Synthesis of Variable Gain Integral Controllers for Linear Motion Systems

Bram Hunnekens, Nathan van de Wouw, Marcel Heerdtjes, and Henk Nijmeijer, Fellow, IEEE

Abstract—In this paper, we introduce a design framework for variable gain integral controllers with the aim to improve transient performance of linear motion systems. In particular, we focus on the well-known tradeoff introduced by integral action, which removes steady-state errors caused by constant external disturbances, but may deteriorate transient performance in terms of increased overshoot. We propose a class of variable gain integral controllers (VGICs), which limits the amount of integral action if the error exceeds a certain threshold, in order to balance this tradeoff in a more desirable manner. The resulting nonlinear controller consists of a loop-shaped linear controller with a variable gain element. The utilization of linear controllers as a basis for the control design appeals to the intuition of motion control engineers therewith enhancing the applicability. For the add-on part of the nonlinear variable gain part of the controller, we propose an optimization strategy, which enables performance-optimal tuning of the variable gain based on measurement data. The effectiveness of VGIC is demonstrated in practice on a high-precision industrial scanning motion system.

Index Terms—Motion control, transient performance, variable gain control, wafer scanners.

I. INTRODUCTION

TRANSIENT performance of a motion control system is often quantified in terms of rising time, overshoot, and settling times in the presence of stepwise setpoint changes and stepwise input disturbances. In general, a controller design will aim at obtaining a fast response with small overshoot, whereas achieving zero steady-state error, which to a large extent can be achieved by designing a controller with proper robustness margins. It is well known that constant external disturbances can be rejected by including integral action in the control design. However, it is also well known that integral action may increase the amount of overshoot in a transient response [1], [10], [12], [25]. To balance the tradeoff between steady-state position accuracy in the presence of constant disturbances and transient performance in terms of overshoot in a more desirable manner, nonlinear variable gain integral control (VGIC) is proposed in this paper.

It is well known that linear controllers are subject to inherent fundamental performance limitations [13], [32], such as the waterbed effect. The idea of variable gain control [also called nonlinear proportional-integral-derivative (NPID) control] for linear motion systems as a means to enhance the performance of the linear system has been used in [4], [3], [9], [14]–[16], [24], [26], [36], [37], and [39]. These papers all modulate the proportional or derivative gains or both in a smart way in order to improve the performance. In addition, in the context of the control of robotic manipulators, the NPID approaches have been used to improve performance [19], [21]. In contrast, this paper will focus on linear motion systems and on modifying the integral action in order to improve the performance. Furthermore, variable gain control for linear motion systems with the aim to balance the tradeoff between high-frequency noise sensitivity and low-frequency disturbance suppression in a more desirable manner, and hence focus on the steady-state performance of the system in the presence of time-varying perturbations was exploited in [14]–[16] and [36]. In this paper, we focus on enhancing transient performance by VGIC.

We propose to limit the integral action if the error exceeds a threshold, thereby limiting the amount of overshoot. Herewith, we introduce the concept of VGIC [18]. Several concepts for improving the transient performance of a control system by modifying the integral action have been proposed in literature.

One particular concept of interest is reset control, of which the so-called Clegg integrator introduced in [10] is an early example. The Clegg integrator resets the state of the integrator to zero upon zero error crossing. Generalizations include first-order reset elements, which reset the controller states, if certain conditions are satisfied [6], [28], [27], [30], [38]. This reset has the capability of improving the transient response of the system, as illustrated in examples given in the references. Note, however, that the reset controller drastically changes the dynamics of the closed-loop system by resetting instantaneously some of the controller states to zero. Such state resets may excite high-frequency resonances typically present in motion systems leading to undesired high-frequency transients. The VGIC proposed in this paper avoids such state resets and the potential problems it associates with. Other approaches modifying the integral action in order to improve the transient performance are, for example, given in [12], where switched integral controllers were used on a plant consisting of an integrator and [25], where switched integral controllers with resets and saturation were used on integrating plants. In [31], gain-modulated PID-controllers are used to improve performance of robotic applications represented by second-order transfer functions. In [33] and [34], the concept of conditional integrators has been introduced in a sliding mode control framework and a more general feedback control framework, which uses Lyapunov redesign and saturated high-gain feedback, to obtain regulation of nonlinear systems...
without chattering behavior. In [20], a continuous sliding mode controller with nonlinear gains has been introduced for nonlinear systems with relative degree one or two, in order to improve the transient performance. The VGIC approach that we present in this paper, focuses on linear plants of arbitrary (finite) order, opposed to the works in [20], [33] and [34], which consider nonlinear plants, and opposed to the works in [12], [25], and [31], which consider linear plants of specific structure. Moreover, we do not consider state-resets in the VGIC approach, opposed to reset controllers [6], [27], [28], [30], [38]. The approach taken in this paper allows for the formulation of easy-to-use graphical conditions to assess global asymptotic stability (GAS) of the nonlinear closed-loop system, on the basis of measured frequency response data. The latter feature greatly enhances its practical applicability. Moreover, because the VGIC is an add-on part to linear (performance-based) loop-shaped controllers, the nonlinear VGIC relates in a clear way to the underlying linear controller designs, further enhancing the practical applicability.

Several benefits of the VGIC concept in terms of practical applicability and stability analysis have been highlighted above. Nevertheless, the main focus of this paper is on achieving improved closed-loop performance, in which performance will be directly related to the transient time-domain error responses. This performance measure can subsequently be used in machine-in-the-loop optimization strategies, see [5] and [11], to optimize the performance by parametric tuning of the VGIC. In this paper, we consider a gradient-based quasi-Newton algorithm [29] in order to find the VGIC controller parameters that optimize the performance. In this approach, the gradients are obtained using a combined model/data-based method, where measured error signals are combined with model-data to obtain the performance measure and its gradients with respect to the controller parameters, using a single experiment. This approach for the determination of the gradients differs from perturbation-based methods using finite-difference methods to estimate gradients, which generally require (at least) two experiments.

The main contributions of this paper can be summarized as follows. First, we introduce the concept of a VGIC motion controller that guarantees robustness against constant disturbances by employing integral control and (at the same time) significantly improves transient performance in terms of overshoot compared with linear motion controllers. Second, a method for VGIC synthesis is proposed in order to facilitate performance-optimal self-tuning of the controller. Third, the proposed VGIC and its performance-optimal tunings are assessed using experiments on a high-precision industrial scanning motion system. This paper extends the preliminary results in [18], in particular by proposing a strategy for the performance-optimal tunings of the VGICs and applying the proposed strategy to an industrial motion system.

The remainder of this paper is organized as follows. In Section II, the VGIC strategy will be introduced. Moreover, stability properties induced by the proposed control scheme will be studied. The method for VGIC tuning for performance optimization will be discussed in Section III. The effectiveness of VGIC and its performance-optimal tuning are demonstrated with experimental results obtained from an industrial scanning motion system in Section IV, which will be followed by a discussion in Section V. Conclusions and recommendations will be presented in Section VI.

II. VARIABLE GAIN INTEGRAL CONTROL DESIGN

In this section, we will introduce the VGIC strategy. We start with a description of the motion control system and, subsequently, elaborate on the design philosophy behind the VGIC. An illustrative example will be given in Section II-B to illustrate the main idea of the nonlinear controller design. Stability analysis of the closed-loop dynamics will be considered in Section II-C.

A. Structure of the VGIC Control System

Consider the single-input-single-output closed-loop variable gain control scheme in Fig. 1, with plant $P$, nominal linear controller $C_{\text{nom}}$, which does not include integral action, constant reference $r$, constant input disturbance $d$, and measured output $y$. Additionally, the variable gain part of the controller consists of the variable gain element $\phi(\cdot)$ ($u = -\phi(e)$), depending on the error $e$, and a weak integrator described by the transfer function

$$C_I(s) = \frac{s + \omega_I}{s}$$

(1)

where $s \in \mathbb{C}$, with $\omega_I > 0$ the zero of the weak integrator. First, consider the situation in which $\phi(e)$ is a linear element and study the following two limits:

1) if $\phi(e) = 0$, we have a linear control scheme with linear controller $\mathcal{C} := C_{\text{nom}}$;
2) if $\phi(e) = e$, we also have a linear control scheme, but with linear controller $\mathcal{C} := C_{\text{nom}}C_I$.

In the first case with $\mathcal{C} = C_{\text{nom}}$, steady-state errors due to constant disturbances cannot be removed, but the amount of overshoot in a transient response will be limited in absence of integrator buffer buildup. In the second case with $\mathcal{C} = C_{\text{nom}}C_I$, zero steady-state error can be achieved, but the amount of overshoot in a transient response may increase due to said buffer buildup (note, in this respect, that also the nominal controller design is important, since this also influences the response of the system [23]). By properly designing the variable gain element $\phi(e)$, we can combine the best of both worlds and obtain both an improved transient response (small overshoot) and zero steady-state error. Consider hereto the following design for the function $\phi(e)$:

$$\phi(e) = \begin{cases} -\delta, & \text{if } e < -\delta \\ e, & \text{if } |e| \leq \delta \\ \delta, & \text{if } e > \delta \end{cases}$$

(2)
which is graphically depicted in Fig. 2. The controller is essentially based on a saturation nonlinearity, and limits the integral action when the error $|e|$ exceeds the saturation length $\delta$. Therefore, it limits the integrator buffer increase, hence the overshoot, whereas inducing full integral control when the error satisfies $|e| \leq \delta$, and at the same time removes steady-state errors. Because the amount of integral action depends through a variable gain $\varphi(e)$ on the error signal $e$, see Fig. 2, we call this controller a VGIC. Note that if $|e| \gg \delta$ (i.e., $\delta$ is very small compared with the error $e$), the overall controller tends to the linear controller $C_{\text{nom}}(s)$. If $|e| \leq \delta$, the linear controller $C_{\text{nom}}(s)C_l(s)$ is active. Moreover, note that the nonlinearity $\varphi(e)$ depends explicitly on the saturation length $\delta$, see Fig. 2, which is the key performance parameter that will be tuned for performance in this paper.

Remark 1: In this paper, the control action is adapted based on the magnitude of the error $e$ (Fig. 2). Other nonlinear control approaches exist, which also use information on the time-derivative of the error $\dot{e}$, which are denoted as phase-based approaches [3], [4].

B. Example

To illustrate the effectiveness and main idea behind the proposed control strategy, consider an elementary motion system depicted in Fig. 3, with $m = 0.01$ kg, $b = 0.03$ Ns/m, $k = 1$ N/m, and control input $F$. A nominal controller $C_{\text{nom}}$ without integrator and a controller $C_{\text{nom}}C_l$ with integrator have been designed using loop-shaping techniques to control the system to the reference $r = 1$; note that the input disturbance $d = 0$ in this case (Fig. 1). $C_{\text{nom}}(s) = k_p(s + \omega_c)/(s + \omega_p)$ is a lead-filter with zero $\omega_c = 10$ rad/s, pole $\omega_p = 100$ rad/s, $k_p = 100$, and the integrator as given by (1), with $\omega_l = 6$ rad/s. Simulations of the step response are depicted in Fig. 4. As can be concluded from the figure, the controller without integrator [thus only based on $C_{\text{nom}}(s)$] has the least amount of overshoot (no integrator buffer), but is not capable of achieving zero steady-state error. The controller with integrator ($C_{\text{nom}}(s)C_l(s)$) is capable of removing the steady-state error, but exhibits the negative effect of increased overshoot (integrator buffer reached 0.074 when crossing $y = 1$). Clearly, the variable gain controller (with $\delta = 0.1$) combines the small overshoot characteristics (integrator buffer reached 0.023 when crossing $y = 1$) with a zero steady-state error response.

C. Stability Analysis

In order to perform a stability analysis of the closed-loop dynamics induced by the VGIC scheme in Section II-A, we firstly observe from Fig. 1 that the system belongs to the class of Lur’e-type systems. These systems consist of a linear dynamical part

$$G_{eu}(s) = \frac{\omega_l}{s} \frac{\mathcal{P}(s)C_{\text{nom}}(s)}{1 + \mathcal{P}(s)C_{\text{nom}}(s)}$$

denoting the transfer function between input $u$ and output $e$, with a nonlinearity $\varphi(e)$ in the feedback loop [22, Ch. 7]. Note that $G_{eu}(s)$ has a simple pole at $s = 0$. A minimal realization of the closed-loop dynamics can be described in state-space form as follows:

$$\dot{x} = Ax + Bu + Br + Bd$$
$$e = Cx + Dr + Dd$$
$$u = -\varphi(e)$$

with state $x$ and $G_{eu}(s)$ in (3) satisfying $G_{eu}(s) := C(sI - A)^{-1}B$. Let us adopt the following two assumptions, which are both natural in a motion control setting.

Assumption 1: The complementary sensitivity, given by transfer function

$$T(s) = \frac{\mathcal{P}(s)C_{\text{nom}}(s)}{1 + \mathcal{P}(s)C_{\text{nom}}(s)}$$

has all poles in the open left-half plane (LHP).

Assumption 2: The complementary sensitivity $T(s)$ in (5) satisfies $T(0) \geq 0$.

Remark 3: Note that Assumptions 1 and 2 are very mild assumptions. The poles of $T(s)$ will lie in the open LHP.
through design of an asymptotically stabilizing linear controller $C_{\text{nom}}(s)$. The condition that $T(0) \geq 0$ is usually satisfied because the complementary sensitivity generally equals one for $\omega = 0$ by design.

Let $x^*$ be defined as the equilibrium point of (4) satisfying $e = 0$. Note that $x^*$ is the only equilibrium point satisfying $e = 0$, due to the fact that the minimal state-space realization (4) implies observability, i.e., the observability matrix has full rank such that the equations $e = 0$, $de/dt = 0, \ldots, (d^{n-1}e)/(dt^{n-1}) = 0$, exhibit a unique solution $x^*$, for $e = 0$.

The following theorem poses sufficient conditions under which GAS of the equilibrium $x^*$ can be guaranteed for the VGIC systems; and hence, under these conditions also the exact tracking of the reference $r$ is guaranteed. Note that other approaches (using, e.g., linear matrix inequalities) can also be used to assess stability of $x^*$ [35]. Here, we propose the condition put forward in Theorem 1 since (6) can easily be checked graphically based on measured plant dynamics, which, in turn, enhances the (industrial) applicability of this approach.

**Theorem 1:** Consider (4) with constant reference $r$ and constant disturbance $d$. If Assumptions 1 and 2 hold, and transfer function $G_{eu}(s)$ in (3) satisfies

$$\text{Re}(G_{eu}(j\omega)) \geq -1 \quad \forall \omega \in \mathbb{R} \quad (6)$$

then the VGIC renders the equilibrium point $x^*$ GAS.

**Proof:** The proof will consist of the following main steps:

1. **Step 1:** employ a coordinate transformation to shift the equilibrium point to the origin;
2. **Step 2:** employ a loop-transformation such that the transformed nonlinearity belongs to the $[0, \infty]$ sector;
3. **Step 3:** employ the positive-real (PR) lemma in order to show stability of the origin;
4. **Step 4:** employ a LaSalle-argument in order to show GAS of the origin.

Note that the essential difference with a standard circle-criterion proof lies in the fact that we cannot employ the strictly positive real (SPR) lemma due to the simple pole at $s = 0$ of $G_{eu}(s)$ in (3). Following similar lines as the proof of the circle criterion in [22, Sec. 7.1], and the proof in [2, Th. 3], the use of the PR lemma in combination with a LaSalle argument will still allow us to conclude GAS of the equilibrium point $x^*$.

**Step 1:** Note that integrator (1) has dynamics that can be described by

$$\dot{x}_I = \varphi(e) \quad (7)$$
$$y_I = \omega_t x_I + \varphi(e) \quad (8)$$

with state $x_I$, input $\varphi(e)$, and output $y_I$. As a result, we conclude that an equilibrium point $x^*$ of the closed-loop system [satisfying $\dot{x}_I = \varphi(e) = 0$] implies $e = 0$, since $\varphi(e) = 0$ only for $e = 0$. Because we consider constant (step) references in $r$ and constant disturbances $d$ in order to assess transient performance, we can employ a coordinate transformation $z = x - x^*$ to study stability of the equilibrium $x^*$ of the closed-loop system (4). The transformed dynamics can be written as

$$\dot{z} = Az + Bu \quad (9a)$$
$$e = Cz \quad (9b)$$
$$u = -\varphi(e) \quad (9c)$$

where we used the fact that $e = 0$ for $x = x^*$. The transfer function between input $u$ and output $e$ for (9) is again given by $G_{eu}(s) = C(sI - A)^{-1}B$. Note that the nonlinearity $\varphi(e)$, see Fig. 2, lies in the sector $\varphi(e) \in [0, 1]$ (i.e., $0 \leq \varphi(e)/e \leq 1$, $\forall e \neq 0$).

**Step 2:** Let us loop-transform (9) such that the transformed nonlinearity belongs to the $[0, \infty]$ sector [22], which is schematically illustrated in Fig. 5. The transfer function $\tilde{G}_{eu}(s)$ from $\tilde{e}$ to $u$, after loop transformation, is given by

$$\tilde{G}_{eu}(s) = G_{eu}(s) + 1 = C(sI - A)^{-1}B + 1 \quad (10)$$

which can be represented in state-space form, see (9), by the minimal state-space realization

$$\dot{z} = Az + Bu \quad (11a)$$
$$\tilde{e} = Cz + u \quad (11b)$$
$$u = -\tilde{\varphi}(\tilde{e}). \quad (11c)$$

**Step 3:** Note that:

1. the poles of $\tilde{G}_{eu}(s)$ are equal to the poles of $G_{eu}(s)$. Due to Assumption 1, the poles of $G_{eu}(s) = \omega_t T(s)/s$ and therefore also the poles of $\tilde{G}_{eu}(s)$, see (3) and (5), lie in the open LHP, except for the simple pole at $s = 0$;
2. due to Assumption 2, by combining (3), (5), and (10), we see that the residue of the simple pole $\text{res}(G_{eu}(s)) = \lim_{s \to 0} s\tilde{G}_{eu}(s) = \omega_t T(0) \geq 0$;
3. due to the condition in the theorem, $\text{Re}(G_{eu}(j\omega)) \geq -1$ $\forall \omega \in \mathbb{R}$. 

---

Fig. 5. Schematic representation of a Lur’e-type system with $\varphi \in [0, 1]$ (left-hand side) and a loop-transformed system with $\tilde{\varphi} \in [0, \infty]$ (right-hand side).
Consequently, the transfer function $\hat{G}_{eq}(s)$ is PR [22] (not SPR, due to the simple pole at $s = 0$). From [22, PR Lemma 6.2], it then follows that there exist matrices $P = P^T > 0$, $L$ and $W$ such that

$$PA + A^TP = -L^TL$$
$$PB = CT - L^T\sqrt{2}$$
$$W^TW = 2.$$  

Using the Lyapunov function candidate $V = 1/2z^TPz$, it then follows that along solutions of (11), the time-derivative $\dot{V}$ satisfies

$$\dot{V} = \frac{1}{2}z^T(PA + A^TP)z + z^TPBu$$
$$= -\frac{1}{2}z^TLz + z^TC^Tu - z^TL^T\sqrt{2}u$$

where we used the PR Lemma. Now let us add and subtract the term $u^2$, such that

$$\dot{V} = -\frac{1}{2}z^TLz + (Cz + u)^T(u - u^2) - z^TL^T\sqrt{2}u$$

$$= -\frac{1}{2}(Lz + \sqrt{2}u)^T(Lz + \sqrt{2}u) + \hat{e}^Tu$$
$$\leq \hat{e}^Tu$$

where it has been used that $\hat{e} = Cz + u$ (11b). Using the fact that the nonlinearity $\hat{\phi}(\hat{e}) \in [0, \infty]$, see Fig. 5, we have

$$\dot{V} \leq \hat{e}^Tu = -\hat{\phi}(\hat{e}) \leq 0.$$  

Inequality (19) implies already stability of $z = 0$, however, not yet asymptotic stability.

**Step 4:** To establish GAS of $z = 0$, we use a LaSalle-type of argument. Note that $\dot{V}$ can only be zero if $\hat{e} = 0$, because $\hat{\phi}(\hat{e})$ has only one unique zero at $\hat{e} = 0$ (Fig. 5). Hence, due to observability of the minimal state-space realization in (11), it then follows that $z = 0$ (i.e., $x = x^*$, with $e = 0$) is GAS.

**Remark 4:** The GAS of $x^*$ can intuitively be understood as follows: the saturation nonlinearity used in the VGIC always (i.e., also outside the band $[-\delta, \delta]$) applies a certain amount of integrating action (Fig. 2), such that the error is always forced to zero. An alternative design of $\phi(e)$ could be obtained by setting $\phi(e) = 0$ for all $|e| > \delta$, such that the integral action would be completely switched off for errors exceeding $\delta$. Although this may improve the transient response in terms of overshoot, a stability analysis shows that this would only yield local asymptotic stability results, see [18] for further details. Because this is undesirable nonrobust behavior in practice, we focus on the use of the VGIC with the saturation nonlinearity as in Fig. 2.

**Remark 5:** Note that in case that the saturation function in Fig. 1 would be placed behind the integrator, the output of the integrator would be limited to a certain value, $\bar{u}$ for example (as sometimes used in antiwindup schemes to avoid integrator windup due to actuator saturation). This would not allow for a GAS result as in Theorem 1, since in that case only disturbances smaller in amplitude than $\bar{u}$ can be compensated.

**Remark 6:** The integral action induced can be increased (decreased) by increasing (decreasing) the frequency of the zero $\omega_i$ of the weak integrator (1).

Using Theorem 1, stability can easily be assessed by checking (6), which can be done graphically using (e.g., measured) frequency response data. Note that stability does not depend on the saturation length $\delta$ of the VGIC, since the nonlinearity lies in the same sector $[0, 1]$ for any $\delta$. This makes $\delta$ a purely performance-based variable. Although the qualitative behavior of the VGIC relates to the underlying linear controllers, a dedicated performance-based tuning of the saturation length $\delta$ of the VGIC is far from trivial. To facilitate such a performance-based tuning, an optimization-based method will be discussed in Section III.

### III. VGIC TUNING FOR PERFORMANCE OPTIMIZATION

To facilitate an automated performance-based tuning procedure for the VGIC, we first have to quantify what we mean with performance in this context. Such performance quantification will be addressed in Section III-A. A gradient-based performance-optimization strategy will be discussed in Section III-B.

#### A. Performance Quantification

The goal of the VGIC, see Section II-B, is to balance the tradeoff between the extra amount of overshoot by including an integrator in the control design with the time needed for the removal of steady-state errors. The following cost function reflects our performance objective:

$$J(\delta) = c_1e(t^*(\delta), \delta) + c_2\int_{t_1}^{t_2} e^2(t, \delta)dt$$

where $t^*(\delta) = \arg\max_{t \in [t_1, t_2]} e(t, \delta)$ is the time of maximal overshoot [note that if the sign of the overshoot is unknown in the application, $t^*$ can alternatively be defined as $t^* := \arg\max_{t \in [t_1, t_2]} e^2(t, \delta)$], and $t_1$ and $t_2$ are the start- and end-time for weighting the integral of squared error (ISE), i.e., the second term in (20), see Fig. 6. The user-defined weighting factors $c_1$ and $c_2$ can be used to balance the importance of overshoot and ISE, and will depend on the specific situation and application at hand. These heuristic factors should be chosen in such a way that a desirable time-domain performance is obtained, which will be illustrated for an industrial application in Section IV.

The performance $J(\delta)$ in (20), and therewith the optimal $\delta$ that minimizes (20), depends on the particular signature of the
external disturbances $r$ and $d$ acting on the system (and on the initial conditions of the system). In many practical situations, such as repetitive tasks in industrial motion systems, the disturbances will be similar from experiment to experiment, which motivates the use of a single $\delta$ for subsequent experiments. The latter statement is further substantiated by the results in an industrial application in Section IV-A. If the disturbance situation does change significantly during operation, a continuous adaptation of $\delta$ will be needed.

B. Gradient-Based Optimization

For the tuning of the VGIC saturation length $\delta$, the optimization strategy considered here is a gradient-based strategy. Naturally, knowledge on the gradients $\partial J / \partial \delta$ is needed when employing such a strategy. Finite-difference approximations could be used for the approximation of the gradients. However, such an approach comes with the difficulty of choosing a suitable parametric spacing for the finite difference approximation: the spacing should be small enough in order to guarantee an accurate approximation of the gradients, but large enough to ensure that the (at least two) experiments employed in the finite difference approximation show significant differences not dominated by noise. Moreover, a finite-difference experiment implies doing multiple experiments in each iteration of the optimization algorithm for obtaining gradient information. To circumvent these difficulties, we will determine the gradients using a mixed model/data-based approach using, on the one hand, a model of the motion system $\mathcal{P}(s)$ and, on the other hand, measured error signals [15]. Note that it is reasonable to assume that accurate plant models for high-tech motion systems are readily available, and that the measured error signals are part of the iterative procedure, such that no additional experiments are required.

1) Computing $\partial J / \partial \delta$: To obtain the gradient $\partial J / \partial \delta$, we differentiate (20) with respect to $\delta$

$$\frac{\partial J}{\partial \delta} = c_1 \frac{\partial e}{\partial \delta} (r^*) + c_1 \frac{\partial e}{\partial \delta} (r^*) \frac{\partial t^*}{\partial \delta} + 2c_2 \int_{t_1}^{t_2} e(t, \delta) \frac{\partial e}{\partial \delta} dt. \quad (21)$$

The second term in (21) contains a term $\partial t^*/\partial \delta$ containing the effect that the time of maximal overshoot $t^*$ depends on the saturation length $\delta$. Although this term is in general unknown, it is multiplied with the term $\partial e/\partial t(r^*)$, which is zero as a result of the definition of maximal overshoot. Hence, the gradient $\partial J / \partial \delta$ can be written as

$$\frac{\partial J}{\partial \delta} = c_1 \frac{\partial e}{\partial \delta} (r^*) + 2c_2 \int_{t_1}^{t_2} e(t, \delta) \frac{\partial e}{\partial \delta} dt. \quad (22)$$

In (22), the error signal $e(t, \delta)$ follows from an experiment and contains the effect of all the (unknown) disturbances acting on the system. The gradient $\partial e/\partial \delta$ is determined as follows.

2) Computing $\partial e/\partial \delta$: In order to compute $\partial e/\partial \delta$, we first differentiate the state equations in (4) with respect to $\delta$. These are known as the sensitivity equations [22, Sec. 3.3]

$$\frac{d}{dt} \left( \frac{\partial x}{\partial \delta} \right) = A \frac{\partial x}{\partial \delta} - B \frac{\partial \varphi}{\partial \delta} (e) \frac{\partial e}{\partial \delta} - B \frac{\partial \varphi}{\partial \delta} (e) \frac{\partial e}{\partial \delta} = \left( A - BC \frac{\partial \varphi}{\partial e} (e) \right) \frac{\partial x}{\partial \delta} - B \frac{\partial \varphi}{\partial \delta} (e) \frac{\partial e}{\partial \delta} \quad (23a)$$

$$\frac{\partial e}{\partial \delta} = C \frac{\partial x}{\partial \delta} \quad (23b)$$

with the state $\partial x/\partial \delta$ and $e(t, \delta)$ an input following (for example) from measurements. Here, we used the fact that the external inputs $d$ and $r$ in (4) do not depend on $\delta$. Note that the controllers $C$ and $C_1$ are known exactly, so that if a plant model $\mathcal{P}$ is available, the matrices $A$, $B$, $C$ can be readily computed. The two terms $\partial \varphi/\partial e(e)$ and $\partial \varphi/\partial \delta(e)$ can be derived from the saturation function in (2) and are depicted in Fig. 7 [note that the derivatives at $|e| = \delta$ do not exist, but in practice, this does not matter since $|e| = \delta$ only occurs incidentally, such that it does not matter which value we use in (23)]. Moreover, experimentally, $|e| = \delta$ will practically never occur.

Numerically, simulation by straightforward forward integration of (23) with state $\partial x/\partial \delta$, with the error signal $e(t, \delta)$ coming from measured data (hence the name mixed model/data-based approach), provides us information on $\partial e/\partial \delta$, which is required to calculate the gradient of the performance map $\partial J/\partial \delta$ in (22).

3) Gradient-Based Optimization Strategy: The optimization method, we consider in this paper, is a second-order gradient-based quasi-Newton algorithm, see Fig. 8, which is used to minimize the cost-function $J$ in (20). Each iteration $k$, an experiment is performed to measure the error signal $e(\delta_{k+1})$, see Step 1 in Fig. 8. With the obtained error signal $e(\delta_{k+1})$, the performance $J$ in (20) can be computed in Step 2. If $J(\delta_{k+1})$ is smaller than $J(\delta_k)$, and $\delta_{k+1}$ lies within a predefined region $[\delta_{\min}, \delta_{\max}]$, see Step 3 in Fig. 8, the point $\delta_{k+1}$ is accepted as the new point. Otherwise, a line-search is performed in the direction $H_k^{-1}(\partial J/\partial \delta(\delta_k))$ until the new point satisfies the above conditions, see Step 4.
If a successful iteration is performed, the iteration index \( k \) is incrementally increased in Step 5 and we proceed to Step 6. The gradient \( \partial J / \partial \delta(\delta_k) \) is determined in Step 6 using (22), and the Hessian estimate \( H_k \) is obtained using gradient information from a Broyden–Fletcher–Goldfarb–Shanno update

\[
H_k = H_{k-1} + \frac{q^Tv}{\partial J / \partial \delta(\delta_k)} - \frac{H_{k-1}^T v}{v^T H_{k-1} v}
\]  

(24)

where \( q = \partial J / \partial \delta(\delta_k) - \partial J / \partial \delta(\delta_{k-1}) \), \( v = \delta_k - \delta_{k-1} \), and the initial Hessian estimate \( H_0 \) is an identity matrix (here \( H_0 = 1 \), because we optimize a scalar variable \( \delta \)), see [29] for more details. The following update is used in Step 7 in the quasi-Newton algorithm [29] to update the parameter \( \delta : \)

\[
\delta_{k+1} = \delta_k - H_k^{-1} \left( \frac{\partial J}{\partial \delta(\delta_k)} \right)^T
\]

(25)

where \( \delta_0 \) is the initial saturation length \( \delta \) of the VGIC.

The algorithm terminates if \( ||\delta_{k+1} - \delta_k|| \leq \epsilon \), with \( \epsilon > 0 \) a prespecified tolerance, or if a fixed maximum number of iterations (experiments) \( N \), with \( N > 0 \), has been performed.

Remark 7: Of course, also other optimization methods can be used to find the optimal parameters minimizing the performance indicator \( J \) in (20). In particular, any gradient-based optimization routine, such as, e.g., Gauss–Newton [15], can directly be employed in combination with the estimated gradients in (22).

In Section IV, the VGIC strategy introduced in Section II will be applied to a high-precision industrial motion stage of a wafer scanner. The controller synthesis method discussed in this section will be used to tune a performance-optimal VGIC.

IV. INDUSTRIAL APPLICATION: WAFER STAGE

The VGIC strategy and tuning procedure will be applied in this section to a high-precision industrial motion stage of a wafer scanner [8]. A wafer scanner is a system used to produce integrated circuits (ICs), see Fig. 9. Light, emitted from a laser, falls on a reticle (mounted on a reticle stage), which contains an image of the chip to be processed. The light is projected onto a wafer (mounted on a wafer stage) by passing through a lens system. The effect of this illumination, in combination with a photo-resist process results in the desired IC pattern being produced. The reticle-stage and wafer-stage both perform high-speed scanning motions in order to efficiently process the wafers.

At the same time at which the wafer is being exposed on the expose-side of the machine, the wafer profile of another wafer is being measured at the measurement-side of the machine. This preparatory measurement involves, for example, height-map measurements (levelling) of the wafer. From the height-map measurements set-point profiles are derived for the vertical directions which will be used at the exposure-side. The transition of the wafer from measurement-side to exposure-side is called the chuck-swap, see Fig. 10, which will be the part of the process that we will focus on in this section. During the chuck-swap, no wafers are being illuminated and, hence, no actual machine throughput is being generated. Therefore, it is important to perform this operation as time-efficient as possible.

Before discussing the experimental results in Sections IV-B and IV-C, we will verify the stability conditions, as posed in Theorem 1, in Section IV-A.

A. Stability Conditions

During the chuck-swap, we will focus on motion in the \( y \)-direction for which an experimentally identified thirty third-order plant-model \( P \) is available. This model will be used for the simulation of (23) to determine \( \partial e / \partial \delta \). The plant will be controlled by a nominal linear controller \( C_{\text{nom}} \) (without integral action), see Fig. 1, consisting of proportional-derivative controller, second-order low-pass filter, and three notches to cope with resonances due to structural flexibilities. This results in the open-loop frequency response function \( P(j\omega)C_{\text{nom}}(j\omega) \) as shown in Fig. 11. The linear controller with integral action is given by \( C_{\text{nom}}C_I \), where the zero of the weak integrator in (1) is set to \( \omega_1 = 5 \cdot 2\pi \text{ rad/s} \) (Fig. 11). Note that the same integral part \( C_I \) is used in the VGIC.
Because a stabilizing tracking controller $C_{\text{nom}}$ has been designed, see Fig. 11, it is easily verified that Assumptions 1 and 2 are satisfied, see also Remark 3. The frequency-domain condition $\text{Re}(G_{\text{eu}}(j\omega)) \geq -1, \forall \omega \in \mathbb{R}$, in (6) is also satisfied, as can be concluded from the Nyquist plot of $G_{\text{eu}}$ in Fig. 12. Therefore, all conditions of Theorem 1 are satisfied, such that the equilibrium point $x^*$ (for which $e = 0$) of the VGIC system (4) is GAS; hence, the steady-state error will converge to zero.

Note once more that the GAS property does not depend on the saturation length $\delta$. This implies that $\delta$ is a purely performance-based variable, and is fully stability-invariant under the condition posed in Theorem 1. We will exploit this fact in the subsequent sections, where we can tune $\delta$ for improved servo performance.

B. VGIC Experiments

From experimental data, it is known that during the chuck-swap, see Fig. 10, a movement in the $x$-direction disturbs the stage module in the $y$-direction (due to crosstalk) around $t = 2.35$ s (Fig. 13). This can be interpreted as a stepwise input disturbance $d$ (Fig. 1). Due to the error buildup, see Fig. 13, the buffer of the integrator of the controller $C_{\text{nom}}C_I$ in $y$-direction also builds up, resulting in overshoot at $t = 2.4$ s (see the red line in Fig. 13 regarding the response of the linear controller with integral action). However, note that the integrator induces removal of the steady-state error at the end of the time-interval. If only the nominal linear controller $C_{\text{nom}}$ without integral action is used, the overshoot decreases, but the steady-state error is no longer removed, as depicted by the blue line in Fig. 13.

In an attempt to shift the tradeoff between overshoot and the removal of steady-state error in a more desirable direction, we apply the VGIC strategy introduced in Section II. The fact that the chuck-swap movement is similar each time it is carried out, allows for the tuning of a single $\delta$ that performs well in subsequent chuck-swaps. To study the influence of the saturation length $\delta$, and to verify the controller tuning results of Section III, we perform experiments for 26 different settings for $\delta$ in the range $[0, 200] \, \mu$m. Note that, see Fig. 2, for $\delta = 0$, the error response will be induced by the linear controller $C_{\text{nom}}$, whereas for $\delta > \max(|e|)$, this response will be induced by the linear controller $C_{\text{nom}}C_I$. For increasing $\delta$, the amount of overshoot increases, but the steady-state error is removed faster. In Section IV-C, we will tune a performance-optimal VGIC using the optimization strategy discussed in Section III.

C. VGIC Tuning

In order to apply the gradient-based optimization strategy of Section III to tune the performance optimal VGIC, we use the performance measure $J(\delta)$ in (20). The interval $[t_1, t_2] = [2.35, 2.65] \, s$ is used for computing the ISE part in (20). Using $c_1 = 10^4$ and $c_2 = 10^9$ as weighting factors for the chuck-swap application strikes a balance between the amount of overshoot and removal of steady-state error in a desirable way. This will be illustrated by means of experimental time-domain error signals later on. Once these weighting factors have been selected, these can be used to automatically tune performance optimal controllers on other wafer scanners to facilitate dedicated machine-dependent performance optimization.

The performance $J$ as a function of the saturation length $\delta$ has been evaluated on the basis of 26 different experiments that have been performed on the wafer scanner (Fig. 13). The resulting performance curve is shown in Fig. 14 by the solid line. Note that in general such a performance curve is...
Fig. 14. Measured performance curve \( J(\delta) \) with the iteration history of the optimization.

Fig. 15. Iteration history of performance \( J \) and saturation length \( \delta \). Note that all iterations (experiments) are shown in this figure, also the iterations leading to a higher \( J \) (which are followed by a line search).

VGIC outperforms the linear controllers with integral action \( (\delta > 150 \, \mu m) \) and without integral action \( (\delta = 0) \) by 59\% and 67\%, respectively.

To compare the optimal VGIC to the linear controller limits with and without integral action, consider the corresponding measured error responses in Fig. 16. Because the buildup of integral action is limited outside \( |e| > \delta \), see Fig. 16, the VGIC has a similar amount of overshoot as the controller without integral action \( C_{\text{nom}} \), but removes the steady-state error in a similar way as the controller with integral action \( C_{\text{nom}}C_I \). This illustrates that the balance between overshoot and steady-state error can be shaped in a more desirable manner using the VGIC, improving the chuck-swap manoeuvre. Moreover, the choice for the weighting factors \( c_1 \) and \( c_2 \) has been such that we indeed obtain the desired balance between overshoot and steady-state error, considering the time-domain evaluation of the optimal VGIC in Fig. 16.

V. DISCUSSION

With a gradient-based optimization strategy, the optimal VGIC is found with an accuracy in the optimal saturation length \( \delta \) in the order of \( \approx 1 \, \mu m \). Note that if the same optimal controller with the same accuracy was to be found using brute-force experiments in the range \([0, 200] \, \mu m\), this would require \( \sim 200 \) experiments. This shows the effectiveness of the proposed optimization strategy. Approximately 10 experiments were needed to converge to the optimal controller.

The fact that only 10 experiments were needed to find the minimum of \( J \), but also the gradients \( \partial J / \partial \delta \), hinges on the fact that we use the mixed model/data-based approach to determine the gradients. If finite-difference approximations would have been used to determine the gradients, 10 additional experiments would have been required. Moreover, finite-difference approximations suffer from the difficulty in selecting suitable step sizes for the approximations. If these step sizes are too small, the difference between two experiments is dominated by noise, making the gradient approximation invaluable. On the other hand, if the steps are too large, the estimated gradient will be inaccurate. The mixed model/data-based approach does not suffer from these difficulties.
As the name indicates, the mixed model/data-based optimization approach requires models of the motion system under study (Section III-B). Next to the controller model \( C(s) \), which is exactly known, note that in the current high-precision motion application, accurate plant models \( P(s) \) can be obtained. If, however, a plant model is not easily assessable, one can choose to estimate the gradients using finite-difference approximations, at the expense of performing additional experiments. Another option, for optimization of a scalar parameter, would be to use a gradient-free method, for example, a sectioning procedure as in [7]. A third option could be to consider iterative feedback tuning [17], a data-based closed-loop parameter-value optimization, which does not use finite-difference approximations, but does require additional experiments.

In the current setting, only optimization of a single parameter (the saturation length \( \delta \)) has been considered. As an extension to the scalar optimization problem considered here, we can consider multiparameter optimization problems. As an example of a situation, where we might want to optimize multiple variables, one can think of optimizing the slope of the saturation function as well. The slope directly affects the amount of integral gain applied. Although the slope of the nonlinearity is not a pure performance-based variable (it is not stability invariant), the graphical frequency domain circle-criterion condition in (6) readily gives a bound for this slope that still guarantees stability. In such a two-parameter constrained optimization, the gradient-based approach from Section III-B can directly be applied. Note that in case more parameters are optimized, a brute-force experimental evaluation of the parameter space becomes cumbersome, and the relative computational and experimental efficiency of an iterative scheme becomes even more apparent.

As a last note, it is worthwhile to stress the intuitive design of the proposed nonlinear VGIC controller. The two linear controller limits, with and without integral controller, see Fig. 11, can be designed using well-known frequency-domain loop-shaping arguments, and form the basis for the nonlinear controller designs. Note that there are no restrictions to the order of the linear plant or nominal controller considered and that only output measurements are used. Moreover, GAS of the equilibrium point can be guaranteed by checking easy-to-check graphical conditions suited for measured frequency response data of the plant. As the name indicates, the mixed model/data-based optimization approach requires models of the motion system under study (Section III-B). Next to the controller model \( C(s) \), which is exactly known, note that in the current high-precision motion application, accurate plant models \( P(s) \) can be obtained. If, however, a plant model is not easily assessable, one can choose to estimate the gradients using finite-difference approximations, at the expense of performing additional experiments. Another option, for optimization of a scalar parameter, would be to use a gradient-free method, for example, a sectioning procedure as in [7]. A third option could be to consider iterative feedback tuning [17], a data-based closed-loop parameter-value optimization, which does not use finite-difference approximations, but does require additional experiments.

As a last note, it is worthwhile to stress the intuitive design of the proposed nonlinear VGIC controller. The two linear controller limits, with and without integral controller, see Fig. 11, can be designed using well-known frequency-domain loop-shaping arguments, and form the basis for the nonlinear controller designs. Note that there are no restrictions to the order of the linear plant or nominal controller considered and that only output measurements are used. Moreover, GAS of the equilibrium point can be guaranteed by checking easy-to-use graphical conditions using solely frequency-response data. These graphical conditions directly provide insight in the robustness of stability to uncertainties in the plant dynamics. The performance-based tuning of the saturation length \( \delta \) can be automated using the method discussed in Section III.

VI. CONCLUSION

In this paper, a nonlinear VGIC strategy for transient performance improvement of linear motion systems has been proposed. In particular, we focused on the tradeoff between overshoot and steady-state errors due to constant disturbances, and proposed a VGIC to balance this tradeoff in a more desirable manner. By limiting the amount of integral control action if the error is large, and only using full integral action if the error is small, the VGIC combines the desired effect of reduced overshoot with the rejection of constant disturbances. Sufficient conditions for the GAS of the setpoint of the resulting closed-loop system have been formulated in terms of easy-to-check graphical conditions suited for measured frequency response data of the plant.

A gradient-based quasi-Newton optimization method has been employed for the automated machine-in-the-loop tunings of the VGIC. The proposed VGIC strategy and its automated performance optimization have been implemented on a wafer scanner in order to improve the so-called chuck-swap manoeuvre. Significant improvements of the transient response have been obtained experimentally in this industrial case-study, compared with the linear controllers either with and without integrator.

REFERENCES
