Data-Based Feed-Forward Control in MIMO Motion Systems

Mark Baggen, Marcel Heertjes and Ramidin Kamidi

Abstract— For MIMO motion systems, a data-based feed-forward control is derived and implemented on a wafer stage of a wafer scanner. On the basis of a quadratic objective function related to a performance-relevant time-frame of the servo error signals, the coefficients of a set of finite impulse response (FIR) filters are optimized using a Gauss-Newton method. Applied to the wafer stage both the optimization algorithm and the FIR filter structure provide the means to significantly improve upon scanning performance. This includes improved settling behavior, the reduced effect of cross talk by MIMO feed-forward forces, and the generalizing properties of these forces at set-points other than for which is optimized.

Index Terms— Data-Based Control, Finite Impulse Response Modelling, Gauss-Newton Optimization, Wafer Scanners.

I. INTRODUCTION

In industry, motion systems performing repetitive tasks are frequently encountered [8]. Examples include pick and place modules in component mounting machines, stages in lithographic machinery [3] or welding robots in assembly lines. Generally speaking these systems require fast positioning with a high position accuracy. Being exposed to large levels of acceleration, a combined feed-forward and feedback design then becomes necessary to achieve performance.

In view of performance, a data-based feed-forward design is presented which forms an extension to the generally applied model-based single-input single-output (SISO) feed-forwards. Having a quadratic objective function, a Gauss-Newton method is used to find the set of coefficients of a multi-input multi-output (MIMO) finite impulse response (FIR) filter in a single iteration (or trial). These coefficients are used to minimize the error response underlying the objective function. Because the objective function is based on measured error responses, whereas system knowledge is obtained through perturbed-parameter experiment, the given approach is strictly data-based. This is different, for example, from [5]. Furthermore, the finite impulse response filter implementation (see also [7]) puts limited constraint on the filter structure or the filter order of the feed-forward control. In fact, design freedom and limited system knowledge are expected advantages as compared to fixed-structure alternatives. Additionally, the data-based approach is expected to provide the means for improved performance as compared to model-based alternatives, see also [1], the latter being strongly limited by uncertainty inherently present in the modelling of complex industrial structures.

Applied to the wafer stage of an industrial wafer scanner, the data-based approach is shown to be effective in achieving performance. Not only for the set-points for which is optimized. But also for other set-points occurring during the process of wafer scanning. To a certain extent, this demonstrates the generalizing - and desired - properties of the optimized feed-forward control; see [2], [4], [9], [10], and [11] for different approaches but with a similar goal.

The paper is further organized as follows. In Section II, the wafer scanning principle in general and wafer stage control in particular is considered. This includes a presentation of the MIMO feed-forward structure based on FIR modelling. Data-based optimization of the FIR filter coefficients using a Gauss-Newton method is discussed in Section III. In Section IV the experimental results are discussed of the optimization approach applied to a wafer stage. The main conclusions of this work are summarized in Section V.

II. WAFER STAGE DYNAMICS AND CONTROL

To outline the wafer scanning problem, the wafer stage dynamics and control are presented in four parts: the wafer scanning principle, the wafer stage dynamics, the SISO control design, and the proposed MIMO feed-forward design.

A. Wafer Scanning Principle

In the production of ICs a wafer scanner is used to illuminate a pattern on a silicon disk called wafer, see also [3], [6]. A schematic representation of a wafer scanner and its main components is shown in Fig.1. Light passes through a mask (reticle), a lens, and onto a die on the wafer. Both reticle and wafer are positioned by accurately controlled stages that perform (repeated) scanning motions while exposing subsequent dies. We will restrict ourselves to the short-stroke of a generally dual-stroke wafer stage and study its dynamics.

B. Wafer Stage Dynamics

To model the wafer stage dynamics, we start with the simplified (horizontal) dynamics in the center of gravity:

$$\mathbf{x}_{cg}(s) = \mathbf{H}(s)\mathbf{f}_{cg}(s), \tag{1}$$

Mark Baggen is with TMC Engineering Professionals, Vestdijk 6, 5611 CA Eindhoven, The Netherlands, e-mail: `mark.baggen@asml.com`

Marcel Heertjes is with the Department of Mechanical Engineering, Eindhoven University of Technology, 5600 MD Eindhoven, The Netherlands, e-mail: `m.f.heertjes@tue.nl` and with ASML, Mechatronic Systems Development, De Run 6501, 5504 DR Veldhoven, The Netherlands, e-mail: `marcel.heertjes@asml.com`

Ramidin Kamidi is with ASML, Mechatronic Systems Development, De Run 6501, 5504 DR Veldhoven, The Netherlands, e-mail: `ramidin.kamidi@asml.com`
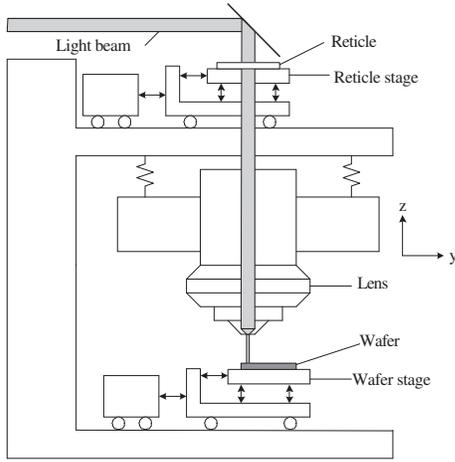
Fig. 1. Schematics of a wafer scanner and its main components.

Fig. 2. Graphical representation of the relation between control coordinates and center-of-gravity coordinates.

with s the Laplace variable, $\mathbf{x}_{\mathrm{cg}} = [x_{\mathrm{cg}}\ y_{\mathrm{cg}}\ r_{z,\mathrm{cg}}]^{\mathrm{T}}$ a column with center-of-gravity positions,

$$\mathbf{H}\,(s) = \begin{vmatrix} \frac{1}{\mathrm{m}s^2} & 0 & 0 \\ 0 & \frac{1}{\mathrm{m}s^2} & 0 \\ 0 & 0 & \frac{1}{\mathrm{J}s^2} \end{vmatrix}, \tag{2}$$

a process matrix which only contains double integrators with m the wafer stage mass and J its moment of inertia, $\mathbf{f}_{\mathrm{cg}} = [f_{x,\mathrm{cg}}\ f_{y,\mathrm{cg}}\ f_{r_z,\mathrm{cg}}]^{\mathrm{T}}$ represents a column with center-of-gravity forces, respectively. Having three actuators in the horizontal plane, it also follows that

$$\mathbf{f}_{\mathrm{cg}} = \mathbf{P}\,\mathbf{f}_{\mathrm{act}}, \tag{3}$$

with $\mathbf{P}$ a matrix containing the actuator geometry, and $\mathbf{f}_{\mathrm{act}} = [f_1\ f_2\ f_3]^{\mathrm{T}}$ a column with actuator forces.

Since the point-to-be-controlled refers to the die position that during illumination is located under the lens, see Fig.1, the center-of-gravity position $\mathbf{x}_{\mathrm{cg}}$ and this controlled die position $\mathbf{x}_{\mathrm{ctrl}} = [x_{\mathrm{ctrl}}\ y_{\mathrm{ctrl}}\ r_{z\,\mathrm{ctrl}}]^{\mathrm{T}}$ are related by

$$\mathbf{x}_{\mathrm{ctrl}} = \mathbf{Q}\,\mathbf{x}_{\mathrm{cg}}, \tag{4}$$

with

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & (y_{\mathrm{setp}} + y_{\mathrm{cg,offset}}) \\ 0 & 1 & -(x_{\mathrm{setp}} + x_{\mathrm{cg,offset}}) \\ 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

a position-dependent (and linearized) matrix with x_{setp} and y_{setp} the wafer stage position set-points and $x_{\mathrm{cg,offset}}$ and $y_{\mathrm{cg,offset}}$ fixed off-sets between the center of the stage and the center of gravity. Graphically, the relations are given in Fig.2. The overall transfer from the actuator forces $\mathbf{f}_{\mathrm{act}}$ to the controlled positions $\mathbf{x}_{\mathrm{ctrl}}$ then becomes

$$\mathbf{H}_{\mathrm{raw}}\,(s) = \mathbf{Q}\mathbf{H}\,(s)\,\mathbf{P}. \tag{6}$$

Because (6) represents a model of the actual plant, it can only be measured through the transfer between the actuator forces $\mathbf{f}_{\mathrm{act}}$ and the controlled positions $\mathbf{x}_{\mathrm{ctrl}}$. From a modelling point of view, however, the relation between the controller

forces $\mathbf{f}_{\mathrm{ctrl}}$ and the actuator forces $\mathbf{f}_{\mathrm{act}}$ can be made explicit:

$$\mathbf{f}_{\mathrm{act}}(s) = \mathbf{P}^{-1}\mathbf{G}\,(s)\mathbf{f}_{\mathrm{ctrl}}(s), \tag{7}$$

with

$$\begin{aligned} \mathbf{G}\,(s) &= \mathbf{H}^{-1}\,(s)\,\mathbf{Q}^{-1}\mathbf{H}\,(s) \\ &= \begin{bmatrix} 1 & 0 & -\frac{\mathrm{m}}{\mathrm{J}}(y_{\mathrm{setp}} + y_{\mathrm{cg,offset}}) \\ 0 & 1 & \frac{\mathrm{m}}{\mathrm{J}}(x_{\mathrm{setp}} + x_{\mathrm{cg,offset}}) \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{8}$$

Herein it is assumed that $\mathbf{x}_{\mathrm{ctrl}}(s) = \mathbf{H}\,(s)\mathbf{f}_{\mathrm{ctrl}}(s)$. The combination of Eqs.(6) and (7) now gives

$$\mathbf{x}_{\mathrm{ctrl}}(s) = \mathbf{H}_{\mathrm{raw}}\,(s)\,\mathbf{P}^{-1}\mathbf{G}\,(s)\mathbf{f}_{\mathrm{ctrl}}(s), \tag{9}$$

hence an expression for the wafer stage dynamics.

The validity of (9) is partly shown in Fig. 3 where the transfer from $\mathbf{f}_{\mathrm{ctrl}}$ to $\mathbf{x}_{\mathrm{ctrl}}$ is depicted in Bode magnitude representation. At the center of gravity, it can be seen

Fig. 3. Bode magnitude representation of the MIMO wafer stage dynamics in the horizontal plane.

that double integrator behavior dominates the low-frequent

diagonal entries. In the high-frequency range, higher-order dynamics appear. Near 180 Hz (the desired bandwidth) the diagonal terms sufficiently exceed the off-diagonal terms in amplitude indicating a proper decoupling. At either the low- or the high-frequency range, however, the dynamics still show a significant expression of coupling. In the low-frequency range this is due to parasitic stiffness and damping contributions. In the high-frequency range, it can be seen that the decoupling, which is based on rigid-body modelling, neglects flexibility. In view of these clear shortcomings, the effect of SISO control is studied.

C. SISO Control Design

A SISO control design is shown in the simplified block diagram of Fig. 4. It consists of three parts: a feedback con-

Fig. 4. Block diagram of the controlled wafer stage dynamics.

troller $\mathbf{C}$, an inertia-based feed-forward controller $\mathbf{F}_{\text{ff}}$, and a MIMO feed-forward controller $\mathbf{F}_{\text{fir}}$. $\mathbf{x}_{\text{setp}}$ and $\mathbf{a}_{\text{setp}}$ represent the position and acceleration set-points, respectively.

The feedback controller $\mathbf{C}$ has a diagonal matrix structure with controllers C_x, C_y, and C_{r_z}. Each of these controllers is based on a series connection of a PID-controller, a notch filter and a low-pass filter, or

$$C_i(s) = \mathcal{PID}(s)\, \mathcal{N}(s)\, \mathcal{LP}(s), \text{ with } i \in \{x, y, r_z\}. \tag{10}$$

The PID part is used for servo tracking and is defined as

$$\mathcal{PID}(s) = \frac{k_{\text{p}}\,(s^2 + \omega_{\text{d}}\,s + \omega_{\text{i}}\omega_{\text{d}})}{\omega_{\text{d}}\,s}, \tag{11}$$

with ω_i the cut-off frequency of the integral action, ω_d the cut-off frequency of the differential action, and k_p the controller gain. A second-order notch filter is used for dealing with the first plant resonances, or

$$\mathcal{N}(s) = \left(\frac{\omega_{\text{p}}}{\omega_{\text{z}}}\right)^2 \frac{s^2 + 2\,\beta_{\text{z}}\omega_{\text{z}}\,s + \omega_{\text{z}}{}^2}{s^2 + 2\,\beta_{\text{p}}\omega_{\text{p}}\,s + \omega_{\text{p}}{}^2}. \tag{12}$$

Herein ω_z and ω_p represent the zero and pole corner frequencies, β_z and β_p the corresponding damping coefficients. A second-order low-pass filter is used to avoid the amplification of noise and reads

$$\mathcal{LP}(s) = \frac{\omega_{\text{lp}}{}^2}{s^2 + 2\,\beta_{\text{lp}}\omega_{\text{lp}}\,s + \omega_{\text{lp}}{}^2}, \tag{13}$$

with ω_{lp} a cut-off frequency and β_{lp} a dimensionless damping coefficient. On the basis of a sampling frequency of $f_{\text{s}} = 5$ kHz, a discrete-time equivalent of these controllers is implemented. The result of which is shown in Fig. 5 through the open-loop characteristics $\text{OL}_y(j\omega) = \text{C}_y(j\omega)\text{H}_y(j\omega)$

Fig. 5. Bode diagram of the measured open-loop frequency response function of a short-stroke wafer stage in scanning y-direction.

of a wafer stage in scanning y-direction. Here a controller bandwidth of ≈ 180 Hz is realized along with a phase margin of ≈ 26 degrees and a gain margin of ≈ 4.6 dB.

The feed-forward controller $\mathbf{F}_{\text{ff}}$ (see Fig.4) also has a diagonal matrix structure with controllers $F_{\text{ff},x}$ and $F_{\text{ff},y}$ but with $F_{\text{ff},r_z} = 0$; this is because the set-points in r_z-direction are zero valued. Motivated by the dominant diagonal rigid-body behavior of Fig. 3, $F_{\text{ff},i}$ is given by $F_{\text{ff},i} = \text{m}$ with $i \in \{x, y\}$. Though inducing a significant performance improvement, the model-based acceleration feed-forwards tend to limit the achievable performance of the system at hand. This is because cross-talk (recall Fig. 3) and flexibility, both being considered as model uncertainty, are not properly accounted for. To improve upon performance in view of these uncertainties, a data-based MIMO feed-forward controller $\mathbf{F}_{\text{fir}}$ is studied. The SISO model-based feed-forward controller, which strictly speaking becomes redundant, is kept as an initial estimate for the required feed-forward commands.

D. MIMO Feed-Forward Design

The MIMO feed-forward controller $\mathbf{F}_{\text{fir}}$ is shown in more detail in Fig. 6. It consists of six finite impulse response (FIR) filters. Input to these filters are the acceleration set-point signals $a_{\text{setp},x}$ and $a_{\text{setp},y}$, respectively. The output is given by the feed-forward forces $f_{\text{fir},x}$, $f_{\text{fir},y}$, and f_{fir,r_z}. Apart from the MIMO properties, the inclusion of rotations r_z in the feed-forward design proves beneficial. This is because the position is not measured at a fixed location on the stage (see Fig. 2). As a result, rotations about the center of gravity will induce position-dependent tracking errors in the translational directions. Improvements in r_z reduce this effect thereby giving improved error reproducibility.

Fig. 6.   Block diagram of the MIMO feed-forward controller structure.

In Fig.6, the relation between $a_{\text{setp},x}$ and $f_{\text{fir},x}$ expressed by F_{xx} is given by a discrete-time linear FIR filter,

$$f_{\text{fir},x}(j) = \sum_{i=-n_1}^{n_2} \theta_{xx}\ (i)a_{\text{setp},x}(j-i). \qquad (14)$$

Herein j denotes a sampling instance, $\theta_{xx} \in \mathbb{R}^{n_1+n_2+1\times 1}$ is a vector containing the FIR filter coefficients, $n = n_1 + n_2 + 1$ is the order of the filter, $n_1 \geqslant 0$ is the look-ahead horizon, and $n_2 \geqslant 0$ is the look-behind horizon. A non-causal mapping is obtained since $f_{\text{fir},x}(j)$ can be a function of $a_{\text{setp},x}(j + n_1)$ future set-point samples. Non-causality enables the feed-forward controller to handle non-minimum phase behavior. Moreover, differentiation of the input $a_{\text{setp},x}$ without introducing delay requires n_1 to be $n_1 \geqslant 1$. The filter order n is obtained by trial-and-error but is bound to the following trade-off: choosing n too small limits the ability to describe the required feed-forward signals. Choosing n too large involves the risk of over-fitting. In this study, a filter order of $n = 15$ with $n_1 = 5$ and $n_2 = 9$ is used. In addition to the simplified structure of Fig.6, a direction-dependent structure is used. Given the sign of the accelerations, two different FIR filters are used giving a total of 12 filters. The contribution of the paper now consists of deriving a set of FIR filter coefficients through data-based optimization.

III. DATA-BASED OPTIMIZATION

In deriving the FIR filter coefficients of the MIMO feed-forward control, a data-based optimization approach is used. The approach consists of three steps: the choice for an objective function, the optimization algorithm needed to minimize this objective function, and the derivation of its gradients used in the process of optimization.

The objective function V is chosen to be a quadratic

function of the measured tracking error, or

$$V(\theta) = e_x^{\text{T}}(\theta)e_x(\theta) + e_y^{\text{T}}(\theta)e_y(\theta) + \gamma e_{r_z}^{\text{T}}(\theta)e_{r_z}(\theta). \quad (15)$$

Hence the sum of squares of the tracking error along a repeated task. To equally distribute the individual error contributions, the scaling factor γ is chosen to be 0.01 m^2/rad^2. This value results from a scaling toward the wafer radius dimension of 0.1 m. Moreover, the tracking error is assumed to be a linear function of the controller parameters $\theta = [\theta_{xx}^{\text{T}} \ldots \theta_{yr_z}^{\text{T}}]^{\text{T}}$, i.e., the FIR filter coefficients, with $\theta \in \mathbb{R}^{m\times1}$. As such the objective function will be a quadratic function of these parameters, thus giving a convex optimization problem. In minimizing the objective function, a second-order Taylor series expansion of V about θ_k is considered:

$$\begin{aligned} V(\theta) \approx V(\theta_k) &+ \nabla V(\theta_k)^{\text{T}}(\theta - \theta_k) \\ &+ \frac{1}{2}(\theta - \theta_k)^{\text{T}}\nabla^2 V(\theta_k)(\theta - \theta_k). \end{aligned} \qquad (16)$$

Minimizing (16) such that $V(\theta_k) \to V(\theta_{k+1})$ follows from applying a (one-trial) Gauss-Newton update law [5],

$$\theta_{k+1} = \theta_k - 2\left(\nabla^2 V(\theta_k) + \nabla^2 V(\theta_k)^{\text{T}}\right)^{-1} \nabla V(\theta_k). \quad (17)$$

Here k refers to the trial number, $\nabla V(\theta) \in \mathbb{R}^{m\times1}$ is the gradient vector of the objective function with respect to the FIR filter coefficients θ, and $\nabla^2 V(\theta) \in \mathbb{R}^{m\times m}$ is the corresponding Hessian matrix.

To obtain both $\nabla V(\theta)$ and $\nabla^2 V(\theta)$, measured data is used from the tracking error signals along a performance-relevant time interval – this is the interval of constant velocity in between acceleration and deceleration where illumination takes place. For each trial k, two experiments are conducted. One experiment is performed under initial parameter conditions. This gives the error signal $e_0(\theta_0)$. A second experiment is conducted under perturbation of filter parameter θ_i with $i \in \mathbb{N}^+$ and gives the error signal $e_p(\theta_0 + \Delta\ \theta_i)$. Note that both e_0 and e_p are time signals with dimensions depending on the time-length of the repeated task. A first-order Euler approximation of the gradient ∇e_i now follows from:

$$\nabla e_i \approx (e_{p,i} - e_0)/\Delta\ \theta_i, \qquad (18)$$

which is used to obtain

$$\begin{aligned} \nabla V(\theta_i) &\approx 2\nabla e_i^{\text{T}}(\theta)e_i(\theta), \\ \nabla^2 V(\theta_i) &\approx 2\nabla^2 e_i^{\text{T}}(\theta)e_i(\theta) + 2\nabla e_i^{\text{T}}(\theta)\nabla e_i(\theta). \end{aligned} \qquad (19)$$

Herein $2\nabla^2 e_i^{\text{T}}(\theta)e_i(\theta)$ is (generally) small and time consuming to obtain such that (19) is further approximated by

$$\nabla^2 V(\theta_i) \approx 2\nabla e_i^{\text{T}}(\theta)\nabla e_i(\theta). \qquad (20)$$

Eqs.(17), (18), (19), and (20) are subsequently used to obtain an updated set of FIR filter coefficients.

The derivation of an updated set of FIR filter coefficients generally requires a large-scale experiment which scales with the number of perturbed parameters times the number of FIR filters. To make the optimization technique more suited for practical application, the choice for a FIR filter structure appears beneficial. This is because the gradient

error responses ∇e_x, ∇e_y, and ∇e_{r_z} for θ_j with $j \in \mathbb{N}^+$ relate to the gradient error responses for θ_1 through $\theta_j = \theta_1 z^{-j+1}$; z^{-1} being a unit sampling delay. The effect is shown in Fig. 7 for two coefficients: θ_1 and θ_{10}. This reduces
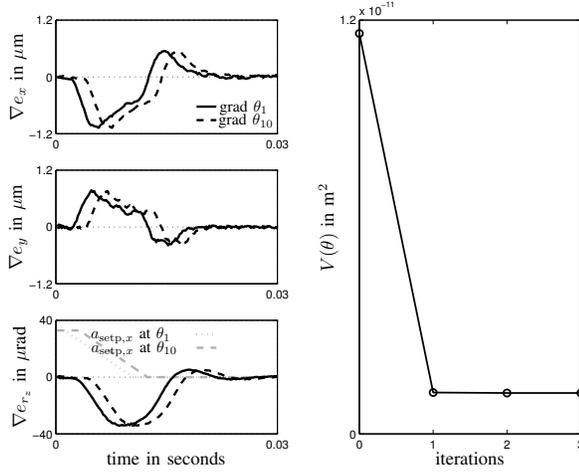
Fig. 7. Time-series measurement of the gradients ∇e_x, ∇e_y, and ∇e_{r_z} (left) and the objective function $V(\theta)$ (right).

the required number of perturbed-parameter experiments to one for each filter. In summary, the following procedure is adopted in obtaining an updated set of FIR filter coefficients.

1) Set the initial controller parameters θ to $\theta_0 = 0$.
2) Execute a repeated task and obtain the error signals e_0 from the time-relevant intervals.
3) Perturb the parameter θ_{xx} (1) of F_{xx} with $\Delta \ \theta_1$, execute the repeated task, and store the signals $e_{p,1}$.
4) Use (18) to compute the error gradient $\nabla e_{xx,1}$.
5) Apply time-shifts and get $\nabla e_{xx,i} \ \forall \ i \in [2 \ldots n+1]$.
6) Restore the parameters and repeat steps 3, 4 and 5 for the remaining filters $[F_{xy}, F_{xr_z}, F_{yx}, F_{yy}, F_{yr_z}]$.
7) Use (17), (19), and (20) to subsequently compute an updated set of filter coefficients.

For gradients obtained from the previously described experiments, the effect of the optimization method on the objective function $V(\theta)$ is shown in the right part of Fig.7. Here a significant reduction of $V(\theta)$ is obtained in a single trial. Additional trials do not seem to improve upon this result, an observation that favors the validity of the assumptions underlying the method. With this optimized set of FIR filter coefficients, performance is studied on a wafer stage.

IV. WAFER STAGE PERFORMANCE ASSESSMENT

At a wafer stage of an industrial wafer scanner performance using data-based MIMO feed-forward control is assessed through experiment. Performance is evaluated before- and after optimization and involves: i) improved wafer scanning performance at a single wafer location, ii) the ability to generalize this improvement toward other locations along the wafer, and iii) a performance deterioration in the interval beyond scanning where performance is no longer required.

At a single wafer location, improved wafer scanning performance is shown in Fig.8. For a repeated acceleration
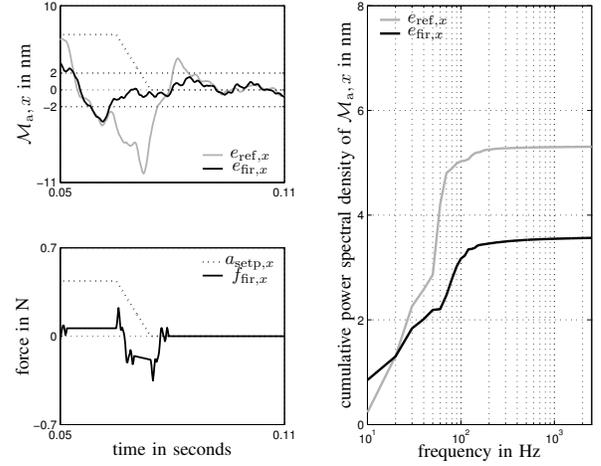
Fig. 8. Time-series measurement of the $\mathcal{M}_a$-filtered tracking errors, forces (left), and the corresponding cumulative power spectral densities (right).

set-point during scanning, the filtered error e_x is depicted before and after feed-forward optimization. Filtering is based on a moving average filter operation, which is defined as

$$\mathcal{M}_a(i) = \frac{1}{n} \sum_{j=i-n/2}^{i+n/2-1} e(j), \ \forall i \in \mathbb{Z}, \qquad (21)$$

with $n \in \mathbb{N}^+$ an application specific time frame. Eq.(21) acts as a low-pass filter operation on the error signal e and expresses the level of position accuracy (scanning overlay) that can be obtained during the process of wafer scanning. Either in time-domain (left) or via cumulative power spectral density (cpsd) analysis (right) it can be seen that the filtered errors during scanning – the interval of constant velocity beyond the acceleration phase – are much smaller than the errors without feed-forward optimization. In particular in terms of settling behavior, scanning performance is significantly improved. Note that the generalized FIR forces (left) can but partly be described in terms of $\mathbf{x}_{setp}$, $\mathbf{a}_{setp}$, and the principle derivatives. This justifies the FIR filter design with respect to the more straightforward fixed structure approaches.

Both the validity and generality of the optimized MIMO feed-forward approach is tested through five representative scans along the wafer, see Fig.9. At different locations
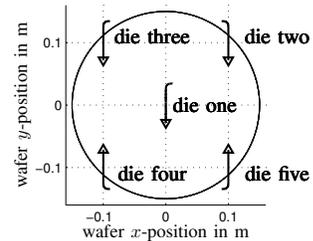
Fig. 9. Schematic overview of five different scan locations along the wafer.

numbered with die one through die five, a combined $\{x, y\}$ scan is performed, the relevant parts of which are shown in Fig.9. An optimized MIMO feed-forward is derived using the error information corresponding to these scans. This feed-forward control is subsequently applied to the wafer stage while performing these five scans, the result of which is depicted in Fig.10. It can be seen that settling behavior
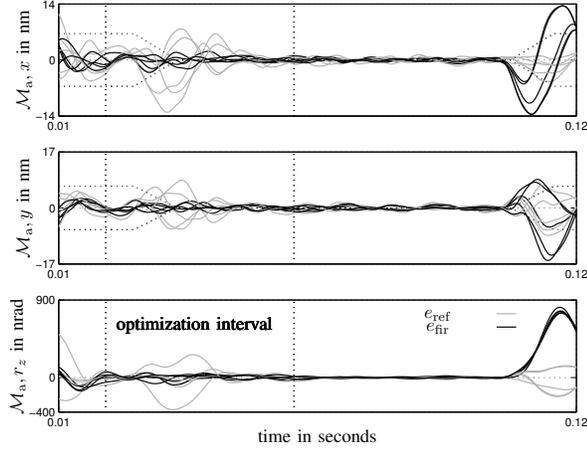
Fig. 10. Time-series measurement of the $\mathcal{M}_{\mathrm{a}}$-filtered error responses e_x, e_y, and e_{r_z} at five different die locations along the wafer.

prior to scanning is significantly improved; note the start of the interval of constant velocity (grey $\rightarrow$ black). This also follows from cumulative power spectral density analysis, see Fig.11 which shows that the rms-values of the corresponding
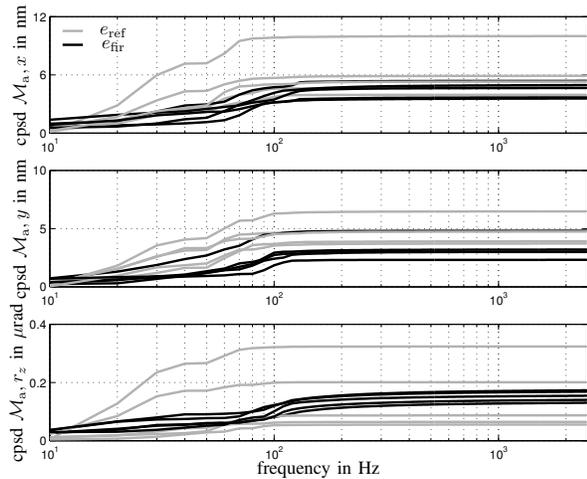
Fig. 11. Cumulative power spectral densities of the $\mathcal{M}_{\mathrm{a}}$-filtered error responses e_x, e_y, and e_{r_z} at five different die locations along the wafer.

error signals are significantly reduced. Additionally, Fig.10 shows a performance deterioration in the deceleration phase beyond the scanning interval. Though not performance limiting, this is a side-effect of the optimization approach in view of direction-dependent behavior and system nonlinearity. Hence the kind of behavior that would hinder any kind of (linear) model-based feed-forward approach. Basically, the

generalizing properties of the obtained feed-forward control appear sensitive to the degree of linearity of the underlying dynamics. It is therefore concluded that large expressions of nonlinearity either come at the price of performance or at the price of robustness to set-point variation.

V. CONCLUSIONS

Using a FIR filter structure, a data-based MIMO feed-forward control is derived with improved wafer scanning properties. The FIR filter coefficients are optimized in a single trial of a data-based optimization using a quadratic objective function on the tracking error. This requires only a single perturbed-parameter experiment for each FIR filter. Wafer scanning performance is achieved in terms of improved settling behavior. Not only for a single acceleration set-point, but in principle generalized to arbitrary set-points along the wafer. The validity of such a generalization is partly studied in view of the direction-dependent (and nonlinear) behavior such as encountered at the wafer stage. Herein performance during scanning is improved but at the cost of deteriorated performance beyond scanning. It is therefore concluded that the strength of the optimization lies in designating performance-relevant time-frames and anticipate to these time-frames at the price of less performance-relevant time-frames. Essentially, this requires a good understanding of what is performance-relevant and what is not. A distinction that appears valid for the scanning wafer stage at hand.

REFERENCES

[1] Fliess M, Lévine J, Martin P, Rouchon P. (1995) Flatness and defect of nonlinear systems: introductory theory and examples. International Journal of Control, 61, pp. 1327-1361.
[2] Gorinevski D, Torfs DE, and Goldenberg AA. (1997) Learning approximation of feed-forward control dependence on the task parameters with application to direct-drive manipulator tracking, IEEE Transactions on Robotics and Automation, 13(4), pp. 567-581.
[3] Heertjes M, and Tso T. (2007) Nonlinear iterative learning control with applications to lithographic machinery, Control Engineering Practice, 15(12), pp. 1545-1555.
[4] Kruif BJ. (2004) Function Approximation for Learning Control, PhD-thesis, University of Twente, ISBN: [90-365-2050-9].
[5] Meulen S van der, Tousain R, and Bosgra O. (2007) Fixed Structure Feedforward Controller Tuning Exploiting Iterative Trials, Applied to a High-Precision Electromechanical Servo System, In Proc. of the 2007 American Control Conference, New York, USA, pp. 4033-4039.
[6] Mishra S, Coaplen J, and Tomizuka M. (2007) Segmented iterative learning control for nonrepetitive disturbances. IEEE Control Systems Magazine, 8, pp. 20-25.
[7] Potsaid B, and Wen J. (2004) High Performance Motion Tracking Control, In Proc. of the 2004 IEEE Transactions International Conference on Control Applications, Taipei, Taiwan, pp. 718-723.
[8] Steinbuch M, and Molengraft van de MJG. (2000) Iterative learning control of industrial motion systems, In Proc. of the 1st IFAC conference on Mechatronic Systems, Darmstadt, Germany, pp. 967-972.
[9] Xu JX. (1996) Direct learning of control efforts for trajectories with different time scales, IEEE transactions on automatic control, 23(7), pp. 1027-1030.
[10] Xu JX. (1997) Direct learning of control efforts for trajectories with different magnitude scales, Automatica, 33(12), pp. 2191-2195.
[11] Lambrechts P, Boerlage M, Steinbuch M. (2005) Trajectory planning and feedforward design for electromechanical motion systems, Control Engineering Practice, 13, pp. 145-157.