Optimal Switching Control of Motion Stages

Marcel F. Heertjes ∗ Bram P.T. van Goch ∗ Henk Nijmeijer ∗

∗ Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands (e-mail: m.f.heertjes@tue.nl,b.p.t.v.goch@student.tue.nl,h.nijmeijer@tue.nl)

Abstract: To avoid the trade-off between improved disturbance rejection and deteriorated noise sensitivity under high-gain feedback, an optimal switching control design is presented. Using a deadzone-based switching gain, large servo error signals induce extra controller gain to improve low-frequency suppression. Small error signals induce no extra controller gain as to maintain a small-gain noise response. In a stability-invariant control context, the choice for the switching length is strictly performance-driven. With an iterative scheme based on the Gauss-Newton method, the optimal switching length is found by minimizing the servo error signals in a time-interval of interest. For a fast and nano-accurate wafer stage improved performance is demonstrated in terms of reduced settling times and improved scanning behavior. Robust stability of the switched system and the iterative scheme is assessed using Lyapunov arguments.

Keywords: Gauss-Newton method, Lyapunov stability, motion systems, optimization, switching control, wafer scanner.

1. INTRODUCTION

Wafer scanners used in the semiconductor industry to produce chips combine nano-scale positioning accuracy with aggressive motion profiles. The motion specifications heavily depend on the feed-forward and feedback design whose possibilities generally are constrained by the waterbed effect: increasing the controller gains gives improved low-frequency tracking of the scanning set-points but increases the sensitivity of the system to high-frequency noises.

To deal with the waterbed effect in view of position-dependent stage dynamics and disturbances a deadzone switching controller is used in Heertjes et al. (2009); see Lin and Antsaklis (2009) for a recent survey on switched linear systems. If the servo error signals exceed a switching length incidentally extra "bumpless" gain is switched on inducing extra low-frequency suppression. Error signals that do not exceed the switching length do not induce extra gain and keep a small noise response. The tuning of the switching length depends on the system properties and disturbances at hand. Since stability is guaranteed irrespective the value of this length, the tuning process is performance-driven. In view of machine-specific variations this validates an automatic tuning procedure.

It is the aim of this paper to present such a procedure and find the optimal switching length of a deadzone switching gain embedded in a nominal PID-based feedback structure; in this respect see also the works of Van der Meulen et al. (2008) and Huusom et al. (2009) in linear automatic control system parameter tuning. The procedure is based on the Gauss-Newton method and utilizes a combined model/data-based approach in obtaining the gradients needed to update the switching length. Stability of both the switched system as well as the constrained optimization scheme is addressed using arguments from Lyapunov theory. Performance is demonstrated in simulation and experiment typically with the aim to minimize error signals in a pre-defined time-interval: the scanning interval taken at different (but representative) scans along the wafer.

The remainder of the paper is organized as follows. First, the nominal control design of a wafer stage is presented along with the additional switching controller. Second, stability of the switched nonlinear system is discussed. Third, the automatic tuning procedure used to obtain the optimal switching length is presented along with a proof of convergence and analysis of its properties using time-domain simulation. Fourth, performances are assessed using measurements from an industrial wafer stage. Fifth, a summary of the main conclusions is given.

2. SWITCHING CONTROL IN WAFER SCANNERS

Wafer scanners are the machinery used to expose wafers through a lithographic process. Light from an excimer laser (having a wavelength of 193 nm) passes a reticle which contains a blueprint of the chip. The resulting image passes a lens system where it is scaled down before it exposes the light sensitive layers of a wafer. Both reticle and wafer are part of two separate motion control systems: the reticle stage and the wafer stage; see Mishra et al. (2008). Each performs highly repetitive and synchronized scanning tasks and is controlled in six degrees of freedom using both feed-forward and feedback control. Feed-forward control requires by far the most control effort. Accerlerating a short-stroke wafer stage requires feed-forward forces of about 1000 N. Contrarily, a peak error level of 10 nanometer under proportional-integrator-derivative (PID) control having a controller gain of $2 \times 10^7 \text{ Nm}^{-1}$ gives rise to feedback forces in the order of 0.2 N, hence 0.02% of the overall forces. Disturbance suppression on a nanometer scale, however, is key to the further improvement of nowadays stage performances. Therefore focus is
on the feedback design of the short-stroke wafer stage; a distinction is made between accurate short-stroke positioning (nanometer scale) and course long-stroke positioning (micrometer scale).

A simplified representation of the feedback control design of a short-stroke wafer stage is shown in the block diagram representation of Fig. 1. The single-input single-output (SISO) feedback design is split in two parts. The nominal discrete-time feedback controller part denoted with the transfer $C_{fb}(z)$ in $z$-domain, and the switching controller part denoted with the series connection of the time-domain switching operation $\phi(e)$ and the transfer $F_1(z)$. Input to the controller is the combined signal $e(t) + u(t)$ in time-domain. The error signal $e(t)$ symbolizes a signal in one of the six principle axes of motion. It is obtained by subtracting the measured output signal $y(t)$ from the corresponding set-point signal $r(t)$. The short-stroke wafer stage plant is represented in $z$-domain by $P(z)$. Additionally, time-domain disturbances are represented by $d(t)$. To obtain improved disturbance rejection in counteracting these disturbances a switching feedback connection is used between $u(t)$ and $e(t)$. Herein the output of the switching controller is represented in time-domain by the signal $e_{cl}(t) = \phi(e(t))e(t)$.

An explicit discrete-time model of the short-stroke wafer stage plant is given in $z$-domain by the transfer

$$P(z) = \frac{c_{d,13}^p z^{13} + \cdots + c_{n,1}^p z + c_{n,0}^p}{z^{14} + c_{d,13}^p z^{13} + \cdots + c_{d,1}^p z + c_{d,0}^p}, \quad \text{(1)}$$

whereas the nominal feedback controller is given by

$$C_{fb}(z) = \frac{c_{d,9}^c z^{13} + \cdots + c_{n,4}^c z + c_{n,0}^c}{z^{14} + c_{d,8}^c z^{13} + \cdots + c_{d,1}^c z + c_{d,0}^c}. \quad \text{(2)}$$

The coefficients in (1) and (2) can be found in table 1. Both controller and plant are represented by the open-loop frequency response functions as shown for the $x$-direction in Fig.2 in Bode representation. Robust stability is guaranteed having a controller bandwidth of 171 Hz, a phase margin of 22 degrees, and a gain margin of 4.2 dB.

3. STABILITY OF THE SWITCHED SYSTEM

In state-space representation the switched system reads

$$\dot{x} = Ax + B_1u + B_2r + B_3d$$

$e = c^T \dot{x}$

$$u = -\phi(e)e,$$

with the state vector $x \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times m}$ Hurwitz, $B_1, B_2, B_3, c \in \mathbb{R}^m$, and the switching function $\phi$:

$$\phi(e) = \begin{cases} 0, & \text{if } |e| \leq \delta_k, \\ \alpha - \frac{\alpha \delta_k}{|e|}, & \text{otherwise}, \end{cases} \quad \text{(4)}$$

with $\alpha > 0$ a gain and $\delta_k \geq 0$ the switching length. The motivation for this deadzone "bumpless" switching gain is shown in Fig.3. Extra gain $\phi(e)$ is induced for large

![Figure 1. Block diagram of the simplified feedback control of a short-stroke wafer stage.](image)

![Figure 2. Bode plot of $C_{fb}(j\omega)P(j\omega)$ in $x$-direction using a Nyquist frequency of 2.5 kHz.](image)

![Figure 3. Motivation for a deadzone switching gain $\phi(e)$.](image)
from this control. Small amplitude (and high-frequency) noises that do not benefit from extra control remain within the switching length \( \delta_k \) and do not induce extra gain \( \phi(e) \). The switched system in (3) is expected to better deal with the non-stationarity contained in the uniformly bounded reference signal \( r = r(t) \) and disturbances \( d = d(t) \).

Stability follows from the next result. System (3) with switching function (4) is input-to-state stable if

\[
\Re \{ c^T(j\omega I - A)^{-1}b_1 \} \geq -\frac{1}{\alpha}.
\]

(5)

The proof is obtained as follows. Through the circle criterion (5) implies the existence of a positive definite \( P \in \mathbb{R}^{n \times n} \) such that the positive real condition

\[
A^T P + PA = -qq^T - \epsilon P
\]

(6)

with \( \epsilon > 0 \) and \( q \in \mathbb{R}^n \). By choosing the Lyapunov function candidate \( V: c_1\|x\|^2 \leq V = x^T P x \leq c_2\|x\|^2 \) it follows that

\[
\dot{V} \leq -\alpha x^T P x - (x^T q + \sqrt{2 \|u\|^2 + 2 x^T P (b_2 r + b_3 d)} \leq -\epsilon V + 2 x^T P (b_2 r + b_3 d),
\]

(7)

where explicit use is made of the so-called sector condition: \( \alpha \epsilon u \leq -u^2 \). Since \( 2\|b_2 r + b_3 d\|\|x\| \leq \|b_2 r + b_3 d\|^2 + (1/\lambda)\|x\|^2 \) for any \( \lambda > 0 \) it now follows for

\[
\|x\|^2 \geq \frac{c_2\lambda^2}{(1 - \sigma \epsilon c_1 \lambda - c_2)} \max\{\|b_2\|, \|b_3\|\} \|r + d\|_\infty,
\]

(8)

with \( 0 < \sigma < 1 \) and \( \|r + d\|_\infty = \sup_{t \geq 0} \|r + d\| \) that

\[
\dot{V} \leq -\epsilon V + c_2\lambda\|b_2 r + b_3 d\|^2 + c_2\|x\|^2,
\]

\[
\leq -\sigma \epsilon V.
\]

System (3) is input-to-state stable (Sontag (1995)) because

\[
\|x\| \leq \rho(\|x(0)\|, t) + \gamma(||r + d||_\infty),
\]

(10)

with

\[
\rho(\|x(0)\|, t) = \sqrt{\frac{c_2}{c_1}} \|x(0)\| \exp \left( \frac{\gamma t}{2} \right)
\]

and

\[
\gamma(||r + d||_\infty) = \sqrt{\frac{c_2\lambda^2}{(1 - \gamma)\epsilon c_1 \lambda - c_1 c_2}} \max\{\|b_2\|, \|b_3\|\}\|r + d\|_\infty.
\]

(11)

Stability thus relates to satisfying (5). A graphical interpretation of which is given in Fig.4. Robust stability is guaranteed for \( \alpha = 3 \) because the frequency response functions obtained from either measurement on the wafer stage or from the wafer stage model remain sufficiently to the right of the point \((-1/3, 0)\). So an equivalent gain margin (or radial distance to the critical straight line through the point \((-1/5, 0)\)) of 4.43 dB is obtained. Herein the fact is used that (5) can be written as:

\[
\Re \{ \frac{F_1(j\omega) C_{\phi}(j\omega) P(j\omega)}{1 + C_{\phi}(j\omega) P(j\omega)} \} \geq -\frac{1}{\alpha},
\]

(12)

where the loop-shaping filter \( F_1 \) is chosen such that a significant amount of extra gain \( \alpha \) is accessible without causing the switched system to become unstable. In the scope of this paper it suffices to merely state its discrete-time representation in \( z \)-domain:

\[
F_1(z) = \frac{c_{f,1} z^2 + c_{f,0} z + c_{f,0}}{z^4 + c_{d,3} z^3 + c_{d,2} z^2 + c_{d,1} z + c_{d,0}},
\]

(13)

see table 2 for numerical values of its coefficients. Note that the choice for the switching length \( \delta_k \) in (4) is stability-invariant, hence strictly performance-driven.

### 4. OPTIMAL SWITCHING LENGTH TUNING

To determine the optimal switching length \( \delta_k \) in (4) consider the following objective function

\[
V_k(\delta_k) = e_k^T(\delta_k) e_k(\delta_k),
\]

(14)

with \( e_k(\delta_k) = [e_1, \ldots, e_n] \in \mathbb{R}^n \) the sampled data column (possibly extended and/or filtered) in a time interval \( t \in [t_1, t_2] \) of interest; for the wafer stage this is the scanning interval. For system (3) and objective function (14) the aim is to find the switching length \( \delta_{\text{opt}} \) that minimizes \( V_k \):

\[
\delta_{\text{opt}} := \arg \min_{\delta_k} V_k(\delta_k).
\]

(15)

Given system (3) for which it is assumed that there exists \( i \in \{1, \ldots, n\} \) for which holds that \( |e_k(i)| \geq \delta_{\text{max}} > \delta_k \) the switching length \( \delta_{\text{opt}} \) that minimizes the objective function \( V_k \) follows from the parameter update law

\[
\delta_{k+1} = \delta_k - \beta \left( \frac{\partial e_k^T}{\partial \delta_k} \frac{\partial e_k}{\partial \delta_k} \right)^{-1} \frac{\partial e_k^T}{\partial \delta_k} e_k,
\]

(16)

with \( 0 < \beta < 1 \). This follows from the following reasoning. System (3) is globally asymptotically stable under the conditions given in the previous section. As a result the sampled error data column \( e_k \in \mathbb{R}^n \) in the time interval \( t \in [t_1, t_2] \) satisfies the algebraic set of equations

\[
e_k = S_r + S d_k - S \varphi(e_k),
\]

(17)

with \( S, S_k \in \mathbb{R}^{n \times n} \) both Toeplitz matrices representing the sensitivity and complementary sensitivity functions, respectively, or

\[
S_r = \frac{C_{\phi}(j\omega) P(j\omega)}{1 + C_{\phi}(j\omega) P(j\omega)}.
\]

(18)
\[
S = \begin{bmatrix}
s_1 & 0 & \cdots & 0 \\
s_2 & s_1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
s_n & s_{n-1} & \cdots & s_1 \\
\end{bmatrix}
\text{and } S_c = \begin{bmatrix}
c_1 & 0 & \cdots & 0 \\
c_2 & c_1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
c_n & c_{n-1} & \cdots & c_1 \\
\end{bmatrix},
\]

(18)

with \(s_1, s_2, \ldots, s_n\) and \(c_1, c_2, \ldots, c_n\); \(s_1\) represents the first error sample \(e_k[1]\) to a unitary set-point impulse \(r[1]\) whereas \(c_1\) represents the first output sample \(r[1] - e_k[1]\), respectively, \(r \in \mathbb{R}^n\) are the (fixed) reference commands and \(d_k \in \mathbb{R}^n\) the corresponding input disturbances. The nonlinear part \(\varphi(e_k) \in \mathbb{R}^n\) can be written as

\[
\varphi(e_k) = \varphi_1(e_k)e_k + \delta_k \varphi_2(e_k),
\]

(19)

with \(\varphi_1(e_k) \in \mathbb{R}^{n \times n}\) a positive semi-definite diagonal matrix

\[
\varphi_1(e_k)[i, i] = \begin{cases} 
0, & \text{if } |e_k[i]| < \delta_k \\
\alpha, & \text{otherwise},
\end{cases}
\]

(20)

and \(\varphi_2(e_k) \in \mathbb{R}^n\) given by

\[
\varphi_2(e_k)[i] = \begin{cases} 
0, & \text{if } |e_k[i]| < \delta_k \\
-\operatorname{sign}(e_k[i]), & \text{otherwise}.
\end{cases}
\]

(21)

From (17), (19), (20), and (21) it follows that

\[
\frac{d}{dt}\frac{e_k}{\delta_k} = -S_c\varphi_1(e_k)\frac{e_k}{\delta_k} - S_c\varphi_2(e_k) = -(I + S_c\varphi_1(e_k))^{-1}S_c\varphi_2(e_k) = -p_k(e_k),
\]

(22)

with \(p_k(e_k) = (I + S_c\varphi_1(e_k))^{-1}S_c\varphi_2(e_k)\) and \(\|p_k(e_k)\| \leq \|p_k\|\) uniformly bounded; (22) exploits the fact that

\[
\frac{\partial \varphi_1(e_k)}{\partial e_k} = \rho(e - \delta_k) - \rho(e + \delta_k) \\
\frac{\partial \varphi_2(e_k)}{\partial e_k} = -\rho(e - \delta_k) - \rho(e + \delta_k),
\]

(23)

with \(\rho(c) = \varphi_1(c) + \varphi_2(c)\delta_k\),

\[
\phi_1(c) = \begin{cases} 
0, & \text{if } |c| < \delta_k \\
\alpha, & \text{otherwise},
\end{cases}
\]

(24)

\[
\phi_2(c) = \begin{cases} 
0, & \text{if } |c| < \delta_k \\
-\operatorname{sign}(c), & \text{otherwise},
\end{cases}
\]

and Dirac’s function \(\rho\) which yields

\[
\frac{\partial \varphi_1(e_k)}{\partial e_k}e_k + \delta_k \frac{\partial \varphi_2(e_k)}{\partial e_k} = 0.
\]

(25)

Substitution of (19) in (17) and using (22) gives

\[
e_k = (I + S_c\varphi_1(e_k))^{-1}(Sr + Sd_k) - \delta_k p_k(e_k),
\]

(26)

which given the update law in (16) results in

\[
e_{k+1} = e_k - \beta p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k + O_k(e_{k+1}, e_k),
\]

(27)

with

\[
O_k(e_{k+1}, e_k) = -\delta_k p_{k+1}(e_{k+1}) - p_k(e_k) \\
+ (I + S_c\varphi_1(e_k))^{-1}Sr \\
-(I + S_c\varphi_1(e_k))^{-1}Sd_k,
\]

(28)

and \(\|O_k(e_{k+1}, e_k)\| \leq \eta\) with the uniform bound \(\eta \geq 0\).

The Lyapunov function difference now becomes

\[
V_{k+1} - V_k = -\beta(1 - \beta)\frac{e_k^T}{\|p_k(e_k)\|^2}p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k \\
- \beta e_k^T p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k \\
+ 2\sigma_{k+1}^T(e_{k+1})\left(1 - \beta p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}\right)e_k \\
+ \sigma_{k+1}(e_{k+1})O_k(e_{k+1}, e_k),
\]

(29)

or using \(2(1 - \beta))\|p_k^T(e_k)\|/\|p_k^T(e_k)\|^2 \leq \alpha\eta^2(1 - \beta)^2 + (1/\alpha)\|p_k^T(e_k)\|^2/\|p_k^T(e_k)\|^2\) for any \(\alpha > 0\) satisfying \(\alpha\beta > 1\),

\[
V_{k+1} - V_k \leq -\beta(1 - \beta)\frac{e_k^T}{\|p_k(e_k)\|^2}p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k \\
+ \frac{1}{\alpha - 1}\|e_k\|^2 + \eta^2(1 + (1 - \beta)^2) \\
- \beta(1 - \beta)\frac{p_k(e_k)p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k,
\]

(30)

for

\[
\|p_k^T(e_k)\| \geq \eta\sqrt{\frac{\alpha + 2\alpha^2(1 - \beta)^2}{\alpha\beta - 1}}\|p_k\|^2.
\]

(31)

Having a positive definite Lyapunov function and a negative definite difference under the conditions in (31) all solutions \(e_k\) converge to a positively invariant set as \(k \to \infty\).

### 4.1 The Autonomous Case

For the autonomous case where \(O_k(e_{k+1}, e_k) = 0\) the difference in (29) reduces to:

\[
V_{k+1}(\delta_{k+1}) - V_k(\delta_k) = -2(1 - \beta)\frac{e_k^T}{\|p_k(e_k)\|^2}p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k,
\]

(32)

which is negative definite for \(0 < \beta < 2\). Hence global asymptotic stability results for the equilibrium \(e_{\text{opt}} = \lim_{k \to \infty} e_k\). This also follows from (27) which now becomes

\[
e_{k+1} = e_k - \beta p_k(e_k)\frac{p_k^T(e_k)}{\|p_k(e_k)\|^2}e_k = e_k - \beta p_k^T e_k.
\]

(33)

For the autonomous case the fact is used that \(p_k(e_k) = p_{k+1}(e_{k+1}) = p_k\) for all \(k \in \mathbb{N}^+\). Namely differentiation of (22) and using (25) gives

\[
\frac{d^2}{dt^2} \frac{e_k}{\delta_k} = -S_c\frac{\varphi_1(e_k)}{\delta_k} \frac{e_k}{\delta_k} - S_c\varphi_1(e_k)\frac{\varphi_2(e_k)}{\delta_k} - S_c\frac{\varphi_2(e_k)}{\delta_k} = -(I + S_c\varphi_1(e_k))^{-1}S_c\left(\frac{\varphi_1(e_k)}{\delta_k} + \frac{\varphi_2(e_k)}{\delta_k}\right)
\]

(34)

\[
= -(I + S_c\varphi_1(e_k))^{-1}S_c\frac{\varphi_1(e_k)}{\delta_k} \left(\frac{\partial}{\partial \delta_k} - \frac{e_k}{\delta_k}\right),
\]

for \(\delta_k \neq 0\). Using (17) and (22) then gives \(e_k = -\delta_k p_k(e_k) = \delta_k \partial e_k / \partial \delta_k\) which (for \(\delta_k \neq 0\)) implies \(\partial e_k / \partial \delta_k = 0\). So the Hessian in (34) is zero yielding the gradient in (22) to be constant. For integer \(m > 1\) (33) can thus be written as

\[
e_{k+m} - e_{k+m-1} = \left(I - \beta \frac{p_k^T}{\|p_k\|^2}\right)^{m-1}(e_{k+1} - e_k).
\]

(35)
which implies exponential convergence of the difference between two consecutive solutions to zero.

4.2 The Non-Autonomous Case with Reference Input

For the non-autonomous case with reference input different from the full non-autonomous case, however, is the fact that \( O_k(\mathbf{e}_{k+1}, \mathbf{e}_k) \) in (28) reduces to:

\[
O_k(\mathbf{e}_{k+1}, \mathbf{e}_k) = \left( p_{k+1}(\mathbf{e}_{k+1}) - p_k(\mathbf{e}_k) \right) \delta_{k+1} + \left( (I + S_{e_2} \varphi_1(\mathbf{e}_{k+1}))^{-1} - (I + S_{e_2} \varphi_1(\mathbf{e}_k))^{-1} \right) Sr. \tag{36}
\]

Without giving a formal proof it follows that if \( \mathbf{e}_k \to \mathbf{e}_{k+1}, \lim_{k \to \infty} \|O_k(\mathbf{e}_{k+1}, \mathbf{e}_k)\| \leq \eta \to 0 \). Consequently the invariant set to which all solutions converge decreases such that \( \mathbf{e}_k \) converges to \( \mathbf{e}_{\text{opt}} \). For this non-autonomous case, of noise. This is shown in the left part of Fig.6 by varying the level of output noise added to each error column \( \mathbf{e}_k \) and using a fixed convergence rate of \( \beta = 0.25 \); see also Eckhard et al. (2009). Under the noise levels \( \hat{n} \in \{0, 10^{-6}, 10^{-3}, 1\} \)

\[ \begin{align*}
\beta = 0.25, \\
\hat{n} = 10^{-6} \text{ nm}, \\
\hat{n} = 10^{-3} \text{ nm}, \\
\hat{n} = 1 \text{ nm}.
\end{align*} \]

Figure 6. Optimal switching length \( \delta_k \) for the non-autonomous case with reference input and noises.

5. WAFFER STAGE PERFORMANCE ASSESSMENT

Performance is assessed by optimizing the filtered servo error signals obtained from the \( x \)-direction of the short-stroke wafer stage. From the initial values \( \beta_1 \in \{0, 10\} \) nm, the right part of Fig.7 shows convergence to the optimal value of \( \delta_{\text{opt}} \approx 1.5 \) nm. Due to the presence

\[ \begin{align*}
\beta = 0.25, \\
\hat{n} = 10^{-6} \text{ nm}, \\
\hat{n} = 10^{-3} \text{ nm}, \\
\hat{n} = 1 \text{ nm}.
\end{align*} \]

Figure 7. Optimal switching length \( \delta_k \) for the \( x \)-direction of a short-stroke wafer stage (\( \beta = 0.25 \)).

of noises, the left part of Fig.7 hardly shows any decrease of the objective function \( V_k \). Basically this validates the model-based approach in obtaining the gradients; a gradient-approximation based approach through perturbed-parameter experiments as in Van der Meulen et al. (2008) generally fails. Small perturbations yield poor signal-to-noise ratios, large perturbations violate linearity.

The time-series measurements of the \( M_A \)-filtered error signals in \( x \)-direction before \( \delta_1 = 10 \) nm and after \( \delta_{\text{opt}} = 1.5 \) nm optimization are depicted in Fig.8. The \( M_A \)-filter is a discrete-time moving average operation defined by \( M_A(\mathbf{e}_k[i]) = (1/T) \sum_{j=-T/2}^{T/2-1} e_k[j], \) with \( T > 0 \) integer and even-valued. Fig.8 shows the results obtained

\[ \begin{align*}
\beta = 0.25, \\
\hat{n} = 10^{-6} \text{ nm}. \\
\hat{n} = 10^{-3} \text{ nm}, \\
\hat{n} = 1 \text{ nm}.
\end{align*} \]
from performing scanning motion at four corner locations on the wafer labelled with die1 to die4. At each die, a significant improvement is obtained through optimization. Not only in terms of settling behavior but also in the scanning interval of constant velocity; this is where the scaled acceleration profile (dashed curves) becomes zero. An extended error signal is used in the optimization which is composed of error signals obtained at these wafer locations. This gives the possibility to effectively deal with position-dependent dynamics and disturbances. Also indicated is the performance interval $t \in [0.03, 0.06]$ s from which the servo error signals are used to construct the objective function $V_k$. Additionally the signals used for optimization are constructed by subtracting the $M_A$-filtered parts from the original signals: $e_k[i] - M_A(e_k[i])$. In so doing weight is put on high-frequency stage behavior. With other linear filter operations, different focus can be given (in the optimization) to the trade-offs between disturbance rejection and noise sensitivity.

Where $M_A$-filtering generally refers to low-frequency behavior, high-frequency behavior is captured by $M_{SD}(e_k[i]) = \sqrt{(1/T) \sum^{T/2-1}_{j=0} (e_k[j] - M_A(e_k[j]))^2}$. The effect of this moving standard deviation operation is shown in Fig.9. Comparing Fig.8 and Fig.9 reveals that both the low-frequency as well as the high-frequency behavior of the wafer stage in $x$-direction improves significantly after optimization. Not only in terms of settling behavior related to the acceleration interval but also in terms of steady-state behavior in the scanning interval of constant velocity; the wafer scanning industry aims at nanometer improvements. The switching controller is shown to effectively deal with the position-dependent stage dynamics and the effect of (non-stationary) disturbances along the wafer.

6. CONCLUSIONS

For a switching controller, an extremum seeking procedure for the tuning of the switching length is presented. Given stability of the switched system, convergence of the optimization is shown regardless the constrained and non-convex nature of the problem. Depending on either the absence or presence of noises, convergence is to the global minimizer or to an invariant set whose size is related to the level of noise. For a wafer stage, optimized switching control demonstrates improved performances in terms of settling behavior and noise sensitivity in the steady-state scanning interval. Key to the success of the procedure is the combined model/data-based gradient approximation. The method is effectively used in the context of position-dependent dynamics and non-stationary disturbances.

REFERENCES


Figure 8. Time-series representation of the $M_A$-filtered error signals before ($\delta_1$) and after ($\delta_{20}$) optimization.

Figure 9. Time-series representation of the $M_{SD}$-filtered error signals before ($\delta_1$) and after ($\delta_{20}$) optimization.