Constrained Iterative Feedback Tuning for Robust Control of a Wafer Stage System

Marcel F. Heertjes, Bart Van der Velden, Tom Oomen

Abstract— **Iterative feedback tuning (IFT) enables the data-driven tuning of controller parameters without the explicit need for a parametric model. It is known, however, that IFT can lead to non-robust solutions. The aim of this paper is to develop an iterative feedback tuning approach with robustness constraints. A constrained IFT problem is formulated that is solved by introducing a penalty function. Essentially, the gradient estimates decompose into (a) the well-known IFT gradients and (b) the gradients with respect to this penalty function. The latter are obtained through a non-parametric model of the controlled system. This guarantees robust stability while only requiring a non-parametric model. Experimental results obtained from the motion control systems of an industrial wafer scanner confirm enhanced performance with guaranteed robustness estimates.**

Index Terms— **data-driven control, high-precision wafer stages, iterative feedback tuning (IFT), robust control.**

I. Introduction

Control systems are often designed using either model-based approaches or data-based approaches. Model-based approaches refer to the system first being modeled, for example through system identification. Modeling, however, can be a time-consuming and difficult process [19], [33]. In this regard, direct data-driven approaches, see, e.g. [2], for an overview, are appealing since they avoid the need for modeling and enable the direct tuning of the controller based on measurement data. This has led to a variety of data-based control design methods including unfalsified control [39], virtual reference feedback tuning [7], and iterative feedback tuning (IFT) [15], [18]. Interestingly, in [19] it is shown that many of the data-based approaches can be interpreted as being model-based, except for IFT. IFT is an iterative optimization approach that aims at obtaining unbiased gradient estimates by conducting multiple experiments. In a gradient-based update scheme, the gradient estimates are used to compute a new set of controller parameters. The optimal set of parameters corresponds to a minimum of a performance-relevant cost function in time domain. IFT has been successfully applied in many applications including process industry [9], [29], robotics [10], [23], [37], mechatronics [1], [28], and stage control [46]. Most of these works involve linear (time-invariant) systems. For nonlinear systems, IFT has been addressed, for example, in [11], [40].

An important field of application where data-based controller tuning is appealing is in high-precision motion control. For feedforward control design, model-based approaches have led to limited performance enhancements due to model uncertainty [4]. This has led to the development of data-based approaches in [30], and [13] for the multivariable case, which induce significant performance enhancements. It has recently been shown in [3] that the IFT algorithm employed in [30] for feedforward control has a direct system identification interpretation. In fact, for the case of feedforward control, the IFT algorithm deals with a closed-loop identification problem, but comes at the expense of efficiency. In [3], more efficient algorithms have been proposed rendering the use of IFT superfluous for feedforward design. Still, IFT seems appealing for feedback design. This is because (a) IFT requires no parametric system model, (b) unlike many model-based synthesis algorithms IFT can deal with a predefined controller structure, and (c) control performance is optimized for the disturbance situation at hand so no disturbance model is needed.

Despite these clear advantages, a direct IFT implementation suffers from robustness issues as there are generally no guarantees that the closed-loop system remains stable during the iterations. This is evidenced by the development of IFT algorithms that take into account robust stability, see for example [5], [34], [43]. These algorithms typically use approximations of robustness measures, including the ν-gap metric [44] and $\mathcal{H}_\infty$ norms of relevant closed-loop transfer functions, which may be rather conservative. Moreover, strong guarantees for robust stability are not provided.

The main contribution of this paper is a new approach for constrained IFT. To ensure robustness, a frequency-domain constraint is added to the objective function that represents the robustness objective. For the considered class of motion systems, non-parametric frequency response function models are relatively inexpensive, fast, and accurate to obtain. Although not directly usable for controller synthesis, since synthesis generally requires a parametric model, these models are well-suited to evaluate robustness margins such as the modulus margin. Preliminary results of this new approach are given in [42]. The present paper extends these results by thoroughly demonstrating the effectiveness of the approach on the wafer

This research is financially supported by the Innovational Research Incentives Scheme under the VENI grant "Precision Motion: Beyond the Nanometer" (no. 13073) awarded by NWO (The Netherlands Organisation for Scientific Research) and STW (Dutch Science Foundation).

M.F. Heertjes is with Eindhoven University of Technology, Department of Mechanical Engineering, Den Dolech 2, 5600 MB Eindhoven, The Netherlands, email: m.f.heertjes@tue.nl. and with ASML Netherlands B.V., Mechatronics System Development, De Run 6501, 5504 DR Veldhoven, The Netherlands, email: marcel.heertjes@asml.com

B.J.C.H. Van der Velden is with TMC, Eindhoven, The Netherlands, email: bart.van.der.velden@asml.com (during the main execution of this work Van der Velden was with Eindhoven University of Technology and ASML).

T.A.E. Oomen is with Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, Den Dolech 2, 5600 MB Eindhoven, The Netherlands, email: t.e.a.oomen@tue.nl.

stage sub-system of a wafer scanner, i.e., a lithography machine. Apart from fine-tuning of the PID controller parameters, this also involves fine-tuning of the notch filter parameters that in [42] are fixed a priori by manual loop-shaping.

The central idea in this paper is to include a constraint, e.g., the modulus margin, in the optimization criterion. By means of a penalty function, the gradient-based IFT scheme can be decomposed into (a) the well-known IFT gradients and (b) the gradients with respect to this penalty function. This is related to the work of [12] where an interior-point method is adopted to take into account signal constraints on the input. In our approach, the numerical differentiation is done off-line based on the identified non-parametric model. During off-line optimization, a modified optimal set of controller parameters is obtained that satisfies the frequency-domain robustness specifications. With this set, a new IFT experiment is performed. Given the accuracy of the non-parametric model, experimental results demonstrate that no experiments are done that involve (significant) violations of the frequency domain robustness specifications, i.e., no controllers are implemented that can possibly lead to unstable control design. This is important for high-precision motion systems like wafer scanners where machine damage should be avoided due to the high costs.

This paper is organized as follows. In Section II, the wafer scanner system and the wafer stage sub-system are discussed. In Section III, a short review of IFT is given in the wafer stage context. In Section IV, the penalty function approach is introduced as a means to impose frequency-domain specifications and the extended optimization algorithm underlying this approach is presented. In Section V, measurement results obtained from the considered wafer stage system are presented. In Section VI, the main conclusions are given.

II. Wafer Scanners and Wafer Stage Control

Wafer scanners are lithography machines used in the process of producing integrated circuits; see [27] for the principles of lithography. In this section, the control tuning problem of wafer scanners in general and wafer stages in specific is introduced; see also [6] for an historical perspective and [33] for an overview of present control challenges. The section is divided into five parts: (a) wafer scanners, (b) motion control context, (c) wafer stage plant description, (d) wafer stage feedback control, and (e) performance of wafer scanners and stages. This completely outlines the setting in which the iterative feedback tuning algorithm is implemented.

A. Wafer Scanners-Stages

A schematic representation of a wafer scanner is depicted in Fig. 1. Extreme ultraviolet light travels from a source (located outside the machine) through projection optics to the light sensitive layers of a wafer (a silicon disk of 300 mm radius). The light contains an image of the integrated circuits to be processed. The image is obtained from the reticle which is part of the reticle stage motion control system. The wafer is part of the wafer stage motion control system. During wafer scanning, i.e., the dose-controlled wafer exposure by an extreme ultraviolet light beam, both the reticle and the wafer stage systems track a series of point-to-point motions in (scanning) y-direction, see also [24]. In this paper, the focus is on the wafer stage system. Due to the close similarity of the two systems, the presented approach directly applies to the reticle stage system.

The wafer stage system consists of two modules: the long-stroke module and the short-stroke module. The positioning accuracy of the long-stroke module is at micrometer level. In scanning an entire wafer, the long-stroke module, which weighs ≈ 67 kg, provides support for the short-stroke module along the different fields of the wafer; typically 100 fields of approximately 30×30 mm need to be exposed sequentially. The positioning accuracy of the short-stroke module, which weighs ≈ 18 kg and which supports the wafer, is at nanometer level. All fields on a wafer are scanned in y-direction by conducting a series of point-to-point motions. Each point-to-point motion has an interval – the scanning interval – of constant velocity. Apart from the scanning interval, the point-to-point motion profiles include acceleration/deceleration intervals in x- and y-direction as to enable full wafer coverage. For the purpose of positioning, the wafer stage actuator systems use voice coil motors whereas the main sensor systems are based on interferometry. The wafer stage modules are controlled in six logical axes: x, y, r_z, z, r_x, and r_y, using a diagonal controller structure in combination with an actuator/sensor decoupling design. In this paper, we only consider the (scanning) y-direction.

B. Motion Control Context

The motion control context of a simplified wafer stage system is depicted in Fig. 2. The linear time-invariant wafer

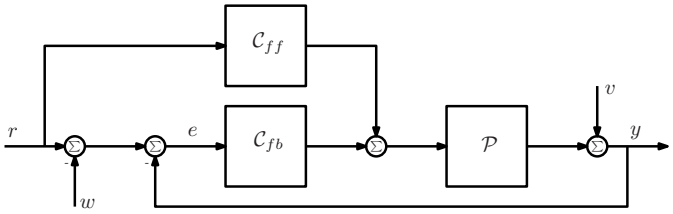

Fig. 2. Block diagram of the simplified wafer stage motion control context.

stage plant $\mathcal{P}$ has output y which is corrupted by the unmeasured disturbances v. These disturbances are assumed to have a stochastic nature and are not correlated with the other inputs r and w. Plant $\mathcal{P}$ is assumed to be single-input single-output and controlled by a linear time-invariant feedforward controller $\mathcal{C}_{ff}$ and feedback controller $\mathcal{C}_{fb}$. Controller $\mathcal{C}_{fb}$ is a function of the parameters $\boldsymbol{\rho} = [\boldsymbol{\rho}(1) \dots \boldsymbol{\rho}(m)]^{\mathrm{T}}$ to be optimized, i.e., $\mathcal{C}_{fb} = \mathcal{C}_{fb}(\boldsymbol{\rho})$, with the number of parameters $m \in \{1, 2, \dots\}$. In conducting point-to-point motion, tracking performance is reflected by the closed-loop error $e = r - w - y$, i.e., the difference between the reference command r, the output y, and an auxiliary input w used for the iterative feedback tuning (IFT) experiments. This auxiliary input is explained in more detail in Section III.

Remark 1 *Though $\mathcal{C}_{fb}$ and $\mathcal{C}_{ff}$ can be optimized jointly, in this paper the feedforward controller $\mathcal{C}_{ff}$ will not be subject*

Fig. 1. Schematics and artist impression of a wafer scanner.

to (constrained) IFT optimization. Basically, $\mathcal{C}_{ff}$ poses no (robust) stability problem. Besides, for feedforward parameter tuning, which often refers to finding an acceleration term, a snap term, and a delay compensation term, more efficient algorithms are available, see for example [3], [30].

C. Wafer Stage Plant Description

In the Bode diagram representation of Fig. 3, it can be seen by frequency response measurements that the short-stroke wafer stage plant description $\mathcal{P}$ essentially boils down to a double-integrator based system with a mass of approximately 17.7 kg. This follows from the -40 dB/dec magnitude slope in the frequency interval between 40 and 400 Hz combined with a -180 degrees phase lag. At low frequencies, a poor signal-to-noise ratio leads to a poor identification result. The main reason is that the closed-loop system identification procedure, as is described in [33, Appendix A], is followed. Essentially, the frequency response function of the open-loop system is obtained by dividing the measured closed-loop process sensitivity data by closed-loop sensitivity data. Since the sensitivity function is typically small at low frequencies due to a high controller gain, the signal-to-noise ratio for an arbitrary and non-optimized experiment design is poor. At high frequencies, several resonances appear that characterize the structural flexibilities of the stage system. For example, around 1 kHz a torsion mode is responsible for the observed behavior. In the frequency response measurements, a sampling time of $T = 10^{-4}$ seconds is used.

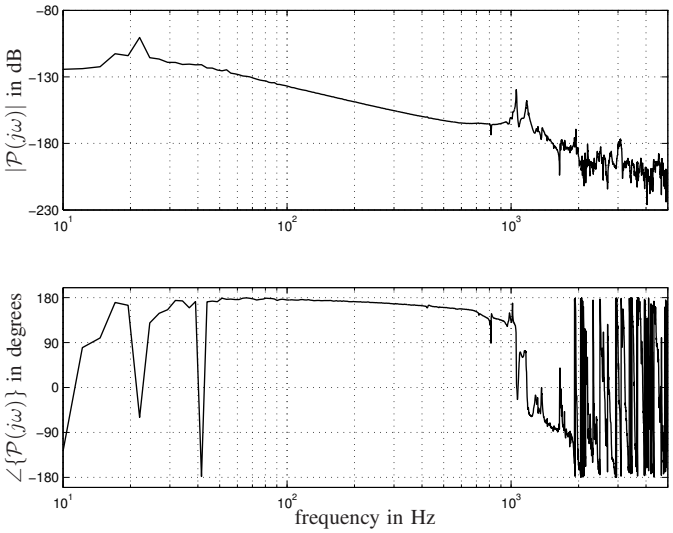

Fig. 3. Bode diagram of the short-stroke wafer stage plant $\mathcal{P}$ in y-direction with measured data.

Remark 2 *The wafer stage system is a multivariable system that through a proper choice of input/output decoupling matrices becomes diagonally dominant. Multi-input multi-output (MIMO) stage control is therefore often done with a diagonal controller structure in the six degrees-of-freedom: x, y, r_z, z, r_x, r_y. For clarity of presentation, only the scanning y-direction is subject to further (constrained) IFT optimization, see [16], [17], [35] for examples of MIMO IFT and [33], [45] for MIMO model-based counterparts.*

D. Wafer Stage Feedback Control

The short-stroke wafer stage plant $\mathcal{P}$ in y-direction is controlled by the feedback controller $\mathcal{C}_{fb}$ which (in continuous-time) has the following structure:

$$\mathcal{C}_{fb}(s) = \mathcal{C}_{PID}(s)\mathcal{C}_{LP}(s)\mathcal{C}_N(s), \tag{1}$$

with s the Laplace variable. The PID-controller part $\mathcal{C}_{PID}$, the low-pass part $\mathcal{C}_{LP}$, and the notch part $\mathcal{C}_N$ are given by

$$\mathcal{C}_{PID}(s) = k_p\left(\frac{s}{\omega_d} + 1 + \frac{\omega_i}{s}\right),$$
$$\mathcal{C}_{LP}(s) = \frac{\omega_{lp}^2}{s^2 + 2\zeta_{lp}\omega_{lp}s + \omega_{lp}^2}, \text{ and } \mathcal{C}_N(s) = \prod_{i=1}^{n}\mathcal{N}_i(s), \tag{2}$$

with the notch filters

$$\mathcal{N}_i(s) = \left(\frac{\omega_{p,i}}{\omega_{z,i}}\right) \frac{s^2 + 2\zeta_{z,i}\omega_{z,i}s + \omega_{z,i}^2}{s^2 + 2\zeta_{p,i}\omega_{p,i}s + \omega_{p,i}^2}. \tag{3}$$

Five notch filters $\mathcal{N}_1, \ldots, \mathcal{N}_5$ are used (a) to deal with the plant resonances and (b) to properly shape the open-loop characteristics as to achieve high bandwidth with sufficient robustness margins. The parameters of $\mathcal{C}_{PID}$, $\mathcal{C}_{LP}$, and $\mathcal{C}_N$ that correspond to the default controller and that are used as starting point for (constrained) IFT are given in Tabel I.

$\mathcal{C}_{PID}$	k_p in Nm^{-1}	ω_i in $\mathrm{rad \cdot s}^{-1}$	ω_d in $\mathrm{rad \cdot s}^{-1}$	
	$1.9 \cdot 10^7$	$2\pi \cdot 80$	$2\pi \cdot 130$	
$\mathcal{C}_{LP}$	ω_{lp} in $\mathrm{rad \cdot s}^{-1}$	ζ_{lp}		
	$2\pi \cdot 1400$	0.80		
$\mathcal{N}_i$	$\omega_{z,i}$ in $\mathrm{rad \cdot s}^{-1}$	$\zeta_{z,i}$	$\omega_{p,i}$ in $\mathrm{rad \cdot s}^{-1}$	$\zeta_{p,i}$
1	$2\pi \cdot 430$	0.01	$2\pi \cdot 430$	0.03
2	$2\pi \cdot 718$	0.01	$2\pi \cdot 718$	0.05
3	$2\pi \cdot 812$	0.01	$2\pi \cdot 812$	0.03
4	$2\pi \cdot 1100$	0.12	$2\pi \cdot 900$	0.20
5	$2\pi \cdot 1950$	0.01	$2\pi \cdot 1850$	0.05

TABLE I
FIXED PARAMETERS OF THE WAFER STAGE FEEDBACK CONTROLLER $\mathcal{C}_{fb}$.

Remark 3 *The controller $\mathcal{C}_{fb}$ is implemented in discrete-time using the δ-operator, see Middleton & Goodwin [31], where $\delta = (z-1)/T$, with z the forward shift operator. The reason is because of the finite (internal) floating point resolution that otherwise induces error in the computation of the poles and the zeros. This especially holds true at low frequencies where these poles and zeros generally lie closer to one than to zero.*

E. Wafer Scanner-Stage Performance

Performance of wafer scanners is basically given by three measures: (a) overlay, (b) imaging, and (c) throughput. Overlay refers to the ability to expose one layer of the wafer atop another; in the process of making chips, typically 20-30 subsequent layers are being exposed. Imaging generally refers to image quality in terms of line-width uniformity. Throughput refers to the number of wafers that are being exposed per hour.

Wafer stage performance is evaluated in time-domain and relates to the servo error signals that are (partly) responsible for overlay, imaging, and throughput. In terms of servo errors, throughput is related to settling times prior to scanning. Overlay and imaging mainly refer to the scanning interval and are related to respectively a moving average ($\mathcal{M}_A$) and a moving standard deviation ($\mathcal{M}_{SD}$) filter operation applied (off-line) to the servo error signals. In continuous-time, these operations are given by:

$$\begin{aligned} \mathcal{M}_A(t) &= \frac{1}{T_e}\int_{t-T_e/2}^{t+T_e/2} e(\tau)d\tau \\ \mathcal{M}_{SD}(t) &= \sqrt{\frac{1}{T_e}\int_{t-T_e/2}^{t+T_e/2} (e(\tau) - \mathcal{M}_A(\tau))^2 d\tau}, \end{aligned} \tag{4}$$

with T_e the exposure time-constant defined by $T_e = s/v$, where s represents the slit height and v the scan velocity. The $\mathcal{M}_A$ filter has low-pass properties whereas the $\mathcal{M}_{SD}$ filter has high-pass properties. Both filter operations are non-causal and therefore can only be applied off-line.

Having set the stage context, iterative feedback tuning (IFT) is considered as a means to fine-tune the controller parameters $\boldsymbol{\rho}$ in $\mathcal{C}_{fb} = \mathcal{C}_{fb}(\boldsymbol{\rho})$, see (1).

III. Iterative Feedback Tuning: The Stage Context

In discussing IFT in the stage context, this section contains two parts: (a) an IFT overview, and (b) IFT experiments and the resulting unbiased gradient estimates.

A. IFT Overview

IFT [15], [16], [17] can be used to find the optimal set of controller parameters $\boldsymbol{\rho}^*$ that minimizes the cost function:

$$J = \frac{1}{2N}\mathbf{e}^T\mathbf{e}, \tag{5}$$

with the data sampled signal $\mathbf{e} = [e(1) \ldots e(N)]^{\mathrm{T}}$, $\mathbf{e} = \mathbf{e}(\boldsymbol{\rho})$, and the number of samples $N > 0$. Equation (5) gives rise to the following unconstrained optimization problem:

$$\boldsymbol{\rho}^* = \min_{\boldsymbol{\rho}} J, \tag{6}$$

which can be solved iteratively using the Gauss-Newton algorithm:

$$\boldsymbol{\rho}_{i+1} = \boldsymbol{\rho}_i - \gamma_i H^{-1} \left.\frac{\partial J}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_i}, \tag{7}$$

where $0 < \gamma_i \leq 1$ is the step size at iteration i, the gradient $\partial J/\partial \boldsymbol{\rho}$ is given by

$$\frac{\partial J}{\partial \boldsymbol{\rho}} = \frac{1}{N}\frac{\partial \mathbf{e}^T}{\partial \boldsymbol{\rho}}\mathbf{e}, \tag{8}$$

and an approximation of the Hessian $H = H(\boldsymbol{\rho})$ is given by

$$H = \frac{1}{N}\frac{\partial \mathbf{e}^T}{\partial \boldsymbol{\rho}}\frac{\partial \mathbf{e}}{\partial \boldsymbol{\rho}}. \tag{9}$$

By seeking the minimum of (5) using a gradient-based search, from (8) and (9) it is clear that finding $\boldsymbol{\rho}^*$ in (6) with (7) requires the gradients $\partial \mathbf{e}/\partial \boldsymbol{\rho}$.

B. IFT Experiments and Gradients $\partial \mathbf{e}/\partial \boldsymbol{\rho}$

In the context of Fig. 2, consider the following three IFT experiments which are conducted under identical (a) point-to-point reference r, (b) feedforward controller $\mathcal{C}_{ff}$, and (c) feedback controller $\mathcal{C}_{fb}$ with parameters $\boldsymbol{\rho}$ [15], [18]:

Experiment I: The result of the first experiment with no auxiliary input, i.e., $w_I = 0$ (the subscript indicates the experiment number) satisfies:

$$\begin{aligned} y_I(\delta) &= v_I(\delta) + \mathcal{P}(\delta)(\mathcal{C}_{ff}(\delta)r(\delta) + \mathcal{C}_{fb}(\delta)(r(\delta) - y_I(\delta))) \\ &= \mathcal{S}(\delta)v_I(\delta) + \mathcal{T}(\delta)\left(\frac{\mathcal{C}_{ff}(\delta)}{\mathcal{C}_{fb}(\delta)} + 1\right) r(\delta), \end{aligned} \tag{10}$$

with the sensitivity function $\mathcal{S}$ and the complementary sensitivity function $\mathcal{T}$ defined in frequency domain by

$$\mathcal{S}(\delta) = \frac{1}{1+\mathcal{C}_{fb}(\delta)\mathcal{P}(\delta)} \text{ and } \mathcal{T}(\delta) = \frac{\mathcal{C}_{fb}(\delta)\mathcal{P}(\delta)}{1+\mathcal{C}_{fb}(\delta)\mathcal{P}(\delta)}. \quad (11)$$

From (10), it follows that

$$\frac{\partial y_I(\delta)}{\partial \boldsymbol{\rho}} = \frac{1}{\mathcal{C}_{fb}(\delta)} \frac{\partial \mathcal{C}_{fb}(\delta)}{\partial \boldsymbol{\rho}} \mathcal{T}(\delta)(r(\delta) - y_I(\delta)). \quad (12)$$

The purpose of experiment I is twofold. First, the data y_I are used (via $e_I = r - y_I$) in the computation of the gradients in (8) regarding the part $\mathbf{e} = \mathbf{e}_I$. Second, these data define the auxiliary input w_{II} for experiment II.

Experiment II: Define the second experiment with the auxiliary input $w_{II} = e_I$, i.e., use the closed-loop error signal from the first experiment as auxiliary input to the second experiment. Note that w_{II} is a signal that is added to the unaltered setpoint signal r (recall Fig. 2). In the implementation, w_{II} therefore only needs to be synchronized in time with the setpoint signal. The second experiment gives:

$$\begin{aligned} y_{II}(\delta) = \mathcal{S}(\delta)v_{II}(\delta) + \mathcal{T}(\delta)\left(\frac{\mathcal{C}_{ff}(\delta)}{\mathcal{C}_{fb}(\delta)} + 1\right) r(\delta) \\ - \mathcal{T}(\delta)e_I(\delta), \end{aligned} \quad (13)$$

The purpose of experiment II is to construct unbiased gradient estimates $\partial J/\partial \boldsymbol{\rho}$ in (8) from sole data. This is done by assuring that the noise contained in $\mathbf{e} = \mathbf{e}_I$ (and obtained from experiment I) does not correlate with the noise contained in the estimates of $\partial \mathbf{e}/\partial \boldsymbol{\rho}$. For this reason, a third experiment is conducted.

Experiment III: From a third experiment, which is similar to the first experiment, i.e., $w_{III} = 0$, we obtain

$$y_{III}(\delta) = \mathcal{S}(\delta)v_{III}(\delta) + \mathcal{T}(\delta)\left(\frac{\mathcal{C}_{ff}(\delta)}{\mathcal{C}_{fb}(\delta)} + 1\right) r(\delta). \quad (14)$$

Subtracting (14) from (13) gives

$$\begin{aligned} &\mathcal{T}(\delta)(r(\delta) - y_I(\delta)) \\ &= y_{III}(\delta) - y_{II}(\delta) + \mathcal{S}(\delta)(v_{II}(\delta) - v_{III}(\delta)). \end{aligned} \quad (15)$$

Hence, in view of the properties of the disturbances v,

$$\text{est}\{\mathcal{T}(\delta)(r(\delta) - y_I(\delta))\} = y_{III}(\delta) - y_{II}(\delta). \quad (16)$$

This demonstrates the ability to construct unbiased gradient estimates $\partial J/\partial \boldsymbol{\rho}$ in (8) from sole data. Namely, substitution of (16) in (12) gives an estimate of the gradients in (12):

$$\text{est}\left\{\frac{\partial y_I(\delta)}{\partial \boldsymbol{\rho}}\right\} = \frac{1}{\mathcal{C}_{fb}(\delta)} \frac{\partial \mathcal{C}_{fb}(\delta)}{\partial \boldsymbol{\rho}} (y_{III}(\delta) - y_{II}(\delta)), \quad (17)$$

which is based on measured data and known model relations regarding the controller $\mathcal{C}_{fb}(\delta)$. From (17) the step toward the gradients $\partial \mathbf{e}/\partial \boldsymbol{\rho}$ is straightforward. Note from Fig. 2 that

$$\frac{\partial e}{\partial \boldsymbol{\rho}} = -\frac{\partial y}{\partial \boldsymbol{\rho}}. \quad (18)$$

With (18) the gradients $\partial \mathbf{e}/\partial \boldsymbol{\rho}$ are defined in time domain by

$$\frac{\partial \mathbf{e}}{\partial \boldsymbol{\rho}} = -[\mathbf{C}_1(\mathbf{y}_{III} - \mathbf{y}_{II}) \dots \mathbf{C}_m(\mathbf{y}_{III} - \mathbf{y}_{II})], \quad (19)$$

with $\mathbf{C}_1 = \mathbf{C}(\boldsymbol{\rho}), \dots, \mathbf{C}_m = \mathbf{C}_m(\boldsymbol{\rho})$ Toeplitz matrices containing the impulse responses of the (stable) filters

$$\mathcal{C}_1(\delta) = \frac{1}{\mathcal{C}_{fb}(\delta)} \frac{\partial \mathcal{C}_{fb}(\delta)}{\partial \boldsymbol{\rho}(1)} \dots \mathcal{C}_m(\delta) = \frac{1}{\mathcal{C}_{fb}(\delta)} \frac{\partial \mathcal{C}_{fb}(\delta)}{\partial \boldsymbol{\rho}(m)}, \quad (20)$$

and the data-sampled signals $\mathbf{y}_{II} = [y_{II}(1) \dots y_{II}(N)]^{\mathrm{T}}$ and $\mathbf{y}_{III} = [y_{III}(1) \dots y_{III}(N)]^{\mathrm{T}}$. In summary, the data $e_I = r - y_I$ from experiment I are used (a) in the computation of the gradients in (8) regarding the part $\mathbf{e} = \mathbf{e}_I$, and (b) to define the auxiliary input w_{II} for experiment II. The gradient estimates in (17) are obtained from the data of experiment II and experiment III, i.e., in view of the uncorrelated disturbances v_{II} and v_{III}. Since v_{II} and v_{III} in $\partial \mathbf{e}/\partial \boldsymbol{\rho}$ do not correlate with v_I in $\mathbf{e}$, $\partial J/\partial \boldsymbol{\rho}$ in (8) becomes unbiased. This would not be the case if, instead of experiment III, the data from (identical) experiment I were used. Note that the approximate Hessian matrix H in (9) may be biased. Unbiased estimates of the Hessian are proposed in [41].

Despite the clear advantages of having unbiased gradient estimates without the use of parametric models, IFT suffers from robustness issues as there are no guarantees that the closed-loop system remains stable during the iterations. In fact, the optimized controller is often close to instability and the danger of closed-loop instability during the iterations may be high, see also [8] for an interesting explanation. Also, strong guarantees for robust stability are not provided. To ensure robustness, a frequency-domain constraint is added to the objective function in (5) that represents the robustness objective. This will be referred to as constrained IFT.

IV. Constrained Iterative Feedback Tuning

In this section, a new constrained IFT approach is presented as a means to impose frequency-domain specifications, e.g., the modulus margin, on the closed-loop stage system. The section is divided in four parts: (a) frequency-domain constraints, (b) unconstrained optimization problem formulation, (c) constrained IFT algorithm (this includes an extra iteration loop to penalize violations of the robustness specifications), and (d) computation of the gradients (with respect to the penalty function). The outcome of the section is an IFT method that ensures robustness.

A. Frequency-Domain Constraints

Constrained IFT will be used to find the optimal set of controller parameters $\boldsymbol{\rho}^*$ that minimizes the cost function $J = J(\boldsymbol{\rho})$ under the constraint

$$g \leq 0, \quad (21)$$

with $g = g(\boldsymbol{\rho})$ defined by

$$g(\boldsymbol{\rho}) = \max_{\omega} \left(|\mathcal{S}(\omega, \boldsymbol{\rho})| - \mathcal{S}_b(\omega)\right), \quad (22)$$

where $\mathcal{S} = \mathcal{S}(\omega, \boldsymbol{\rho})$ is the closed-loop sensitivity function, see (11), and $\mathcal{S}_b = \mathcal{S}_b(\omega) > 0$ is a frequency-dependent function that for each frequency ω specifies the amplitude constraints imposed on $\mathcal{S}$. Note that the particular choice of g in (22) is made to facilitate the exposition. The theory that is presented is general and can be applied to more arbitrary choices of g.

B. Unconstrained Optimization Problem Formulation

To transform the above-described constrained optimization problem into an unconstrained optimization problem formulation similar to (6), the cost function in (5) is extended with a penalty $J_g = J_g(\omega, \boldsymbol{\rho})$ that is based on the constraint function in (22), or

$$J = \underbrace{\frac{1}{2N}\mathbf{e}^T\mathbf{e}}_{J_e} + \alpha \underbrace{\frac{1}{2}\phi(g)g^2}_{J_g}, \tag{23}$$

with $\alpha > 0$ a scaling factor and ϕ a piecewise linear function

$$\phi(g) = \begin{cases} 0, \text{ if } g \leq 0 \\ 1, \text{otherwise.} \end{cases} \tag{24}$$

This unconstrained optimization problem can be solved iteratively using the Gauss-Newton algorithm in (7). Different from (8), however, the gradient $\partial J/\partial \boldsymbol{\rho}$ is given by

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\rho}} &= \frac{\partial J_e}{\partial \boldsymbol{\rho}} + \alpha \frac{\partial J_g}{\partial \boldsymbol{\rho}} \\ &= \frac{1}{N}\frac{\partial \mathbf{e}^T}{\partial \boldsymbol{\rho}}\mathbf{e} + \alpha\phi(g)g\Bigg(\underbrace{\frac{\partial \phi(g)}{\partial g}\frac{\partial g}{\partial \boldsymbol{\rho}}}_{0}g + \phi(g)\frac{\partial g}{\partial \boldsymbol{\rho}}\Bigg), \end{aligned} \tag{25}$$

whereas an approximation of the Hessian $H = H(\boldsymbol{\rho})$ is given by

$$H = \underbrace{\frac{1}{N}\frac{\partial \mathbf{e}^T}{\partial \boldsymbol{\rho}}\frac{\partial \mathbf{e}}{\partial \boldsymbol{\rho}}}_{H_e} + \alpha \underbrace{\phi^2(g)\left(\frac{\partial g}{\partial \boldsymbol{\rho}}\right)^2}_{H_g}. \tag{26}$$

From (25) and (26), it follows that finding $\boldsymbol{\rho}^*$ in (6) with (7) requires the additional gradients $\partial g/\partial \boldsymbol{\rho}$; the gradients $\partial \mathbf{e}/\partial \boldsymbol{\rho}$ are obtained from the IFT experiments as explained in Section III.

C. Constrained Iterative Feedback Tuning Algorithm

Before deriving the gradients $\partial g/\partial \boldsymbol{\rho}$, consider the resulting constrained IFT algorithm as depicted in Fig. 4. The algorithm consists of the following basic steps:

1) Conduct three IFT experiments in the context of Fig. 2 with the most recent controller parameters $\boldsymbol{\rho}$.
2) Obtain the gradients $\partial J(\rho_i)/\partial \boldsymbol{\rho}$ and the approximate Hessian $H(\rho_i)$. When coming from step (1), this involves $\partial J_e/\partial \boldsymbol{\rho}$ and $H_e(\rho_i)$ only. When coming from step (4), first, $\partial J_e/\partial \boldsymbol{\rho}$ and $H_e(\rho_i)$ need to be recomputed and, second, the gradients $\partial J_g/\partial \boldsymbol{\rho}$ and the approximate Hessian $H_g(\rho_i)$ are required.
3) Using (7), compute the candidate parameter set $\boldsymbol{\rho}_{i+1}$ and update the controller $\mathcal{C}_{fb} = \mathcal{C}_{fb}(\boldsymbol{\rho}_{i+1})$.
4) On the basis of a non-parametric model, check if $g(\boldsymbol{\rho}_{i+1}) \leq \epsilon$ with $\epsilon > 0$. If satisfied, accept $\boldsymbol{\rho}_{i+1}$ as controller parameter set and either return to step (1) if $i \leq i_{\text{max}}$ or terminate the algorithm otherwise. If not satisfied, proceed with step (2) to find a candidate parameter set $\boldsymbol{\rho}_{i+1}$ that does satisfy $g \leq \epsilon$. However, if $g(\boldsymbol{\rho}_{i+1})$ exceeds the user-defined threshold value indicated by $\psi > 0$, first the step size γ_i is reduced as to damp the convergence process. The motivation for this specific stopping criterion will be given in the next sub-section.

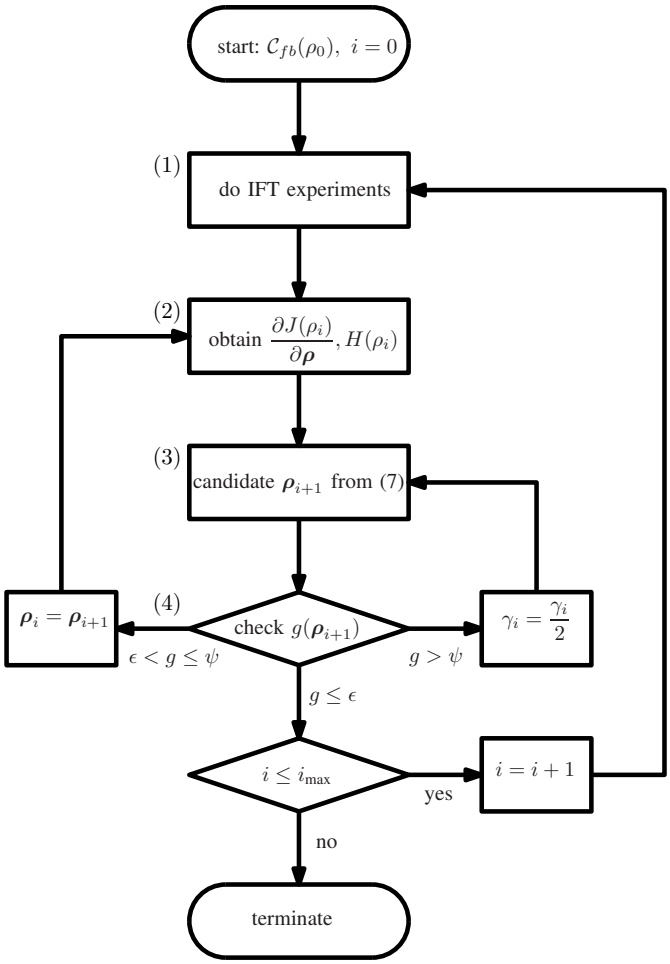

Fig. 4. Overview of the constrained IFT algorithm.

D. Computing the Gradients $\partial g/\partial \boldsymbol{\rho}$

In conducting the iterative feedback tuning algorithm in Fig. 4, computing the gradients $\partial g/\partial \boldsymbol{\rho}$ is done upon violation of the constraint g in (21), i.e., $g > 0$. To detect such a constraint violation, a non-parametric model based on frequency response data is used for the sensitivity function $\mathcal{S} = \mathcal{S}(\omega, \boldsymbol{\rho})$ with the candidate parameter set $\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1}$. It is important to realize that no extra experiments are conducted to obtain $\partial g/\partial \boldsymbol{\rho}$. Furthermore, note that frequency response data measurements for motion systems are inexpensive and quickly obtained, but cannot be used directly for controller synthesis, as model-based controller synthesis requires a parametric model. By imposing constraints, machine damage as a result of controller parameters that possibly induce severe violations of the closed-loop frequency-domain specifications will be avoided. In fact, without reasonable assurance that the candidate parameter set $\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1}$ found in step (3) of the IFT algorithm in Fig. 4 avoids such violations, no IFT experiments will be conducted.

To obtain the gradients $\partial g/\partial \boldsymbol{\rho}$, i.e., the constrained part g, the center difference scheme is used:

$$\begin{aligned}\left.\frac{\partial g}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} &= \left.\frac{\Delta g}{\Delta \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} + \mathcal{O}(\|h\|^2) \text{ for } \|h\| \to 0 \\ &\approx \left[\left.\frac{\Delta g}{\Delta \boldsymbol{\rho}(1)}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} \cdots \left.\frac{\Delta g}{\Delta \boldsymbol{\rho}(m)}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}}\right]^T,\end{aligned} \tag{27}$$

i.e., a perturbation method with step sizes $h = [h_1 \dots h_m]^{\mathrm{T}}$ where h is related to the controller parameters obtained per iteration i, $h_i = 0.01\rho_i$, and

$$\left.\frac{\Delta g}{\Delta \boldsymbol{\rho}(j)}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} = \frac{g(\boldsymbol{\rho}(1)\dots\boldsymbol{\rho}(j)+h_j\dots\boldsymbol{\rho}(m)) - g(\boldsymbol{\rho}(1)\dots\boldsymbol{\rho}(j)-h_j\dots\boldsymbol{\rho}(m))}{2h_j} \tag{28}$$

To estimate the effect of the gradient $\partial J_e/\partial \boldsymbol{\rho}_i$ and approximate Hessian H_e coming from the unconstrained part e (and for which we need to re-compute the candidate parameter set $\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1}$) the fact is used that

$$\begin{aligned}\left.\frac{\partial J_e}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} &= \\ \left.\frac{\partial J_e}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_i} &+ H_e(\rho_i)\Delta\boldsymbol{\rho} + \mathcal{O}(\|\Delta\boldsymbol{\rho}^2\|) \text{ for } \|\Delta\boldsymbol{\rho}\| \to 0,\end{aligned} \tag{29}$$

with $\Delta\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1} - \boldsymbol{\rho}_i$, which implies

$$H_e|_{\boldsymbol{\rho}=\boldsymbol{\rho}_{i+1}} = H_e|_{\boldsymbol{\rho}=\boldsymbol{\rho}_i}. \tag{30}$$

In [9] the validity of such a parameter approximation is investigated. With (29), (30), and (27) substituted in (25) and (26) a new candidate set of the controller parameters $\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1}$ is computed with (7). In an iterative way, the set $\boldsymbol{\rho} = \boldsymbol{\rho}_{i+1}$ is sought that satisfies $g \leq \epsilon$ with $\epsilon > 0$ a constant that can be chosen arbitrary small. Typically we choose $\epsilon = 0.5$ dB.

Remark 4 *The unconstrained optimization problem in* (6) *is generally non-convex, irrespective of adding J_g to the original cost function J_e in* (23)*. With* (7)*, it is therefore not realistic to expect more than local convergence, i.e., a sufficiently good initial parameter set $\boldsymbol{\rho}_0$ is generally required to reach the global minimum; in this paper, $\boldsymbol{\rho}_0$ is the result of an industrial (and performance-relevant) robust controller tuning process. To ensure convergence $g \leq \epsilon$ rather than the more generally-known stopping criterion $\partial J/\partial \boldsymbol{\rho} \leq \bar{\epsilon}$, with $\bar{\epsilon}$ a threshold, is used. The reason for this is that $g \leq \epsilon$ is a direct measure of the performance objectives in the IFT algorithm of Fig. 4. We thus accept a controlled violation of the frequency-domain robustness specifications by the maximum amount of ϵ. An alternative would be to use a barrier function approach.*

Remark 5 *In the unconstrained optimization problem of* (23)*, J_g is added to J_e. To scale both parts with respect to each other, the scaling factor α is chosen at $\alpha = 10$ nm^2. This is because $e \sim$ nm and $g \sim 1$ such that both parts J_e and J_g become equally weighted in* (23)*. Similarly, the optimization window, i.e., the sampled data interval used to define J_e in* (23)*, should be chosen with the application specifics in mind. For the stage control case, the data sampled signals* $\mathbf{e}$ *have $N = 500$ samples with 250 samples in the acceleration phase and 250 samples in the scanning phase of constant velocity, see [13] for more details on the choice of optimization window.*

Remark 6 *An important aspect in the IFT algorithm of Fig. 4 is the choice of the iteration factor γ_i in* (7)*. The strategy that is being followed is to start with $\gamma_0 = 1$ (recall step (1) in Fig. 4) and then do step refinement depending on the amount of violation that is encountered with respect to the constant $\psi > 0$; see also [20]. Typically we choose ψ which is a measure of violation at ψ = 1 dB. Constraint violations $g > \psi$ will induce step refinement. Constraint violations $\epsilon < g \leq \psi$ do not induce step refinement as to avoid too slow convergence rates. Constraint violations $g \leq \epsilon$ (in the case that $i \leq i_{max}$) induce a re-initialization of the step size to $\gamma_i = 1$. A representative sequence for the step size γ_i, such as encountered in the experiments during 5 iterations, is given by $\{1, 0.25, 0.25, 0.0625, 0.001953125\}$. Note that reducing γ_i comes with the risk of effectively halting the optimization. In reducing this risk, one could decide to choose a larger ψ.*

V. Measurements Results

In this section measurement results are presented that are obtained with constrained IFT on a wafer scanner. Since constrained IFT does not require a parametric plant model, it is more convenient to apply constrained IFT directly to the wafer stage system rather than doing simulation studies first. For the experiments, the motion control structure from Fig. 2 is used including the auxiliary input w. In presenting the results, the section is divided into four parts: (a) parameter convergence, (b) optimized controller settings, (c) robust stability aspects, and (d) time-domain performance.

A. Parameter Convergence

Given the structure of the wafer stage feedback controller $\mathcal{C}_{fb} = \mathcal{C}_{fb}(\boldsymbol{\rho})$ in (1) constrained IFT refers to tuning the PID-controller parameters k_p, ω_i, ω_d, see also [25], [26], [42], and the additional notch filter frequencies $\omega_{z,i}, \omega_{p,i}$, or

$$\boldsymbol{\rho} = \begin{bmatrix} k_p & \omega_d & \omega_i & \omega_{z,1} & \omega_{p,1} & \dots & \omega_{z,5} & \omega_{p,5} \end{bmatrix}^{\mathrm{T}}. \tag{31}$$

Note that machine-in-the-loop optimization of the notch filters is relevant for the following reason. Limitations in current machine performance due to unknown variations from machine-to-machine in terms of disturbances or frequency characteristics (resonances) are likely to be overcome by machine-dedicated tunings that properly address these variations.

Remark 7 *In optimizing the feedback controller parameters $\boldsymbol{\rho}$ in $\mathcal{C}_{fb} = \mathcal{C}_{fb}(\boldsymbol{\rho})$ with constrained IFT, the zero and pole damping coefficients $\zeta_{z,i}, \zeta_{p,i}$ are fixed, see also Table I in Section II. This rather arbitrary choice yields a reduced number of parameters to be optimized, which reduces numerical*

complexity, but (potentially) comes at the cost of performance. It is clear that such design choices strongly relate to the (performance) specifics of the application at hand. The low-pass filter $\mathcal{C}_{LP}$ in (1) *is fixed as to assure high-frequency roll-off.*

The results of the IFT algorithm in terms of parameter convergence are depicted in Fig. 5. Given an identical set of controller parameters $\boldsymbol{\rho}_0$ at iteration number $i = 0$, it can be seen that convergence occurs in basically five iteration steps. The differences between the initial parameter set and the final

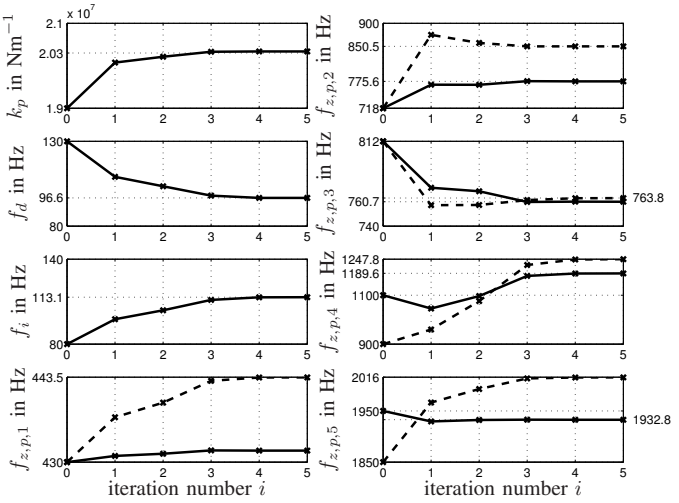

Fig. 5. Measured parameter convergence diagram; for the notch filters $\mathcal{N}_i$ convergence of the zero frequencies $f_{z,i} = \omega_{z,i}/2\pi$ is indicated by the solid curves while convergence of the pole frequencies $f_{p,i} = \omega_{p,i}/2\pi$ is indicated by the dashed curves; $\gamma_0 = 1$, ϵ = 0.5 dB.

parameter set obtained with constrained IFT are mainly caused by the variation in disturbances and (plant) dynamics from machine-to-machine, hence the need for a machine-dedicated tuning of the notch filters. Note the subtle differences (for example) in notch filter $\mathcal{N}_1$ which becomes a skew notch filter in order to obtain more disturbance suppression around 400 Hz, while notch filter $\mathcal{N}_3$ demonstrates lowering of the notch frequency but leaving the notch shape unchanged as to avoid noise amplification around 760 Hz. Both frequencies relate to turbo pumps that through excitation of the metrology frame disturb the interferometer measurement system, i.e., disturbances that corrupt the plant output, see also Fig. 2. Interestingly, Fig. 5 shows that convergence in the parameter space is non-monotonic. This indicates that J is probably non-quadratic in the parameters, necessitating several iterations of (7).

B. Optimized Controller Settings

Given the controller parameter sets $\boldsymbol{\rho}_0$ and $\boldsymbol{\rho}_5$ found in Fig. 5 at $i = 0$ and $i = 5$, respectively, Fig. 6 shows the Bode diagrams of the resulting controllers $\mathcal{C}_{fb}$. From the figure, it can be seen that constrained IFT induces ≈ 3.5 dB extra low-frequency disturbance rejection below 30 Hz but at the cost of $\approx$ 12 dB high-frequency noise amplification around 5 kHz. From the perspective of error minimization,

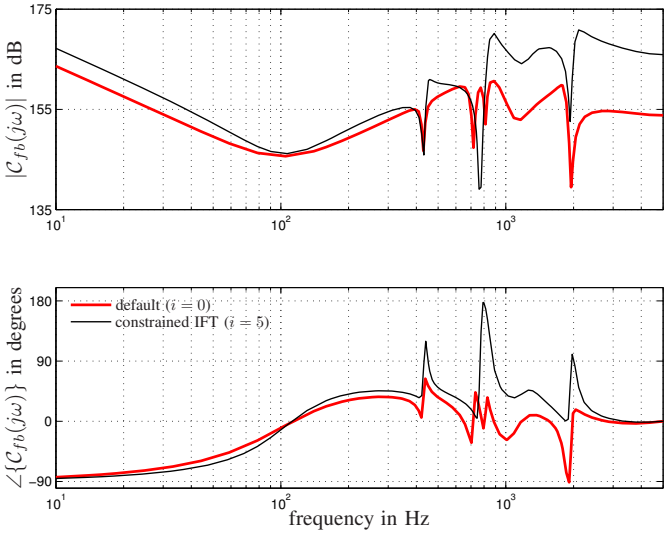

Fig. 6. Bode diagrams of the controllers before ($i = 0$) and after ($i = 5$) optimization.

this result is clear. The spectral contents of the error signals is (relatively) small at the high-frequency interval thereby justifying the amplifications. Formulation of a performance-relevant cost function can therefore be seen as the crucial step in the control design, see also [32]. Apart from the more subtle differences in notch filter design beyond 800 Hz, a 90 degrees phase lead is induced with respect to the default parameter set $\boldsymbol{\rho}_0$. This involves the first plant resonance around 1kHz, recall Fig. 3, and will be discussed in the next sub-section.

C. Robust Stability Aspects

The comparison of the default parameter set $\boldsymbol{\rho}_0$ in Fig. 5 with the constrained IFT parameter set $\boldsymbol{\rho}_5$ in terms of the closed-loop sensitivity function $\mathcal{S}$ in (11) is shown in Fig. 7. It can be seen that with constrained IFT the final parameter

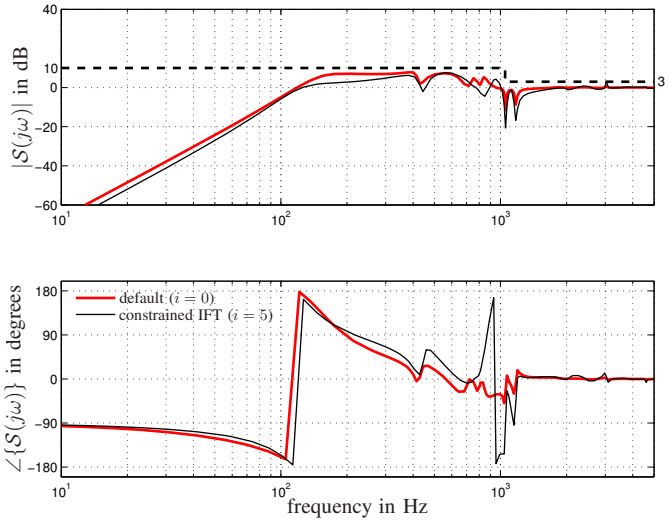

Fig. 7. Bode diagrams of the sensitivity function $\mathcal{S}$ based on measured data before ($i = 0$) and after ($i = 5$) constraint IFT optimization; $\mathcal{S}_b$ = 9.5 dB on the interval $f \in [0\ 1.05)$ kHz and 2.5 dB on the interval $f \in [1.05\ 5]$ kHz, $\alpha = 10$ nm^2, $\gamma_0 = 1$, ϵ = 0.5 dB.

set ρ_5 induces improved low-frequency performance (below 80 Hz but also between 110 and 500 Hz) while satisfying the indicated amplitude constraints $\mathcal{S}_b(\omega) + \epsilon$ (dashed curve).

Remark 8 *Note that by using a weighted modulus margin, robust stability is obtained against inverse multiplicative uncertainties. Moreover, the non-parametric weighting is different from the standard approaches used in robust control.*

Remark 9 *As to avoid machine damage, IFT (without constraints) is not applied to the machine. The reason for this is shown in Fig. 8 which contains simulation results only: IFT*

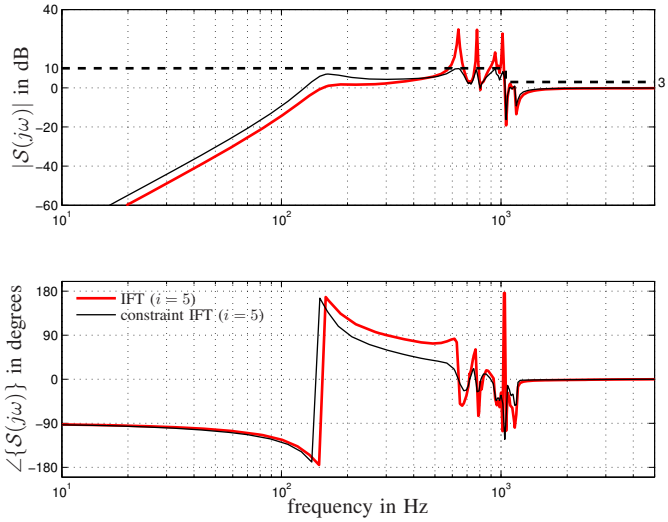

Fig. 8. Bode diagrams of the sensitivity function $\mathcal{S}$ based on simulated data with IFT ($i = 5$), i.e., no constraints, and with constraint IFT ($i = 5$); $\mathcal{S}_b$ = 9.5 dB on the interval $f \in [0\ 1.05)$ kHz and 2.5 dB on the interval $f \in [1.05\ 5]$ kHz, $\alpha = 10$ nm^2, $\gamma_0 = 1$, ϵ = 0.5 dB.

induces unacceptable amplifications at high frequencies. With constrained IFT, it can be seen that the amplitude constraints are merely touched upon at distinct frequency points. The latter observation could favor the introduction of frequency-varying weights as done in [14].

Robust stability properties by constraining the sensitivity (amplitude) characteristics also follow from the Nyquist plot evaluation in Fig. 9. In the figure, it can be seen that the modulus margin of 10 dB is met (either before and after optimization) by proper encirclement of the disk with radius of ≈ 0.32 around the point -1. Moreover, the earlier mentioned phase lead of 90 degrees, which is induced by constrained IFT around the plant resonance frequency of 1 kHz, allows for amplification around this frequency, see also Fig. 7, and leads to $|\mathcal{S}(\omega = 2\pi 1000)| \approx$ -20 dB.

Remark 10 *The optimized controller obtained through constrained IFT clearly outperforms the default controller. This, however, is partly the result of different optimization objectives. Namely, the default controller tuning is the result of a frequency-domain (and loopshaping-based) optimization process in which frequency response data are used from several machines. As such, a controller tuning is obtained that*

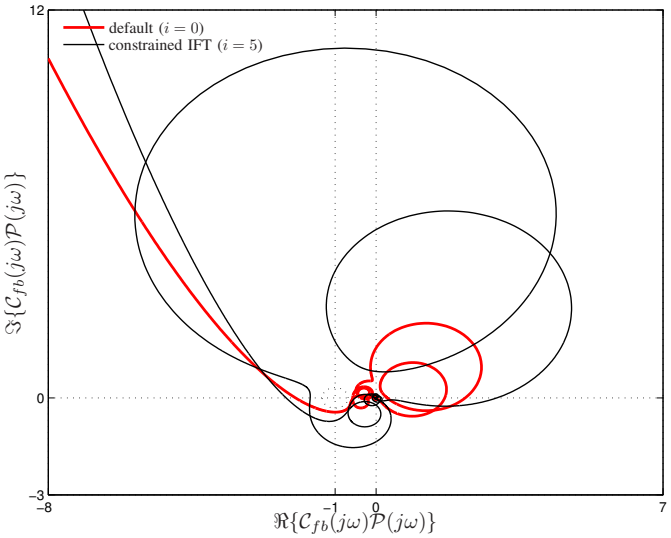

Fig. 9. Nyquist plots of the open-loop frequency response functions before ($i = 0$) and after ($i = 5$) optimization.

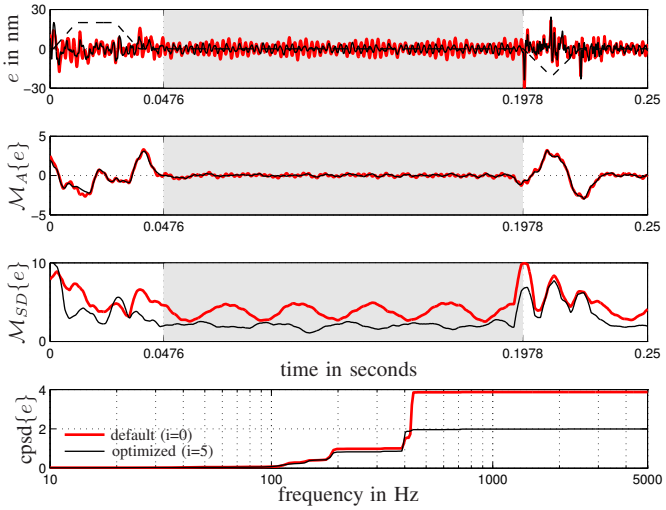

Fig. 10. Time-series representation of experimental data at iteration $i = 0$ and iteration $i = 5$ together with the results from cumulative power spectral density analysis; the scaled acceleration set-point profile is indicated with dashed lines; the scanning performance window is indicated in grey.

is robust against a certain degree of machine variation. But this tuning is not very likely to achieve optimal performance when applied to specific machines. In particular, when in the tuning process machine-specific disturbances have not been taken into account.

D. Time-Domain Performance

The effect of constrained IFT in terms of time-domain performance is depicted in Fig. 10. Performance is evaluated for representative point-to-point scanning motion at a corner part of the wafer, see the dashed curves for the scaled acceleration set-point profile. In the top part of the figure unfiltered error responses e are shown before ($i = 0$) and after ($i = 5$) optimization. The middle part of the figure depicts these responses after either using a moving average filter operation

$\mathcal{M}_A$ – as a measure for machine overlay – or a moving standard deviation filter operation $\mathcal{M}_{SD}$ – as a measure for machine imaging, recall the explanation of (4). The lower part of the figure shows a cumulative power spectral density analysis of the error responses e. From Fig. 10, it is clear that in terms of error responses constrained IFT mainly contributes to the removal of the frequency contributions around 400 Hz, which are caused by the earlier mentioned exogenous disturbances from turbo pumps. These disturbances mainly induce $\mathcal{M}_{SD}$-filtered errors. This machine-specific advantage was not fully known in advance and therefore could not be properly anticipated for in the (manual) loop shaping tunings. This confirms the potential benefit of IFT to account for actual disturbance spectra in a control-relevant manner.

VI. CONCLUSIONS

In this paper, a new approach for constrained IFT has been presented and experimentally demonstrated on the high-precision stage systems of a lithography machine. The IFT algorithm is extended with a frequency-domain penalty to penalize violations of the frequency-domain specifications imposed on the closed-loop sensitivity function. Using a non-parametric model, an adaptation is made to the candidate parameter set obtained from IFT. The prediction involves a straightforward Taylor series expansion part using the resulting unbiased gradient estimates. Additionally, it involves a part that uses the perturbation method to obtain the gradients with respect to the penalty. With the adaptation, a parameter set is obtained that from a model point of view meets the frequency-domain specifications. As a result, constrained IFT is done (a) under the assurance that frequency-domain specifications are satisfied, or (b) violations of the specifications stay within controllable levels. For both the PID and notch filter controller parameters of a generally more advanced controller structure, it is demonstrated that machine-specific performance, which is clearly disturbance-related, can be obtained with constrained IFT. In time-domain, constrained IFT is used to create the machine-specific fine-tunings of the feedback control design in addition to the nominal loop shaping tuning process that is done in frequency domain. Ongoing work involves the MIMO case, effective constraint formulation, and the development of efficient algorithms.

VII. ACKNOWLEDGEMENT

We would like to thank the contributions by dr. D. Kostić and dr. B. de Jager for their contribution in an early stage of this research.

REFERENCES

[1] A. Al Mamun, W.Y. Ho, W.E. Wang, and T.H. Lee. Iterative Feedback Tuning (IFT) of Hard Disk Drive Head Positioning Servomechanism. *In Proceedings of the Conference of the IEEE Industrial Electronics Society*, Taipei, Taiwan, 769-774, 2007.

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

[3] F. Boeren, T. Oomen, and M. Steinbuch. Iterative motion feedforward tuning: A data-driven approach based on instrumental variable identification. *Control Engineering Practice*, To appear.

[4] M. Boerlage, M. Steinbuch, P. Lambrechts, and M. Van de Wal. Model-based feedforward for motion systems. *In Proceedings of the Conference on Control Applications*, 2003.

[5] F. De Bruyne, and L.C. Kammer. Iterative feedback tuning with guaranteed stability. *In Proceedings of the American Control Conference*, San Diego, California, USA, 3317-3321, 1999.

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

[7] M.C. Campi, A. Lecchini, and S.M. Savaresi. Virtual reference feedback tuning: A direct method for the design of feedback controllers. *Automatica*, 38:1337–1346, 2002.

[8] S. Garatti, M.C. Campi, and S. Bittanti. Iterative robust control: SPeeding up improvement through iterations. *Systems & Control Letters*, 59:139–146, 2010.

[9] M. Gevers. A decade of progress in iterative process control design: from theory to practice. *Journal of Process Control*, 12(4):519-531, 2002.

[10] S. Gunnarsson, O. Rousseaux, and V. Collignon. Iterative feedback tuning applied to robot joint controllers. *Journal of Process Control*, 12(4):519-531, 2002.

[11] K. Hamamoto, T. Fukuda, and T. Sugie. Iterative feedback tuning of controllers for a two-mass-spring system with friction. *Control Engineering Practice*, 11(9):1061–1068, 2003.

[12] A. Hansson, K. El-Awady, and B. Wahlberg A primal-dual interior-point method for iterative feedback tuning. *In Proceedings of the IFAC World Congress*, Beijing, China, 36–41, 1998.

[13] 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.

[14] M.F. Heertjes, M. Galluzzo, and L. Kuindersma. Robust Data-Driven Control for the Stage Synchronization Problem. *IFAC World Congress*, Cape Town, South Africa, 1–6, 2014.

[15] H. Hjalmarsson, S. Gunnarsson, and M. Gevers A convergent iterative restricted complexity control design scheme. *In Proceedings of the Conference on Decision and Control*, Lake Buena Vista, Florida, USA, 1735-1740, 1994.

[16] H. Hjalmarsson, and T. Birkeland. Iterative feedback tuning of linear time-invariant MIMO systems. *In Proc. Conference on Decision and Control*, Tampa, Florida, USA, 3893-3898, 1998.

[17] H. Hjalmarsson. Efficient tuning of linear multivariable controllers using iterative feedback tuning. *Journal of Adaptive Control and Signal Processing*, 13(7):553-572, 1999.

[18] H. Hjalmarsson. Iterative feedback tuning - An overview. *Journal of Adaptive Control and Signal Processing*, 16(5):373-395, 2002.

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

[20] J.K. Huusom, N.K. Poulsen, and S.B. Jørgensen. Improving convergence of iterative feedback tuning. *Journal of Process Control*, 19(4):570-578, 2009.

[21] J.K. Huusom, H. Hjalmarsson, N.K. Poulsen, and S.B. Jørgensen. A design algorithm using external perturbation to improve iterative feedback tuning converence. *Automatica*, 47:2665-2670, 2011.

[22] S. Kissling, Ph. Blanc, P. Myszkorowski, and I. Vaclavik. Application of iterative feedback tuning (IFT) to speed and position control of a servo drive. *Control Engineering Practice*, 17:834-840, 2009.

[23] D. Kostić. Data-driven robot motion control design. *Ph.D. thesis*, Eindhoven University of Technology, 2004.

[24] P. Lambrechts, M. Boerlage, and M. Steinbuch. Trajectory planning and feedforward design for high performance motion systems. *In Proc. American Control Conference*, Boston, Massachusets, USA, 4637-4642, 2004.

[25] O. Lequin. Optimal closed loop PID tuning in the process industry with the Iterative Feedback Tuning scheme. *In Proceedings of the European Control Conference*, Bruxelles, Belgium, Paper TH-A-H6, 1997.

[26] O. Lequin, M.Gevers, M. Mossberg, E. Bosmans, and L. Triest. Iterative feedback tuning of PID parameters: comparison with classical tuning rules. *Control Engineering Practice*, 11(9):1023-1033, 2003.

[27] H.J.Levinson. Principles of lithography - Third Edition. *SPIE PRESS*, Bellingham, Washington, 2010.

[28] D. Liu, A.J. McDaid, K.C. Aw, and S.Q. Xie. Position control of an ionic polymer metal composite actuated rotary joint using iterative feedback tuning. *Mechatronics*, 21(1):315-328, 2011.

[29] K. Mahathanakiet, M.A. Johnson, A. Sanchez, and M.A. Wade. Iterative feedback tuning and an application to a wastewater treatment plant. *In Proceedings of the Asian Control Conference*, Singapore, 256-261, 2002.

[30] S. Van der Meulen, R.L. Tousain, and O. Bosgra. Fixed structure feedforward controller design exploiting iterative trials: Application to a wafer stage and a desktop printer. *Journal of Dynamical Systems, Measurement, and Control*, 130:051006–1, 2008.

[31] R.H. Middleton, and G.C. Goodwin. Improved finite word length characteristics in digital control using delta operators. *IEEE Transactions on Automatic Control*, 31(11):1015–1021, 1986.

[32] M. Mossberg. Controller tuning via minimization of time weighted absolute error. *In Proceedings of the American Control Conference*, Boston, Massachusetts, 2740-2744, 2004.

[33] T. Oomen, R. Van Herpen, S. Quist, M. Van de Wal, O.H. Bosgra, and M. Steinbuch. Connecting system identification and robust control for next-generation motion control of a wafer stage. *IEEE Transactions on Control Systems Technology*, 22(1):102-118, 2014.

[34] 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.

[35] M.B. Radac, R.E. Precup, S. Preitl, and C.A. Dragos. Iterative feedback tuning in MIMO systems. Signal processing and application. *International Symposium on Applied Computational Intelligence and Informatics*, Timisoara, Romania, 77-82, 2009.

[36] Z.P. Rico, A. Lecchini-Visintini, and R.Q. Quiroga. Position control of an ionic polymer metal composite actuated rotary joint using iterative feedback tuning. *Mechatronics*, 21(1):315-328, 2011.

[37] Z.P. Rico, A. Lecchini-Visintini, and R.Q. Quiroga. Iterative feedback tuning for the joint controllers of a 7-DOF whole arm manipulator. *Conference on Decision and Control*, Maui, Hawaii, USA, 544-549, 2012.

[38] D. Rupp, and L. Guzzella. Iterative tuning of internal model controllers with application to air/fule ratio control. *Transactions on Control Systems Technology*, 18(1), 177-184, 2010.

[39] M.G. Safonov and T.C. Tsao. The unfalsified control concept and learning. *IEEE Transactions on Automatic Control*, 42(6):843–847, 1997.

[40] J. Sjöberg, F. De Bruyne, M. Agarwal, B.D.O. Anderson, M. Gevers, F.J. Kraus, and N. Linard. Iterative controller optimization for nonlinear systems. *Control Engineering Practice*, 11(9):1079–1086, 2003.

[41] G. Solari, and M. Gevers. Unbiased estimation of the Hessian for iterative feedback tuning. *In Proceedings of Conference on Decision and Control*, Paradise Island, Bahamas, 1759-1760, 2004.

[42] 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, 4915–4920, 2014.

[43] 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.

[44] G. Vinnicombe. Frequency domain uncertainty and the graph topology. *IEEE Transactions on Automatic Control*, 38(9): 1371-1383,1993.

[45] M. Van de Wal, G. Van Baars, F.B. Sperling, and O.H. Bosgra. Multivariable $\mathcal{H}_\infty/\mu$ feedback control design for high-precision wafer stage motion. *Control Engineering Practice*, 10(7):739–755, 2002.

[46] P-H Yang, and S-L Koo. Control systems and methods applying iterative feedback tuning for feed-forward and synchronization control of microlithography stages and the like. *United States Patent*, US 8,451,431 B2, 2013.