Performance-improved design of N-PID controlled motion systems with applications to wafer stages
Heertjes, M.F.; Schuurbiers, X.G.P.; Nijmeijer, H.

Published in:
IEEE Transactions on Industrial Electronics

DOI:
10.1109/TIE.2009.2012420

Published: 01/01/2009

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Dec. 2018
Performance-Improved Design of N-PID Controlled Motion Systems With Applications to Wafer Stages

Marcel François Heertjes, Xander G. P. Schuurbiers, and Henk Nijmeijer, Fellow, IEEE

Abstract—A nonlinear filter design is proposed to improve nanopositioning servo performances in high-speed (and generally linear) motion systems. The design offers a means to adapt fundamental control design tradeoffs—like disturbance suppression versus noise sensitivity—which are otherwise fixed. Typically, performance-limiting oscillations in the feedback system that benefit from extra control are temporarily upscaled and subjected to nonlinear weighting. For sufficiently large amplitudes, this nonlinear filter operation induces extra controller gain. Oscillations that do not benefit from this extra control (typically because they represent noise contributions that should not be amplified) remain unscaled and, as such, do not induce extra controller gain. The combined usage of linear weighting filters with their exact inverses renders this part of the nonlinear filter design strictly performance based. The effective means to improve servo performance is demonstrated on a short-stroke wafer stage of an industrial wafer scanner. Since the nonlinear filter design is largely based on Lyapunov arguments, stability is guaranteed along the different design steps.

Index Terms—Disturbance rejection, loop shaping, Lyapunov stability, motion control systems, noise sensitivity, nonlinear proportional-integral-differential (N-PID) control.

I. INTRODUCTION

ALMOST a century after the pioneering work of authors like Minorsky [18], nowadays, motion control systems are still dominated by proportional–integral–differential (PID) control. Examples include component-mounting machines used to place electronic components on printed circuit boards, stages in the semiconductor industry needed to position an image on a silicon wafer [22], and the objective lens in a disk drive for reading, and writing information on a disk [20]. The combination, which is the contribution of this paper, is evaluated in the experiment. Section VI summarizes the main conclusions.

In this paper, N-PID control refers to the application of variable (nonlinear)-gain elements embedded in a nominal (linear) PID-based control design (see [5]). Aside from the variable-gain elements, performance and stability of the nonlinear feedback loop involve a proper choice of filtering. Here, special attention is given to the introduction of a weighting filter design. Different from a loop-shaping argument, the weighting filter design solely aims at performance. It is stability invariant by design, because the filter operations are used in series connection with their exact inverses. The idea is to temporarily lift performance-limiting parts from the error response and to decide—on the basis of the lifted signal amplitudes—whether to apply extra gain. Since the weighting filter operation is frequency dependent, the choice of variable gains also becomes frequency dependent. This provides extra design freedom in achieving servo performance. Small-amplitude (and low-frequency) disturbances can be discriminated from large-amplitude but high-frequency noise and effectively be dealt with; clearly, this kind of discrimination is impossible on the sole basis of signal amplitudes or using linear control.

By monitoring the signals at hand and acting accordingly, the N-PID control design combines three functions: monitoring by lifting the frequency content of interest temporarily from the error signals at hand, selection by (nonlinear) weighting of the lifted error signals through a frequency-dependent controller switch, and loop shaping to keep the closed-loop stability valid. Different from this loop shaping, which was recently introduced [13], the monitoring function is long used in the field of audio applications [20]. Also, the selection function is known [1], [6], [7]. The combination, which is the contribution of this paper, is unknown to the authors.

In the remainder of this paper, the following organization is employed. In Section II, the considered class of motion systems is briefly discussed along with the nonlinear control design methodology. This involves a closed-loop-shaping, variable-gain, and weighting filter design in the context of Lyapunov stability. In Section III, N-PID control is considered in view of the wafer scanning problem, where different expressions of error behavior along the wafer during the process of wafer scanning require different control actions so as to improve scanning performance. In Section IV, the design of weighting filters and their choice of parameters is discussed in the context of servo performance. In Section V, N-PID control is evaluated in the experiment. Section VI summarizes the main conclusions.
II. NONLINEAR MOTION CONTROL SYSTEMS

In mechatronics, motion control systems often refer to electromechanical systems performing controlled motion tasks (see, for example, [3], [17], and [21]). Being part of a generally more complex system, the plant to be controlled and the control design used to meet the specifications are key in obtaining (sub)system performances. In this regard, both the plant and the control design are considered in more detail.

A. Nominal Controlled System

A schematic representation of a nominal controlled motion system is shown in the block diagram of Fig. 1. Given a reference command $r$, a servo error signal $e$ is constructed using the relation $e = r - y$, where $y$ represents the output of plant $P$. This plant generally represents a double integrator extended with low-frequency stiffness and damping properties and high-frequency flexibilities. The error signal $e$ is fed into a stabilizing controller $C_{fb}$ which is usually a single-input single-output (SISO) PID-based controller. The controller $C_{fb}$ aims at disturbance rejection in view of set-point disturbances $r$’s and force disturbances $f$’s. To obtain sufficient tracking accuracy, an inertia-based (and generally also) SISO feedforward controller $C_{ff}$ is added.

B. Nonlinear Controlled System

Toward improved low-frequency tracking without the usual amplification of high-frequency noise, the linear feedback connection in Fig. 1 is given an extra nonlinear path (see Fig. 2). Extra controller gain is induced through the occurrence of performance-limiting oscillations contained in the error $e$. Because of the nonstationary and incidental nature in which these oscillations are assumed to occur, the amplification of noise is kept limited. In the nonlinear path, three functions, all being related to filter operations, are distinguished: monitoring, selection, and loop shaping.

A monitoring function is given to the filter structure $F_2$ which aims at lifting (performance-limiting) oscillations contained in $e$. A logical choice for such a filter structure would typically be a series connection of notch filter operations, or

$$F_2(s) = \mathcal{N}_1(s), \ldots, \mathcal{N}_n(s), \quad i \in \{2, \ldots, n\}$$  

(1)

where $\beta_{z,i} > \beta_{p,i} > 0$ and $\omega_{z,i} = \omega_{p,i} > 0$ representing the dimensionless damping coefficients and the zero and pole frequencies, respectively. By choosing $\beta_{z,i}$ and $\beta_{p,i}$ small, a narrow-band amplification is obtained about $\omega_{z,i} = \omega_{p,i}$. From a performance perspective, this filter is used to lift a specific frequency contribution from the error response and act on this contribution accordingly. It is to be understood that other choices for $F_2$ are possible and, sometimes, even preferable, given the specific nature of the problem at hand.

A selection function is given to the (nonlinear) weighting $\phi(\cdot)$ of the lifted error signals. This memoryless operation is given by a dead-zone nonlinearity, or

$$\phi(e_1) = \begin{cases} 0, & \text{if } |e_1| \leq \delta \\ 1 - \frac{\delta}{|e_1|}, & \text{if } |e_1| > \delta \end{cases}$$  

(3)

with $\delta$ being a dead-zone length. $\phi(e_1)$ is sector bounded in the sense that $0 \leq \phi(e_1)e_1^2 \leq e_1^2$. The idea of selection can be explained as follows (see Fig. 3). If the absolute value of the lifted error response $e_1$ exceeds the value of $\delta$, then the extra controller gain $\phi(e_1)$ is induced. Through feedback, this extra gain is used to reduce the error response. If $e_1$ in absolute value remains inside a bound defined by $\delta$, then no extra gain is induced, thereby avoiding the extra amplification of noise by feedback. The rather arbitrary choice for $\phi(\cdot)$ in (3) stems from preserving the nominal controller and related properties at small levels of $e_1$. Other choices are possible.

A loop-shaping function is given to the filter structure $F_1$ so as to improve the stability and robustness properties of the control design. This is discussed in [13] (see [15] for its parameter tuning), from which we adopt

$$F_1(s) = \alpha \mathcal{N}_1(s), \ldots, \mathcal{N}_n(s) \mathcal{L}_{p,1}(s), \ldots, \mathcal{L}_{p,m}(s)$$  

(4)

with $\alpha > 0$ being an extra gain ratio, $n, m \in \mathbb{N}^+$, and

$$\mathcal{L}_{p,m}(s) = \frac{\omega_{lp,m}}{s + \omega_{lp,m}}$$  

(5)

with $\omega_{lp,m} > 0$ being the breakpoint of a first-order low-pass filter.
C. Design for Stability

In designing a nonlinear feedback loop, stability of the closed-loop system generally requires careful analysis. For the system in Fig. 1 extended with the nonlinear path of Fig. 2, closed-loop stability is sufficiently guaranteed on the basis of the next result.

**Theorem:** Assume the strictly proper \( \mathcal{P} \) in Fig. 1 that is stabilized under bounded disturbances \( r \) and \( f \) by the strictly proper and stable \( \mathcal{C}_h \). Also assume \( \mathcal{F}_2 \) in Fig. 2 to be stable and proper. Then, any controller of the form \((1 + \mathcal{F}_1 \phi(\cdot) \mathcal{F}_1^{-1} \mathcal{F}_2) \mathcal{C}_h \), with \( 0 \leq \phi(\cdot) \leq 1 \), stabilizes \( \mathcal{P} \) if

\[
\Re \{ \mathcal{F}_1^{-1}(j\omega) \mathcal{F}_2(j\omega) \mathcal{S}_c(j\omega) \mathcal{F}_1(j\omega) \} \geq -1
\]

(6)

with

\[
\mathcal{S}_c(j\omega) = \frac{\mathcal{C}_h(j\omega) \mathcal{P}(j\omega)}{1 + \mathcal{C}_h(j\omega) \mathcal{P}(j\omega)}.
\]

(7)

**Proof:** Since \( \mathcal{P} \) is stabilized by \( \mathcal{C}_h \), which is also strictly proper, \( \mathcal{S}_c \) and, hence, \( \mathcal{F}_2 \mathcal{S}_c \) are both stable and strictly proper functions. The remainder of the proof now directly follows from the properties of \( G' = \mathcal{F}_2 \mathcal{S}_c \) evaluated through the circle criterion (see, for example, [13] and [23]).

D. Stability Invariance

The application of \( \mathcal{F}_1 \) with its exact inverse is invariant under the result in (6). This means that a strictly performance-based filter operation is obtained to effectively deal with performance-limiting oscillations. In practice, however, the construction of an inverse filter operation and its discrete-time implementation usually differs from its mathematical description. As a result, a residue remains to which sufficient robustness should be included in the design of the loop-shaping filter \( \mathcal{F}_2 \). Now, suppose

\[
\mathcal{F}_1 \left( G_1 + \Delta G_1 \right) = 1
\]

(8)

with \( \Delta G_1 \) being a mismatch in the inverse operation. The circle condition in (6) then becomes

\[
\Re \left\{ \mathcal{F}_2(j\omega) \mathcal{S}_c(j\omega) - \mathcal{F}_1(j\omega) \mathcal{F}_2(j\omega) \Delta G_1(j\omega) \mathcal{S}_c(j\omega) \right\} \geq -1.
\]

(9)

Thus, (9) can be reevaluated in view of this mismatch as long as it is described by a proper and stable transfer function. If this transfer function is unstable, it generally renders the nonlinear closed loop unstable. The design for performance related to the choice of \( \mathcal{F}_1 \) is addressed in more detail in Section IV. Prior to this design, however, the wafer scanning problem is considered in Section III. This serves as a benchmark toward demonstrating the effectiveness of the approach.

III. WAFER SCANNING PROBLEM

In the semiconductor industry, wafer scanners [19] are used to produce integrated circuits. The wafer scanning problem can be summarized as follows. On the one hand, the chip dimensions shrink, giving tighter specifications on servo position accuracy. On the other hand, wafer throughput increases with each new machine release, resulting in increased motion requirements. In terms of servo performance, both demands are generally conflicting.

A. Wafer Scanning Principle

The wafer scanning principle is best explained from an industrial example of a wafer scanner (see Fig. 4). Light from a laser passes a mask (or reticle) through a lens, which reduces the desired image by a factor of four, and onto a wafer.

The wafer is represented by a silicon disk of 300 mm in diameter. Both reticle and wafer are part of two separate controlled motion subsystems: the reticle stage and the wafer stage. Each employs a dual-stroke strategy. A long stroke is used for coarse positioning (micrometer accuracy), whereas a short stroke is used for fine positioning (nanometer accuracy). The short-stroke stages are represented by floating masses which are controlled in 6 DOF.

B. Wafer-Stage Dynamics and Control

To study the dynamics and control of a short-stroke wafer stage, we remark that explicit usage is made of the dynamics in the \( z \)-direction. This choice is rather arbitrary but basically captures features that are also found in the remaining directions. Furthermore, explicit reference is made to numerical values in the filter design so as to provide an intuitive feeling for the problem at hand. For the \( z \)-direction of the short-stroke wafer stage, the SISO wafer-stage plant is given by the following simplified fourth-order model:

\[
\mathcal{P}(s) = \frac{m_1 s^2 + b_{12} s + k_{12}}{m_1 m_2 s^4 + b_{12}(m_1 + m_2) s^3 + k_{12}(m_1 + m_2) s^2}
\]

(10)

with \( m_1 + m_2 \approx 22.5 \) kg being the wafer-stage mass (see the block diagram representation of Fig. 1 in Section II). According to the principle of measurement, which is on the motor side instead of the load side [21], a high-frequency resonance is included through the distinction between \( m_1 \approx 5 \) kg and \( m_2 \approx 17.5 \) kg, an interconnected stiffness coefficient...
The feedback controller $C_{fb}$ in Fig. 1 is modeled as a series connection of a PID controller which aims at disturbance rejection and robust stability, a second-order low-pass filter to avoid high-frequency noise amplification, and three notch filters designed to counteract higher order plant resonances. In transfer function notation, the simplified model reads

$$C_{fb}(s) = F_{pid}(s)F_{ip}(s)F_{n,1}(s)F_{n,2}(s)F_{n,3}(s)$$  \hspace{1cm} (11)

with

$$F_{pid}(s) = \frac{k_p}{\omega_d s}$$  \hspace{1cm} (12)

where $k_p = 6.9 \times 10^6 \text{ N} \cdot \text{m}^{-1}$ is a loop gain, $\omega_d \approx 3.8 \times 10^2 \text{ rad} \cdot \text{s}^{-1}$ is the cutoff frequency for a differential operation, and $\omega_l \approx 3.14 \times 10^2 \text{ rad} \cdot \text{s}^{-1}$ is the cutoff frequency for an integral operation

$$F_{ip}(s) = \frac{\omega_{lp}^2}{s^2 + 2\beta \omega_{lp} s + \omega_{lp}^2}$$  \hspace{1cm} (13)

where $\omega_{lp} \approx 3.04 \times 10^3 \text{ rad} \cdot \text{s}^{-1}$ is the cutoff frequency of a second-order low-pass filter, $\beta \approx 0.08$ is a dimensionless damping coefficient, and three notch filters having a second-order filter structure

$$F_{n,i}(s) = \left(\frac{\omega_{ni}}{\omega_{z,i}}\right)^2\frac{s^2 + 2\beta_{ni}\omega_{z,i}s + \omega_{z,i}^2}{s^2 + 2\beta_{pi}\omega_{pi}s + \omega_{pi}^2}.$$  \hspace{1cm} (14)

The parameter values are given in Table I. Note that, in the actual implementation, a discrete-time version of this controller is used on the basis of a sampling frequency of 5 kHz.

Both plant $P$ and controller $C_{fb}$ are characterized by the open-loop frequency response functions, as shown in Fig. 5. In Bode representation, it shows $O_L = C_{fb}P$ via the characteristics reconstructed from closed-loop measurement (black) along with the characteristics of a model (gray). Robust stability is sufficiently guaranteed with a controller bandwidth of $\approx 160$ Hz along with a phase margin of $20^\circ$ and a gain margin of $-4.5$ dB near 300 Hz. This industry is characterized by a high level of reproducibility in the mechanics of its motion systems. The kind of reproducibility that justifies a single controller tuning with fairly small robustness margins to guarantee robust stability over a series of machines.

### C. Closed-Loop Performance

Given the linear control design, closed-loop performance is assessed at five distinct positions on the wafer: at four wafer corner points and at the wafer center point [8]. At each position, labeled die one to die five, an identical scan in terms of its $x$- and $y$-acceleration set points is performed, whose result is shown in Fig. 6. Through coupling, the $z$-error response shows large deviations at the considered wafer positions, particularly within the indicated interval of constant velocity. This is the interval where scanning performance should be achieved. The differences in error response are emphasized by applying the weighting filter operation $F_1$ [see (1)], with $\beta_{z,1} = 1$, $\beta_{p,1} = 0.02$, and $\omega_{z,1} = \omega_{p,1} = 75\pi \text{ rad} \cdot \text{s}^{-1}$. From the figure, it is clear that the controlled wafer stage shows position-dependent (nonstationary) behavior. The error responses significantly differ over the wafer in both magnitude and frequency distribution. This provides the motivation for the application of the nonlinear feedback, as introduced in Section II.

### D. Motivation

The potential of N-PID control to achieve performance in the presence of position-dependent (and generally nonstationary) oscillations is shown in Fig. 7. Under linear feedback conditions

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\omega_{z,i}$ in rad s$^{-1}$</th>
<th>$\beta_{z,i}$</th>
<th>$\omega_{p,i}$ in rad s$^{-1}$</th>
<th>$\beta_{p,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$90\pi$</td>
<td>0.4</td>
<td>$96\pi$</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>$1456\pi$</td>
<td>1.53 $\times$ $10^{-6}$</td>
<td>$1624\pi$</td>
<td>0.88</td>
</tr>
<tr>
<td>3</td>
<td>$2207\pi$</td>
<td>0.19</td>
<td>$1459\pi$</td>
<td>0.18</td>
</tr>
</tbody>
</table>

$k_{12} = 7.5 \times 10^2 \text{ N} \cdot \text{m}^{-1}$, and a damping coefficient $b_{12} = 9 \times 10^1 \text{ N} \cdot \text{s} \cdot \text{m}^{-1}$.
(the extra nonlinear controller in Fig. 2 of Section II is not yet activated) and at two scan positions, namely, die one and die two, the upper part of the figure shows the result of the monitoring function, i.e., the filter operation $F_1$ on the error signal $e$ in the $z$-direction. Also shown is a scaled representation of the $y$-acceleration set point. The user-defined value of the dead-zone length $\delta = 40$ nm is indicated by means of the dotted lines. At die one, the resulting signal almost entirely falls within the indicated bounds, whereas at die two, it significantly exceeds these bounds. The effect of selection through the nonlinear filter operation $\phi(\cdot)$ [see (3)] is shown in the middle part of the figure. For die two, it can be seen that the sinusoidal-based behavior largely remains unaffected, whereas for die one, almost all oscillations are removed from the error response. Combined with the inverse filter operation $F_1^{-1}$—this is the lower part of the figure—the nonlinear filter operation gives a strong reduction of the original error signal $e$ (black versus gray) at die one. Contrarily, it roughly shows the same signal behavior at die two, which provides a means to obtain improved disturbance suppression at this die (extra gain is induced), while at the same time, it keeps the amplification of noise at die one small (no extra gain is induced). Having a nonlinear mechanism to deal with position-dependent (nonstationary) behavior, filtering and its choice of parameters require a proper design and tuning.

IV. DESIGN FOR PERFORMANCE

Given the N-PID control of Section II, design for performance refers to the design and tuning of two filter operations: the weighting filter operation $F_1$ and the nonlinear filter operation $\phi(\cdot)$, particularly the choice of dead-zone length $\delta$. This is because neither $F_1$ nor $\delta$ affects the stability result in (6) which renders both strictly performance-based variables.

A. Tuning the Nonlinear Filter

In tuning the nonlinear filter $\phi(\cdot)$ in (3) of Section II, the choice for the dead-zone length $\delta$ aims at keeping a small noise response. By choosing

$$\delta = \lim_{t \to \infty} \sup |e(t)|$$

(15)

for all $e$ in a performance-relevant time interval and by assuming that $F_1 = 1$, it is obtained that all error signals, within this interval, remain inside the dead-zone length. As a result, no gain $\phi(e)$ is induced, thus maintaining a low-gain noise response.

B. Design and Tuning of the Weighting Filters

Design and tuning of the weighting filter $F_1$ involves knowledge of the servo errors (and underlying disturbances) in terms of performance-limiting oscillations. To this end, a time–frequency analysis is conducted, whose result is shown in Fig. 8. For two die positions on the wafer (die four and die five), the error signals (solid curves) in the $x$-direction subjected to an equal acceleration set-point profile (dotted curves) are shown in the lower part of the figure. The main part of the figure depicts the magnitudes of the frequency contributions contained in these error signals during the considered time intervals. As such, a clear view is obtained on the occurrence (and the nonstationary) of performance-limiting oscillations during scanning. The dashed line at 190 Hz represents the critical frequency, below which extra gain potentially induces improved disturbance rejection; beyond this frequency, control merely contributes to an increased noise response. In view of this critical frequency, it is concluded that the temporary occurrence of large-amplitude oscillations near 130 Hz can be suppressed through extra control. This is the aim of shaping the filter characteristics of $F_1$ in a manner done in the left part of the figure. By choosing a single notch filter $F_1 = N_1$ [see (1)], performance-limiting oscillations occurring between 10 and 80 ms at die four are lifted from the error signal $e$ so as to induce extra gain $\phi$. At die five, where these oscillations beyond 30 ms are much less present, the amplification induced by $F_1$ is generally too small to lift the error signal past the dead-zone length. Hence, no gain $\phi$ is induced, and the low-gain noise response at die five is preserved. Contrarily, at die
four, this noise response is affected as soon as extra gain \( \phi \) is induced by the lifted error response. This is shown in Fig. 9 for die two. To minimize the increased noise response in both time and frequency domains, a second notch filter \( N_2 \) is used in \( F_1 \), whose result is shown in the right part of Fig. 9. By choosing the notch filter frequency of \( N_2 \) at exactly three times the first notch filter frequency of \( \approx 63.3 \) Hz, the noise response roughly remains equal to the case of low-gain feedback (see the left part of the figure), but low-frequency disturbance rejection shows major improvement (see Table II for the weighting filter parameters). It should be noted that the effectiveness of the notch filter tuning strongly depends on the disturbance situation at hand. For the wafer-stage example, the fixed setpoint conditions along with the high level of reproducibility of the stage dynamics render a fixed tuning sufficiently robust.

C. On the Occurrence of Higher Harmonics

The positive effect of \( F_1 = N_1 N_2 \) in Fig. 9 with the notch frequency of \( N_2 \) tuned at three times the first notch filter frequency is related to the occurrence of higher harmonics induced by the nonlinear operation \( \phi(\cdot) \). This can be explained as follows. Assume the input \( e_1(t) = \sin(2\pi f_1 t) \) (see also Fig. 2 of Section II). The output \( \phi(e_1) e_1 \) is then given by

\[
\phi(e_1) e_1 = \begin{cases} 0, & \text{if } |e_1| \leq \delta \\ g(t), & \text{if } |e_1| > \delta \end{cases}
\]

with

\[
g(t) = \sin(2\pi f_1 t) - \frac{4\delta}{\pi} \sum_{n=1}^{\infty} \sin \left( \frac{(2n-1)2\pi f_1 t}{2n-1} \right).
\]

An approximation of (17) with \( n \in \{1, 2\} \) and \( \delta = \pi/4 \) reads

\[
g(t) \approx -\frac{1}{3} \sin(6\pi f_1 t).
\]

Assume that the notch frequency of \( N_2 \) is tuned at \( 3f_1 \), the ratio between its damping coefficients is one-third of the ratio corresponding to \( N_1 \) [this is related to the amplitude in (18)] but still sufficiently large, and the notch width is chosen sufficiently small. Then, the third harmonic is largely removed from the additional error signal \( e_2 \). As a result, the error that is fed into the controller \( e_{\text{in}} \), i.e., \( e + e_2 \), is given by

\[
\mathcal{L}\{e + e_2\} \approx F_1^{-1} \mathcal{L}\{\sin(2\pi f_1 t)\} e
\]

so no significant error \( e_2 \) is induced such that the linear feedback path of Fig. 2 shows no significant contribution of higher harmonics. Without the second notch filter, the error becomes

\[
\mathcal{L}\{e + e_2\} \approx F_1^{-1} \mathcal{L}\{\sin(2\pi f_1 t)\} - F_2 F_1^{-1} \mathcal{L}\left\{ \frac{1}{3} \sin(6\pi f_1 t) \right\} e
\]

It is clear that the linear feedback path of Fig. 2 does contain third harmonics of significant amplitude. It is therefore concluded that \( e \), which is input to the low-gain feedback loop, may—in a limited sense—sustain higher harmonics if \( F_1 = N_1 N_2 \) but not in a way comparable to the case of \( F_1 = N_1 \). It is important to note that higher harmonics in (17) have smaller amplitudes. Moreover, the closed-loop transfer from the output of the nonlinearity to its input, which is based on a complementary sensitivity function, has low-pass characteristics. Thus, higher harmonics located sufficiently beyond the controller bandwidth (160 Hz) decrease substantially in amplitude. Both arguments validate the approximation in (18) and the choice to limit the weighting filter \( F_1 \) by only including a notch filter \( N_2 \) at the third harmonic. Having sufficient arguments toward the design and tuning of the weighting filters, a performance assessment is conducted on a wafer stage.

V. WAFER-STAGE PERFORMANCE ASSESSMENT

Prior to a wafer-stage performance assessment, two servo measures are discussed from the wafer scanning industry: the moving average filter operation and the moving standard-deviation filter operation (see also [15]).

A. Performance Measures

In quantifying wafer scanning performance, the wafer scanning industry expresses performance in terms of overlay and fading. Overlay [9] is a measure for position accuracy, hence the ability to perform a new scan atop a previous one. In terms of servo control performance, overlay can be assessed through

\[
\mathcal{L}\{e + e_2\} \approx \frac{1}{3} \sin(6\pi f_1 t).
\]

\[
(19)
\]

\[
(20)
\]
the moving average filter operation. For a time-sampled error signal $e(i)$, the moving average filter operation is defined as

$$M_a(e(i)) = \frac{1}{n} \sum_{j=i-n/2}^{i+n/2-1} e(j) \quad \forall i \in \mathbb{Z} \quad (21)$$

where $n \in \mathbb{N}^+$ represents a time frame which relates to the scanning diaphragm (the so-called slit) and the scanning speed; (21) is a low-pass filter operation on $e$. Fading relates to image quality, in particular linewidth uniformity, which results from the ability to focus light in the path from lens to wafer. Here, a moving standard-deviation filter operation is used, or

$$M_{sd}(e(i)) = \sqrt{\frac{1}{n} \sum_{j=i-n/2}^{i+n/2-1} (e(j) - M_a(i))^2} \quad \forall i \in \mathbb{Z} \quad (22)$$

hence an rms-based high-pass filter operation.

Both the low-pass characteristics of (21) and the high-pass characteristics of (22) are shown in Fig. 10. In magnitude representation, it can be seen that the $M_{sd}$-filter operation approximates the characteristics of a first-order high-pass filter. At high frequencies, however, it tends toward the rms value of $\sqrt{2}/2 \approx 0.7$. Contrarily, the $M_a$ filter approximates a first-order low-pass filter. Its zeros are expressed at multiples of $f_s/n$, with $f_s$ being the sampling frequency.

**B. Performance Evaluation**

For die two and die three of a scanning wafer stage in the $x$-direction (see also Fig. 9), Fig. 11 shows time-series measurements of the $M_a$- and $M_{sd}$-filtered error signals along a fixed acceleration set-point profile. The figure shows the low-gain case with $\phi = 0$, the high-gain case with $\phi = \alpha$, and the nonlinear case. It is clear that the case of nonlinear control gives rise to best disturbance suppression (see the upper part of the figure) not only during the acceleration phase but also during the interval of constant velocity, i.e., the interval with zero acceleration where scanning performance should be obtained.

In terms of noise response, the lower part of the figure shows that during the acceleration phase, the response tends toward the high-gain response. That is, a more preferable response is obtained than the high-amplitude response under low-gain feedback. Contrarily, during the interval of constant velocity, the response tends toward the low-gain response, thus giving favorable characteristics in comparison with the increased noise.
response under high-gain feedback. In fact, the response induced by N-PID control tends toward the best response at hand. This can sequentially be a high-gain response or a low-gain response (see also [10] for a similar claim on improved transient response).

The previous observation also follows from the settling-based representation in Fig. 12, where

$$
\begin{align*}
\mathcal{f}(\|\mathcal{M}_a(e(i))\|_\infty) &= \max \|\mathcal{M}_a(e(i))\|_\infty, \ldots, \|\mathcal{M}_a(e(k))\|_\infty]^T \\
\mathcal{f}(\|\mathcal{M}_{sd}(e(i))\|_\infty) &= \max \|\mathcal{M}_{sd}(e(i))\|_\infty, \ldots, \|\mathcal{M}_{sd}(e(k))\|_\infty]^T
\end{align*}
$$

(23)

with $i \leq k$, where $k$ indicates the last sample in the scanning interval of constant velocity. In this representation, the settling effects at the beginning of the constant velocity interval become clearly visible. Moreover, the nonlinear response shows favorable settling properties in the beginning of the scan as compared to the low-gain response. Further on up the scan, the nonlinear response matches with the high-gain response in terms of improved disturbance suppression (see the upper part of the figure) but with the low-gain response in terms of a small noise response (see the lower part of the figure). The nonlinear design has these favorable performance properties at different scan locations (or dies), each having different expressions in terms of error response. Compare, for example, the error amplitudes at die two and die three prior to scanning.

The ability of N-PID control to balance between disturbance rejection and noise sensitivity in favor of servo performance is considered in Fig. 13 for the $z$-direction (see also Figs. 6 and 7 of Section III). The example is of interest because of its clear distinction between a low-frequency $\mathcal{M}_a$ problem at die two and a high-frequency $\mathcal{M}_{sd}$ problem at die one. At die two, both nonlinear and high-gain linear feedback induce equally good servo performance in terms of keeping the peak values of the $\mathcal{M}_a$ error small. Significant extra low-frequency disturbance rejection is obtained in comparison with the low-gain response. The nature of the $\mathcal{M}_a$ problem is expressed by the fact that the $\mathcal{M}_{sd}$ responses (see the lower left part of the figure) do not show significant differences among the considered designs. Similarly, die one shows a clear $\mathcal{M}_{sd}$ problem. Both the nonlinear and the low-gain feedback show a favorable error response in comparison with the high-gain response, whereas the $\mathcal{M}_a$ responses do not differ significantly among the considered alternatives. It is therefore concluded that adapting toward the disturbance problem at hand (potentially) induces significant improvements in servo performance.

In terms of cumulative power spectral density analysis, Fig. 14 shows the results of the five considered dies. At each die, it can be seen that N-PID control induces an error response that tends toward the smallest rms values obtained with either low- or high-gain feedback. This makes N-PID control suitable in achieving performance under position-dependent conditions and provides a clear motivation for its application within the context of linear motion systems.

VI. C ONCLUSION

For linear motion systems, an N-PID control design is proposed which aims at improved disturbance rejection under equal noise response. What is important is the distinction between stability and performance. While nonlinear closed-loop stability is guaranteed in the presence of loop-shaping filters, performance is effectively dealt with by the introduction of a stability-invariant weighting filter connection. For a wafer stage of an industrial wafer scanner, the nonlinear control design relates to improved performance in the presence of position-dependent behavior. Measured at distinct (and sufficiently
distributed) positions along the wafer, improved low-frequency disturbance rejection is obtained without the necessary transmission of high-frequency noise, the kind of transmission that is inherently present under high-gain feedback. It is therefore concluded that N-PID control can significantly contribute to improved performance of linear motion systems.

References


