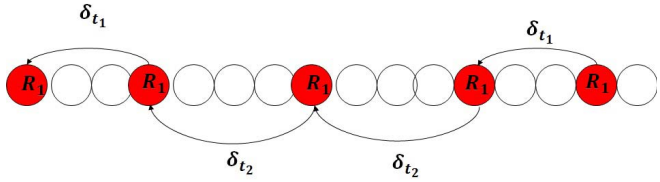

Fig. 3. Mixed-recipe schedule system with random drift.

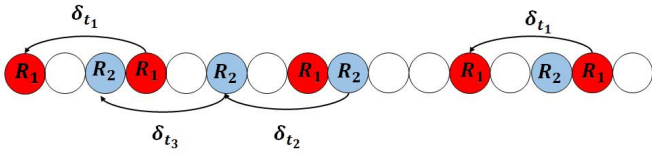

Fig. 4. Mixed-recipe schedule system with random drift and random sequence.

intercept or random offset into the linear or nonlinear regression model. Therefore, to optimize the performance of the controller, a random effect LS-SVR algorithm has been applied when the polynomial kernel function based on Lyapunov stability condition is employed as a mapping function to enhance the performance of the proposed controller.

A. Review of LS-SVR

The dynamic model proposed in this paper compensates for the majority of process dynamics and noise-based disturbances, such as gain and offsets of multivariate processes [29], [30], the quadratic effects of process variables [31], autocorrelation and deterministic drifting effects [32], stochastic metrology delay [33], and nonstationary disturbances [34]. For these purposes, consider a set of training points $\{(\mathbf{u}_1+\mathbf{d}_1, T+\varepsilon_1), \ldots, (\mathbf{u}_{t-1}+\mathbf{d}_{t-1}, T+\varepsilon_{t-1})\}$. The model carries out the optimization and identification of the empirical risk of approximation as follows:

$$\begin{aligned} &\min\ R_{\text{emp}} = \frac{1}{t}\sum_{l=1}^{t} C_l(T+\varepsilon_l, \widehat{\mathbf{Q}}_l) \\ &\text{s.t. } \widehat{\mathbf{Q}}_t = \mathbf{h}(\widehat{\mathbf{Q}}_{t-1}, \ldots, \widehat{\mathbf{Q}}_1, \mathbf{u}_{t-1}, \ldots, \mathbf{u}_1, \mathbf{d}_{t-1}) \end{aligned} \tag{2}$$

where both $C_t(.)$ and $\mathbf{h}(.,.)$ are assumed to be twice continuously differentiable. This model structure is intended for modeling a dynamic system with input $\mathbf{u}_t \in R^n$ and output $\mathbf{Q}_t \in R^n$.

The control objective is designed to provide the control signal, based on the system and an adaptation law, for adjusting control parameters. Therefore, the state vector of the approximator function in the presence of an external disturbance follows the desired trajectory state (Target). The tracking error in (1) converges to zero, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2 \approx 10^{-15})$. Appropriately, ε_t joins to the constant value of target T to model a stochastic target variable.

Now, subject to the equality constraint in (2), consider a class of a MISO system in the following form:

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}) + g(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})\mathbf{u}_{t-1} + \mathbf{d}_{t-1} \\ \widehat{\mathbf{Q}}_t &= \mathbf{x}_t \end{aligned} \tag{3}$$

where unknown f and g functions are bounded and no prior knowledge is required for bounding. The state vector of the system can be estimated through the optimization process of the control loop. To have a controllable system for the model mentioned in (3), the following assumptions are considered:

Assumption 1: An external disturbance is required to be bound by an unknown constant, which is equivalent to $\lim_{t\to\infty} E(\widehat{\mathbf{Q}}_t) = T + E(\varepsilon_t) = T$.

Assumption 2: It is required $g(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$ to be positive, which is equivalent to $\lim_{t\to\infty} Var\ (\widehat{\mathbf{Q}}_t) < \infty$.

In this paper, to optimize the regret function in (2) regarding the state-space model in (3), the LS-SVR is trained to map from the input space to the feature space in the presence of the disturbance $\mathbf{d}_t$. The given kernel function $\kappa(u_t, \mathbf{u})$ handles the mapping model in the feature space. The linear equation in (4) is called the LS solution of (2) (for details, see [35])

$$\begin{bmatrix} 0 & \mathbf{I}' \\ \mathbf{I} & \kappa(u,u') + \dfrac{\mathbf{I}}{\mathbb{C}} \end{bmatrix} \begin{bmatrix} 0 \\ (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Q} \end{bmatrix}. \tag{4}$$

In the optimal control problem, the aim is to solve the problem (2). To find the optimal control law, one could construct the corresponding approximation function of (3) by using the kernel function as follow

$$\widehat{\mathbf{Q}}_t = \sum_{l=1}^{t} \big(-\alpha_l + \alpha_l^*\big)\kappa(u_l, \mathbf{u}) + \widehat{b} \tag{5}$$

where $\widehat{b} = -\sum_{l=1}^{t}(-\alpha_l + \alpha_l^*)\kappa(u_l, \mathbf{u}) + T + \varepsilon_l$, and

$$C_t = (\widehat{\mathbf{Q}}_t - (T+\varepsilon_t))^2 \tag{6}$$

is the cost function.

B. Random Effect LS-SVR

Now, consider a set of training points $\{(u_{1j} + d_{1j}, T+\varepsilon_1), \ldots, (u_{t-1j} + d_{t-1j}, T+\varepsilon_{t-1})\}$. Based on the theory of best linear unbiased prediction [36], the optimization model in (2) with a random effect for jth factor is formulated as

$$\begin{aligned} &\min\ R_{\text{emp}} = \frac{1}{t}\sum_{l=1}^{t} C_l(T+\varepsilon_l, \widehat{Q}_{lj}) \\ &\text{s.t. } \widehat{Q}_{tj} = \mathbf{h}(\widehat{Q}_{t-1j}, \ldots, \widehat{Q}_{1j}, u_{t-1j}, u_{t-1j} \ldots, \\ &\qquad z_{t-1j}, \ldots, z_{1j}, d_{tj}). \end{aligned} \tag{7}$$

When including the random effects into the LS-SVR model, the single regularization parameter changes into two regularization parameters with one parameter for random error and one parameter for random effect. The random effect parameter vector has $N(0, \Sigma_{z_{tj}})$ and the error vector $N(0, \Sigma_{C_t})$ distribution where $\Sigma_{z_{tj}}$ and Σ_{C_t} are the covariance matrices and are

known or could be estimated. Hence, the corresponding dual form of (7) represents the model (8) with random effect (see the Appendix)

$$\begin{bmatrix} 0 & \mathbf{I}' \\ \mathbf{I} & \kappa(u_t,\mathbf{u})+\dfrac{\mathbf{I}}{\mathbb{C}_1}z'_{t_j}\Sigma^{-1}_{z_{tj}}z_{tj}+\dfrac{\mathbf{I}}{\mathbb{C}_2}\Sigma^{-1}_{C_t} \end{bmatrix}\begin{bmatrix} 0 \\ (\boldsymbol{\alpha}-\boldsymbol{\alpha}^{*}) \end{bmatrix}=\begin{bmatrix} 0 \\ \mathbf{Q} \end{bmatrix}. \tag{8}$$

The solution for linear model in (8) optimizes the bias $\widehat{b}$ and the support vector $(\alpha_{lj}-\alpha^{*}_{lj})$. Then, the optimal regression function for the given $\mathbf{u}$ and $\mathbf{z}$ is obtained by

$$\widehat{\mathbf{Q}}_t=\sum_{l=1}^{t}\sum_{j=1}^{N}\big(-\alpha_{lj}+\alpha^{*}_{lj}\big)\kappa(u_{lj},\mathbf{u}) +\frac{\mathbf{I}}{\mathbb{C}_1}\sum_{l=1}^{t}\sum_{j=1}^{N}\big(-\alpha_{lj}+\alpha^{*}_{lj}\big)\Sigma^{-1}_{z_{lj}}z_{lj}+\widehat{b}. \tag{9}$$

In the system dynamic model in (3), the input value of the current run $\mathbf{u}_t$ will be updated by (for more details, see [37])

$$\widehat{u}_{tj}=u_{t-1\,j}+\frac{E_{t-1\,j}}{\frac{\partial\kappa(u_{t-1\,j},\mathbf{u})}{\partial u_{t-1\,j}}\sum_{l=1}^{t}\big(-\alpha_{lj}+\alpha^{*}_{lj}\big)\kappa(u_{lj},\mathbf{u})}. \tag{10}$$

C. Stability of Random-Effect LS-SVR

The basic idea of stability is that the result of a learning-based system with a full sample should not be very different from the result obtained by removing only one observation. More precisely, for any two subsets of empirical data, the Euclidean distance between the corresponding loss function should be bounded by $M \geq 0$.

σ-admissibility [38] condition is considered to impose the stability of a robust algorithm through the learning-based cross-validation scenario.

Definition 1: σ-admissibility condition, a cost function C_t is σ-admissible with respect to the output class $\mathbf{Q}$ if it will be differentiable almost everywhere and there exists $\sigma \in R^{+}$ such that for any two outputs $Q'_{tj}, Q''_{tj} \in \mathbf{Q}_j$ and all label information $(T+\varepsilon_t)$

$$\big|C_t\big(Q'_{tj},T+\varepsilon_t\big)-C_t\big(Q''_{tj},T+\varepsilon_t\big)\big|\leq\sigma\big|Q'_{tj}-Q''_{tj}\big|. \tag{11}$$

This assumption holds for the quadratic cost functions where the set of output and target values is bounded by $M \in R^{+}$: $\forall Q_{tj}\in\mathbf{Q}_j, |Q_{ij}|<M$ and $|T+\varepsilon_t|<M$.

The LS-SVR model with quadratic cost function meets σ-admissible condition with $\sigma=2$ for any two outputs $Q'_{tj}, Q''_{tj}\in\mathbf{Q}_j$ [38, Corollary 11.1, p. 255].

D. Nonlinear Lyapunov-Based Kernel Function

Inserting a time delay in a loop of the control system causes a reduction in the performance. Moreover, larger values of time delays change the stability of the system. For a stability analysis of time-delay systems (TDSs), the Lyapunov method [39] is known as the most efficient technique. The following Lyapunov theory is validated for all linear TDSs.

Definition 2: Assume the following Lyapunov function of the system in (3)

$$V_{\mathrm{Lyap}}(t)=\mathbf{x}_t^{T}\mathbf{P}\mathbf{x}_t \tag{12}$$

the stability of the linear form of dynamic system in (3) can guarantee if and only if

$$\forall\mathbf{A}\in\mathbf{x}_t=\mathbf{A}\mathbf{x}_{t-1}+\mathbf{B}\mathbf{u}_{t-1}$$

$$A_{11}\mathbf{P}+\mathbf{P}A_{11}=-\mathbf{P}^{*} \tag{13}$$

where $\mathbf{P}$ and $\mathbf{P}^{*}$ are positive symmetric definite matrices and $(\partial/\partial t)V_{\mathrm{Lap}}(t)<0$. For the feedforward control system, $\mathbf{P}^{*}=\mathbf{I}$ [40].

For the stability condition of a control system, the characterization of kernel functions and matrices is used in the following way.

Proposition 1 [41]*:* "Every positive semidefinite and symmetric matrix is a kernel matrix. Conversely, every kernel matrix is symmetric and positive semidefinite."

Therefore, the linear mapping function of Lyapunov condition in (12) is invertible, and the stability property for the linear time-invariant system in (3) is granted under the Lyapunov function where a positive-definite symmetric matrix $\mathbf{P}$ exists, and

$$\big\langle\mathbf{x},\mathbf{x}'\big\rangle_{\mathbf{p}}=\mathbf{P}, \tag{14}$$

can be considered as the inner product of the system in (3) [42].

Generally, in the lithography process, a linear model is modified to illustrate the performance of the exposure tool. Nevertheless, the wafer surface structure abundantly concedes the real distribution of the overlay error as nonlinear and in the curve shape (e.g., due to the upstream process of the exposure step). To deal with this phenomenon, the Lyapunov kernel mapping function of the LS-SVR can achieve the form of the polynomial kernel function as follows:

$$\mathbf{P}=\kappa(u_t,\mathbf{u})=\big(\langle\mathbf{x},\mathbf{x}'\rangle_{\mathbf{p}}+m\big)^{n} \tag{15}$$

where $m=0$, and n are estimated through the tuning procedure. Regarding the dynamic model in (3), an additional condition is required for asymptotic stability of a nonlinear system as follows [39, p. 130]:

$$|f(\mathbf{x}_1,\ldots,\mathbf{x}_{t-1})|^2\leq\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_{t-1}\end{bmatrix}\mathbf{P}\begin{bmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_{t-1}\end{bmatrix}. \tag{16}$$

IV. Simulation Study

The following steps illustrate the performance of the proposed random effect LS-SVR control system in the presence of metrology delay and high-mixed recipe system in a simulation scenario.

Step 1: Consider $R_1,\ldots,R_5$ as five different recipes and u_{ji}, for $j=1,\ldots,10$, $i=1,\ldots,5$ as an input variable for jth overlay factor and ith recipe.

Step 2: Generate 1000 samples for each recipe and overlay factors from multivariate Normal distribution as

$\mathbf{u}_1, \ldots, \mathbf{u}_5 \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_5$ are a mean vector of $\vec{\mathbf{0}}, \vec{\mathbf{1}}, \vec{\mathbf{2}}, \vec{\mathbf{3}}, \vec{\mathbf{4}}$ (to represent the effect of impulse shift), respectively, with length 10 and $\boldsymbol{\Sigma}$ is a 10×10 diagonal matrix with diag $\{10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005\}$ to represent the effect of disturbance on the system.

Step 3: Set stochastic target value almost surely to zero generated by $T + \varepsilon_t \sim N(0, 10^{-15})$.

Step 4: Mix all recipe data randomly to build up a new data set with 5000 instances and ten overlay factors.

Step 5: $z_{tj} = i$ is defined for random effect at tth run such that for jth overlay factor recipe i is used.

Step 6: Simulate stochastic metrology delay from ZIP($\lambda = 3, \gamma = 0.9$) distribution where a delay happens with the maximum length of 8 at each lot (25 runs).

Step 7: Generate a sequence of 25 dummy data for $T + \varepsilon_t' \sim T + \varepsilon_t$ for the minimum requirement of the LS-SVR learning algorithm.

Step 8: Maintain five lots of historical data for compensating the overlay error through the learning algorithm.

Step 9: Employ the EWMA and tEWMA controllers as the baseline for the performance comparison. A linear form of EWMA in (17) is assumed for the process output and input estimation at run t

$$\begin{aligned} \widehat{\mathbf{Q}}_t &= a + \tau \mathbf{u}_{t-1} + d_t \\ \mathbf{u}_t &= \mathbf{u}_{t-1} - \frac{\theta}{\hat{\tau}} E_{t-1} \end{aligned} \tag{17}$$

where the unknown parameters a and τ can be estimated by the tuning algorithm. Based on the engineers' domain knowledge, the discount factor sets to $\theta = 0.3$. Similarly, for tEWMA [19], the process input and output represent as follows:

$$\begin{aligned} \widehat{\mathbf{Q}}_t &= a_t + \tau \mathbf{u}_{t-1} + \mathbf{d}_t \\ \mathbf{u}_t &= \frac{T - a_t}{\hat{\tau}} \\ a_t &= (1-\theta) a_{t-1} + \theta(\widehat{\mathbf{Q}}_t - \hat{\tau}\mathbf{u}_{t-1} - a_{t-1}). \end{aligned} \tag{18}$$

Step 10: Initiate the tuning algorithm to estimate the best parameter setting of random-effect LS-SVR in (8) with the Lyapunov-based kernel function in (15). The regularization parameters $\mathbb{C}_1$ and $\mathbb{C}_2$ are modified to the interval [0,10], with a lag of 1. m and n in (15) are confirmed to the intervals [-1,1] and [1,10] with a lag of 0.1 and 1, respectively.

Step 11: Adopt the higher first pass rule for R2R-MISO controller in Algorithm 1, in which the input overlay factor with the highest variation from the targeted setpoint enter into the c_1 controller, second highest to the c_2, and so on. The system reaches the steady-state condition if the root-mean-square error (RMSE) condition in (19) between two control systems exists

$$|c_j(\text{RMSE}) - c_{j-1}(\text{RMSE})| \leq 10^{-15} \tag{19}$$

where

$$\text{RMSE} = \sqrt{\frac{\sum_{l=1}^{t} [\widehat{\mathbf{Q}}_l - (T + \varepsilon_l)]^2}{t}}.$$

Step 12: Initialize the setting of the control system at a current run for each recipe with the setting of the last run when the same recipe was applied to the system.

Step 13: Merge the ten-fold cross-validation learning algorithm to check the optimal setting stability in the LS-SVR controller.

Figs. 5–9 illustrate the comparison of the estimated input and output between the LS-SVR and EWMA and tEWMA controllers. Consequently, the LS-SVR has smoother variations and improved compensation performance (e.g., reduced variance and closer to the target) than both EWMA and tEWMA.

The results have shown that the proposed LS-SVR controller tightens up the excellent performance bound, and eventually achieves a lower cost, together with an extensive disturbance, in comparison with the EWMA and tEWMA control system. When the variation increases, the result is more tangible. When the unmeasurable disturbance makes a tangible shift in overlay factors, LS-SVR can competently deal with process shift, while EWMA and tEWMA are inaccurate, although tEWMA has a better performance than EWMA.

On average, after the fifth overlay factor enters into the model, the system reaches the steady-state condition, which means that the proposed LS-SVR controller can compensate the effect of disturbance, noise and impulse shift smaller than 0.1. However, the result shows that the convergence rate strongly depends on the disturbance or impulse shift.

According to the sensitivity analysis implemented by the tuning procedure, the most frequent result for the parameter setting of the kernel function in (15) achieved at $m = 0$ and $n = 3$. $\mathbb{C}_1$ and $\mathbb{C}_2$ parameters of LS-SVR in (8) are set as free parameters.

To test how the Lyapunov stability condition and polynomial kernel function effectively enhance the performance of the LS-SVR method, three types of kernel functions merged with the mixed-effect LS-SVR method (the polynomial kernel function upgraded with the Lyaponuv stability condition, simple polynomial kernel function, and simple linear kernel function). The result is summarized in Table I which shows that both the polynomial kernel function and Lyapunov stability condition are constructive in compensating the overlay factors. It is apparent that when disturbance and impulse shifts are increasing, the performance of the proposed mixed-effect LS-SVR controller is significantly better than the two other LS-SVR controllers.

V. Empirical Study

An empirical study is conducted to estimate the validity of the proposed approach. To speed up the analysis of huge empirical data, the most frequent result from simulation experiment has been used as the initial setting for the optimization model. The general kernel function is considered as $\mathbf{P} = (\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathbf{p}})^3$.

The empirical data include four recipes connected to the reticle of the scanner. Among ten overlay factors in [17], asymmetric rotation and asymmetric magnification have not

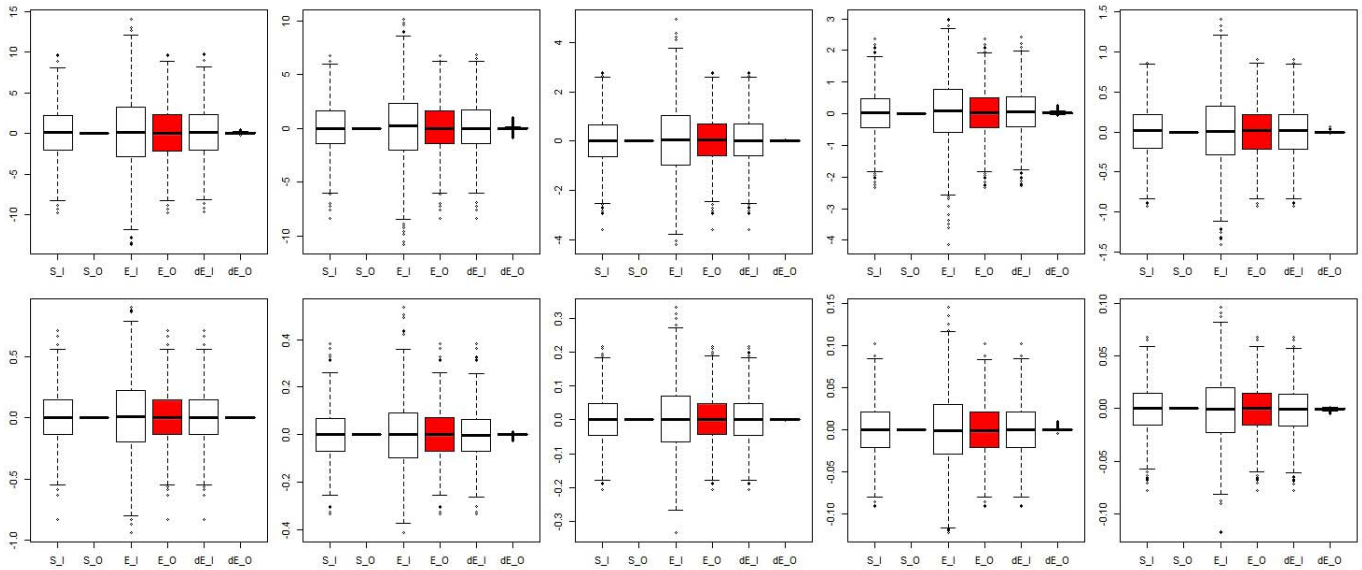

Fig. 5. Simulation result of estimated input and output of ten process variables for random-effect LS-SVR versus EWMA and tEWMA controller from the first recipe (R_1) (S_I denotes to LS-SVR input, S_O to LS-SVR output, E_I to EWMA input, E_O to EWMA output, dE_I to tEWMA input, and dE_O to tEWMA output).

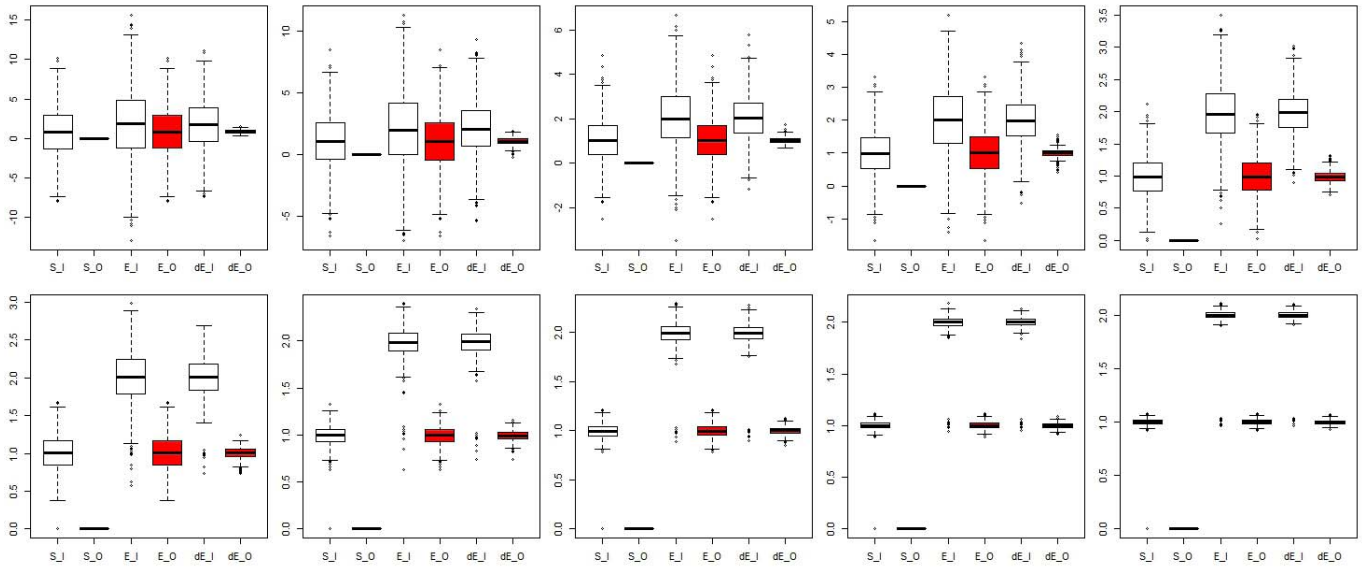

Fig. 6. Simulation results for estimated input and output of ten process variables for random-effect LS-SVR versus EWMA and tEWMA controller from the second recipe (R_2).

been controlled in this fab. The range (20) and RMSE of empirical data are summarized in Table II.

$$\text{Range} = \max_t \widehat{Q}_t - \min_t \widehat{Q}_t. \tag{20}$$

In the measurements, 30% of lots received delay from the metrology tools. The maximum length of the delay is calculated as 80 lots, and the average length is estimated as four lots.

The EWMA controller with $\theta = 0.3$ is used for the feedback control in this fab. The target value of each overlay factor and total overlay errors set to zeroe. According to the result of the simulation study, variables with variations smaller than 0.1 from the target set point are eliminated from the MISO system. Table III summarizes the improvement of range and RMSE for each overlay factor between EWMA and LS-SVR.

The overlay model proposed in [17] is used to approximate the compounded overlay error on the x-axis and the y-axis. At each run, only the maximum value of measurement error for the x-axis and the y-axis is considered. The result is summarized in Table IV and shows how the error compensation has improved by using the proposed mixed-effect LS-SVR controller for each recipe. Our results show that the proposed LS-SVR controller has achieved an improvement of a minimum 32% on the indices.

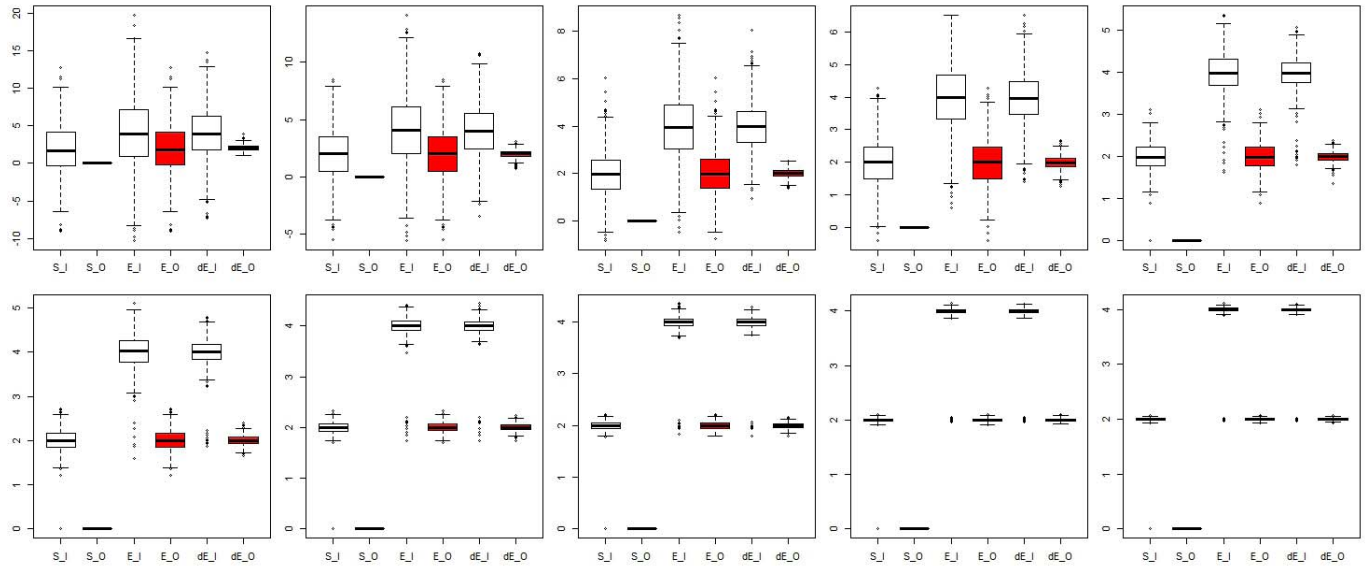

Fig. 7. Simulation result of estimated input and output of ten process variables for random-effect LS-SVR versus EWMA and tEWMA controller from the third recipe (R_3).

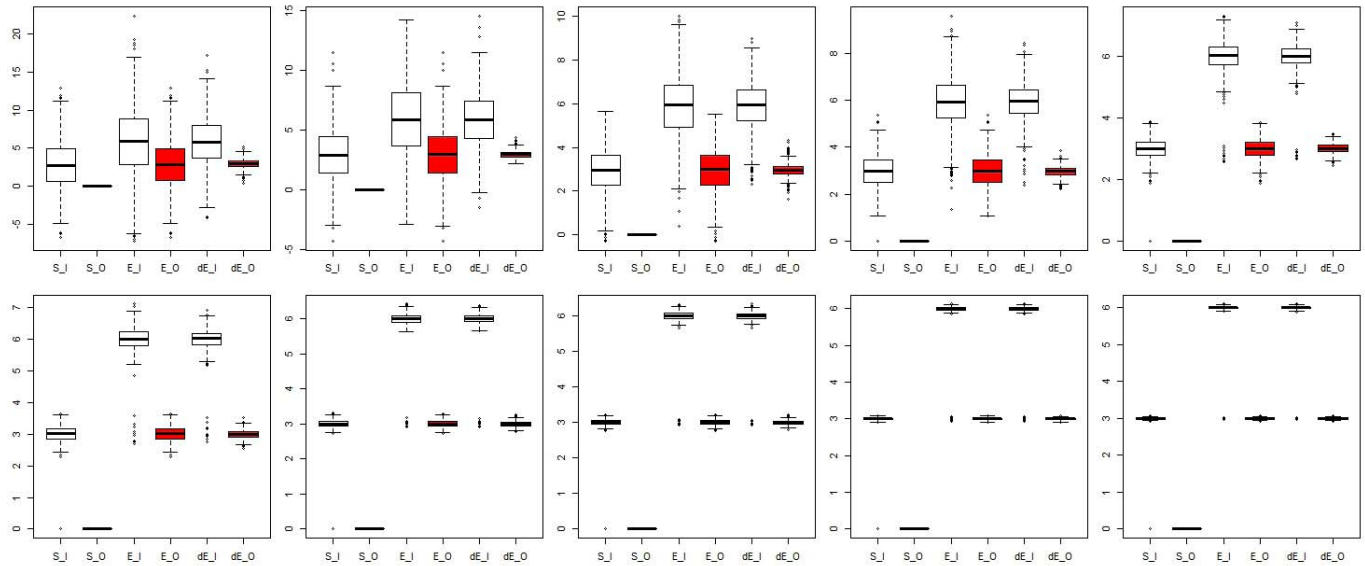

Fig. 8. Simulation result of estimated input and output of ten process variables for random-effect LS-SVR versus EWMA and tEWMA controller from the fourth recipe (R_4).

TABLE I

SIMULATION RESULT OF RANGE AND RMSE IMPROVEMENT COMPARED WITH MIXED-EFFECT LS-SVR WITH LINEAR KERNEL FUNCTION (MIXED-EFFECT LS-SVR WITH POLYNOMIAL KERNEL WITHOUT LYAPUNOV STABILITY CONDITION)

Overlay Factors	**Recipe**									
	I		**II**		**III**		**IV**		**V**	
	RMSE	*Range*	*RMSE*	*Range*	*RMSE*	*Range*	*RMSE*	*Range*	*RMSE*	*Range*
V.1	13%(3%)	1%(1%)	14%(4%)	2%(2%)	20%(5%)	5%(4%)	59%(7%)	6%(6%)	61%(9%)	9%(9%)
V.2	4%(2%)	0%(0%)	12%(3%)	0%(0%)	21%(4%)	2%(2%)	23%(7%)	5%(4%)	23%(8%)	6%(6%)
V.3	3%(2%)	0%(0%)	4%(2%)	0%(0%)	7%(3%)	2%(2%)	23%(6%)	4%(3%)	23%(7%)	6%(6%)
V.4	3%(2%)	0%(0%)	4%(2%)	0%(0%)	7%(3%)	2%(1%)	16%(5%)	3%(3%)	17%(6%)	5%(5%)
V.5	3%(2%)	0%(0%)	3%(2%)	0%(0%)	7%(3%)	1%(1%)	13%(4%)	3%(2%)	14%(5%)	5%(4%)
V.6	3%(2%)	0%(0%)	3%(2%)	0%(0%)	7%(3%)	1%(1%)	9%(3%)	2%(2%)	14%(5%)	5%(2%)
V.7	3%(2%)	0%(0%)	3%(2%)	0%(0%)	7%(3%)	0%(0%)	9%(3%)	2%(1%)	10%(4%)	2%(2%)
V.8	3%(2%)	0%(0%)	3%(2%)	0%(0%)	7%(3%)	0%(0%)	9%(3%)	1%(0%)	9%(3%)	1%(1%)
V.9	3%(2%)	0%(0%)	3%(2%)	0%(0%)	4%(2%)	0%(0%)	7%(2%)	0%(0%)	7%(2%)	1%(1%)
V.10	1%(1%)	0%(0%)	1%(1%)	0%(0%)	3%(1%)	0%(0%)	6%(1%)	0%(0%)	6%(1%)	1%(1%)

VI. CONCLUSION

This paper developed an accurate, high efficient optimization technique for overlay error minimization during the high-mixed photolithography process validated by simulation analysis and an empirical study. Further study can be done as the basis for productivity improvements in digital communications and industrial transformation [43]. The proposed

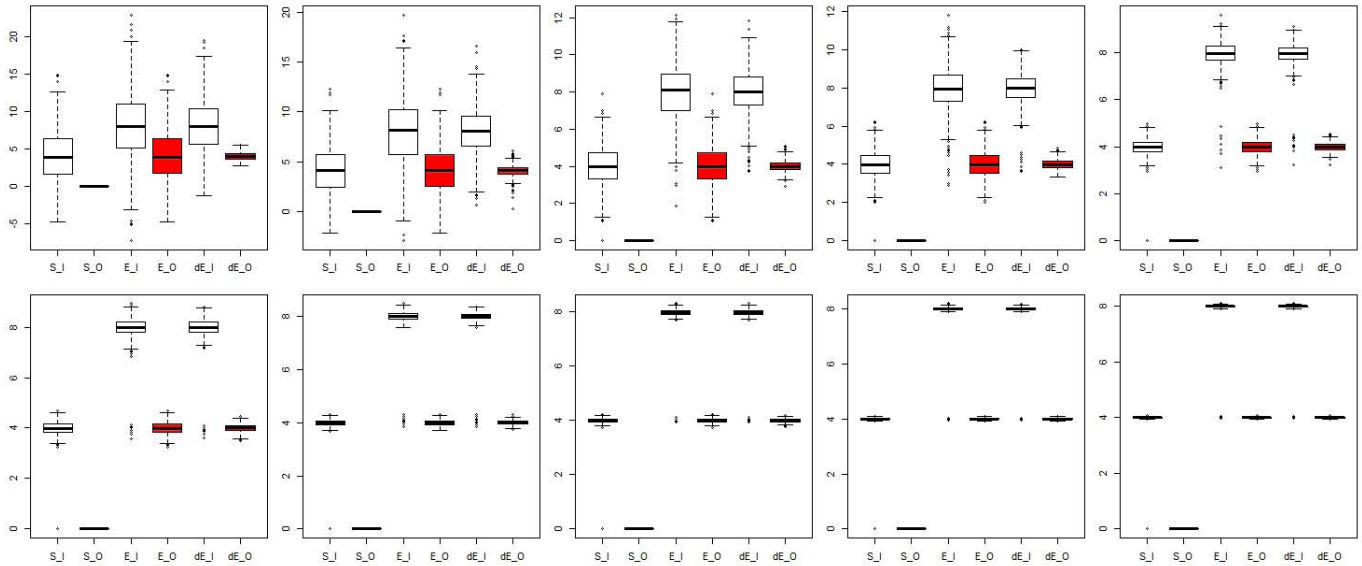

Fig. 9. Simulation result of estimated input and output of ten process variables for random-effect LS-SVR versus EWMA and tEWMA controller from the fifth recipe (R_5).

TABLE II
SUMMARY OF EMPIRICAL DATA

Overlay Factors	Recipe (number of lot)							
	A (154)		B (25)		C (3)		D (2)	
	RMSE	*Range*	*RMSE*	*Range*	*RMSE*	*Range*	*RMSE*	*Range*
Wafer rotation	7.02	68.71	7.32	59	6.47	37.1	6.8	29.54
Non-orthogonality	2.79	33.42	2.49	16.61	1.89	9	2.17	11.89
Scaling along the X axis	6.71	52.9	6.91	68.45	5.79	25.2	7.45	31.47
Scaling along the Y axis	7.61	62.5	7.45	62.22	6.97	40.2	7.79	32.03
Translation along the X axis	7.89	54.6	7.7	41.5	6.78	30.6	8.72	48.2
Translation along the Y axis	8.37	74	8	48.8	6.94	41.3	7.15	30.1
Reticle rotation	3.23	30.04	3	19.7	2.47	13.99	3.32	16.48
Isotropic magnification	2.98	37.57	2.66	19.03	2.41	12.34	4.5	27.68

TABLE III
RANGE AND RMSE IMPROVEMENT

Overlay Factors	Recipe (number of lot)							
	A (154)		B (25)		C (3)		D (2)	
	RMSE	*Range*	*RMSE*	*Range*	*RMSE*	*Range*	*RMSE*	*Range*
Wafer rotation	92%	84%	100%	90%	83%	139%	157%	171%
Non-orthogonality	51%	90%	64%	59%	40%	74%	41%	104%
Scaling along the X axis	54%	53%	52%	71%	52%	69%	43%	40%
Scaling along the Y axis	54%	65%	56%	61%	70%	123%	37%	43%
Translation along the X axis	57%	49%	39%	46%	50%	37%	35%	60%
Translation along the Y axis	57%	60%	46%	49%	29%	44%	23%	26%
Reticle rotation	43%	60%	102%	77%	23%	38%	34%	37%
Isotropic magnification	29%	49%	25%	45%	59%	35%	52%	93%

TABLE IV
IMPROVEMENT FOR OVERLAY ERROR ON x-AXIS AND y-AXIS

Overlay Factors	Recipe (number of lot)							
	A (154)		B (25)		C (3)		D (2)	
	RMSE	*Range*	*RMSE*	*Range*	*RMSE*	*Range*	*RMSE*	*Range*
X-axis	11%	8%	-8%	0%	15%	14%	14%	26%
Y-axis	17%	23%	16%	21%	18%	11%	18%	30%

mixed-effect LS-SVR controller with a self-tuning algorithm combined with polynomial Lyapunov-based kernel function has shown its robust capability to compensate the overlay errors. The assumed Lyapunov condition composes a stable controller to deal with the lack of process information caused by the metrology delay.

The LS-SVR algorithm may take a long time for tuning the input parameters (m, n, $\mathbb{C}_1$ and $\mathbb{C}_2$) to converge to the

best parameter setting. In particular, there are two sensitive parameters ($\mathbb{C}_1$, and $\mathbb{C}_2$, the regularization parameters) that should be estimated. For evaluating the robustness of the proposed model, three mixed-effect LS-SVR controllers with the polynomial Lyapunov-based kernel, a simple polynomial kernel, and simple linear kernel functions have been compared.

The result of the simulation study shows that both the polynomial kernel and the Lyapunov stability function enhance the performance of the proposed control system.

To validate the proposed approach, the empirically collected data from a leading semiconductor fab in Taiwan have been used to validate the proposed LS-SVR controller.

In comparison to the EWMA or tEWMA, the LS-SVR method is enforced via the historical data for enhancing the learning process algorithm, such that more training information will be affected positively by the performance of the control system. In this study, the information of at least last five lots is used as the training data. The mixed-recipe process may be directly applicable to the mixed-product process, and with a few adjustments is relevant to the mixed-tool process.

Future research can be done to enhance the proposed approach to deal with the effects owing to multilayer lithography process by considering an additional random effect in each layer. Also, more studies can be done to extend the proposed approach for similar issues in other semiconductor manufacturing processes such as dry-etching (DE) and chemical-mechanical polishing (CMP). Furthermore, the proposed approach can be employed for high-mixed production in similar industries such as TFT-LCD, and solar cell. In addition, due to the complexity and emergence in process control of semiconductor industry, the application of other learning-based optimization models [44] can be an interesting topic in this field.

APPENDIX

Regards to the model with random-effects, the optimization problem for LS-SVR can be equivalently written as

$$\begin{aligned}\min_{w,C_l,b,\boldsymbol{\beta}_l} & \frac{1}{2}\|w\|^2+\frac{\mathbb{C}_1}{2}\sum_{l=1}^{t}\beta_l'\Sigma_{z_{lj}}^{-1}\beta_l \\ & +\frac{\mathbb{C}_2}{2}\sum_{l=1}^{t}C_l'\Sigma_{C_l}^{-1}C_l \\ \text{s.t. } & w\phi(\boldsymbol{u}_l)+\boldsymbol{\beta_l}\boldsymbol{z}_l+C_l+b=T+\varepsilon_{lj} \quad \forall l=1,\ldots,t. \end{aligned} \tag{21}$$

This is a convex quadratic program with affine constraints. (22) is introducing the corresponding Lagrangian function

$$\begin{aligned}L(w,C_t,b,\boldsymbol{\beta}_t;\boldsymbol{\alpha},\boldsymbol{\alpha}^*) &= \frac{1}{2}\|w\|^2+\frac{\mathbb{C}_1}{2}\sum_{l=1}^{t}\beta_l'\Sigma_{z_{lj}}^{-1}\beta_l+\frac{\mathbb{C}_2}{2}\sum_{l=1}^{t}C_l'\Sigma_{C_l}^{-1}C_l \\ &-\sum_{l=1}^{t}\boldsymbol{\alpha}_l(T+\boldsymbol{\varepsilon}_l-w\phi(\boldsymbol{u}_l)-\boldsymbol{\beta}_l\boldsymbol{z}_l-b-C_l) \\ &-\sum_{l=1}^{t}\boldsymbol{\alpha}_l^*(T+\boldsymbol{\varepsilon}_l-w\phi(\boldsymbol{u}_l)-\boldsymbol{\beta}_l\boldsymbol{z}_l-b-C_l). \end{aligned} \tag{22}$$

The Karush–Kuhn–Tucker (KKT) optimization conditions for a solution can be achieved by partially differentiating concerning $w, C_t, b, \boldsymbol{\beta}_t, \boldsymbol{\alpha}$, and $\boldsymbol{\alpha}^*$

$$\begin{cases}\dfrac{\partial L}{\partial w}=0\rightarrow & w=\displaystyle\sum_{l=1}^{t}(\alpha_l-\alpha_l^*)\phi(u_l) \\ \dfrac{\partial L}{\partial b}=0\rightarrow & \displaystyle\sum_{l=1}^{t}(\alpha_l-\alpha_l^*)=0 \\ \dfrac{\partial L}{\partial C_t}=0\rightarrow & \mathbb{C}_2\Sigma_{C_t}^{-1}C_t-(\boldsymbol{\alpha}_t-\boldsymbol{\alpha}_t^*)=0 \\ \dfrac{\partial L}{\partial \alpha_t}=0\rightarrow & \displaystyle\sum_{l=1}^{t}(T+\varepsilon_l-w\phi(u_t)-\boldsymbol{\beta}_l z_l-b-C_l)=0 \\ \dfrac{\partial L}{\partial \alpha_t^*}=0\rightarrow & \displaystyle\sum_{l=1}^{t}(T+\varepsilon_l-w\phi(u_l)-\boldsymbol{\beta}_l z_l-b-C_l)=0 \\ \dfrac{\partial L}{\partial \boldsymbol{\beta}_t}=0\rightarrow & \mathbb{C}_1\Sigma_{z_{tj}}^{-1}\beta_t-\alpha_t z_t=0\end{cases} \tag{23}$$

which leads to the following equivalent dual problem regarding the kernel matrix $\kappa(u_t,u)=\phi(u_t)\phi^T(u_t)$

$$\begin{bmatrix}0 & \mathbf{I}' \\ \mathbf{I} & \kappa(u_t,\mathbf{u})+\dfrac{\mathbf{I}}{\mathbb{C}_1}z_{t_j}'\Sigma_{z_{tj}}^{-1}z_{tj}+\dfrac{\mathbf{I}}{\mathbb{C}_2}\Sigma_{C_t}^{-1}\end{bmatrix}\begin{bmatrix}0 \\ (\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)\end{bmatrix}=\begin{bmatrix}0 \\ \mathbf{Q}\end{bmatrix}. \tag{24}$$

REFERENCES

[1] 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.

[2] 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.

[3] Y. Zheng, D. S.-H. Wong, Y.-W. Wang, and H. Fang, "Takagi–Sugeno model based analysis of EWMA RtR control of batch processes with stochastic metrology delay and mixed products," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1155–1168, Jul. 2014.

[4] A.-C. Lee, J.-H. Horng, T.-W. Kuo, and N.-H. Chou, "Robustness analysis of mixed product run-to-run control for semiconductor process based on ODOB control structure," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 2, pp. 212–222, May 2014.

[5] C. E. Chemali, J. Freudenberg, M. Hankinson, and J. J. Bendik, "Run-to-run critical dimension and sidewall angle lithography control using the PROLITH simulator," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 3, pp. 388–401, Aug. 2004.

[6] Y. Jiao and D. Djurdjanovic, "Stochastic control of multilayer overlay in lithography processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 404–417, Aug. 2011.

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

[8] 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. Practice*, vol. 25, no. 6, pp. 828–842, 2018.

[9] H. Narita, "Semiconductor device manufacturing method and semiconductor device," U.S. Patent 9 269 671, Feb. 23, 2016.

[10] K. Nara and T. Hamada, "Method for manufacturing display element, manufacturing apparatus of display element and display device," U.S. Patent 9 310 656, Apr. 12, 2016.

[11] 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. 103901–103906, 2015.
[12] 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.
[13] H. Fuyun and Z. Zhang, "State space model and numerical simulation of overlay error for multilayer overlay lithography processes," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 1123–1127.
[14] D. MacMillen and W. D. Ryden, "Analysis of image field placement deviations of a 5×microlithographics reduction lens," *Proc. SPIE*, vol. 0334, pp. 78–89, Sep. 1982.
[15] 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.
[16] 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, 2006.
[17] 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, 2011.
[18] J. Park *et al.*, "Exact and reliable overlay metrology in nanoscale semiconductor devices using an image processing method," *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 13, no. 4, pp. 041409-1–041409-7, 2014.
[19] M.-D. Ma, C.-C. Chang, D. S.-H. Wong, and S.-S. Jang, "Threaded EWMA controller tuning and performance evaluation in a high-mixed system," *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 4, pp. 507–511, Nov. 2009.
[20] A.-C. Lee, T.-W. Kuo, and S.-W. Chiang, "Run-to-run mixed product overlay process control: Using tool based disturbance estimator (TBDE) approach," *Int. J. Eng. Technol.*, vol. 5, no. 3, p. 349, 2013.
[21] B. Ai, Y. Zheng, Y. Wang, S.-S. Jang, and T. Song, "Cycle forecasting ewma (CF-EWMA) approach for drift and fault in mixed-product run-to-run process," *J. Process Control*, vol. 20, no. 5, pp. 689–708, 2010.
[22] 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.
[23] B. J. de Kruif and T. J. A. de Vries, "On using a support vector machine in learning feed-forward control," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, vol. 1, Jul. 2001, pp. 272–277.
[24] J. A. Suykens, J. Vandewalle, and B. De Moor, "Optimal control by least squares support vector machines," *Neural Netw.*, vol. 14, no. 1, pp. 23–35, 2001.
[25] S. K. Lahiri and K. C. Ghanta, "Support vector regression with parameter tuning assisted by differential evolution technique: Study on pressure drop of slurry flow in pipeline," *Korean J. Chem. Eng.*, vol. 26, no. 5, pp. 1175–1185, 2009.
[26] C. L. Smith, *Advanced Process Control: Beyond Single Loop Control*. New York, NY, USA: Wiley, 2010.
[27] J. Rico-Azagra, M. Gil-Martínez, and J. Elso, "Quantitative feedback control of multiple input single output systems," *Math. Problems Eng.*, vol. 2014, Apr. 2014, Art. no. 136497.
[28] T. Liu and F. Gao, "Cascade control system," in *Industrial Process Identification and Control Design*. London, U.K.: Springer, 2011, pp. 321–347.
[29] B. Lu *et al.*, "Improving process monitoring and modeling of batch-type plasma etching tools," Ph.D. dissertation, Dept. Chem. Eng., Univ. Texas, Austin, TX, USA, 2015.
[30] Y. E. Shao and Y.-T. Hu, "Using the ANN classifier to recognize the disturbance patterns for a multivariate system," in *Proc. 5th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2016, pp. 630–634.
[31] X. Cheng, X. Ren, S. Ma, and S. Li, "A modeling method based on artificial neural network with monotonicity knowledge as constraints," *Chemometrics Intell. Lab. Syst.*, vol. 145, pp. 93–102, Jul. 2015.
[32] T. Sun and J. Liu, "Predicting MEMS gyroscope's random drifts using LSSVM optimized by modified PSO," in *Proc. IEEE Chin. Guid., Navigat. Control Conf.*, Aug. 2016, pp. 146–149.
[33] B. Ahmadi, H. Nourisola, and S. Tavakoli, "Robust adaptive sliding mode control design for uncertain stochastic systems with time-varying delay," *Int. J. Dyn. Control*, vol. 6, no. 1, pp. 413–424, 2018.
[34] S. R. Mohanty, P. K. Ray, N. Kishor, and B. Panigrahi, "Classification of disturbances in hybrid DG system using modular PNN and SVM," *Int. J. Elect. Power Energy Syst.*, vol. 44, no. 1, pp. 764–777, 2013.
[35] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.
[36] G. K. Robinson, "That BLUP is a good thing: The estimation of random effects," *Statist. Sci.*, vol. 6, no. 1, pp. 15–32, 1991.
[37] 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.
[38] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.
[39] E. Fridman, *Introduction to Time-Delay Systems: Analysis and Control*. Basel, Switzerland: Birkhäuser, 2014.
[40] W. Zhang, M. S. Branicky, and S. M. Phillips, "Stability of networked control systems," *IEEE Control Syst. Mag.*, vol. 21, no. 1, pp. 84–99, Feb. 2001.
[41] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Jan. 2004.
[42] 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.
[43] M. Khakifirooz, D. Cayard, C. F. Chien, and M. Fathi, "A system dynamic model for implementation of industry 4.0," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, Jun. 2018, pp. 1–6.
[44] M. Khakifirooz, "A framework for intelligent decision support system for smart manufacturing to empower industry 4.0: The illustration of semiconductor manufacturing," Ph.D dissertation, Dept. Ind. Eng. Eng. Manag., Nat. Tsing Hua Univ., Hsinchu, Taiwan, Oct. 2018.

Marzieh Khakifirooz (S'15) received the M.S. degree in industrial statistics and the Ph.D. degree in industrial engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2014 and 2018, respectively.

She is currently a Post-Doctoral Researcher with the Artificial Intelligence for Intelligent Manufacturing Systems Research Center, NTHU, sponsored by the Ministry of Science and Technology, Taiwan. Her research interests include smart manufacturing, game theory, big data analytics, statistical inferences, and operational research.

Chen-Fu Chien (M'03) received the B.S. degree (Phi Tao Phi 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 operations research and decision sciences from the University of Wisconsin–Madison, Madison, WI, USA, in 1994 and 1996, respectively.

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 with the Industrial Engineering Division, Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu. He received the PCMPCL Training from the Harvard Business School, Boston, MA, USA, in 2007. He is currently a Tsinghua Chair Professor and a Micron Chair Professor with NTHU. He is also the Director of the Artificial Intelligence for Intelligent Manufacturing Systems Research Center sponsored by the Ministry of Science and Technology, the NTHU-TSMC Center for Manufacturing Excellence, and the Principal Investigator for the Semiconductor Technologies Empowerment Partners (STEP) Consortium. He holds eight U.S. invention patents on semiconductor manufacturing. He has publised four books, more than 170 journal papers, and a number of case studies in Harvard Business School. He proposed Industry 3.5 as a hybrid strategy between the existing Industry 3.0 and to-be Industry 4.0, empowered by AI and big data analytics for disruptive innovations. His book on Industry 3.5 (ISBN 978-986-398-380-4) is a bestselling book. His research efforts focus on decision analysis, modeling and analysis for semiconductor manufacturing, manufacturing strategy, and manufacturing intelligence.

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 the Ministry of Science and Technology, the Distinguished University-Industry Collaborative Research Award from the Ministry of Education, the University Industrial Contribution Awards from the Ministry of Economic Affairs, the TECO Award, 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, the Distinguished Engineering Professor by Chinese Institute of Engineers in Taiwan, the 2012 Best Paper Award of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and the 2015 Best Paper Award of the IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. He is an Area Editor of the *Flexible Services and Manufacturing Journal*, an Editorial Board Member of *Computers and Industrial Engineering*, and an Advisory Board Member of *OR Spectrum*. He has been invited to give keynote lectures at several conferences, including APIEMS, C&IE, FAIM, IEEM, IEOM, IML, ISMI, and leading universities worldwide.

Mahdi Fathi (M'09) received the B.S. and M.S. degrees from the Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran, in 2006 and 2008, respectively, and the Ph.D. degree from the Iran University of Science and Technology, Tehran, in 2013.

Dr. Fathi received three post-doctoral fellowships at Industrial Engineering Laboratory, Ecole Central Paris, Châtenay-Malabry, France, in 2014; Stochastic Modeling and Analysis of Communication Systems Group at the Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium, in 2015; Department of Industrial and Systems Engineering, Mississippi State University, Starkville, MS, USA, in 2018. Moreover, he was a Visiting Scholar at the Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA, from 2017 to 2018 and at the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2018. His research interests are queuing theory and its applications, stochastic process, optimization, artificial intelligent, uncertainty quantification, smart manufacturing, and reliability.