Everything is under control
Everything is under control

Presented on March 6, 2020
at Eindhoven University of Technology
Introduction

‘Everything is under control’ refers to the sheer number of controls being used nowadays in high-precision systems. One example is stage systems in wafer scanners (the tool used to produce microchips), which have hundreds of control loops for their motion control [1]. This number is expected to increase as more and more (sub)systems are connected by control in a system-of-systems manner [2]. ‘Everything is under control’ also refers to the task that control engineering has in meeting system specifications and performances. Though control technology is often not visible from the outside as it is hidden within algorithms (hence the hidden technology [3]), high-precision systems can by no means meet specifications without this technology [4]. Lastly, ‘everything is under control’ imparts a feeling of ease, often against better judgement. Despite the impressive progress in mechatronics and hardware design, and regardless of the promotion of modern control technologies [27], the high-precision industry is still dominated by classical control concepts.

In the inaugural lecture, the state-of-the-art in industrial motion control will first be highlighted using an example from a stage control system. This section aims to provide a view of where we stand in terms of motion control performance in industry and what room there is for the progression of motion control technologies. Secondly, two control scenarios will be discussed that aim for reflection on what future motion control may look like: (machine) learning in feedforward control and hybrid integrator-gain systems in feedback control. Both scenarios have been chosen in line with worldwide trends regarding data-based control [5] and nonlinear control for linear systems [6,7] and illustrate the context of industrial nonlinear control for high-precision systems, the area in which I am a part-time professor. Lastly, some thoughts are shared on the interplay between industry and the university, especially regarding research and education.
Motion control technology - a stage example

In motion control, one basically distinguishes between two problems: the servo/tracking problem and the regulator problem [8]. In the servo/tracking problem, we want the output of the controlled system to stay as close as possible to a reference trajectory. Think of the commanded meander pattern used in wafer scanners to enable exposure along the wafer [1]. In the regulator (or disturbance rejection) problem, the output of the controlled system subject to disturbances should be as close as possible to a predetermined value. The servo/tracking problem is often tackled by feedforward control, whereas the regulator problem is dealt with by feedback control.

For many motion control systems, a significant source of servo error stems from the servo problem, i.e. the setpoint profiles that we essentially provide ourselves. As a testament to this claim, consider Figure 1, which depicts an industrial wafer stage system on its left. This shows the short-stroke stage during exposure of one of its many fields on the wafer.

![Figure 1. (a) Wafer stage system. (b) Servo tracking performance of a stage system.](image)

On the righthand side, the figure shows the pointwise mean (in red) of eight measured servo error signals which resulted from applying eight identical setpoint profiles, as indicated by the scaled, dashed curve. The figure also shows the spread of the eight signals in gray. A distinction is made between the time interval of scanning, i.e. where exposure takes place under constant velocity motion, and the time interval at which the system accelerates or decelerates in order to prepare (among other things) for exposure of the next field. As the data underlying the figure were obtained from a recently-produced stage system, one may expect state-of-the-art industrial motion control and the resulting tracking performance.

**Observation 1: error signals inside the scanning interval are smaller in magnitude than error signals outside of the scanning interval.**

This observation is in line with the fact that (servo) performance in wafer scanners is mainly relevant when exposure takes place, i.e. in the scanning interval. A closer look, however, reveals that the error at the beginning of the scanning interval is significantly larger than the error closer to the end of the scanning interval, hence the presence of a settling effect.

**Observation 2: error signals both inside and outside of the scanning interval show reproducible behavior.**

At the larger error levels, the spread among the different error signals becomes relatively small (the signals look alike), indicating the presence of reproducible behavior. There is a strong correlation between the setpoint (acceleration) profile and the system’s response. As the setpoint is known beforehand and as we have a good understanding of the system through identification [9], we should be able to avoid this setpoint-induced part from occurring in the system’s response.

For wafer scanners, the importance of both observations stems from its relationship with key performance indicators like machine overlay and throughput. Consider, for example, the time needed to expose one field, which in Figure 1 equals 0.0476 seconds. Exposing 105 fields per wafer would thus lead to an upper bound on throughput of about 720 wafers per hour, which exceeds the scanner’s specifications by a rough factor of three. Imagine the increase in throughput if one could simply expose while still accelerating (or decelerating). From a control point of view, however, one would experience a large hinderance due to the increase of error magnitudes outside of the scanning interval, as previously discussed.

But what if we could avoid the setpoint-induced error responses, leaving the magnitudes both inside and outside of the scanning interval equally small?

For industrial high-precision mechatronics, the stage example considered does not stand on its own. In fact, its system performance extrapolates well to many other application areas for the obvious reason of similarity in both system and control.
design. Be it a CD player, wafer stage, component mounter or photocopier, all are represented by double integrator-based systems according to Newton’s second law and all use PID-based control designs [25].

So, let’s continue with the stage example by considering the regulator problem as well. Take Figure 2, which shows the frequency content of the previously-discussed error responses through cumulative power spectral density analysis.

The indicated value of 2.8 nanometers represents the root-mean-square (RMS) value as a measure of the power of the servo error responses in the scanning interval. In addition, the value of 2.1 nanometers is associated with the frequency denoted by $f_{\text{bw}_3\text{dB}}$. This frequency can be interpreted as the maximum frequency below which the closed-loop system takes sole advantage from feedback control, i.e., the point at which disturbance suppression is obtained without inducing any amplifications. The figure shows that 76% of the frequency content of the error responses that could benefit from feedback is not further suppressed by state-of-the-art industrial control. Why does feedback control (particularly linear PID-based control) fail to further suppress these mostly non-reproducible contributions? Suppose we had controls at our disposal that could (a) avoid the reproducible parts from occurring in the error response and (b) further suppress the frequency content below $f_{\text{bw}_3\text{dB}}$. What would these controls bring to future high-precision systems?

Take a look at Figure 3, which shows a sample of 30 recently-produced stage systems. In the sample, three different types are considered. Within each type, ten different stage systems have been randomly selected. The figure shows the estimated benefit of the mean error responses in RMS values by (a) removing the reproducible part and (b) removing the frequency content below $f_{\text{bw}_3\text{dB}}$. This shows room for improvement by factors of $2.4 \times 3.1 = 7.4$ (type 1), $2.9 \times 1.9 = 5.5$ (type 2) and $4.2 \times 3.8 = 16.0$ (type 3), respectively. Also, the response outside of the scanning interval no longer significantly differs from the response inside the scanning interval if one treats the reproducible part with appropriate care.

In the remainder of this lecture, two scenarios will be outlined: (a) learning to deal with reproducible errors in feedforward control and (b) hybrid integrator-gain systems in feedback control to deal with non-reproducible errors. Both scenarios help to shape a vision on how to progress with motion control technology.
Learning in feedforward control

Feedforward control is key to achieving the tracking performance of motion control systems [10]. To visualize this, consider a floating mass that one wants to move according to a setpoint trajectory. Given Newton’s second law of classical mechanics, the force that you need would be the mass itself multiplied by the setpoint acceleration. For a 20 kg stage system with a setpoint profile that has a maximum absolute acceleration level of 50 meters per square second, this implies a feedforward force of $20 \times 50 = 1000$ newtons. Suppose we used feedback to steer the motion of the stage toward the desired setpoint trajectory. Through the application of Hooke’s law, the required control force equals the controller gain multiplied by the tracking error. In generating the 1000 newtons with a static controller gain of $2 \cdot 10^7$ newtons per meter, servo errors are induced in the order of $1000 / 2 \cdot 10^7 = 0.056$ micrometers, which is a rough but not unrealistic estimate that largely exceeds specifications, hence the need for feedforward control. A reasonable estimate of the actual feedback force applied in a stage system follows from the earlier RMS error value of 2.8 nanometers, specifically $2 \cdot 10^7 \times 2.8 \cdot 10^{-9} = 0.056$ newtons. Feedforward does 99.99% of the job when it comes to control forces.

Feedforward control benefits from an accurate description of what is often called the inverse of the plant. This is the dynamic relationship between setpoint trajectory as an input and the resulting feedforward control force as an output.

One powerful data-based control technique is iterative learning control (ILC) [13], [24], which is capable of removing undesired reproducible behavior from the error response. Can ILC be used to predict (and subsequently compensate for) the mean error response from the earlier stage example? The answer is affirmative, as will be shown later. So why is ILC not widely embraced in the wafer scanner industry?

Essentially, the learned feedforward force sequence $y(k)$ with time samples $k$ is given by the weighted sum of the setpoint trajectory sequence $x(k)$ and its $n$ discrete-time derivatives. Learning involves finding the weights $\rho_0, ..., \rho_n$ through data-based optimization. So how does a feedforward force obtained from a learned filter compare to the force signal obtained with ILC?

In Figure 5, the learned forces on the left side and the resulting error response on the right side are shown for both ILC and FIR feedforward.
As stated before, industrial feedforward controllers generally match the inverse of the plant, i.e. $C_{ff} \rightarrow P^{-1}$. However, what we define as the plant depends on our choice of system boundaries. In applying the setpoint trajectory $r$, systems other than plant $P$ may be excited too, such as the indicated plant $X$ (possibly controlled by $C_{xx}$). As a result of these excitations, the setpoint-induced disturbances $F_{d,d}$ may in turn affect the system output $y$. These disturbances may contain frequency content related to dynamics involved in the (controlled) plant $X$ and are most likely not compensated by any feedforward design based solely on inverting plant $P$, hence the earlier inability of FIR feedforward control to predict the error responses. In the stage example, a specific structural mode occurring in the support frame appears to be responsible for the observed error response. How can this be learned and how can we include this essential domain knowledge in the feedforward design?

As an example, consider the nonlinear autoregressive exogenous (NARX) neural network representation in Figure 7.

The learned feedforward force sequence $y(k)$ with time samples $k$ is again given by the weighted sum of the setpoint trajectory sequence $x(k)$ and its time derivatives. Additionally, a second-order damped oscillator is added as basis function $f(z)$. Learning involves finding the weights of all neurons $n_i$ with $i \in \{1, \ldots, 20\}$ and $j \in \{1,2,3\}$ through data-based optimization.

The results of such an approach are shown in Figure 8.
Hybrid integrator-gain systems in feedback

Hybrid integrator-gain systems are essentially nonlinear control blocks. In the control of motion systems, nonlinear blocks are traditionally used to linearize the feedback loop, such as in the presence of plant nonlinearities like actuator saturation or friction. In my research, nonlinearity is also used to intentionally de-linearize the feedback loop, i.e. to control a linear plant with a nonlinear feedback controller in order to better deal with design trade-offs [6]. The motivation stems from the fact that most linear control designs suffer from inherent design limitations, a well-known example being the waterbed effect that imposes limitations on achievable control bandwidths and specifications [16]. The waterbed effect was initially described by Hendrik Bode [17]; simply stated, the benefit one has from linear feedback control in certain frequency intervals comes with an equal (or higher) cost in other frequency intervals. In this sense, control engineers typically shift problems.

At the core of these limitations lies a causality argument: feedback is reactive, so there must be an error before corrective actions are taken [3]. This reactiveness comes with time aspects like phase lag and time delay, which appear to be the true performance limiters. To illustrate this, consider Figure 9, which shows the response of a linear integrator (gray curve) to a multi-sine input (dashed black curve).

<table>
<thead>
<tr>
<th>time in seconds</th>
<th>0</th>
<th>0.0378</th>
<th>0.0854</th>
<th>0.1</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>force in Newton</td>
<td>0</td>
<td>0.0178</td>
<td>0.0854</td>
<td>0.1</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Note that the inclusion of the basis function $f(z)$ (i.e. the blue curve) is crucial. The predicted error response with this learned feedforward approach matches well with the measured error response and leads to the following conclusions. For industrial stage control, the reproducible part of the error response (a) highly correlates to the applied setpoint trajectories, (b) is well-predicted using simple plant/controller models and (c) can thus be largely prevented by appropriate feedforward compensation.

Obviously, this lecture would be of limited value if the ultimate solution could simply be found in (machine) learning feedforward control. However, as the example shows, learning control in general and machine learning in particular can be instrumental in increasing the physical understanding needed to design appropriate feedforward controls. The fact that (machine) learning mainly involves increasing our own understanding should come as no surprise: deep domain knowledge is often key to providing effective control solutions. For motion control, the challenge thus lies in developing generic solution concepts that still offer room for application specifics. On the detailed level, many hurdles clearly need to be dealt with involving point-of-control versus point-of-interest, spatial and temporal dependencies and robustness to model uncertainties, to name a few. However, as the stage example shows, there seems to be no technical argument supporting the presence of setpoint-induced reproducible parts in the error response of high-precision motion systems other than inappropriate feedforward controls. In the remainder of the lecture, let’s take a look at what can be done about the non-reproducible parts via (nonlinear) feedback control.
For a given initial condition, one can observe that the output of the linear integrator – a basic block used in feedback control – roughly mimics the input signal but is shifted (or delayed) in time. As a result of this shift, known as phase shift, the sign of the integrator output signal is only equal to the sign of the input signal half of the time. The key point is as follows: if one uses this output signal as a corrective action to the input (error) signal, this corrective action would point in the wrong direction half of the time. The figure also shows the output of a hybrid integrator-gain system, abbreviated to HIGS (solid red curve). The adjective ‘hybrid’ refers to the alternating switching that occurs between integrator mode and gain mode. Note that the output of the HIGS mimics the input signal, seemingly with no delay. This raises the question of whether HIGS-based control designs can lift linear design limitations and thus offer solutions that enable better design trade-offs.

Consider a specific result from linear control theory which essentially states that any stable linear time-invariant unity feedback system with a real-valued open-loop right-half-plane pole must include overshoot in its step response [18]. Under such conditions, a HIGS-based controller enables a step response with no overshoot and can thus beat any linear controller. Why is this relevant to the field of motion control? Recall from the earlier stage example that a large portion of the frequency content of the error response which could benefit from feedback control remains unsuppressed under linear PID-based control. If design limitations experienced under such controls keep us from further suppressing non-reproducible error contributions, we should challenge the control designs themselves as these may unnecessarily prevent us from achieving enhanced performance, therefore posing less tight specifications. Since linear PID-based control is (not without reason) the industrial standard for high-precision motion systems, HIGS-based control which is not necessarily bound to similar limitations may become particularly useful in progressing the field of industrial motion control. However, one difficulty emerges. As motion control design is mainly done in the frequency domain, a frequency-domain framework is desirable for nonlinear motion control in general [26] and therefore for HIGS-based motion control in particular.

Frequency-domain representations of nonlinear control systems generally exist through approximation. For HIGS-based control design, practically-relevant approximations can be obtained (for example) by describing function analysis [19]. This was originally developed by Krylov and Bogoliubov in the 1930s. Consider the describing function description of a HIGS-based integrator in comparison to a linear integrator, such as the one shown in Figure 10.

While the linear integrator (indicated by the dashed gray curve) has a minus 90 degree phase lag, the HIGS-based integrator can be tuned close to a minus 38 degree phase lag, depending on the choice of extra tuning parameter $\omega_h$. In other words, a significant phase advantage is obtained under equal magnitude characteristics. Such a phase advantage opens up many opportunities for motion control in terms of increased gains, bandwidths and performance.

Consider the stage example once more but now think of the possibilities of a HIGS-based PID control design in dealing with the non-reproducible part of the error response. For a brief sketch of how such a design would look, first consider a PID controller with output signal $y(t) = y_p(t) + y_i(t) + y_d(t)$. The output signal is composed of a proportional part $y_p(t)$, an integral part $y_i(t)$ and a derivative part $y_d(t)$ according to

$$\begin{cases}
y_p(t) = k_p u(t) \\
y_i(t) = k_i \int_0^t u(\tau)d\tau \\
y_d(t) = \frac{k_d}{\omega_d} \frac{du(t)}{dt}
\end{cases}$$
with proportional gain $k_p$, integrator frequency $\omega_i$, derivative frequency $\omega_d$, input signal $u(t)$ and appropriate initial conditions at time $t = 0$. Given the same input $u(t)$, the HIGS-based PID controller output $y(t) = y_p(t) + y_i(t) + y_d(t)$ is typically obtained by replacing the linear integrator output signal $y_i(t)$ with a HIGS integrator output signal $y^h_i(t)$. The latter follows the application of a matched proportional-integral operation on the signal $\nu(t)$, or

$$y^h_i(t) = \frac{k_p \omega_i^{-1}}{1 + 4j/\pi} \nu(t) + k_p \omega_i \int_0^t \nu(t) \, dt,$$

where $\nu(t)$ is the output of the HIGS operation on $u(t)$, which is defined as

$$\begin{align*}
\frac{dx(t)}{dt} &= \omega_h u(t), \quad \text{if } (u(t),x(t),\nu(t)) \in F_{\text{int}}, \\
\frac{du(t)}{dt} &= x(t), \quad \text{if } (u(t),x(t),\nu(t)) \in F_{\text{gain}}.
\end{align*}$$

$F_{\text{int}}$ refers to the region in which the HIGS operates in integrator mode, while $F_{\text{gain}}$ refers to the region in which the HIGS operates in gain mode. A HIGS-based PID control design now refers to a combination of a PID controller, low-pass filter and notch filters in which several integrators are replaced by HIGS integrators in order to exploit the HIGS’ phase advantages.

In assessing HIGS-based PID control, an identical linear PID-based control structure - but with linear instead of HIGS integrators - is used for comparison purposes.

Both controllers are subject to equal design constraints and both are tuned using state-of-the-art auto-tuners. The result is shown at the top of Figure 11 in terms of magnitude sensitivity functions.

The figure indicates that the HIGS-based design induces a factor of two improvement in low-frequency disturbance suppression. At the bottom of the figure, a power spectral density analysis is shown using measurement results obtained from a state-of-the-art industrial stage system. At the critical frequency $f_{3\text{dB}}$, the results confirm the factor of two improvement. As a result, the non-reproducible part in the error response can be significantly reduced by HIGS-based PID control when compared to what current industrial control can offer.

The design of hybrid integrator-gain systems is still in its infancy and requires extensive research and development in order to reach its full potential. Also, HIGS-based PID control is just one example of the rich field of nonlinear control, which contains many other solutions that challenge design trade-offs that occur in classical linear control. Though nonlinear control design in general dates back at least to early examples from Anatolii Lur’e in the 1940s [20], many open problems exist today. Among the most interesting for industrial nonlinear control for high-precision systems is the development of robust control designs in the frequency domain [21] with rigorous stability and performance guarantees.

This also broadens the discussion on de-linearizing the feedback loop through nonlinear control. More specifically, frequency-domain tools for control systems with nonlinear plant characteristics are increasingly expected to be required as well. One can think of new actuator concepts which are being considered for future high-precision systems and which are highly nonlinear, such as piezoelectric actuators that have hysteresis, thermal actuators using Peltier elements that are input non-affine and electrostrictive actuators that possess a quadratic relationship between voltage and displacement.

At this point in the lecture, I will briefly reflect on how research and education on industrial motion control for high-precision systems can be shaped given the interplay between industry and the university.
Industry and the university

Having been involved in industry for the past 20 years and having been affiliated part-time to the university for the past 13 years, industry and the university seem closely related to me. So, what are the gaps between industry and the university - or, more specifically, between practice and theory - that people worry about, and what can be done?

Research

Interestingly, a gap between practice and theory, or industry and the university, did not seem to exist in the early days of automatic control. Just consider the affiliations of its pioneers, Hendrik Wade Bode and Harry Nyquist [22]. Both worked at Bell Laboratories on real-life problems in telecommunications until their retirements. If a gap exists nowadays, both industry and the university should take appropriate actions on closing it. A starting point could be to further exploit individual strengths, such as the long-term vision and the unique vantage point that control researchers have in developing new control technologies or the deep domain knowledge and insights into industrial control problems which are generally owned by engineers in industry. Regarding the latter, the promotion of domain knowledge and industrial control challenges can be instrumental in maximizing the impact of modern control technologies. In this regard, low-hanging fruit involves: (a) a commitment from industry on the participation/attendance of people from industry at control conferences, (b) starting the discussion with editorial boards of journals in order to find a more adequate balance between practice and theory, particularly by stimulating people from industry to scientifically report on their research as part of their own development and competence skills and (c) arranging the managerial conditions in order to facilitate departmental members in doing research in industry.

Education

“The gap is being caused by the difference in the expectations of industry and the academic preparations that graduates receive from universities” [23]. In this regard, one of the mission statements of the recently-installed IFAC industry committee which I have joined is “to establish the core competencies and key skills that industry expects for entry-level control positions.” As a control community, we should periodically monitor and discuss what qualifies as core competencies and should therefore be offered by the university and what qualifies as domain knowledge and should therefore be offered by industry. Regarding the latter, several courses and training programs for graduates have been installed in recent years with the aim of efficiently expanding the domain knowledge of graduates. As an example, two control courses have been developed that connect deep domain knowledge and ASML-specific control designs with current standards in control technology, though an appropriate nonlinear control course is still lacking. At the same time, we need to realize that the field of Systems & Control is rapidly evolving. To be able to innovate industrial controls, we therefore also need to provide the managerial conditions for senior engineers and architects to upgrade their core competencies, especially regarding control advances.

A lot more can be said and more can certainly be done (or already has been). “In theory, there is no difference between theory and practice. But, in practice, there is” - a quote from Manfred Eigen.
Acknowledgements

Alongside the Executive Board of the university and the Department Board of Mechanical Engineering for giving me this opportunity and ASML for giving me full support in this endeavor (and with the obvious risk of forgetting someone), I would like to acknowledge those that inspired, motivated and helped me throughout my career.

First, my mentors Henk Nijmeijer and Maarten Steinbuch, who inspired, coached, challenged and supported me over 20 years at the many crossroads between industry and the university.

Second, my co-authors (and co-applicants) on over 100 scientific papers (and over 20 patents), who helped me to define ‘industrial nonlinear control for high-precision systems’.

Third, my students – currently over 150 in total – who helped me with many of the ideas I developed over the years, starting from 1996.

Fourth, my colleagues from Philips (particularly the ‘Centrum voor Fabrikage Technieken (CFT)’), ASML (particularly the control groups in D&E and ASML Research) and Eindhoven University of Technology (particularly the research groups ‘Dynamics & Control’ and ‘Control System Technologies’) for help in finding a balance between theory and practice.

Fifth, my colleagues from other universities, institutions, and companies worldwide for the many fruitful discussions, control conferences, invited sessions, workshops, meetings, discussions and exams that I participated in.

Sixth, the supporting staff at Eindhoven University of Technology, particularly Roos van Otterdijk, for helping me with the organization.

Seventh, the reviewers of the inaugural lecture booklet, particularly Henk Nijmeijer (TU/e), Marc van de Wal (ASML), Wilbert Simons (ASML) and Jan Nouwens (ASML) for the suggestions content wise that helped improve the text.

Last but not least, my family and friends for their support, particularly my wife Ciska and my daughters Josefien & Marlieke for their unconditional agreement to the terms that a life in control brings.

Ik heb gezegd.
References

Curriculum Vitae

Prof.dr.ir. Marcel Heertjes was appointed part-time professor of Industrial nonlinear control for high-precision systems at the Department of Mechanical Engineering at Eindhoven University of Technology (TU/e) on February 1, 2019.

Marcel Heertjes received both his MSc and PhD in mechanical engineering from Eindhoven University of Technology (TU/e) in 1995 and 1999, respectively. In 2000, he joined the Philips Centre for Industrial Technology (CFT) in Eindhoven. In 2007 he joined ASML, Department of Mechatronic Systems Development, in Veldhoven, and also TU/e (part-time) as a visiting scientist. In 2011 he was appointed part-time assistant professor in the group Dynamics & Control as well as the group Control Systems Technology. In 2012 he became part-time associate professor. His primary affiliation is with ASML in the role of principal engineer (since 2007) and control competence leader (since 2012). He served as guest editor of International Journal of Robust and Nonlinear Control (2011) and IFAC Mechatronics (2014) and is currently associate editor of IFAC Mechatronics (since 2016). He is (co-)recipient of the IEEE Control Systems Technology award (2015) for variable gain control and its applications to wafer scanners. His main research interests are with industrial control for high-precision mechatronic systems with specific focus on nonlinear control, feedforward and learning control, and data-driven optimization.
Visiting address
Building 1, Auditorium
Groene Loper, Eindhoven
The Netherlands

Navigation
De Zaale, Eindhoven

Postal address
PO Box 513
5600 MB Eindhoven
The Netherlands
Tel. +31 (0)40 247 9111
www.tue.nl/map