Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Robust Data-Driven Control for the Stage Synchronization Problem

Marcel F. Heertjes [*,***] **Marco Galluzzo** [**] **Lucas Kuindersma** [***]

[*] *Mechanical Engineering Department, Eindhoven University of Technology, The Netherlands (e-mail corresponding author: m.f.heertjes@tue.nl).*
[**] *Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy*
[***] *Mechatronics Technology Development, ASML, The Netherlands*

Abstract: This paper discusses a method for introducing frequency-domain robustness constraints in a data-driven control design. The method is based on adaptive weighting filters used in a time-domain cost function to be minimized. In an iterative way, violation of the frequency-domain robustness constraints is checked by model prediction and accounted for by adapting the weighting filters. The method is applied to the stage synchronization problem in wafer scanners, which involves synchronization of the outputs of two motion systems, and will be discussed in view of both simulation results and experimental results.

1. INTRODUCTION

High-precision motion systems are electro-mechanical systems that perform commanded (or reference) motion often within nanometer accuracy. Examples include the pick-up units in storage drives like blu-ray discs (BD) or hard-disk drives HDD [Park et al., 1997], pick-and-place machinery in the component mounting industry and robotics [Kostić et al., 2004], stages in electron microscopes [Van Bree et al., 2010], and the high-acceleration stages in wafer scanners used to produce integrated circuits [Heertjes et al., 2010]. Most of these systems have two or more subsystems that need to be coordinated or synchronized. In this paper we focus in particular on the synchronization between the wafer and the reticle stage of a wafer scanner [Wang et al., 2009].

To deal with the stages synchronization a synchronization controller is used [Butler, 2011]. In a feed-forward sense, the synchronization controller aims at tracking the measured wafer stage input signals such that the synchronization error between the wafer and the reticle stage motion systems is minimized. The synchronization controller suffers from a causality problem, however. The wafer stage input signals need to be measured before they can be compensated for by the synchronization controller. As a result of this causality problem, proper tracking of the wafer stage input signals at frequencies of interest generally comes at the price of deteriorated tracking at other frequencies, i.e. a waterbed effect which is formalized using the so-called feed-forward sensitivity function [Heertjes et al., 2013]; the latter being the transfer between wafer stage error input and synchronized error output. This will be further referred to as the stage synchronization problem.

Minimizing the synchronized error in view of the stage synchronization problem will be done with a data-driven approach, see also Mishra et al. [2008] for an approach based on iterative learning control. In Hjalmarsson [2005] it is shown that many of the data-driven approaches [Bazanella et al., 2012] can be interpreted as being model-based, except for iterative feedback tuning (often abbreviated with IFT); IFT is an iterative optimization approach that aims at obtaining unbiased gradient estimates by conducting multiple experiments. IFT, however, suffers from robustness issues as there are generally no guarantees for closed-loop stability along the iterations, hence the development of IFT algorithms that include robust stability measures [Veres and Hjalmarsson, 2002, Prochazka et al., 2005]. For the stage synchronization problem, robustness is of paramount importance: (a) low-frequency disturbance suppression is required as to meet robust performance, and (b) high-frequency amplification should be contained in view of robust stability. It thus makes sense to reside to (non-parametric) plant modeling as a means to impose robustness constraints on the feed-forward sensitivity function, hence a model-based interpretation.

In a data-driven context, this paper describes a new approach toward including robust stability measures. For a candidate set of optimized synchronization controller parameters, which will be of the finite impulse response (FIR)-type, each iteration it will be checked if the feed-forward sensitivity function satisfies the frequency-domain robustness constraints. This is done by computing the feed-forward sensitivity function using a non-parametric plant model. If satisfied, the candidate set is accepted and the optimization algorithm proceeds to the next iteration. If not satisfied, weighting filters will be derived that aim at counteracting the feed-forward sensitivity function at the specific frequency intervals of violation. The weighting filters are based on notch filter design. By filtering the error responses with the weighting filters in the indicated frequency intervals of violation, solutions that would otherwise lead to these violations will be penalized. With the weighting filters a new candidate set is derived. In general, the weighting filters are adapted until the robustness constraints are fully met. In this way, local convexity properties of the constrained optimization problem are preserved. See also Van der Velden et al. [2014] where the constrained optimization problem in an IFT context is transferred into an unconstrained optimization problem but at the cost of convexity properties.

The remainder of the paper is organized as follows. In Section 2, the stage synchronization problem will be further expedited. In Section 3 the data-driven optimization approach used to obtain the synchronization controller parameters, i.e. the FIR coefficients, will be discussed. This includes the main contribution of this paper: imposing frequency-domain robustness constraints via adaptive weighting filter design. Section 4 discusses experimental results obtained from an industrial wafer scanner. In Section 5, the main conclusions will be summarized.

2. STAGE SYNCHRONIZATION PROBLEM

In discussing the stage synchronization problem, consider the schematics of a wafer scanner in Fig.1. (Extreme) ultraviolet

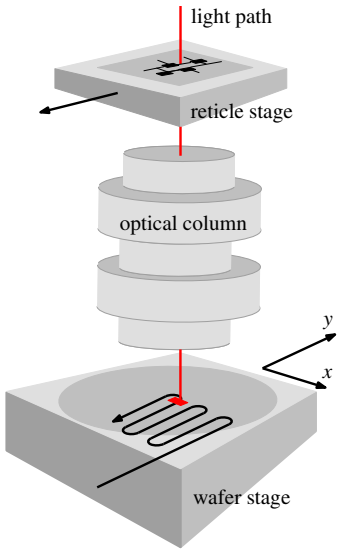

Fig. 1. Schematics of a wafer scanner.

light containing an image of the integrated circuits to be processed travels via a light path through an optical column to expose the light sensitive layers of a wafer. The image is obtained from the reticle which is part of the reticle stage motion control systems. Similarly, the wafer is part of the wafer stage motion control systems. During wafer scanning both the reticle and the wafer stage systems track a series of point-to-point motions in (scanning) y-direction.

In terms of control, consider the simplified block diagram of Fig.2, see also Mishra et al. [2008], Wang et al. [2009], Xiao & Zhu [2006] for other configurations. In this figure, $y_{\rm ws}$ and $y_{\rm rs}$

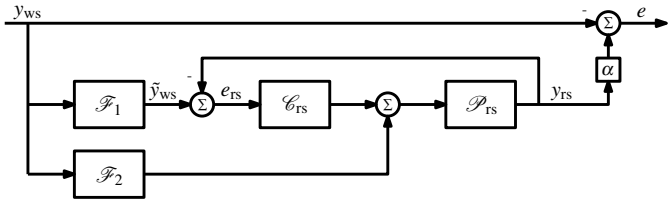

Fig. 2. Simplified stage synchronization design.

represent the output signals of respectively the wafer and the reticle stage. With the optical factor α these outputs define the synchronization error via

$$e = \alpha y_{\rm rs} - y_{\rm ws}. \tag{1}$$

The reticle stage plant is given by $\mathscr{P}_{\rm rs}$ whereas its feedback controller is given by $\mathscr{C}_{\rm rs}$. The error signal $e_{\rm rs}$ is defined by

$$e_{\rm rs} = \tilde{y}_{\rm ws} - y_{\rm rs}, \tag{2}$$

with $\tilde{y}_{\rm ws}$ the shaped wafer stage output signal. The synchronization controller consists of two parts: $\mathscr{F}_1$ and $\mathscr{F}_2$.

Let $\mathscr{P}_{\rm rs}$ be a continuous-time double integrator plant (stages are of the double integrator type or extensions thereof):

$$\mathscr{P}_{\rm rs}(s) = \frac{1}{ms^2}, \tag{3}$$

with m the reticle stage mass and s the Laplace variable. By choosing the synchronization controller as

$$\mathscr{F}_1(s) = \frac{1}{\alpha} \text{ and } \mathscr{F}_2(s) = \frac{ms^2}{\alpha}, \tag{4}$$

it follows from (2), (3), and (4) that

$$e_{\rm rs}(s) = \mathscr{F}_1(s) y_{\rm ws}(s) - \mathscr{P}_{\rm rs}(s)\mathscr{F}_2(s) y_{\rm ws}(s) = 0. \tag{5}$$

Moreover, in combination with (1) it follows that:

$$e(s) = \alpha \mathscr{F}_1(s) y_{\rm ws}(s) - y_{\rm ws}(s) = 0. \tag{6}$$

Zero error tracking $e = 0$ is thus obtained with the synchronization controller in (4), the latter being the solution to the stage synchronization problem. In practice this solution is not feasible, however. The reason is because of causality.

Consider $\mathscr{P}_{\rm rs}$ in (3) but sampled using a zero-order hold circuit. In discrete-time, it follows that

$$\mathscr{P}_{\rm rs}(z) = (1-z^{-1})\mathscr{Z}\left\{\frac{\mathscr{P}_{\rm rs}(s)}{s}\right\} = \frac{\frac{1}{2}z^{-1} + \frac{1}{2}z^{-2}}{c_0(1-2z^{-1}+z^{-2})}, \tag{7}$$

with constant $c_0 = mT^{-2}$, z^{-1} a unit time delay, and $\mathscr{Z}\{\cdot\}$ denoting the z-transform. It can be checked from Fig.2 that

$$\mathscr{F}_1(z) = \frac{1}{2\alpha}z^{-1} + \frac{1}{2\alpha}z^{-2} \text{ and } \mathscr{F}_2(z) = \frac{c_0}{\alpha}\left(1-2z^{-1}+z^{-2}\right), \tag{8}$$

yields zero error tracking $e_{\rm rs} = 0$ irrespective of the discrete-time feedback controller $\mathscr{C}_{\rm rs}$. The inverse dynamics problem resulting from inverting the proper plant $\mathscr{P}_{\rm rs}$ in (3) is thus circumvented by delaying the input signal $y_{\rm ws}$ to the reticle stage system via $\mathscr{F}_1$ in (8). The synchronized error e follows from (1), (2), (7), and (8) with $e_{\rm rs} = 0$, and is given by

$$e(z) = \left(-1 + \frac{1}{2}z^{-1} + \frac{1}{2}z^{-2}\right) y_{\rm ws}(z) = \mathscr{S}_{\rm ff}(z) y_{\rm ws}(z), \tag{9}$$

with $\mathscr{S}_{\rm ff}$ the so-called feed-forward sensitivity function. Note that e is non-zero for $y_{\rm ws} \neq 0$. Due to time delay, the signal $y_{\rm rs}$ is generally too late in compensating for $y_{\rm ws}$ in (1) and therefore zero synchronized error tracking is no longer achieved.

This also follows from the feed-forward sensitivity function $\mathscr{S}_{\rm ff}$ in (9), which is shown in Fig.3 in Bode representation. It can be

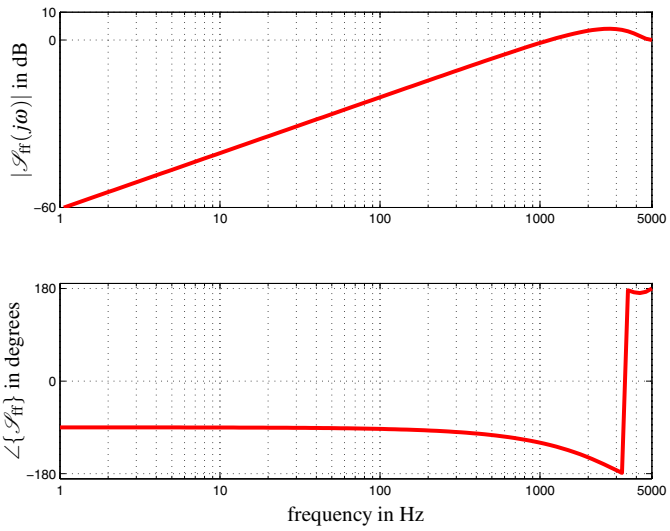

Fig. 3. Bode diagram of the feed-forward sensitivity function.

seen that up to 1 kHz the synchronized error e is suppressed with respect to the wafer stage signal $y_{\rm ws}$. The zero cross-over frequency is referred to as the bandwidth of the synchronization controller. Beyond this bandwidth, $y_{\rm ws}$ is amplified in e. This waterbed effect is described in Heertjes et al. [2013] and is formalized by

$$\int_0^\infty \ln(|\mathscr{S}_{\rm ff}(j\omega)|)d\omega = \pi \sum_{i=1}^{N_z} \Re\{z_i\}, \tag{10}$$

with N_z being the number of right-half plane zeros of $\mathscr{S}_{\rm ff}$.

In discrete-time, the stage synchronization problem thus involves minimizing the synchronized error e, which often implies maximizing the synchronization controller bandwidth, by a proper choice of filters $\mathscr{F}_1$ and $\mathscr{F}_2$. Given the structure in (8), it seems quite natural to choose these filters of the finite impulse response (FIR)-type:

$$\begin{aligned}\mathscr{F}_1(z) &= a_0 + a_1 z^{-1} + \cdots + a_n z^{-n} \\ \mathscr{F}_2(z) &= b_0 + b_1 z^{-1} + \cdots + b_n z^{-n},\end{aligned} \tag{11}$$

with a_i, b_i, $i \in \{1 \ldots n\}$, the FIR coefficients, see also Jouaneh & Anderson [2006]. Furthermore, minimization should be done in view of the waterbed effect in (10). That is, we want to avoid undesired (high-frequency) amplifications that compromise machine performance. We therefore seek the set of FIR coefficients $\{a_i, b_i\}$ in (11) through data-driven optimization that minimize a cost function $J(e)$ while satisfying frequency-domain robustness constraints on $\mathscr{S}_{\text{ff}}$.

3. DATA-DRIVEN OPTIMIZATION APPROACH

In finding the FIR filter coefficients through data-driven optimization, consider the $\mathscr{H}_2$ cost function

$$J = \mathbf{e}^{\mathrm{T}}\mathbf{e}, \tag{12}$$

which represents the quadratic sum of the synchronized error signals $\mathbf{e} = [e(1) \ldots e(k)]^{\mathrm{T}}$ with k the number of data samples within a relevant performance interval. Furthermore, let the FIR filter parameters to-be-optimized be given by

$$\mathbf{p}(\kappa) = [a_0(\kappa) \ldots a_n(\kappa)\ b_0(\kappa) \ldots b_n(\kappa)]^{\mathrm{T}}, \tag{13}$$

with κ the iteration number. Firstly, it should be remarked that the FIR filter optimization procedure is considered to be part of a machine calibration start-up sequence. As a result, machine throughput is not compromised by the number of iterations even in the case where κ is allowed to become large. Secondly, for the simple example in (8), where the physical interpretation of the FIR coefficients is straightforward, it follows that y_{ws} generally leads to non-zero (and non-minimal) synchronization errors e, hence the feedforward sensitivity function in (3). Minimizing (12) therefore comes at the cost of the physical interpretation of the FIR coefficients and strongly relates to the frequency contents of y_{ws} which is generally unknown.

With the Gauss-Newton method [Bazanella et al., 2012], the FIR coefficients $\mathbf{p}$ in (13) that minimize (12), or

$$\tilde{\mathbf{p}} = \arg\min_{\mathbf{p}(\kappa)} J, \tag{14}$$

are generally found through

$$\mathbf{p}(\kappa+1) = \mathbf{p}(\kappa) - \zeta(\nabla^{\mathrm{T}}\mathbf{e}\nabla\mathbf{e})^{-1}(\nabla^{\mathrm{T}}\mathbf{e})\mathbf{e}, \tag{15}$$

with $\mathbf{p}(0) = \mathbf{0}$, damping coefficient $0 < \zeta \leq 1$, and gradients

$$\nabla\mathbf{e} = \left[\frac{\partial \mathbf{e}}{\partial a_0} \cdots \frac{\partial \mathbf{e}}{\partial a_n}\ \frac{\partial \mathbf{e}}{\partial b_0} \cdots \frac{\partial \mathbf{e}}{\partial b_n}\right]. \tag{16}$$

In a data-driven context finding $\tilde{\mathbf{p}}$ in (14) using (15) essentially boils down to determining two gradient error signals: $\partial\mathbf{e}/\partial a_0$ and $\partial\mathbf{e}/\partial b_0$. For each iteration κ, $\mathbf{e}$ in (15) can be obtained directly from measurement whereas the remaining gradient error signals in (16) are obtained using the fact that within the FIR filter structure, each FIR coefficient is delayed one sample with respect to its predecessor, or

$$\frac{\partial \mathbf{e}}{\partial a_n} = \frac{\partial \mathbf{e}}{\partial a_0} z^{-n} \text{ and } \frac{\partial \mathbf{e}}{\partial b_n} = \frac{\partial \mathbf{e}}{\partial b_0} z^{-n}. \tag{17}$$

To derive the gradient error signals: $\partial\mathbf{e}/\partial a_0$ and $\partial\mathbf{e}/\partial b_0$, it can be derived using Fig.2 that

$$y_{\text{rs}}(z) = \mathscr{S}_{\text{rs}}(z)\left(\mathscr{P}_{\text{rs}}(z)\mathscr{C}_{\text{rs}}(z)\mathscr{F}_1(z) + \mathscr{P}_{\text{rs}}(z)\mathscr{F}_2(z)\right) y_{\text{ws}}(z), \tag{18}$$

where $\mathscr{S}_{\text{rs}}$ is the reticle stage sensitivity function, or

$$\mathscr{S}_{\text{rs}}(z) = (1 + \mathscr{C}_{\text{rs}}(z)\mathscr{P}_{\text{rs}}(z))^{-1}, \tag{19}$$

with $\mathscr{P}_{\text{rs}}(z)$ a non-parametric model of the reticle stage plant that can be obtained both quickly and accurately from closed-loop frequency response function measurements. From (1), (11), and (18) it follows that

$$\begin{aligned}\frac{\partial e(z)}{\partial a_0} &= \alpha\mathscr{S}_{\text{rs}}(z)\mathscr{P}_{\text{rs}}(z)\mathscr{C}_{\text{rs}}(z)y_{\text{ws}}(z) := \alpha\mathscr{S}^c_{\text{rs}}(z)y_{\text{ws}}(z) \\ \frac{\partial e(z)}{\partial b_0} &= \alpha\mathscr{S}_{\text{rs}}(z)\mathscr{P}_{\text{rs}}(z)y_{\text{ws}}(z) := \alpha\mathscr{S}^p_{\text{rs}}(z)y_{\text{ws}}(z),\end{aligned} \tag{20}$$

with $\mathscr{S}^c_{\text{rs}}$ the complementary sensitivity function of the reticle stage system and $\mathscr{S}^p_{\text{rs}}$ the corresponding process sensitivity function. In time-domain, the impulse responses corresponding to $\mathscr{S}^c_{\text{rs}}$ and $\mathscr{S}^p_{\text{rs}}$ can be used to construct the corresponding Toeplitz matrices $\mathbf{S}^c_{rs}$ and $\mathbf{S}^p_{rs}$ as to obtain the following lifted system representation [Bamieh et al., 1991]:

$$\frac{\partial \mathbf{e}}{\partial a_0} = \mathbf{S}^c_{rs}\mathbf{y}_{ws} \text{ and } \frac{\partial \mathbf{e}}{\partial b_0} = \mathbf{S}^p_{rs}\mathbf{y}_{ws}, \tag{21}$$

with $\mathbf{y}_{ws} = [y_{\text{ws}}(1) \ldots y_{\text{ws}}(k)]^{\mathrm{T}}$.

It should be remarked that stability and convergence of (15) strongly relates to: (a) the FIR coefficients being affine in the synchronized error signal e, and (b) plant $\mathscr{P}_{\text{rs}}$ being stabilized by controller $\mathscr{C}_{\text{rs}}$, both of which are linear time-invariant. In view of these properties, minimizing (12) renders the optimization problem convex which in combination with (15) yields global asymptotic convergence, see Heertjes et al. [2013].

To find the optimized set of FIR coefficients $\tilde{\mathbf{p}}$ in (14) but at the same time to constrain the feed-forward sensitivity function toward undesired amplifications, hence the stage synchronization problem from Section 2, we adopt the optimization algorithm as shown in Fig.4, which includes the following steps:

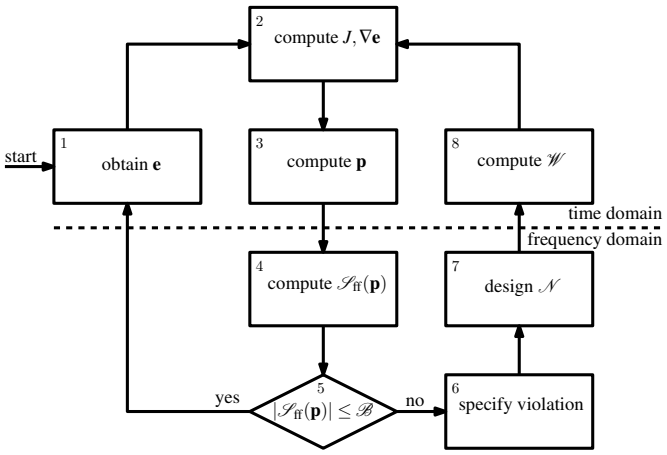

Fig. 4. Overview of the optimization algorithm.

(1) Obtain the synchronized error signals $\mathbf{e}$ and (re)set the optimization parameters/filters.
(2) Compute the (possibly weighted) cost function J with (12) and the gradient error signals with (16).
(3) Compute the candidate set of FIR coefficients in (13) with the Gauss-Newton update law in (15).
(4) Compute the feed-forward sensitivity function:

$$\mathscr{S}_{\text{ff}}(z) = \alpha\mathscr{S}^p_{\text{rs}}(z)(\mathscr{C}_{\text{rs}}(z)\mathscr{F}_1(z) + \mathscr{F}_2(z)) - 1, \tag{22}$$

using (11); (22) results from substitution of (18) in (1).
(5) Check if $|\mathscr{S}_{\text{ff}}(j\omega)|$ from (22) does not violate the user-defined magnitude characteristics $\mathscr{B}(j\omega)$. If no violation

occurs accept the candidate set from step (3) and return to step (1) for a possible next iteration $\kappa+1$. If violation occurs, reject the candidate set and proceed with step (6).
(6) Specify the frequency intervals where violation occurs.
(7) Design a set of notch filters $\mathscr{N}$ needed to account for the violation in each of the intervals found in step (6).
(8) Translate the notch filters $\mathscr{N}$ into a weighting filter $\mathscr{W}$ to penalize the (gradient) error signals in the direction of increased feed-forward sensitivity and return to step (2).

In the adaptive weighting filter design of steps (6)-(8) we distinguish three parts: violation specification, notch filter design, and weighting filter design. The former two are performed in frequency-domain, whereas the latter is done in time-domain.

Step (6). Violation specification starts by defining the amplitude constraints $\mathscr{B}$ on $\mathscr{S}_{\text{ff}}$. Let $\omega^s = \{\omega_0, \omega_0+\Delta\omega, \ldots, \omega_N\}$ with ω_0 the lowest frequency in the frequency response of $\mathscr{S}_{\text{ff}}$ and $\omega_N = \omega_0 + N\Delta\omega$ the Nyquist frequency. For the subset $\omega^b = \{\omega_0, \ldots, \omega_{N-i}, \omega_N\} \subset \omega^s$ define $\mathbf{b}$ as

$$\mathbf{b} = [b_0 \ \ldots \ b_{N-i} \ b_N], \tag{23}$$

where each frequency in ω^b associates with a gain $b_0, \ldots, b_{N-i}, b_N > 0$. Given (23) let $\mathscr{B} = \mathscr{B}(\omega)$ on the frequency interval $\omega = [\omega_0, \omega_N)$ be defined as

$$\mathscr{B}(\omega) = \sum_{i=0}^{k-1} \left(\mathbf{b}(\omega_i) + \frac{\omega_{i+1}-\omega}{\omega_{i+1}-\omega_i} \mathbf{b}(\omega_{i+1}) \right) \phi_i(\omega), \tag{24}$$

with $\mathscr{B}(\omega_N) = b_N$ and

$$\phi_i(\omega) = \begin{cases} 1, & \text{if } \omega_i \le \omega < \omega_{i+1} \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

For (22) and (24) we seek the subset $\omega^v \subset \omega^s$ (see step (5) in Fig.4) that satisfies

$$|\mathscr{S}_{\text{ff}}(j\omega_i^v)| > \mathscr{B}(\omega_i^v), \tag{26}$$

with ω_i^v indicating an arbitrary element taken from the set ω^v. If $\omega^v \neq \emptyset$, i.e. the bounds on the feed-forward sensitivity function are being violated, then $\omega^v = \{\omega^{v,1}, \ldots, \omega^{v,p}\}$ with $p \ge 1$ and each subset $\omega^{v,i} = \{\omega_1^{v,i}, \ldots, \omega_1^{v,i} + m_i\Delta\omega\} \subset \omega^v$ containing a sequence of consecutive frequencies with $m_i \ge 1$ and $i \in \{1, \ldots, p\}$. Note that each subset $\omega^{v,i}$ thus defines a closed frequency interval of violation.

Step (7). The notch filter design aims at finding the series connection of notch filters $\mathscr{N}_k = \mathscr{N}_1\mathscr{N}_2\ldots\mathscr{N}_p$ where each notch filter $\mathscr{N}_i$ relates to the violation of the amplitude constraints in (24) corresponding to subset $\omega^{v,i}$. For this purpose, define the center frequency $\omega_{c,i}$ and the magnitude of violation γ_i as

$$\begin{aligned} \omega_{c,i} &= \omega_1^{v,i} + \frac{m_i\Delta\omega}{2} \\ \gamma_i &= \max\{\varphi(\omega_1^{v,i}), \ldots, \varphi(\omega_1^{v,i} + m_i\Delta\omega)\}, \end{aligned} \tag{27}$$

with

$$\varphi(\omega) = \begin{cases} \dfrac{\mathscr{B}(\omega)}{\left|\dfrac{\mathscr{S}_{\text{ff}}(j\omega)}{\mathscr{S}_{\text{ff}}(j\omega)}\right|}, & \text{if } |\mathscr{S}_{\text{ff}}(j\omega)| \ge 1 \\ \dfrac{\mathscr{S}_{\text{ff}}(j\omega)}{\mathscr{B}(\omega)}, & \text{otherwise.} \end{cases} \tag{28}$$

The set of notch filters is then given by

$$\mathscr{N}_k(j\omega) = \prod_{i=1}^{p} \frac{-\omega^2 + 2j\zeta_1^i\omega_{c,i}\omega + \omega_{c,i}^2}{-\omega^2 + 2j\zeta_2^i\omega_{c,i}\omega + \omega_{c,i}^2}, \tag{29}$$

with damping coefficients ζ_1^i, ζ_2^i that are related via

$$\zeta_2^i = \gamma_i\zeta_1^i. \tag{30}$$

With (30), the height of the notch filter is determined by the maximum amount of violation that is encountered in each subset $\omega^{v,i}$. The values for the damping coefficients ζ_1^i and ζ_2^i are subject to manual tuning. In this paper, we scale these values to the different sizes of the p subsets in ω^v, i.e. the frequency intervals of violation. Let $\Delta\omega^{v,i} = m_i\Delta\omega$, then a typical scaling $\eta_i \in \eta = \{\eta_1, \ldots, \eta_p\}$ is found by:

$$\eta_i = \frac{\Delta\omega^{v,i}}{\min\{\Delta\omega^{v,1}, \ldots, \Delta\omega^{v,p}\}} = \frac{m_i}{\min\{m_1, \ldots, m_p\}}, \tag{31}$$

This scaling is used to set $\zeta_1^i = 0.01\eta_i$, which in combination with (30) gives $\zeta_2^i = 0.01\gamma_i\eta_i$.

Step (8). The weighting filter design aims at penalizing the (gradient) error signals in the direction of increased feed-forward sensitivity. Given the notch filters in (29), an adaptive weighting filter $\mathscr{W}$ is constructed:

$$\mathscr{W}_{k+1}(z) = \mathscr{W}_k(z)\mathscr{N}_k(z), \ \mathscr{W}_1(z) = 1, \ k \in \{1, \ldots, q\}, \tag{32}$$

with k the number of sub-iterations required for the feed-forward sensitivity function in (22) to satisfy the constraints $\mathscr{B}$ in (24) and q the maximum number of iterations. For each iteration k, a weighted error signal $\tilde{\mathbf{e}}$ is derived via

$$\tilde{\mathbf{e}} = \mathbf{W}\mathbf{e}, \tag{33}$$

with $\mathbf{W}$ a Toeplitz matrix of appropriate dimensions and derived from $\mathscr{W}$ in (32). By using the weighted error signal $\tilde{\mathbf{e}}$ from (33) instead of $\mathbf{e}$ in (12), (15), and (16), we effectively penalize violation of the frequency-domain constraints $\mathscr{B}$ in (24) in the choice of the candidate set of FIR coefficients.

It should be remarked that global convexity properties in the constrained optimization problem underlying Fig.4 are generally lost such that global stability and convergence properties are no longer guaranteed. A poor choice of $\mathscr{B}$, e.g. one that conflicts with the waterbed effect in (10), may lead to no convergence whatsoever. Furthermore, upon acceptance of a candidate set, the (sub-)iteration parameter k of the adaptive weighting filter design will be reset to $k = 1$ this to ensure that $\mathscr{W}_1 = 1$ at the next iteration of the optimization scheme $\kappa+1$, i.e. each iteration κ starts under unconstrained conditions.

4. WAFER SCANNING RESULTS

In this section we will discuss the results obtained from several experiments taken from an industrial wafer scanner. The discussion is started with the derivation of non-parametric models for the controlled reticle stage. These models are used to derive the gradient error signals in (16) and (20), which form the basis of the Gauss-Newton optimization scheme in (15).

To model the plant $\mathscr{P}_{\text{rs}}$ consider the Bode diagrams in Fig.5. The figure shows frequency response functions (FRFs) either reflecting measured plant data or a 12-th order (non-parametric) model fit obtained from these data; note that the measured plant data are not obtained from direct measurement but from reconstruction using closed-loop measurement data. Between 40 and 1 kHz both FRFs show double integrator behavior apart from a time delay, see (7). Below 40 Hz, deviation from this behavior is the result of poor closed-loop identification. Beyond 1kHz, resonances determine the frequency response. The 12th-order model is expected to be sufficiently accurate to derive the gradient error signals in (20). To derive the gradient error signals, we also need the characteristics of the feedback controller $\mathscr{C}_{\text{rs}}$, which in principle are known.

Fig. 6 shows the Bode diagram of the 13-th-order PID-based controller $\mathscr{C}_{\text{rs}}$, which was designed by manual loop shaping. $\mathscr{C}_{\text{rs}}$ includes five notch filters to deal with the plant resonances and a second-order low-pass filter to limit high-frequency amplification. The controller in Fig.6 with the plants in Fig.5 typically

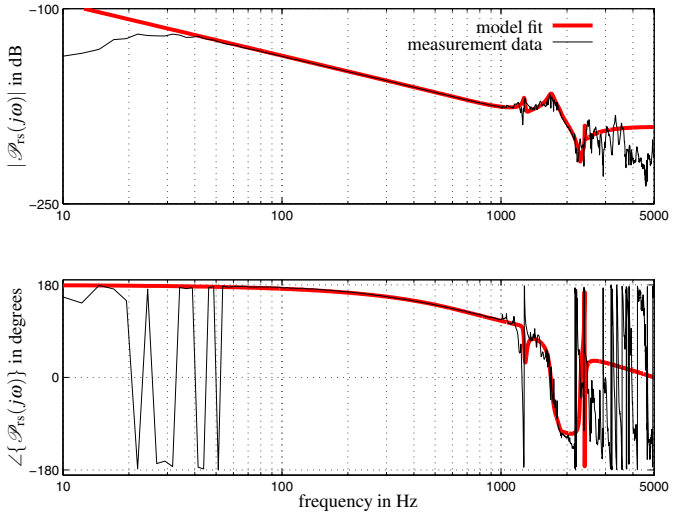

Fig. 5. Bode diagrams of the frequency response functions of $\mathscr{P}_{\rm rs}$ either by measurement data (black) or by a 12-th order (non-parametric) model fit (red).

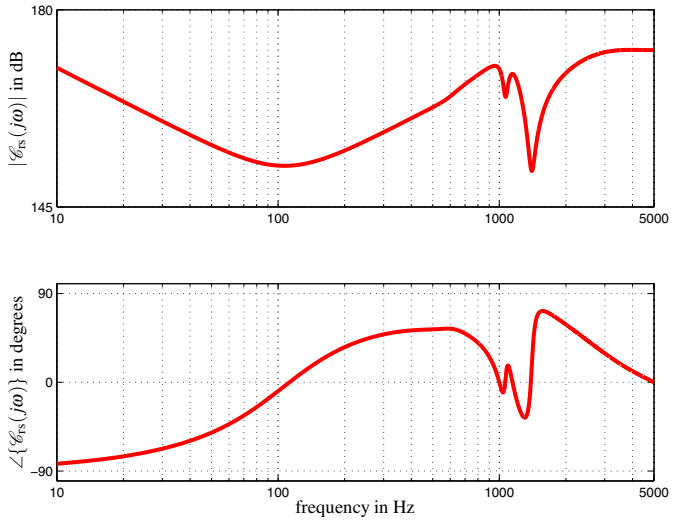

Fig. 6. Bode diagram of the frequency response function of the reticle stage controller $\mathscr{C}_{\rm rs}$.

induce a controller bandwidth of 412 Hz with a phase margin of 20.5 degrees and a gain margin of 5.8 dB.

With the optimization algorithm as discussed in Fig.4 and the reticle stage models for plant $\mathscr{P}_{\rm rs}$ in Fig.5 and controller $\mathscr{C}_{\rm rs}$ in Fig.6, a series of optimization experiments have been conducted. In terms of the feed-forward sensitivity function $\mathscr{S}_{\rm ff}$, recall (9) and (22), Fig.7 shows the effect of three different experiments. In the first experiment, see the top part of Fig.7, the case is considered where no frequency-domain constraints are used in the optimization algorithm, i.e. steps (5)-(8) of Fig.4 are skipped. For this case, the optimal set of FIR coefficients in $\mathscr{F}_1$ and $\mathscr{F}_2$ in (11) is basically found in one iteration. The resulting feed-forward sensitivity function $\mathscr{S}_{\rm ff} = \mathscr{S}_{\rm ff}(\mathscr{F}_1, \mathscr{F}_2)$ shows that a synchronization controller bandwidth is obtained of 588.2 Hz. Also the resemblance is shown with $\mathscr{S}_{\rm ff}$ from (9) which is merely the result of time delay, recall the discussion in Section 2. At the high-frequencies, $\mathscr{S}_{\rm ff}$ from (22) shows a significant amplification according to the waterbed effect in (10). For example, almost 17 dB amplification is induced at 1.5 kHz. To limit these amplifications two bounds $\mathscr{B}_1$ and

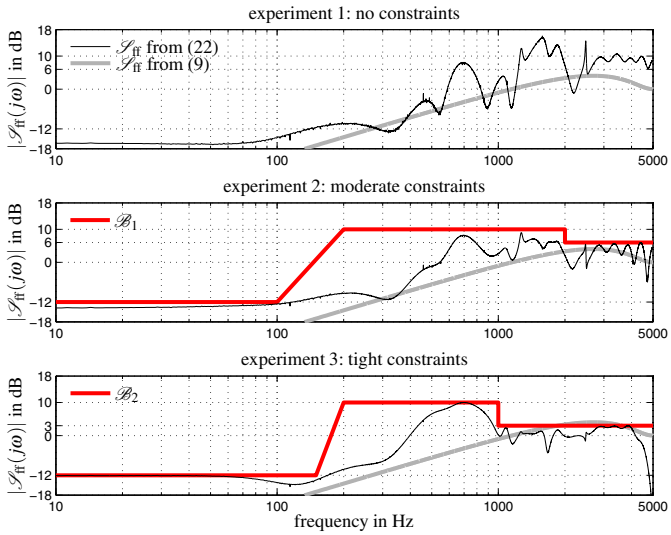

Fig. 7. Bode magnitude plots of the reconstructed feed-forward sensitivity functions $\mathscr{S}_{\rm ff}(j\omega)$ for different experiments: no constraints, moderate constraints, and tight constraints.

$\mathscr{B}_2$ are introduced according to (24); the solid red curves in respectively the center and bottom parts of the figure. Bound $\mathscr{B}_1$ is a moderate constraint that is mainly designed to limit the high-frequency amplifications. In doing so, also a low-frequency constraint of -12 dB is used as to avoid significant low-frequency deterioration. It is clear from the center part of the figure that the frequency-domain constraints are nicely met whereas performance is hardly compromised; the synchronization controller bandwidth drops from 588.2 Hz to 543.7 Hz. For the more tight bound $\mathscr{B}_2$ in the bottom part of the figure, it can be seen that the synchronization controller bandwidth is significantly reduced, however, to 397.7 Hz; both constrained optimizations were done with $\kappa \leq 5$ and $k \leq 10$.

In time-domain, the effect of a reduced synchronization bandwidth is shown in Fig.8 in terms of filtered error responses. In

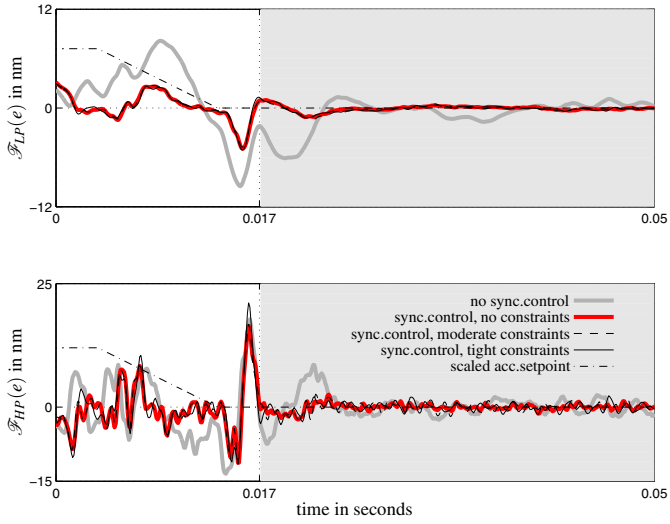

Fig. 8. Measured synchronized error responses for different experiments: without synchronization controller or with synchronization controller and no constraints, moderate constraints, and tight constraints.

the top part of the figure, a first-order low-pass filter operation

(which has resemblance to a moving average filter operation as used in the wafer scanning industry) is used according to:

$$\mathscr{F}_{LP}(s) = \frac{\omega_{LP}}{s+\omega_{LP}}, \tag{34}$$

with $\omega_{LP} = 2\pi \cdot 70$ rad $\cdot$ s^{-1} the cross-over frequency. In the bottom part of the figure a first-order high-pass filter operation is used which is defined as $\mathscr{F}_{HP}(s) = 1 - \mathscr{F}_{LP}(s)$; this filter operation has some of the features encountered with a commonly-used moving standard deviation filter operation [Butler, 2011]. For a representative point-to-point motion, see the scaled acceleration set-point, the experimental data of four experiments are shown. By taking the mean of four repeats, these data give a fair representation of the recurring parts of the synchronized error signals for the given set-point. The four experiments involve: (a) no synchronization controller, i.e. $\mathscr{F}_1 = \mathscr{F}_2 = 0$, (b) an optimized synchronization controller, but without frequency-domain constraints, (c) an optimized synchronization controller under moderate constraints $\mathscr{B}_1$, and (d) under tight constraints $\mathscr{B}_2$. In the top part of Fig.8, it can be seen that all optimized sets perform significantly better than the case with no synchronization controller. In the bottom part of Fig.8, it is shown that higher bandwidths induce better disturbance rejection properties in the region $t \in [0.017\ 0.05]$ seconds, i.e. the indicated region of constant velocity where wafer scanning takes place.

In frequency-domain, Fig.9 in terms of cumulative power spectral density (CPSD) analysis confirms the correlation between improved disturbance rejection properties and increased bandwidths. The RMS values for e of 2.08 nm, 2.10 nm, 2.33 nm,

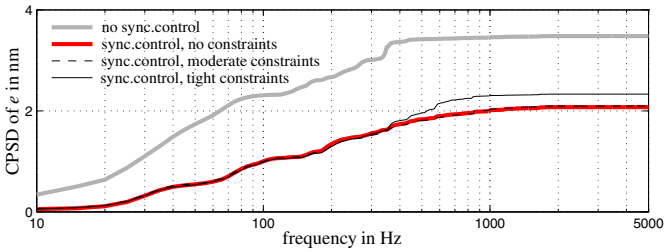

Fig. 9. Measured cumulative power spectral densities of the synchronized error responses for different experiments: without synchronization controller or with synchronization controller and no constraints, moderate constraints, and tight constraints.

and 3.48 nm are obtained with synchronization controller bandwidths of respectively 588.2 Hz, 543.7 Hz, 397.7 Hz, and 0 Hz.

5. CONCLUSIONS

In this paper, data-driven control is applied to the stage synchronization problem. By imposing frequency-domain robustness constraints on the feed-forward sensitivity function, it is demonstrated that synchronization performance can roughly be maintained while meeting frequency-domain specifications. Too tight constraints, however, induce a significant reduction of the synchronization controller bandwidth in view of the waterbed effect and thus compromise performance. Introducing frequency-domain constraints clearly renders the constrained optimization problem, which includes both time-domain and frequency-domain aspects, globally non-convex. As a result, global stability and convergence properties corresponding to the unconstrained (and convex) optimization problem are generally lost. The adaptive weighting filter design as a means to impose frequency-domain constraints provides a heuristic approach with similarities to manual loop shaping. For specific frequency contributions that cause violation of the constraints fairly good results have been demonstrated with this approach. For broad-band violations one may want to reside to more rigorous approaches like the one considered in the companion paper Van der Velden et al. [2014] where the constrained optimization problem is transformed into an unconstrained optimization problem but at the cost of convexity properties.

REFERENCES

B. Bamieh, P.J. Boyd, B.A. Francis, and A. Tannenbaum. A lifting technique for linear periodic systems with applications to sampled-data control. *Systems and Control Letters*, 17:79-88, 1991.

A.S. Bazanella, L. Campestrini, and D. Eckhard. Data-driven controller design - the $\mathscr{H}_2$ approach. *Springer*, 2012.

P.J. van Bree, C.M.M. van Lierop, and P.P.J. van den Bosch. On hysteresis in magnetic lenses of electron microscopes. *In Proc. Industrial Electronics*, Bari, italy, 268–273, 2010.

H. Butler. Position control in lithographic equipment; an enabler for current-day chip manufacturing. *IEEE Control Systems Magazine*, 11: 28-47, 2011.

M.F. Heertjes, D. Hennekens, and M. Steinbuch. MIMO feed-forward design in wafer scanners using a gradient approximation-based algorithm. *Control Engineering Practice*, 15(5): 495–506, 2010.

M.F. Heertjes, B. Temizer, and M. Schneiders. Self-Tuning in Master-Slave Synchronization of High-Precision Stage Systems. *Control Engineering Practice*, 21:1706-1715, 2013.

H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41:393-438, 2005.

M.K. Jouaneh, and E. Anderson. Input shaping using finite impulse response filters. *In Proceedings of the Conference on Decision and Control*, San Diego, California, USA:6525-6530, 2006.

D. Kostić, A.G. de Jager, and M. Steinbuch. Motion control by linear feedback methods. *In Robotics and Automation Handbook, editor: T.R. Kurfess*, 15:1–23, 2004.

S. Mishra, W. Yeh, and M. Tomizuka. Iterative learning control design for synchronization of wafer and reticle stages. *In Proceedings of the American Control Conference*, Seattle, WA, USA, 3908:3913, 2008.

M.S. Park, Y. Chait, and M. Steinbuch. Inversion-free design algorithms for multivariable quantitative feedback theory: An application to robust control of a CD-ROM. *Automatica*, 33(5): 915–920, 1997.

H. Prochazka, M. Gevers, B.D.O. Anderson, C. Ferrera. Iterative feedback tuning for robust controller design and optimization. *In Proceedings of the Conference on Decision and Control, and the European Control Conference*, Seville, Spain, 3602-3607, 2005.

B. Van der Velden, T. Oomen, and M.F. Heertjes. Constrained Iterative Feedback Tuning for Robust High-Precision Motion Control. *IFAC World Congress*, Cape Town, South Africa, 1–6, 2014.

S. Veres, and H. Hjalmarsson. Tuning for robustness and performance using iterative feedback tuning. *In Proceedings of the Conference on Decision and Control*, Las Vegas, Nevada, 4682-4687,2002.

C. Wang, K. Hu, Y. Zhu, and W. Yi. Optimal synchronous trajectory tracking control of wafer and reticle stages. *Tsinghua Science and Technology*, 14(3): 287-292, 2009.

Y. Xiao Y, and K.Y. Zhu. Optimal synchronization control of high-precision motion systems. *IEEE Transactions on Industrial Electronics*, 53(4): 1160-1169, 2006.