CIRP Annals - Manufacturing Technology 72 (2023) 429–432

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

Robust control of maximum photolithography overlay error in a pattern layer

Noah Graff[a], Grani A. Hanasusanto[b], Dragan Djurdjanović (2)[a,*]

[a] Walker Department of Mechanical Engineering, The University of Texas, Austin, TX, USA
[b] Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, IL, USA

ARTICLE INFO

Article history:
Available online 2 May 2023

Keywords:
Photolithography overlay
Process control
Robust control

ABSTRACT

This paper presents a novel method for control of overlay errors in photolithography processes in semiconductor manufacturing. It minimizes the largest overlay error across all measurement markers on a pattern layer, and this minimization is done for the worst-case scenario regarding bounded process bias and modeling noise terms. This large-scale robust optimization problem was formulated as a linear program which can be solved within seconds to generate optimal control commands. Simulations based on wafer data obtained from a major 300 mm semiconductor fab illustrate consistent and significant advantages of this approach over the benchmark control strategies.

© 2023 CIRP. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In the field of semiconductor manufacturing, photolithography processes are used to successively transfer patterns from masks onto wafers and thus build circuits up layer by layer [1]. Adequate alignment between consecutive pattern layers is critical to the production of functioning microelectronic devices [2]. Therefore, precise control of misalignments between pattern layers on a wafer, referred to as overlay errors, is a problem of great importance to the industry [3].

Due to wave behavior of light at sub-nanometer scales relevant to overlay errors, their characterization errors on a given wafer requires measurements of overlay errors from a number of overlay measurement markers dispersed across the wafer [4]. There could be hundreds of such markers on a wafer and their locations are predetermined and known. In the foundation of all overlay control approaches are Zernike-polynomial based models that relate overlay errors in those markers to controllable parameters on the photolithography tool on which the process is performed [5]. At such small scales, significant uncertainties in the behavior of controllable parameters on the tool are inevitable and because of them, there is always a difference between what a control parameter on a lithography tool is commanded to do and what the tool actually executes. Those differences are commonly referred to as *biases* in controllable parameters of the tool and their uncertain, time-varying effects on overlay errors need to be countered through the control process.

The most commonly utilized overlay control strategy in the industry is the so-called run-to-run (R2R) strategy. R2R control methods utilize historical records of overlay errors and Zernike-polynomial based overlay models to estimate behavior of biases in tool parameters over time, based on which various prediction techniques are used to predict those biases for the next wafer. Predicted biases are then countered by pre-compensating control parameters for the next wafer in the opposite direction of predicted biases in those control parameters. A comprehensive review of R2R control strategies and their use in semiconductor manufacturing can be found in [6] and references therein.

Regardless of the underlying estimation and prediction approaches, R2R control strategies naturally deal with overlay errors of an individual pattern layer only. Nevertheless, production of a functioning circuit requires fabrication and alignment of dozens of pattern layers. In the recent years, Stream of Variation (SoV) type models [7] were used to model this inherent multistage character of lithography processes and augment R2R control paradigm with considerations of stack-up overlay errors across non-neighboring pattern layers. In [8], an algorithm for optimal stochastic control of overlay and stack-up overlay errors was developed under Bayesian assumptions of perfectly known model parameters, with model noise and bias prediction errors assumed to be normal, independent, and identically distributed (NIID) random variables. However, the tractability and convincing nature of results from [8] are predicated on the aforementioned assumptions, which are overly restrictive, especially in the area of semiconductor photolithography, where extremely minute scale of the relevant phenomena implies that the corresponding models are rarely (never) perfect, while model noise terms are rarely (never) NIID.

This problem is recognized in [9], where the authors introduced a control approach robust to imperfections in the knowledge of the variance characteristics of modeling noise and errors in bias predictions. Results in [9] showed significant benefits, especially when the bounds on the unknown parameters become wide. These results were further expanded in [10], where SoV-form models were used to devise a generalized method for robust control of multistage manufacturing quality in the presence of bounded noise terms and uncertainties in model parameters. Applications of this control approach in simulations based on the data and models of lithography overlay processes from a real fab demonstrated consistent and significant benefits of robust, distribution-agnostic control paradigm from [10] over traditional R2R approach [6], or Bayesian multistage method from [8].

* Corresponding author.
E-mail address: dragand@me.utexas.edu (D. Djurdjanović).

https://doi.org/10.1016/j.cirp.2023.03.015
0007-8506/© 2023 CIRP. Published by Elsevier Ltd. All rights reserved.

Nevertheless, in all overlay control research thus far [6,10], characterization of overlays in any given pattern layer is approached using the standard Euclidean norm of the vector consisting of overlay errors measured in all available markers. On the other hand, performance of the resulting microelectronic device is often (usually) driven by the worst overlay in the wafer area relevant to that device. Therefore, rather than minimizing the Euclidean L^2 norm of the vector of overlay errors in a given pattern layer, as was done in [6,10], one needs to minimize the L^∞ norm (maximum element) of that vector. This is a much more difficult, less tractable problem and it will be addressed in this paper.

Specifically, this paper presents a control strategy that minimizes the maximum of the overlay errors across all measurement markers in the entire pattern layer, and we seek to do that robustly, for the worst-case scenario of bounded modeling noise terms and errors in bias predictions. The remainder of this paper is structured as follows. Section 2 describes the formulation of the underlying min-max-max optimization problem, as well as its transformation into a tractable form that can be efficiently solved. Section 3 evaluates the performance of the newly proposed control strategy and compares it with that of the conventional R2R control, as well as with the overlay control approach from [10], which robustly minimizes the L^2 norm of overlay errors. All comparisons are done using data and models from a major 300 mm semiconductor manufacturing fab. Finally, Section 4 offers conclusions of the research presented in this paper and details directions for possible future work.

2. Methods

As demonstrated in [10], the vectors $\boldsymbol{o}_k^x$ and $\boldsymbol{o}_k^y$ consisting of overlay errors in the x and y directions in all measurement markers across pattern layer k can be expressed as:

$$o_k^x = D^x \cdot (u_k^x + c_k^x) + r_k^x; o_k^y = D^y \cdot (u_k^y + c_k^y) + r_k^y \tag{1}$$

where matrices $\mathbf{D}^x$ and $\mathbf{D}^y$ denote regression matrices that are fully defined by locations of measurement markers on the wafer, vectors $\boldsymbol{u}$ consist of control commands issued to the tool, vectors $\boldsymbol{c}$ denote random vectors of process biases which inherently exist due to actuator uncertainties and minute scale of overlay errors, while vectors $\boldsymbol{r}$ contain residuals that account for unmodeled effects and process noise. The goal of the control strategy presented in this paper is to *minimize the largest magnitude of overlay errors present on a pattern layer*, subject to the worst-case scenario regarding process bias terms and model residuals, which are unknown, but are assumed to reside within known limits.

The magnitude of overlay errors at measurement marker j of pattern layer k is naturally defined using the Euclidean L^2 norm as $\| o_{k,j} \| = \sqrt{o_{k,j}^x 2 + o_{k,j}^y 2}$. However, minimization of the maximum of such Euclidean overlay norms across all markers j, which is also robust to bounded uncertainties in model noise terms and bias prediction errors ends up being an intractable, large-scale nonlinear optimization problem. Instead, standard inequality

$$\| \boldsymbol{o}_{k,j} \| = \sqrt{\left(\boldsymbol{o}_{k,j}^x\right)^2 + \left(\boldsymbol{o}_{k,j}^y\right)^2} \le \left|\boldsymbol{o}_{k,j}^x\right| + \left|\boldsymbol{o}_{k,j}^y\right|$$

can be used to establish an upper bound on the Euclidean magnitude of overlay errors. This allows for the robust optimal control problem for pattern layer k of wafer t to be formulated as

$$\left(\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}\right) = \underset{\substack{\boldsymbol{u}_{t,k}^x \in \mathbb{R}^{N_x},\\ \boldsymbol{u}_{t,k}^y \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \quad \max_{\substack{\boldsymbol{c}_{t,k}^{\prime x}, \boldsymbol{c}_{t,k}^{\prime y}\\ \boldsymbol{r}_{t,k}^{\prime x}, \boldsymbol{r}_{t,k}^{\prime y}}} \max_{j \in \{1,2,\ldots J\}} \left|\boldsymbol{o}_{t,k,j}^x\right| + \left|\boldsymbol{o}_{t,k,j}^y\right|$$

$$s.t. \cdot \begin{cases} \boldsymbol{o}_{t,k}^x = \left[\mathbf{D}^{cx} \cdot \left(\boldsymbol{u}_{t,k}^x + \boldsymbol{c}_{t,k}^{\prime x}\right) + \boldsymbol{r}_{t,k}^{\prime x}\right] \\ \boldsymbol{o}_{t,k}^y = \left[\mathbf{D}^{cy} \cdot \left(\boldsymbol{u}_{t,k}^y + \boldsymbol{c}_{t,k}^{\prime y}\right) + \boldsymbol{r}_{t,k}^{\prime y}\right] \\ \boldsymbol{c}_{t,k}^{lb,x} \le \boldsymbol{c}_{t,k}^{\prime x} \le \boldsymbol{c}_{t,k}^{ub,x};\ \boldsymbol{c}_{t,k}^{lb,y} \le \boldsymbol{c}_{t,k}^{\prime y} \le \boldsymbol{c}_{t,k}^{ub,y} \\ \boldsymbol{r}_{t,k}^{lb,x} \le \boldsymbol{r}_{t,k}^{\prime x} \le \boldsymbol{r}_{t,k}^{ub,x};\ \boldsymbol{r}_{t,k}^{lb,y} \le \boldsymbol{r}_{t,k}^{\prime y} \le \boldsymbol{r}_{t,k}^{ub,y} \end{cases} \tag{2}$$

where $(\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*})$ are resulting vectors of optimal control commands that are to be given to the lithography tool to fabricate pattern layer k on wafer t, limits $\boldsymbol{c}_{t,k}^{lb,x}$, $\boldsymbol{c}_{t,k}^{lb,y}$, $\boldsymbol{r}_{t,k}^{lb,x}$ and $\boldsymbol{r}_{t,k}^{lb,y}$ denote the lower bounds on the bias vectors $\boldsymbol{c}$ and modeling noise terms $\boldsymbol{r}$, limits $\boldsymbol{c}_{t,k}^{ub,x}$, $\boldsymbol{c}_{t,k}^{ub,y}$, $\boldsymbol{r}_{t,k}^{ub,x}$ and $\boldsymbol{r}_{t,k}^{ub,y}$ denote the corresponding upper bounds, while integers J, N_x and N_y respectively denote the number of markers on the wafer and the numbers of controllable tool parameters in the x and y directions of the wafer.

Let us now transform formulation (2) into a tractable linear program (LP), which can be solved in times comparable to cycle-times relevant to the underlying process (seconds). To simplify notation, we will observe that optimization (2) is performed for a given pattern layer on a given wafer and hence the subscripts t and k can be neglected. Let us now define vector ξ as

$$\xi = \left[\boldsymbol{c}_x^{\prime T}\ \boldsymbol{c}_y^{\prime T}\ \boldsymbol{r}_x^{\prime T}\ \boldsymbol{r}_y^{\prime T}\right]^T$$

i.e., as the vector containing all uncertain terms in the optimization (2). Also, let us define matrices

$$\mathbf{M}_{cx} = \left[\mathbf{I}_{N_{cx}}\ \ 0_{N_{cx}\ x\ N_{cy}}\ 0_{N_{cx}\ x\ N_{rx}}\ 0_{N_{cx}\ x\ N_{ry}}\right]$$
$$\mathbf{M}_{cy} = \left[0_{N_{cy}\ x\ N_{cx}}\ \ \mathbf{I}_{N_{cy}}\ 0_{N_{cy}\ x\ N_{rx}}\ 0_{N_{cy}\ x\ N_{ry}}\right]$$
$$\mathbf{M}_{rx} = \left[0_{N_{rx}\ x\ N_{cx}}\ 0_{N_{rx}\ x\ N_{cy}}\ \mathbf{I}_{N_{rx}}\ \ 0_{N_{rx}\ x\ N_{ry}}\right]$$
$$\mathbf{M}_{ry} = \left[0_{N_{ry}\ x\ N_{cx}}\ 0_{N_{ry}\ x\ N_{cy}}\ 0_{N_{ry}\ x\ N_{rx}}\ \mathbf{I}_{N_{ry}}\right]$$

where $\mathbf{I}_n$ denotes an n-by-n identity matrix, while $0_{p,q}$ denotes a p-by-q matrix of zeros. Then, if we define matrix $\mathbf{S}$ as

$$\mathbf{S} = \left[\mathbf{M}_{cx}^T\ -\mathbf{M}_{cx}^T\ \mathbf{M}_{cy}^T\ -\mathbf{M}_{cy}^T\ \mathbf{M}_{rx}^T\ -\mathbf{M}_{rx}^T\ \mathbf{M}_{ry}^T\ -\mathbf{M}_{ry}^T\right]^T$$

and vector $\boldsymbol{g}$ as

$$g = \left[c^{ub,xT} - c^{lb,xT}\ c^{ub,yT}\ -c^{lb,yT}\ r^{ub,xT}\ -r^{lb,xT}\ r^{ub,yT}\ -r^{lb,yT}\right]^T$$

constraints

$$c_{t,k}^{lb,x} \le c^{\prime x}{}_{t,k} \le c_{t,k}^{ub,x};\ c_{t,k}^{lb,x} \le c^{\prime y}{}_{t,k} \le c_{t,k}^{ub,y}$$
$$r_{t,k}^{lb,x} \le r^{\prime x}{}_{t,k} \le r_{t,k}^{ub,x};\ r_{t,k}^{lb,y} \le r^{\prime y}{}_{t,k} \le r_{t,k}^{ub,y}$$

on the uncertain bias vectors $\boldsymbol{c}$ and model residuals $\boldsymbol{r}$ in (2) can be compactly expressed as $\mathbf{S} \cdot \xi \le \boldsymbol{g}$, and the uncertainty set ξ can be defined as $\boldsymbol{\xi} = \{\boldsymbol{\xi} \in \mathbb{R}^{N_\xi} : \mathbf{S} \cdot \boldsymbol{\xi} \le \boldsymbol{g}\}$, where N_ξ is the number of elements in ξ.

Following [11], the two maximizations in (2) can be brought into the constraints, which allows us to reformulate (2) as

$$\left(\boldsymbol{u}_x^*, \boldsymbol{u}_y^*\right) = \underset{\substack{\boldsymbol{u}_x \in \mathbb{R}^{N_x},\\ \boldsymbol{u}_y \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \ \min_{\tau \in \mathbb{R}} \tau \tag{3}$$

$$s.t. \cdot \tau \ge \max_{j \in \{1,2,\ldots J\}} \ \max_{\boldsymbol{\xi} \in \boldsymbol{\xi}} \left|o_j^x\right| + \left|o_j^y\right| \tag{4}$$

$$s.t. \cdot \begin{cases} \boldsymbol{o}^x = \mathbf{D}^{cx} \cdot (\boldsymbol{u}_x + \mathbf{M}_{cx} \cdot \boldsymbol{\xi}) + \mathbf{M}_{rx} \cdot \boldsymbol{\xi} \\ \boldsymbol{o}^y = \mathbf{D}^{cy} \cdot (\boldsymbol{u}_y + \mathbf{M}_{cy} \cdot \boldsymbol{\xi}) + \mathbf{M}_{ry} \cdot \boldsymbol{\xi} \end{cases} \tag{5}$$

Constraint (4) is equivalent to

$$\tau \ge \max_{\boldsymbol{\xi} \in \boldsymbol{\xi}} |o_j^x| + \left|o_j^y\right| \ \ \forall\ j \in \{1, 2, \ldots, J\} \tag{6}$$

which can be used to replace the nested max-max optimization in (4) with J constraints (6), each containing a single optimization within it. By using this new set of constraints and defining vectors $\boldsymbol{e}_j$ as a J-dimensional vectors with a 1 in the j^{th} position and zeros in all other positions, optimization (3)-(5) can be rewritten as

$$\left(u_x^*, u_y^*\right) = \underset{\substack{u_x \in \mathbb{R}^{N_x},\\ u_y \in \mathbb{R}^{N_y}}}{\arg\min} \ \min_{\tau \in \mathbb{R}} \tau$$

$$s.t. \cdot \tau \ge \max_{\xi \in \xi} \left\{ \left| \boldsymbol{e}_j^T \cdot \mathbf{D}^{cx} \cdot (\boldsymbol{u}_x + \mathbf{M}_{cx} \cdot \boldsymbol{\xi}) + \boldsymbol{M}_{rx} \cdot \boldsymbol{\xi} \right| +

$$+\left|e_j^T \cdot D^{cy} \cdot \left(u_y^+ M_c y \cdot \xi\right) + M_r x \cdot \xi\right|\}, \forall j \in \{1,2,\ldots,J\} \quad (7)$$

For overlay errors o_j^x and o_j^y in marker j, expression $|o_j^x| + |o_j^y|$ can take on the form of either $o_j^x + o_j^y$, $o_j^x - o_j^y$, $-o_j^x + o_j^y$, or $-o_j^x - o_j^y$. Thus, each absolute value in the constraints on the variable τ in (7) can be replaced with four constraints - one for each possible permutation of the signs. Then, by defining indices $s_{x,i}$ and $s_{y,i}$ as follows:

$$s_{x,i} = 1 \quad \forall\ i \in \{1,2,\ldots,2J\};\ \cdot s_{x,i} = -1 \quad \forall\ i \in \{2J+1, 2J+2, \ldots, 4J\}$$

$$s_{y,i} = 1 \ \forall\ i \in \{1,2,\ldots,J\} \cup \{2J+1, 2J+2, \ldots, 3J\}$$

$$s_{y,i} = -1 \ \forall\ i \in \{J+1, J+2, \ldots, 2J\} \cup \{3J+1, 3J+2, \ldots, 4J\}$$

as well as vectors

$$\boldsymbol{a}_i^T = s_{x,i} \cdot \boldsymbol{e}_i^T \cdot (\mathbf{D}^{cx} \cdot \mathbf{M}_{cx} + \mathbf{M}_{rx}) + s_{y,i} \cdot \boldsymbol{e}_i^T \cdot \left(\mathbf{D}^{cx} \cdot \mathbf{M}_{cy} + \mathbf{M}_{ry}\right)$$

and functions

$$b_i(\boldsymbol{u}_x, \boldsymbol{u}_y) = s_{x,i} \cdot \boldsymbol{e}_i^T \cdot \mathbf{D}^{cx} \cdot \boldsymbol{u}_x + s_{y,i} \cdot \boldsymbol{e}_i^T \cdot \mathbf{D}^{cy} \cdot \boldsymbol{u}_y$$

for $i \in \{1,2,\ldots,4J\}$, optimization (7) can be rewritten as

$$\left(\boldsymbol{u}_x^*, \boldsymbol{u}_y^*\right) = \underset{\substack{\boldsymbol{u}_x \in \mathbb{R}^{N_x}, \\ \boldsymbol{u}_y \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \ \min_{\tau \in \mathbb{R}} \tau$$

$$s.t. \cdot \tau \geq \max_{\xi \in \xi}\{a_i^T \cdot \xi + b_i(u_x, u_y)\} \cdot \forall\ i \in \{1,2,\ldots,4J\} \quad (8)$$

Functions b_i do not depend on ξ and hence, (8) can be rewritten as

$$\left(\boldsymbol{u}_x^*, \boldsymbol{u}_y^*\right) = \underset{\substack{\boldsymbol{u}_x \in \mathbb{R}^{N_x}, \\ \boldsymbol{u}_y \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \ \min_{\tau \in \mathbb{R}} \tau \quad (9)$$

$$\mathrm{s.t} \cdot \tau - \mathrm{b_i}\left(\mathbf{u_x}, \mathbf{u_y}\right) \geq \max_{\boldsymbol{\xi} \in \boldsymbol{\xi}}\{\mathbf{a_i^T} \cdot \boldsymbol{\xi}\} \ \forall\ \mathrm{i} \in \{1,2,\ldots,4J\} \quad (10)$$

Following [12], one can dualize optimization problems (10), transforming them into

$$\tau - b_i\left(\boldsymbol{u}_x, \boldsymbol{u}_y\right) \geq \min_{\boldsymbol{\theta}_i \in \mathbb{R}_+^{N_\xi}} \{\boldsymbol{\theta}_i^T \cdot \boldsymbol{g}\} \quad \forall i \in \{1,2,\ldots,4J\} \quad (11)$$

$$s.t.\ a_i^T - \theta_i^T \cdot S = 0. \quad (12)$$

where $\boldsymbol{\theta}_i$ is the vector of dual variables corresponding to the constraint $\boldsymbol{\xi} \in \boldsymbol{\xi}$ in the i^{th} optimization problem described by (10).

Let us observe that (11) is always satisfied as long as we have

$$\tau - b_i(\boldsymbol{u}_x, \boldsymbol{u}_y) \geq \boldsymbol{\theta}_i^T \cdot \boldsymbol{g}$$

This more stringent constraint can now be used to eliminate the minimization in (11) the final formulation of the optimization problem can be expressed as the following LP

$$\left(\boldsymbol{u}_x^*, \boldsymbol{u}_y^*\right) = \underset{\substack{\boldsymbol{u}_x \in \mathbb{R}^{N_x}, \\ \boldsymbol{u}_y \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \ \min_{\substack{\tau \in R \\ \theta_i \in \mathbb{R}_+^{N_\theta}}} \tau$$

$$s.t. \begin{cases} \tau - b_i(\boldsymbol{u}_x, \boldsymbol{u}_y) \geq \boldsymbol{\theta}_i^T \cdot \boldsymbol{g} \ \forall i \in \{1,2,\ldots,4J\} \\ \boldsymbol{a}_i^T - \boldsymbol{\theta}_i^T \cdot \mathbf{S} = 0 \ \forall i \in \{1,2,\ldots,4J\} \end{cases} \quad (13)$$

This LP can be solved within seconds using standard solvers. Hence, one can use (13) to obtain control vectors $\boldsymbol{u}_x^*$ and $\boldsymbol{u}_y^*$ for the lithography tool which will minimize the maximum of overlay errors in the relevant pattern, under the worst-case scenario of bounded uncertainties in the bias terms and modeling residuals.

3. Results

The min-max-max robust control method described in the previous section is evaluated using lithography overlay error models and measurements of overlay errors obtained from 4 consecutive pattern layers of more than 80 wafers processed in a major 300 mm fab. Due to its proprietary nature, the raw data was altered via random, but consistent scaling, which removed physical units from the data, but still allowed relative comparison of controller performances in terms of outgoing overlay errors.

Performance of the newly proposed control algorithm was compared to that of R2R control approach commonly utilized by industrial practitioners, as well as with performance of the overlay control method described in [10], where Euclidean L^2 norms of overlay errors rather than their worst magnitudes are controlled. Gaussian process regression (GPR) based prediction [13] was used to predict process biases and thus realize the R2R control approach. Even though other prediction methods, such as Kalman Filtering or Exponentially Weighted Moving Average based prediction, are more commonly used in the industry [6], GPR was used in this paper because it allows for the modeling of non-stationary, heteroscedastic processes typically observed in a fab. It also naturally yields predictions of process biases, along with the corresponding uncertainty bounds needed for the underlying robust optimization.

Following suggestion from fab engineers, predictions of process biases and corresponding uncertainty bounds were obtained using a moving window of 30 wafers which was sliding through the actual fab data. For each of those windows of 30 wafers, stochastic behavior of overlays for the next wafer was simulated by the following procedure. We drew 100 samples of process biases from the corresponding GPR-predicted normal distributions, and for each marker location, we drew 100 samples of residuals from actual, location-specific distributions of those residuals, modelled based on the data from the fab. Then, for each of those 100 samples of vectors of biases and residuals, we calculated the overlays at each measurement marker using Zernike polynomial models of the form (1), with control signals $\boldsymbol{u}$ given by the R2R controller (labelled as *R2R control strategy*), by the control method from [10] (labelled as Robust L^2 control strategy) and by the novel robust control method introduced in this paper (labelled as L^∞ control strategy). This produced 3 distributions of vectors of overlay errors: one for R2R control strategy, one for the L^2 control strategy and one for the L^∞ control strategy. Those distributions were then compared and results of those comparisons are presented here. Please note that the actual values have been removed from all figures due to the proprietary nature of the data.

Fig. 1 and Table 1 show results from 100 simulations of the 31st wafer, which correspond to simulations based on the data records from the first window of 30 wafers. These simulations produced 100 values for the mean, standard deviation, median, and maxima of overlay errors in all measurement markers across the 31st wafer. Fig. 1 shows box-and-whisker plots of those distributions for the three controllers, while Table 1 offers a summarized results of comparisons of several metrics of those distributions. Namely, top entries in each cell of Table 1 show percent reductions in the average, maximum and standard deviation of the distribution of overlay error magnitudes produced by the newly proposed L^∞-robust controller, when compared to the corresponding metrics in the distribution of overlay error magnitudes produced by the R2R controller. Similarly, bottom entries in each cell of Table 1 show the corresponding percent reductions given by the L^2-robust controller from [10], when compared to the distribution metrics produced by the R2R controller. It is clearly visible that on this wafer, the newly proposed L^∞-robust control strategy outperformed the benchmark control strategies in every metric, with that improvement being even larger relative to the R2R controller.

Fig. 2 shows averages, standard deviations, medians and maxima of distributions of overlay magnitudes obtained when the above-described simulations were performed for 20 consecutive wafers (i.e. with 20 consecutive windows of 30 wafers sliding through the historical data from the fab). Visual inspection of these results shows that the novel L^∞-robust control strategy consistently outperforms the benchmark strategies in every metric, with that improvement being quite remarkable when compared to the R2R control performance. This is confirmed by the data shown in Table 1, which lists statistical metrics of the distributions shown in Fig. 1. Furthermore, the novel control scheme significantly outperforms both benchmark controllers when it comes to the maximal overlay

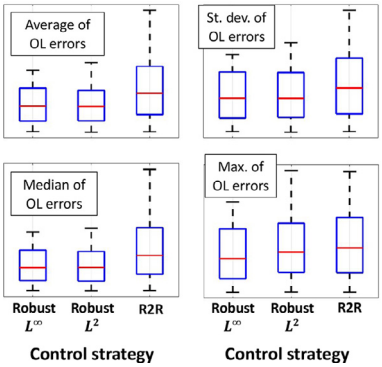

Fig. 1. Average, standard deviation, median, and maximum of overlay errors across 100 simulations of the 31st wafer under the R2R, L^2-robust *and* L^∞-robust control strategies.

Table 1
Percent reductions in metrics of the distributions of statistical measures of overlay errors across 100 simulations of the 31st wafer.

	Metric of the distribution % Red. of Average	% Red. of Median	% Red. of Maximum	% Red. of St. Deviation	
Distribution of	**Average OL Errors**	**52.3** (L^∞) **40.6** (L^2)	**47.9** (L^∞) **45.9** (L^2)	**61.0** (L^∞) **44.9** (L^2)	**57.0** (L^∞) **42.6** (L^2)
	Median OL Errors	**56.1** (L^∞) **43.3** (L^2)	**50.5** (L^∞) **47.0** (L^2)	**64.0** (L^∞) **47.0** (L^2)	**60.3** (L^∞) **44.9** (L^2)
	Maximum OL Errors	**32.6** (L^∞) **16.6** (L^2)	**30.4** (L^∞) **29.5** (L^2)	**34.2** (L^∞) **8.7** (L^2)	**36.2** (L^∞) **12.7** (L^2)
	St. Dev. of OL Errors	**29.2** (L^∞) **24.4** (L^2)	**30.8** (L^∞) **34.2** (L^2)	**32.9** (L^∞) **28.3** (L^2)	**36.2** (L^∞) **25.2** (L^2)

error magnitudes on each wafer, which is not surprising given that the new controller was indeed designed to robustly control the worst-case overlay magnitudes. Finally, in terms of computational speed, the novel control scheme generates control commands within a few seconds, while the iterative optimization in the foundation of the L^2-robust strategy from [10] truncated its iterations after about 15 min per pattern layer. This demonstrates yet another clear advantage of the newly proposed control strategy and explains why it kept outperforming the L^2-robust strategy even in terms of the average and median of overlay error magnitudes.

Finally, let us note that the R2R methods rely on the assumption that the distributions of process biases and residuals are NIID. Since simulations conducted in this paper generated process biases by sampling from a normal distribution, one can infer that the R2R controller performed better in these simulations than it would have under a more realistic distribution of process biases. Therefore, implementation of the newly proposed controller in a real system would likely lead to even more pronounced performance improvements over the traditional R2R controller.

4. Conclusions and future work

This paper presents a novel method for control of overlay errors in photolithography processes which minimizes the maximum of the overlay errors across all measurement markers in the entire pattern layer, and does that robustly, for the worst case of uncertainties in bias prediction errors and overlay modeling residuals. When compared with the industrial standard of R2R control strategy, simulations of this method produce pattern layers with significantly smaller average and maximum overlay errors, as well as smaller standard deviations of overlay errors. These improved metrics could lead to higher yields and smaller minimum feature sizes for semiconductor manufacturers.

There are several directions for future work that could build upon the results of this paper. Firstly, this paper does not consider stack-up overlay errors, i.e. errors across non-neighboring pattern layers. Stack-up overlay errors are highly impacting on the performance metrics of the resulting products and future adaptations of the approach presented in this paper are needed to account for them. However, straightforward expansion of the methods presented in this paper would not be possible. Namely, the use of absolute values in the objective function (2), which does not account for stack-up overlay errors, required $4 \cdot J$ constraints in Eq. (5) to account for each permutation of the signs. Following such an approach to include an additional layer into (2) and thus account for stack-up overlay errors would require $16 \cdot J$ constraints and this number would be multiplied by 4 with each additional layer. This exponential scaling becomes unfeasible with modern semiconductor manufacturing which requires dozens of layers to be aligned properly. Future inclusion of stack-up overlay errors would require some reformulation of the problem (2) which would allow for more manageable scaling. Finally, the newly proposed control strategy is ready for implementation in an industrial setting where its true effects on overlay errors and yield rates can be observed.

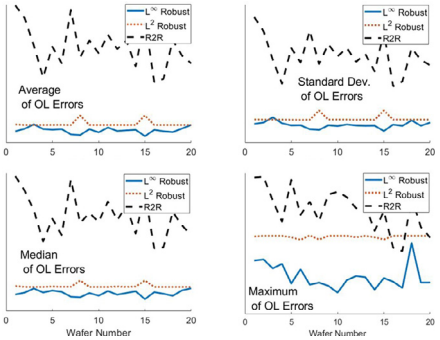

Fig. 2. Average, standard deviation, median, and maximum of overlay errors across 100 simulations of each wafer over 20 wafers under the R2R, wafer under the R2R, L^2-robust *and* L^∞-robust control strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported in part by Samsung Advanced Institute of Technology (SAIT) through the Global Research Outreach (GRO) program. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of SAIT.

References

[1] Geng H (2018) *Semiconductor Manufacturing Handbook*, McGraw-Hill Educ.New York, NY, USA.
[2] Ghaida RS, Gupta M, Gupta P (2013) Framework for Exploring the Interaction Between Design Rules and Overlay Control. *Journal of Micro/Nanolithography, MEMS, and MOEMS* 12(3). 033014–033014.
[3] Mack CA (2012) *Fundamental Principles of Optical Lithography: The Science of Microfabrication*, WileyChichester, U.K.
[4] Chien CF, Chang KH, Chen CP (2003) Design of a Sampling Strategy for Measuring and Compensating for Overlay Errors in Semiconductor Manufacturing. *Int'l. Jour. of Prod. Research* 41(11):2547–2561.
[5] v. d. Brink MA, de Mol CGM, George RA (1988) Matching Performance for Multiple Wafer Steppers Using an Advanced Metrology Procedure. *SPIE Proc. Integrated Circuit Metrology, Inspection, and Proc. Control II* 921:180–197.
[6] Tan F, Pan ZLT, Chen S (2015) Survey on Run-to-Run Control Algorithms in High-Mix Semiconductor Manufacturing Processes. *IEEE Trans. on Ind. Informat.* 11 (6):1435–1444.
[7] Hu SJ (1997) Stream of Variation Theory for Automotive Body Assembly. *CIRP Annals* 46(1):1–6.
[8] Jiao Y, Djurdjanovic D (2011) Stochastic Control of Multilayer Overlay in Lithography Processes. *IEEE Trans. on Semicond. Manuf.* 24(3):404–417.
[9] Djurdjanovic D, Jiao Y, Majstorovic V (2017) Multistage Manufacturing Process Control Robust to Inaccurate Knowledge about Process Noise. *CIRP Annals* 66 (1):437–440.
[10] Djurdjanovic D, Ul Haq A, Magnanini M, Majstorovic V (2019) Robust Model-Based Control of Multistage Manufacturing Processes. *CIRP Annals* 68(1):479–482.
[11] Gorissen BL, Hertog D (2013) Robust Counterparts of Inequalities Containing Sums of Maxima of Linear Functions. *European Journal of Operational Research* 227(1):30–43.
[12] Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization*, Princeton University PressPrinceton, NJ.
[13] Le QV, Smola AJ, Canu S (2005) Heteroscedastic Gaussian Process Regression. In: *Proc. of the 22nd Int. Conf. on Mach. Learn. (ICML)*, 489–496.