2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Control of Wafer Scanners: Methods and Developments

M.F. Heertjes, H. Butler, N.J. Dirkx, S.H. van der Meulen *, and
R. Ahlawat †, and
K. O'Brien, J. Simonelli, K-T. Teng, Y. Zhao ‡
*ASML, Veldhoven, The Netherlands, email: marcel.heertjes, hans.butler, nic.dirkx, stan.van.der.meulen {@asml.com}
†Rahul Ahlawat was formerly with Cymer, San Diego, California
‡Cymer, San Diego, California, email: kevin.obrien, james.simonelli, kuo-tai.teng, yingbo.zhao {@asml.com}

***Abstract*— In this tutorial paper, control design aspects of wafer scanners used in the semiconductor industry will be highlighted. At the same time, challenges for control design development as to meet the ever increasing demands on accuracy and speed are presented. Mechatronic systems that will be discussed are: (a) the light source needed to generate the ultraviolet light that is used for wafer exposure, (b) the optical and metrology systems needed for accurate measurement and imaging, and (c) the reticle and wafer stage systems needed for accurate and fast positioning. The control challenges associated with these systems mainly involve dealing with: (a) rejection of high frequency aliased disturbances, (b) large-scale or fast-updated (state) reconstruction, (c) vibration control and isolation in view of structural vibrations and disturbances, (d) inherent design tradeoffs like Bode's sensitivity integral and gain-phase relationships, (e) multivariable plant identification of (quasi-static) deformations and structural dynamics for point-of-interest control, and (f) thermal modelling, model reduction, and the control of (local) time-varying deformation. Results will be discussed using representative examples.**

***Index Terms*— advanced motion control, feedback/feedforward control, light source, nano-precision mechatronics, observer design, thermal control, vibration isolation & control, wafer scanners, wavelength control.**

I. INTRODUCTION

Today's information economy including paradigms like the internet-of-things and the big data era is built on half a century of technological development done in the semiconductor industry according to Moore's law [49]. The main technological enabler underlying these developments is widely-acknowledged to be the lithography process as it provides the current standard for producing microchips in a cost effective manner [40]. Lithography is the key process in wafer scanners, *i.e.*, the machinery used to make microchips.

Wafer scanners whose list prices are in the range of 10,000,000-150,000,000 USD consist of many highly-complex mechatronic systems that together combine high throughput with high precision [50]. Regarding throughput, modern wafer scanners process around 280 wafers per hour where less than 10 seconds are needed to expose a 300 mm wafer that contains around 100 exposure fields. Each field allows for constructing a complex processor chip through a scanning process. Scanning is done by commanding a sequence of concatenated point-to-point motions during which the tracking specifications of the scanner's motion systems mostly lie in the (sub-)nanometer range.

Wafer scanners consist of several main subsystems. For example, the light source, an illumination optical system, reticle and wafer stage systems, and material handling robots, see also the overview in [15]. These subsystems are high-precision mechatronics that make extensive use of advanced control to meet specifications. In this tutorial paper, a few of these subsystems will be discussed in more detail where the focus will be on (a) control designs, and (b) control challenges associated with these designs. In this sense, the main contributions of this paper lie in providing (a) an overview of the wafer scanner and its subsystems, (b) the state-of-practice from a control engineering perspective, and (c) potential developments needed to support future demands.

The paper is organized as follows. In Section II an overview of a wafer scanner is provided along with a discussion of its control performance indicators. In the subsequent sections, three key sub-systems will be discussed in more detail: source in Section III, optics in Section IV, and stage systems in Section V and VI. Possible directions for control research and development will be summarized in Section VII.

II. WAFER SCANNERS

Wafer scanners exploit the principles of photolithography, which is the method that is used for patterning almost all integrated circuits fabricated today [40]. As such, it provides a key step in the recurrent process of making microchips such as illustrated in Fig. 1. Specifications of wafer scanners are typically given in terms of overlay, resolution, focus, and throughput. Overlay, be it Single-Machine Overlay (SMO) measured on the same system or Matched-Machine Overlay

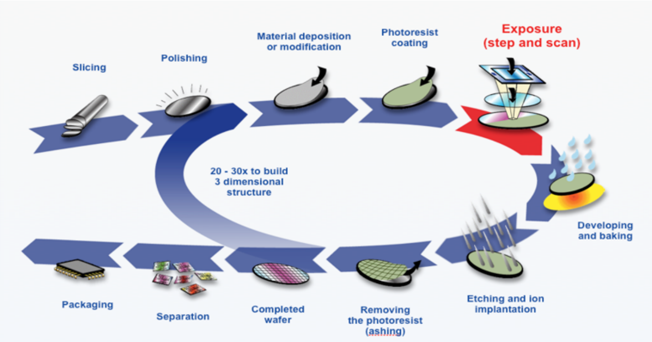

Fig. 1: Photolithography in the recurrent process of making microchips.

978-1-5386-8266-1/$31.00 ©2020 AACC

(MMO) measured between different systems, defines the accuracy to print a pattern directly atop the pattern of a previous layer [47]. Resolution, often represented by Critical Dimension (CD), describes the smallest feature size that can be printed. Resolution follows from the Rayleigh criterion,

$$CD = k_1 \frac{\lambda}{NA}, \tag{1}$$

with k_1 the k-factor for a given process, λ the wavelength, and $NA > 0$ the numerical aperture. Typical values for lithography today are $k_1 = 0.25$, $\lambda = 193$ nm, and $NA = 1.35$ leading to 38 nm line widths [26]. These numbers reflect the so-called DUV (deep ultraviolet) technology. Focus, often represented by Depth-of-Focus (DoF), refers to the peak intensity of a point object focused by a lens to remain within 20% of the peak value for best focus [40], and is written as

$$DoF \approx k_2 \frac{\lambda}{NA^2}, \tag{2}$$

with $k_2 \geq 0.5$. Throughput represents the scanner's output, *i.e.*, the amount of wafers processed per time unit, which for modern wafer scanners is around 280 wafers per hour.

A system overview of a wafer scanner is shown in Fig. 2. The example features the latest EUV (extreme ultraviolet) technology [3]. Differences between DUV and EUV, though apparent, do not limit us from discussing the principles of wafer scanning in general and their control aspects in specific. In fact, both technologies will be described with some randomness throughout the paper. The EUV system from Fig. 2, that weighs 64,380 kilograms and has a list price around 115,000,000 USD, consists of four key modules: light source (of which substantial parts are placed outside the scanner), projection optics box, reticle stage, and wafer stage. The functional aspects of these modules within the wafer scanner can be explained as follows. EUV light with a wavelength of 13.5 nm, and generated in the source, falls onto a mask called reticle that contains a blueprint of (part of) the microchip pattern that needs to be printed. The reticle is mounted on a reticle stage system that repetitively moves forward and backward in scanning y-direction and under constant velocity, typically with 1.2 m/s. The image that is created by reflecting (and partly absorbing) the light beam from the reticle travels through the projection optics box and is demagnified by a factor of four. To do so, the projection optics box contains multiple individually controlled mirror systems that are placed in a high vacuum chamber. A vacuum environment is needed to avoid EUV absorption otherwise occurring in air. Upon leaving the projection optics box, the scaled-down image is projected onto a light-sensitive layer of a wafer, which is represented by a 300 mm (diameter) silicon disk that is typically subdivided into 100 fields. Note that the lithographic process involves printing one layer at a time, so the wafer typically re-enters the machine 30-150 times, depending on the manufacturing process: memory (DRAM) or logic. The wafer is supported by the wafer stage, which has two main functions: exposure and measure. During exposure, the wafer on the expose side is positioned underneath the optics such that each field on the wafer can be exposed in a scanning manner while maintaining full synchronicity with the reticle stage. During measure, the wafer on the measure side is positioned underneath the level sensor as to measure (while scanning) the full wafer topology. This topology, captured in a so-called wafer map, will be provided as wafer setpoints in the z, r_x, and r_y directions during exposure to meet image quality. The wafer stage exposure and measure functions are conducted in parallel as can be seen in Fig. 2. That is, while exposing one wafer, which is shown on the right hand side of the figure, another wafer is being measured, which is shown on the left hand side of the figure.

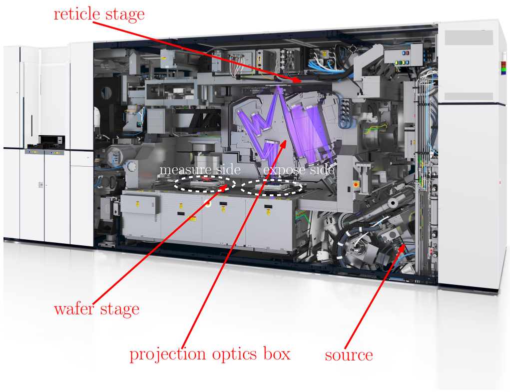

Fig. 2: Artist impression of an EUV wafer scanner.

Apart from these key modules, the wafer scanner has many more functional modules. For example, (a) transport of wafers and reticles in and out the scanner by wafer and reticle handling robots, (b) reticle masking needed to get light in a designated area called slit on the wafer, (c) metrology and isolation needed to provide a vibration-free reference for measurement, (d) power electronics, (e) systems dealing with cooling fluids, conditioning gasses, and thermal aspects. In the remainder of this paper, however, we will focus on the control designs of source, optics, and stages only.

III. SOURCE: GENERATION AND CONTROL OF LIGHT

The light source is a complex, nonlinear, and Multi-Input Multi-Output (MIMO) laser system. In this system, light is produced in the form of bursts of pulses at several kHz referred to as repetition rate of the laser. Bursts are followed by quiescence during which no light is produced, referred to as inter-burst intervals. As evident from (1), the wavelength of the light source directly dictates the size of the feature that can be printed. Basically, two light sources are used: a DUV light source that produces 193 nm wavelength and an EUV light source that produces 13.5 nm wavelength, see also [1] for details on laser-produced plasma light sources. This section will focus on a DUV light source in particular.

State-of-the-art DUV light sources contain two plasma chambers: a Master Oscillator (MO) chamber and a Power Amplifier (PA) chamber, see also [10]. Both chambers are filled with pressurized gas that contains precise quantities of Fluorine and Xeon, with Argon and Neon as buffer gasses. Light is produced when applied voltage on the electrodes in the respective chambers results in a discharge. The MO chamber first generates a seed light that has relatively low energy. After filtering the wavelength—or equivalently, narrowing the spectral bandwidth—the seed light is sent to the PA chamber for power amplification.

The light source is subject to narrow-band, stochastic, and often aliased disturbances that degrade wavelength, spectral bandwidth stability, and energy and hence must be actively rejected. Maintaining the stability of these key performance characteristics in the light source is paramount for imaging performance. As such, light output measurements are provided to the control systems after each pulse by various measurement devices. The outputs of interest are subsequently modified by various actuators like solenoids, piezoelectric transducers, and stepper motors. In this section, an overview is presented of the control philosophy regarding wavelength and spectral bandwidth. Advanced energy control has also been implemented in the source, but space concerns preclude its discussion here.

A. *Wavelength control*

The light source wavelength is controlled in the MO chamber using the Line-Narrowing-Module (LNM) [10]. Fig. 3 is a schematic of the LNM: the module housing the wavelength actuators. As light enters the LNM (upper-left), it travels through four different prisms before reaching the grating, which disperses the incoming light, reflecting it in different directions based on the wavelength of the photons. The LNM is designed such that only photons travelling back along the same path will be able to make their way out of the module. Conceptually, the LNM acts as a bandpass filter—photons with wavelength outside a selected range will not be able to exit the module. Since the wavelength in the output light is a function of the angle of incidence on the grating, the position of all four prisms affects the output wavelength, though not equally; the prisms have a lessened effect on the final pointing angle as the optical path is followed away from the grating. Therefore, the closer the prism is to the grating, the larger its gain from prism angle to wavelength. Driven by a stepper motor, the prism that is closest to the grating is thus chosen to act as coarse wavelength control, whereas actuated by a piezoelectric transducer (PZT), the second prism from the grating is chosen as fine wavelength control. The piezoelectric transducer benefits both from the lower gain associated with its position in the optical setup and from its analog positioning capability; the coarse wavelength actuator accuracy is limited by the combination of its larger DC gain and quantized positioning.

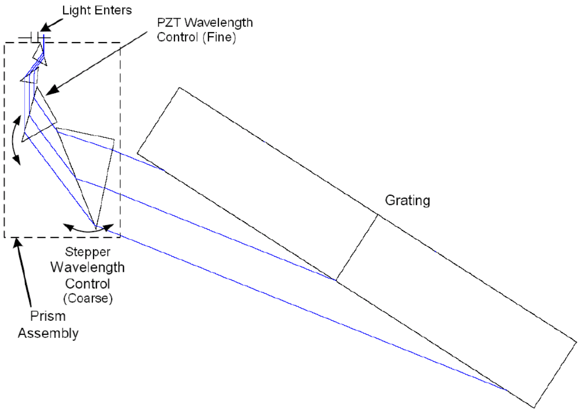

Fig. 3: Schematics of the LNM used for wavelength control.

As wavelength is a function of prism position, any vibration acting on the prisms and the grating induce disturbances to the output wavelength. For example, there are blowers in the chamber that circulate and homogenize the gas mixture and cause acoustic waves inside the chamber, coupling into the optics and inducing wavelength disturbances. These disturbances are tones that, due to the wide range of repetition rates supported (i.e. sampling frequencies), may appear as aliases anywhere. They must be attenuated by the wavelength control algorithm in order to provide DUV light with the lithography application's requisite extreme stability. Typical specifications require a wavelength of 193.368 nm, which has to be controlled within the setpoint range of 193.368 ± 0.00012 nm along the entire wafer with an accuracy of 10^{-15} m, see also [14].

Fig. 4 shows a block diagram of the wavelength control algorithm presented here, designed to achieve wavelength stability. A fundamental challenge in designing the control strategy for the DUV light source is the unknown measurement update rate. As it is impossible to measure wavelength without having fired the laser, measurements are available only on each individual laser pulse, the rate of which varies from 500 - 6000 Hz depending on the customer's wafer exposure recipe, of which *a priori* knowledge cannot be assumed. This motivates the adoption of a continuous-discrete Kalman filter in the control design to estimate system states—including modeled, known disturbances [54] — at a much higher rate than it is possible for measurements to come in, while actuating based on an optimal control law at a fixed rate irrespective of the measurement rate.

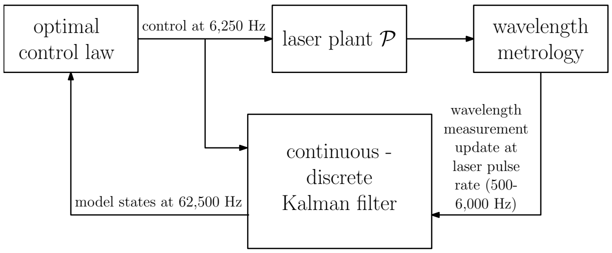

Fig. 4: Block diagram of a wavelength control algorithm.

B. *Linewidth control*

While the wavelength control system maintains the stability of the central wavelength, the linewidth control system is responsible for maintaining the linewidth or spectral bandwidth of the laser. Linewidth is usually reported as E95 which represents 95% integral energy of the spectrum. Poor linewidth control will cause loss of contrast in printed features on the wafer, see Fig. 5, in particular the blue boxes. To avoid loss of contrast, a linewidth control system is used [17],

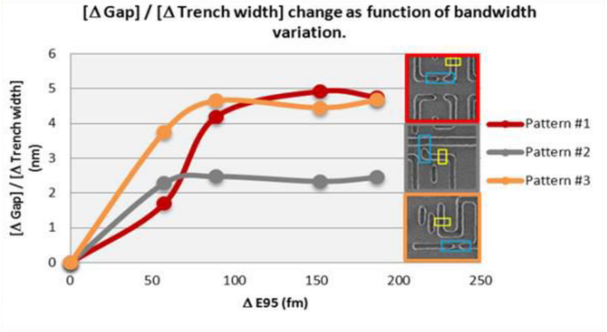

Fig. 5: Example of loss of contrast on printed features [2].

[22]. Consider the linewidth model depicted in Fig. 6, which includes plant $\mathcal{P}$, sensor $\mathcal{B}$, and filter $\mathcal{M}$, the latter computes figure of merit f. The plant has two control inputs u_1 and u_2, and three outputs, namely linewidth b, wavelength λ, and energy n. Disturbances d comprise laser usage variations like periodic transients, coupling from laser light's pulse energy, and other intrinsic properties of the production of excimer laser light. The linewidth measurement process $\mathcal{B}$ is corrupted by laser wavelength λ. Moreover, it is susceptible to output noise (and possibly bias) v_b and results in final linewidth measurement m. The energy measurement process is modeled as additive white noise v_n.

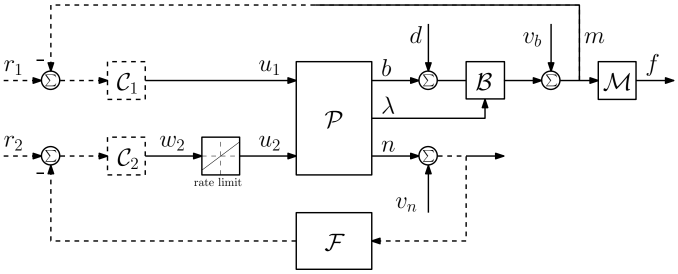

Fig. 6: Block diagram of linewidth (control) system.

The laser plant $\mathcal{P}$ can be modeled mostly as a static nonlinear mapping from the control inputs u_1, u_2 to the outputs b, λ, n, or

$$b = g_1(u_1, u_2), \tag{3}$$

$$\lambda = g_2\left(\frac{du_2}{dt}\right) + R, \tag{4}$$

$$n = g_3(u_1, u_2). \tag{5}$$

Herein, λ is a function of the rate of input u_2, though the actuator suffers from a limit on this rate resulting in the difference between w_2 and u_2. λ is also affected by exogenous effects represented by R. g_1, g_2, and g_3 are subject to variation and uncertainty. For example, the function g_1 can be depicted as in Fig. 7, where each color represents a different laser or a given laser at different time instances.

The control objective underlying Fig. 6 is to maintain figure of merit f at a target while satisfying specifications on stability margins, variance, and transient response under noise, disturbances and uncertainties.

The linewidth control system consists of two controllers to compute the control signals u_1 and u_2. The rate limit on u_2 imposes a significant performance penalty, so u_1 is often chosen as primary linewidth actuation. Due to the limited control range of u_1, see Fig. 7, u_2 is instead used to center u_1. The two controllers are designed in tandem, see the dashed lines in Fig. 6. The main purpose of u_1 is to regulate figure of merit f to reference r_1 [1], which is done by derivative control

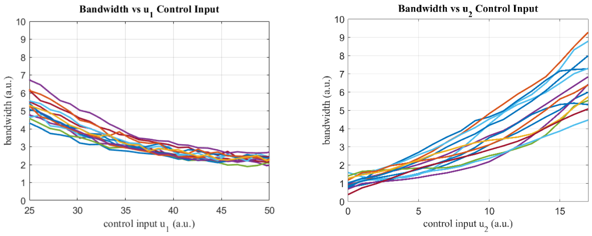

Fig. 7: Variation of spectral bandwidth: u_1 (left) and u_2 (right).

$$\mathcal{C}_1(z) = k_1 \frac{z-1}{z}. \tag{6}$$

with gain k_1 tuned to limit linewidth transients below an allowed threshold, and z^{-1} a unit step delay. This implies an upper limit on how fast a drift rate can be rejected from disturbance d, and therefore poses a control design trade-off. This tuning is exemplified in Fig. 8. The left part of Fig. 8 shows the noise suppression in figure of merit f over the controller gain k_1; each curve represents different slopes taken from the curves in Fig. 7, where the dashed line represents the selected value of k_1. This value yields good noise suppression and limits the possibility of excessive noise sensitivity. The right part of Fig. 8 shows the root locus of this u_1 control loop, highlighting the expected region of plant gain variation from Fig. 7, and indicating good robust stability and performance margins. Controller $\mathcal{C}_2$ is designed to maintain u_1 near reference centering position r_2. The idea is to adjust the linewidth through u_2 such that controller $\mathcal{C}_1$ must re-adjust u_1 toward its center position thereby maintaining linewidth at r_1. The actuators u_1 and u_2 are distinct from the wavelength control actuators, but the effects of u_2 do couple into wavelength. To prevent this coupling from having undesirable wavelength performance effects, the actuation rate of u_2 is intentionally limited so as to remain in a frequency band where the wavelength controller's sensitivity is low. This allows the wavelength

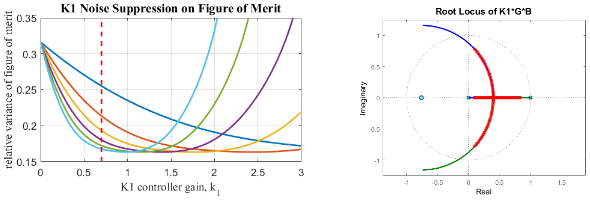

Fig. 8: Noise suppression for different control sensitivities (left) versus root locus highlighting expected operational region (right).

---

[1] $\mathcal{M}$ has unity DC gain and r_1 changes infrequently, so regulating f to r_1 is tantamount to regulating m to r_1.

controller to treat the actuation of u_2 as a disturbance and thus ensures its rejection.

IV. OPTICS: ISOLATION AND CONTROL OF VIBRATION

The projection optics forms the heart of the lithographic tool since it forms an image of the original pattern on the mask onto the wafer. In DUV tools, the projection lens often contains multiple refractive lens elements whereas in EUV tools multiple multi-layer reflective mirrors are used [4]. In the optical system, generally a frame serves as position reference for the optical elements, and also for the stages. Either the optical elements are actively controlled with respect to this frame, or physically connected to it. In any case, this optical frame should be free of vibrations to enable stable optical element positions during operation. Additionally, low-frequency motion should be avoided to limit deformations of the frame itself. The first part of this section will be devoted to active vibration isolation, whereas the second part considers deformation control.

A. *Vibration isolation architecture and approach*

The scanner's base frame is connected to the floor and typically is subject to significant reaction forces induced by stage motions; this is schematically shown in Fig. 9. Consequently, dynamic isolation toward the projection optics is important. The first vibration isolation step is often formed by so-called airmounts, *i.e.*, vibration isolators that create a pneumatic spring supporting the system's metroframe. This frame is used to hold the stage position measurement system (either interferometer or encoder) and also supports the projection optics. The second vibration isolation step is the support of the projection optics itself.

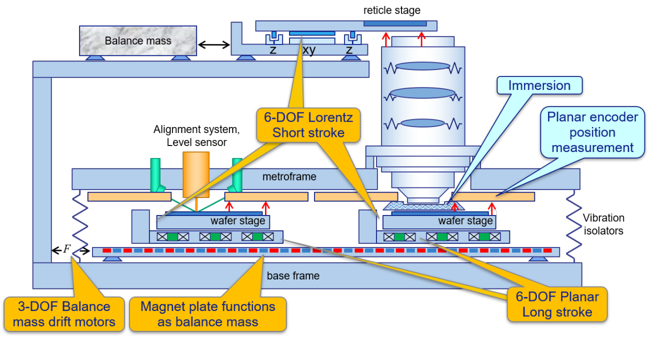

Fig. 9: An example of a DUV wafer scanner's dynamical layout [4].

This two-step approach is illustrated in Fig. 10. Here, the base frame is connected to the floor by an as-high-as-possible stiffness k_b. The base frame supports the metroframe by the airmount isolator, indicated by stiffness k_m. The lens is connected to the metroframe by a connection k_l. To minimize lens accelerations, the transmissibility, or the transfer from x_b to x_l, should be minimized in the frequency range where base frame disturbances are dominantly present. Simultaneously, the compliance function, *i.e.*, the transfer from F_l to x_l, should be minimized in the frequency range where direct disturbances like acoustics or cooling water flow are present.

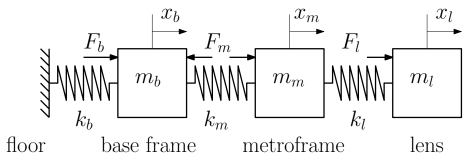

Fig. 10: Simplified representation of a two-step isolation approach.

The airmount isolation frequency is selected as low as possible to optimally block base frame disturbances, typically 0.5 Hz. However, the lens support stiffness (indicated by k_l) is chosen differently for DUV systems than for EUV systems. In DUV systems, resistance to direct lens forces by acoustics is important, leading to a relatively stiff lens to metroframe connection. However, since EUV systems operate in vacuum, a lower lens suspension frequency can be admitted to further minimize disturbances transmitted from base frame through metroframe to the lens. This is indicated in the frequency response of Fig. 11. By decreasing the lens suspension frequency, in the example of Fig. 11 from 90 to 30 Hz, an improved transmissibility can be obtained, be it at the cost of a degraded compliance. Note that for the purpose of illustration, Fig. 11 does not yet include proper damping of the lens suspension frequency.

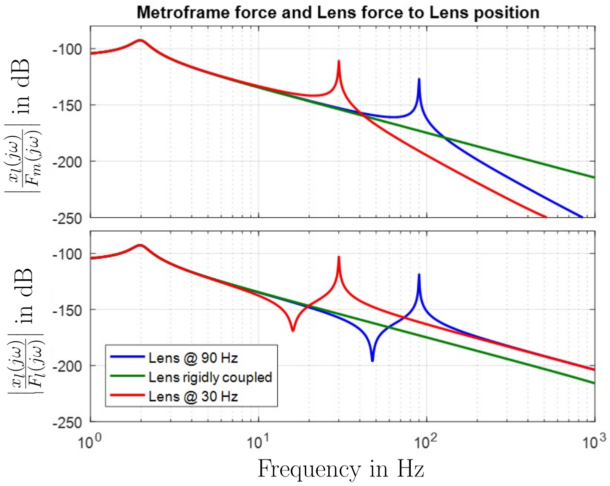

Fig. 11: Bode magnitude plots of compliance from forces F_m to displacement x_l (top) versus F_l to x_l (bottom).

B. *Vibration isolation step one: airmount control*

The airmount vibration isolation system provides a low-stiffness connection by the use of pneumatic springs. Schematically this is shown in Fig. 12. A pneumatic volume is pressurized to create a lifting force which is equal to the payload. The internal pressure is controlled based on the vertical position of the payload. In addition to pneumatic control, position sensors, accelerometers and Lorentz actuators are present to control the frame position in 6 degrees of freedom, and in particular to damp the resonance associated with the suspension frequency. This damping can be achieved (a) by absolute velocity feedback, known as skyhook damping [34], using an accelerometer or geophone, or (b) by

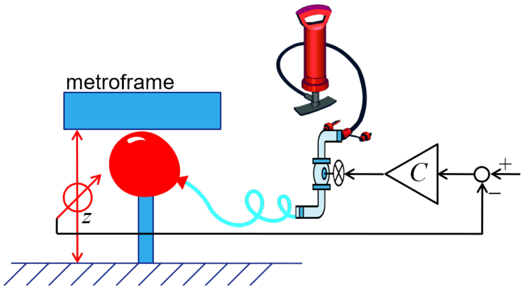

Fig. 12: Principles of an active pneumatic system.

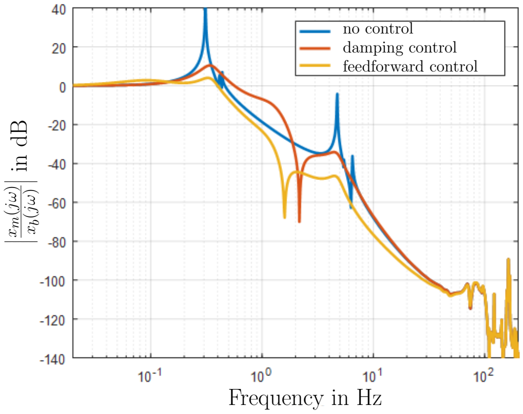

Fig. 13: Simulated transmissibility using relative damping (red) and including base frame disturbance feedforward control (yellow).

relative position feedback in combination with a lead filter. In the latter case, additional low-pass filtering is required to avoid negative impact on high-frequency transmissibility together with an integrator needed for positioning. Although absolute velocity feedback provides a better transmissibility, its use is limited in lithographic tools, because:

- The suspension frequency is often below 1 Hz, where sensor noise and accuracy become problematic;
- Accelerometers and geophones suffer from inherent crosstalk from rotation to translation [46].

For these reasons, relative position feedback is often used.

In addition to feedback control, one could reside to feedforward control using the base frame acceleration as input [5]. Consider the equation of motion of the system from Fig. 10, which after further simplification yields:

$$(ms^2 + d_m s + k_m)x_m(s) = (d_m s + k_m)x_b(s), \tag{7}$$

with $m = m_m + m_l$ and damping coefficient d_m. Suppose a control force $F_m(t) = \mathcal{C}_{ff}(p)\ddot{x}_b(t)$ is applied that gives

$$\begin{aligned} &(ms^2 + d_m s + k_m)x_m(s) \\ &= (\mathcal{C}_{ff}(s)s^2 + d_m s + k_m)x_b(s). \end{aligned} \tag{8}$$

From (8) it can be seen that the disturbance feedforward controller

$$\mathcal{C}_{ff}(s) = -\frac{d_m s + k_m}{s^2}, \tag{9}$$

minimizes frame motion x_m; (9) is known as the Wiener-solution [51]. Key is the input to the feedforward controller being a measured signal that is *a priori* unknown, see [5] for the control design including the feedback part.

The effect of disturbance feedforward control applied on a FE model is shown in Fig. 13 through simulation results. It can be seen that the feedforward controller can compensate for the 'bump' effect of relative damping. This technique may apply to all 6 degrees of freedom, either in multi-loop SISO or full MIMO construction. Remark that, in case of the latter, the stiffness constant k_m transforms to a full 6×6 stiffness matrix. Also remark that pneumatic isolators, providing a low suspension frequency, use large pressurized air tanks that lead to acoustical dynamics in the pneumatic circuits, thereby creating a dynamic stiffness [64]. If captured by a filter, this dynamic stiffness could be (partly) compensated by disturbance feedforward control, see for example [6].

In lithographic tools, multiple airmount-supported modules may be present, for example the metroframe and the illuminator. Since their relative position becomes increasingly important, one suspended system can be made to track the other, as shown schematically in Fig. 14. Note that in the figure the position of the frame to-be-tracked is used as setpoint (position) for the tracking frame, which in this case is the illuminator. Some airmount control research directions are:

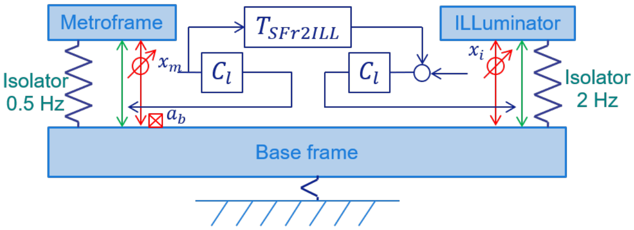

Fig. 14: Illuminator tracking the metroframe.

- High-order accelerometer filter design to limit low-frequency noise impact;
- Minimizing accelerometer crosstalk from rotation to translation;
- Automatic tuning of dynamic feedforward filters using low-energy existing floor motions, see also [7];
- Integrated optimization in the case of multiple suspension systems coupled by control.

C. Vibration isolation step two: PALM

Active damping in the form of Piezo-Active Lens Mounts, abbreviated as PALM, is a technique often found in DUV tools [29]. PALM ensures that internal lens elements are not excited and the stage can track the lens accurately. This second isolation step occurs at relatively high frequencies (around 90 Hz) as shown in Fig. 11. PALM schematics are shown in Fig. 15. Here, the lens is supported by three PALMs, each capable of motion in two degrees of freedom, namely tangential and vertical. A piezo actuator element, creating a displacement, is connected in series with a piezo sensing element, measuring a force. Integral force feedback [60] can be used via a 6 degree-of-freedom decoupled feedback controller, aiming at damping the 90 Hz resonance.

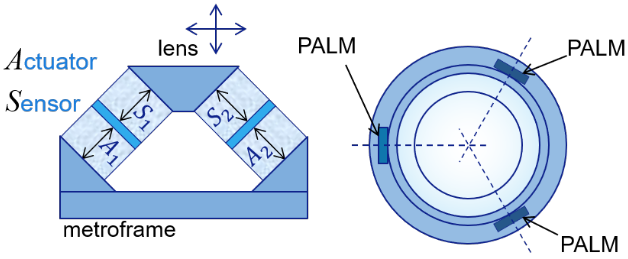

Fig. 15: Schematics of Piezo-Active Lens Mounts (PALM).

In EUV systems, due to the vacuum environment, direct disturbances on the lens are limited and hence the suspension frequency can be significantly lowered. As a result, relative motion between the metroframe and the projection optics increases, making the PALM system unusable. Instead, a conventional spring-like suspension may be used, *i.e.*, damping the associated resonance by inertial feedback [53]. The suspension frequency of 5 - 30 Hz allows for using conventional accelerometers for control. Both concepts are illustrated in Fig. 16. Here, $d(s)$ represents the damping controller, which consists of an integrator extended with low-pass and high-pass filters.

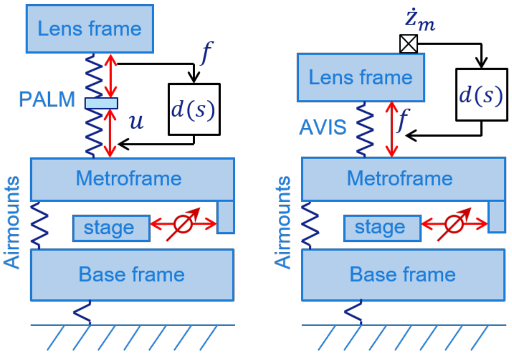

Fig. 16: PALM for DUV (left) and active vibration isolation (AVIS) for EUV systems (right).

D. Deformation control

Deformation control, *i.e.*, controlling the shape of an object through over-actuation and over-sensing, provides an addition to the earlier discussed methods. Frames tend to increase in mass and size and thus may require four supporting points instead of three.

A basic example to illustrate the concept of deformation control is shown in Fig. 17. The four supports create a torsional stiffness between the frames. Close to each support there is an accurate position sensor measuring relative displacement z_i and an actuator providing a force f_i. If the bottom frame experiences torsional deformation, through a torsional moment, this also induces torsional deformation of the suspended top frame. The top frame deformation is in the sub-nanometer range and therefore difficult to measure. However, the torsion of the bottom frame can be measured using the position sensors. Control can reduce the torsion of the top frame by creating an active (negative) torsional stiffness [16]. This is similar to applying negative control stiffness to compensate for a positive spring stiffness [48], but in a torsional degree of freedom:

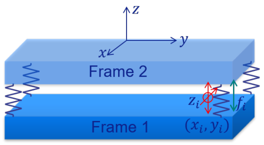

Fig. 17: Frame supported by four isolators.

- From the four z-sensors, a torsion measurement s_T is derived, *i.e.*, relative torsion between the frames;
- Using four z-actuators, a force F_T is created acting on the top frame with respect to the bottom frame.

From the torsion measurement value, a compensation torsional force can be created that (partially) compensates the torsion stiffness k_T which is the result of the 4 vibration isolation springs, and which is shown in Fig. 18. The torsion measurement system T_S and actuator system T_A depend on the definition of the unit of torsion.

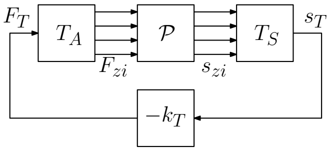

Fig. 18: Schematics of torsion shape control.

We define this unit according to a general torsional shape, relating a local z displacement to its (x, y)-location, or $z_i = s_T(x_i y_i)$, with $i \in \{1, 2, 3, 4\}$. The torsion measurement system T_S is defined by collecting the relation between (a) the rigid-body displacements in z, Rx, Ry direction together with the unit of torsion s_T, and (b) the four z-sensor values $z_1 \dots z_4$, or

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & y_{s1} & -x_{s1} & (x_{s1}y_{s1}) \\ 1 & y_{s2} & -x_{s2} & (x_{s2}y_{s2}) \\ 1 & y_{s3} & -x_{s3} & (x_{s3}y_{s3}) \\ 1 & y_{s4} & -x_{s4} & (x_{s4}y_{s4}) \end{bmatrix}}_{T_S^{-1}} \begin{bmatrix} z \\ Rx \\ Ry \\ s_T \end{bmatrix}. \tag{10}$$

The values in the last column indicate how the sensor observes the torsion mode shape. Note that the first three columns relate a frame motion to individual sensor measurements, and are less relevant for torsion. By inverting the above matrix, the sensor matrix T_S appears. Similarly, to find the actuator matrix T_A, the relation between individual motor forces and corresponding logical forces and torques is found through

$$\begin{bmatrix} f_z \\ T_x \\ T_y \\ f_T \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ y_{a1} & y_{a2} & y_{a3} & y_{a4} \\ -x_{a1} & -x_{a2} & -x_{a3} & -x_{a4} \\ (x_{a1}y_{a1}) & (x_{a2}y_{a2}) & (x_{a3}y_{a3}) & (x_{a4}y_{a4}) \end{bmatrix}}_{T_A^{-1}} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}.$$

The contribution of each of the four vibration isolation springs to the torsional stiffness k_T is related to their location, giving the total torsional stiffness

$$k_T = \sum_i \left(x_{pi} y_{pi}\right)^2 k_i, \tag{11}$$

in which x_{pi}, y_{pi} are the horizontal xy-positions of spring i and k_i the corresponding vertical stiffness coefficients.

As a measure of success, the relation between a bottom frame torsional force and the top frame torsional shape is shown in Fig. 19. In the example, finite element models of both frames are used. Without torsion control, a -246 dB value indicates that 10 N of torsional input to the bottom frame creates a torsional shape of 5 pm. This is reduced by at least a factor of ten using the above strategy.

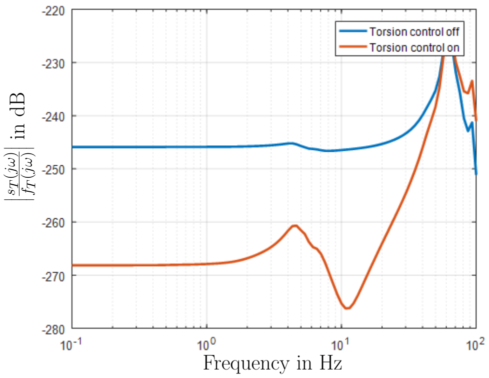

Fig. 19: Magnitude plot showing the effect of torsion shape control.

E. Outlook

Damping of suspension frequencies is a critical feature of isolation performance, which aims at keeping the acceleration levels of the optics low. Conventional feedback techniques are being expanded with feedforward control. Extending these to a lower-frequency region becomes a challenge, and so is the inclusion of dynamic behaviour for example originating from acoustics.

Apart from damping, quasi-static image positioning errors caused by deformation of frames and connections shows increased relevance. In this regard, torsion shape control is just an example of a possible future direction. Prediction of the deformation shapes by existing sensors, *e.g.*, using observer techniques, is expected to become relevant.

V. STAGES: PART ONE - CONTROL OF MOTION

Wafer and reticle stage systems are fast and accurate positioning systems that conduct (concatenated) point-to-point motions. In doing so, these systems depend heavily on control in achieving both tracking performance and disturbance rejection. In Section V-A, a brief description of stage systems will be given mainly from electro-mechanical perspective. In Section V-B, two control problems will be discussed: the servo (or tracking) control problem, and the disturbance rejection problem. Current control methods dealing with these problems will be discussed in Section V-D. The focus will be, first, on feedforward control and, second, on feedback control. In Section V-E an outlook will be given on control developments.

A. System description

A detailed representation of a wafer stage system is shown in Fig. 20. The wafer stage system as shown in Fig. 20 exploits a dual stroke concept including balance mass [4]. The short-stroke (SS) is controlled at (sub) nanometer level within a limited stroke of micrometers. The long-stroke (LoS), which basically follows the short-stroke, is controlled at the micrometer level and has a larger stroke. Actuation of the long-stroke is based on the Lorentz principle [50], which is done in combination with the magnet plate. The magnet plate combines two functions. First, it serves as actuator part for the long-stroke motor. The accompanying part, *i.e.*, the part containing the coils, is connected to the floating long-stroke structure. Second, the magnet plate avoids the long-stroke reaction forces to enter the base frame of the wafer scanner, recall Section IV. In the remainder of this section, we will focus on the short-stroke wafer stage only.

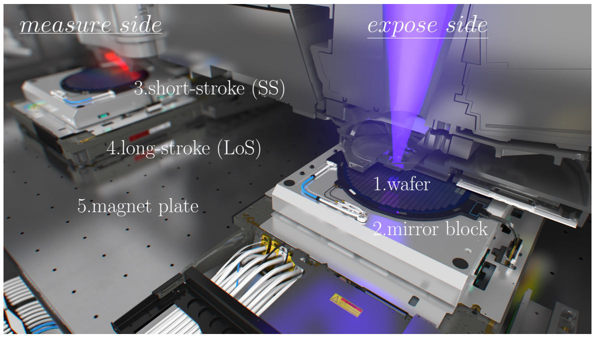

Fig. 20: Modern wafer stage system.

The short-stroke wafer stage, as indicated in Fig. 20, consists of the following main components: (a) the wafer, (b) the wafer table that flattens and supports the wafer, (c) the mirror block that supports the wafer table and provides the means for position measurement by interferometry, and (d) the magnet yoke (not visible in Fig. 20) that together with the coils mounted on the long-stroke provides the means for short-stroke actuation via Lorentz' principle.

Short-stroke wafer stage control is done in six degrees-of-freedom: x, y, rz, z, rx, and ry in a centralized manner. For this purpose, an appropriate static decoupling strategy is used that (largely) enables a multi-loop single-input-single-output (SISO) control design. The resulting 6×6 plant characteristics in essence describe double-integrator behavior on the diagonal entries only, *i.e.*, short-stroke stages mainly behave as floating mass systems.

Short-stroke wafer stage plant characteristics are shown in Fig. 21. Through frequency response measurement, the Bode diagram up till 500 Hz shows -40dB/dec magnitude slope combined with -180 degrees phase lag. Below 30 Hz, reduced measurement quality follows from closed-loop identification with random noise (for advancements see [57]), whereas at higher frequencies the additional phase lag shown

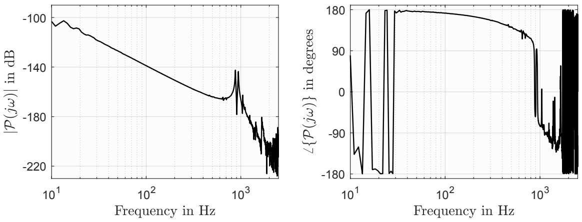

Fig. 21: Bode diagram of measured frequency response data of a short-stroke wafer stage in y-direction.

is due to the sampling frequency of 5 kHz. It should be noted that beyond 500 Hz, the rigid body assumption of the wafer stage plant is no longer valid. In fact, the non-collocated behavior occurring around 1 kHz largely follows from the magnet yoke being decoupled from the mirror block.

B. Control problem formulation

For stage control we basically distinguish between two problems: the servo problem and the regulator problem [61]. In the servo (or tracking) problem, we desire the output of the controlled system to stay as close as possible to a reference trajectory, for example the wafer meander pattern. In the regulator (or disturbance rejection) problem, we desire the output of the controlled system subjected to (unknown) disturbances to be as close as possible to zero.

For stages, the servo problem is generally solved by feedforward control, whereas the remaining regulator problem is dealt with by feedback control. In a motion control context this is illustrated by the simplified motion control scheme of the wafer stage system as depicted in Fig. 22. In this figure, the double integrator-based plant $\mathcal{P}$ is controlled by both feedback controller $\mathcal{C}_{fb}$ and feedforward controller $\mathcal{C}_{ff}$. The outputs of these controllers, *i.e.*, u_{fb} and u_{ff}, respectively, together with the input disturbances d form the input $u = u_{fb} + u_{ff} + d$ to $\mathcal{P}$. The output y is subtracted from the reference signal r to form the servo error signal $e = r - y$.

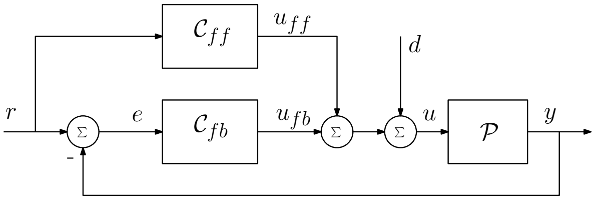

Fig. 22: Block scheme of a simplified controlled wafer stage.

Consider the closed-loop servo error e that from straightforward algebra (in Laplace domain) yields

$$e(s) = \mathcal{S}(s)(I - \mathcal{P}(s)\mathcal{C}_{ff}(s))r(s) - \mathcal{S}_p(s)d(s), \quad (12)$$

with $\mathcal{S}(s) = (1 + \mathcal{P}(s)\mathcal{C}_{fb}(s))^{-1}$ the input sensitivity function and $\mathcal{S}_p(s) = \mathcal{P}(s)\mathcal{S}(s)$ the process sensitivity function. The importance of (12) in view of both the servo and regulator problem is essentially the following. In finding $\mathcal{C}_{fb}, \mathcal{C}_{ff}$ that minimize e in some sense, we have several options. First, by feedforward control, zero error tracking could be achieved if one is able to design $\mathcal{C}_{ff} \to \mathcal{P}^{-1}$, hence the feedforward controller matches with the plant inverse. Second, by feedback control, $\mathcal{S}, \mathcal{S}_p$ need to be designed as small as possible, at least in the frequency interval of interest. Third, one may reduce the effect of the inputs r and d themselves, for example by input (or setpoint) shaping and/or disturbance compensation-like schemes [11], [58]. We will focus on the feedforward and feedback options.

C. Control methods: feedforward control

To deal with structural dynamics in the servo problem of stages, both acceleration-snap feedforward control and compliance compensation offer a solution for the quasi-static deformation occurring in these structures under acceleration. To explain the differences, consider the fourth-order (two-mass-spring) plant model as shown in Fig. 23. In the figure, u presents the plant input, which is a force (or a torque) in case of the short-stroke wafer stage. The plant $\mathcal{P}$ in the simplified representation of Fig. 23 is assumed to consist of two point masses: m_1 roughly representing the magnet yoke and m_2 representing the mirror block and remaining stage mass. The stiffness connecting the magnet yoke to the mirror block is modelled by k_{12}. The (non-measured) output at the motor side is represented by y_1, the measured output at the measured side is given by y, whereas the (non-measured) center-of-gravity position is denoted by y_{cog}.

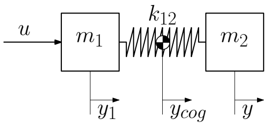

Fig. 23: Fourth-order plant model.

For this simplified system, it follows that

$$\frac{y(s)}{u(s)} = \frac{k_{12}}{m_1 m_2 s^4 + k_{12}(m_1 + m_2)s^2}, \quad (13)$$

$$\frac{y_1(s)}{u(s)} = \frac{m_2 s^2 + k_{12}}{m_1 m_2 s^4 + k_{12}(m_1 + m_2)s^2}, \quad (14)$$

with (13) referring to the non-collocated case of actuation on m_1 and measurement on m_2, and (14) referring to the collocated case where both actuation and measurement refer to the same point mass m_1. If we assume

$$y_{cog}(s) = \frac{m_1 y_1(s) + m_2 y(s)}{m_1 + m_2}, \quad (15)$$

substitution of Eqs. (13) and (14) in (15) gives

$$\frac{y_{cog}(s)}{u(s)} = \frac{1}{(m_1 + m_2)s^2}. \quad (16)$$

As a result, an obvious choice for y_{cog} to tracking r is

$$u_{ff}(s) = (m_1 + m_2)s^2 r(s), \quad (17)$$

which is known as acceleration feedforward and which for $u(t) \to u_{ff}(t)$ results (under appropriate initial conditions) in $y_{cog}(t) \to r(t)$. Acceleration feedforward enables perfect tracking if the wafer stage system behaves as a rigid body. As already shown in Fig. 21, this is only true for low frequencies. For high frequencies, structural flexibilities dominate the system response. One way to cope with these flexibilities is by accounting for quasi-static deformation. This can easily be seen by substituting (15) and (16) back in (13) using $m^* = m_1 m_2/(m_1 + m_2)$ to obtain

$$\begin{aligned} y(s) &= \frac{1}{\frac{m^*}{k_{12}}s^2 + 1} y_{cog}(s) = \left(1 - \frac{m^* s^2}{m^* s^2 + k_{21}}\right) y_{cog}(s) \\ &= \left(\frac{1}{(m_1 + m_2)s^2} - \frac{\frac{m^*}{(m_1 + m_2)}}{m^* s^2 + k_{12}} \right) u(s), \qquad (18) \end{aligned}$$

i.e., y partly moves as a rigid body, but is held back at low frequencies with respect to the center of gravity y_{cog} by the amount of $m_1 m_2/((m_1 + m_2)^2 k_{12})$, which represents the quasi-static deformation of (in principle) all flexible modes. To counteract the quasi-static deformation, the feedforward force u_{ff} in (17) should therefore be multiplied with factor $(m^*/k_{12})s^2 + 1$, which resembles the relation between y and y_{cog} in (18), and which gives

$$\begin{aligned} u_{ff}(s) &= \left(\frac{m^*}{k_{12}}s^2 + 1\right)(m_1 + m_2)s^2 r(s) \\ &= (m_1 + m_2)s^2 r(s) + \frac{m_1 m_2}{k_{12}} s^4 r(s). \qquad (19) \end{aligned}$$

This is known as acceleration-snap feedforward control [18], [12]; similar reasoning applies to the collocated case, *i.e.*, controlling y_1, which is left to the reader. Furthermore, substitution of (19) in (18) with $u_{ff}(s) = u(s)$ gives $y(s) = r(s)$, which in this specific undamped situation apart from the quasi-static deformation also deals with the resonant behavior. Also note that (19) in time-domain requires four times differentiation of the setpoint $r = r(t)$, which is possible with a fourth (or higher) order setpoint. More remarks can be made regarding acceleration-snap feedforward control, but the fact remains that (19) does induce significant high-frequency amplification of the setpoint r. This may pose a severe disadvantage especially when the short-stroke wafer stage plant contains highly undamped resonances, typically being the result of the vacuum environment in which it operates, recall Section IV. For this reason, the wafer scanning industry frequently resorts to other methods for dealing with quasi-static deformation, for example compliance compensation [65], [37].

In compliance compensation, the quasi-static (but time-dependent) deformation $m_1 m_2/((m_1 + m_2)k_{12})\ddot{r}(t)$, which follows from (18) with (17), is added as signal y_{cc} (in software) to the measured plant output signal $\bar{y} = \bar{y}(t)$ to form the signal $y = \bar{y} + y_{cc}$ for servo control. The signal y_{cc} in Laplace domain is given by

$$y_{cc}(s) = \frac{m_1 m_2}{(m_1 + m_2)k_{12}} s^2 r(s). \qquad (20)$$

The difference between acceleration-snap feedforward control and compliance compensation in dealing with quasi-static deformation is illustrated in Fig. 24. In the figure, two paths are indicated (dashed lines): the upper-left path represents the snap feedforward branch with $k_s = m_1 m_2/k_{12}$, the upper-right path represents the extra term in compliance compensation with $k_c = m_1 m_2/((m_1 + m_2)^2 k_{12})$. Note that both schemes use acceleration feedforward as in (17) with $k_a = m_1 + m_2$. At low frequencies, by loop transformation, the snap contribution to u_{ff} can be transformed to the compliance contribution in y_{cc} and vice versa. The difference between both approaches (not directly apparent from the figure) is in the physics. Snap feedforward control as in (19) adds a branch to the feedforward force u_{ff} forcing the true position $\bar{y}$ of $\mathcal{P}$ to overcome the quasi-static deformation during acceleration. Compliance compensation adds a branch (in software) to the measured output signal of the plant $\bar{y}$ instead. With compliance compensation the true plant position $\bar{y}$ still lacks behind by the amount of the quasi-static deformation. Note that with compliance compensation the true output $\bar{y}$ only differs from the (manipulated) servo output y in the time interval of non-zero accelerations $\ddot{r} \neq 0$, where the correction y_{cc} can be seen as a modification of the setpoint r toward $\tilde{r} = r - y_{cc}$. Such a modification is allowed by the grace of scanning under constant velocity, namely during scanning $y_{cc} = 0$, *i.e.*, $\tilde{r} = r$.

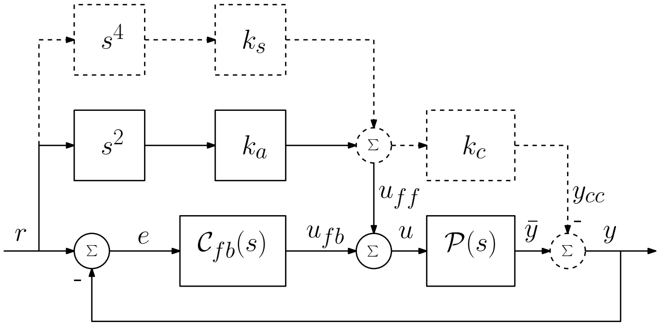

Fig. 24: Simplified block diagram to illustrate acceleration-snap feedforward versus compliance compensation.

D. Control methods: feedback control

To deal with the regulator problem, two feedback control designs will be discussed. Both designs are based on PID control, which is the current standard in the wafer scanning industry. While one design is strictly linear, the other design contains nonlinear integrators, so-called hybrid integrator-gain systems [21]. These enable the user in dealing in a different manner with linear control design limitations like the Bode sensitivity integral in frequency domain or in time domain with overshoot, rise and settling times. Both designs are cast in a robust control framework.

Robust control design with a PID-based controller essentially aims at finding the controller parameters ρ that maximize the controller bandwidth $\omega_b = \omega_b(\rho)$ while satisfying constraints on closed-loop frequency response functions like the input sensitivity $\mathcal{S}(j\omega) = (I + \mathcal{C}_{fb}(j\omega)\mathcal{P}(j\omega))^{-1}$ and the process sensitivity function $\mathcal{S}_p(j\omega) = \mathcal{P}(j\omega)(I + \mathcal{C}_{fb}(j\omega)\mathcal{P}(j\omega))^{-1}$. Bandwidth is defined at the first unity crossing of the open-loop (magnitude) frequency response, *i.e.*, $|\mathcal{C}_{fb}(j\omega_b)\mathcal{P}(j\omega_b)| = 1,\ \omega_b > 0$. The effectiveness of a frequency-domain approach stems from the fact that for wafer stage systems frequency-domain identification is both accurate as well as fast [57]. Another supporting fact is the use of manual loop shaping [59], which is based on a clear physical interpretation regarding control parameter tuning that is understood by the large community of engineers involved in wafer stage systems. Two enablers for manual loop shaping are (a) a sufficient level of decoupling of the MIMO wafer stage plant, thus enabling multi-loop SISO control, and (b) the stage design philosophy that aims at having $1/ms^2$ behavior for as high-frequencies as possible and thus enabling a relatively low-order controller.

The PID-based controller $\mathcal{C}_{fb}$ that is generally used in stage control, has three components: (a) a PID-controller $\mathcal{C}_{pid}$, (b) a low-pass filter $\mathcal{C}_{lp}$, and (c) a set of five notch filters $\mathcal{C}_{n,1}, \dots, \mathcal{C}_{n,5}$. The control parameters are basically chosen according to heuristic design rules that reflect the experiences built within this industry. Both controller components and design rules for the linear control design (as function of bandwidth ω_b) are given as follows:

$$\begin{aligned} \mathcal{C}_{pid}(s) &= k_p \cdot \left(\frac{\omega_i}{s} + 1 + \frac{s}{\omega_d} \right), \\ \mathcal{C}_{lp}(s) &= \frac{\omega_{lp}^2}{s^2 + 2\beta_{lp}\omega_{lp}s + \omega_{lp}^2}, \\ \mathcal{C}_{n,i}(s) &= \frac{\omega_{p,1}^2}{\omega_{z,1}^2} \cdot \frac{s^2 + 2\beta_{z,i}\omega_{z,i}s + \omega_{z,i}^2}{s^2 + 2\beta_{p,i}\omega_{p,1}s + \omega_{p,1}^2}, \end{aligned} \tag{21}$$

with $k_p = 0.52k_a\omega_b^2$ the controller gain, $\omega_i = 0.35\omega_b$ the integrator frequency, $\omega_d = 0.55\omega_b$ the differentiator frequency, $\omega_{lp} = 6.4\omega_b$ the low-pass frequency, $\beta_{lp} = 0.83$ the dimensionless damping coefficient, $\omega_{p,i}, \omega_{z,i}$ the pole and zero frequency of the notch filters, respectively, with $\beta_{p,i}, \beta_{z,i}$ the dimensionless damping coefficients, and $i \in \{1, 2, \dots, 5\}$.

The specific choice of notch filters (and parameters) very much depends on the plant characteristics. Several notch filters are tuned at selected resonance frequencies of $\mathcal{P}$ otherwise limiting bandwidth. In that case, both pole and zero frequencies are identical $\omega_{p,i} = \omega_{z,i}$. Often, one skew notch filter is used with its pole frequency preceding the zero frequency, *i.e.*, $\omega_{p,i} < \omega_{z,i}$. This is done to rotate a resonance occurring in the open-loop frequency response at ω_r with $\omega_{p,i} < \omega_r < \omega_{z,i}$ for some i to the righthand side of the complex plane with the effect of extra closed-loop disturbance suppression at ω_r, *i.e.*, $|\mathcal{S}(j\omega_r)| << 0dB$. In other cases, a skew notch filter in combination with $\beta_{n,i} << 1$ is used to give extra lowpass filtering.

With these controller components and tuning rules, the linear PID-based controller becomes

$$\mathcal{C}_{fb}(s) = \mathcal{C}_{pid}(s)\mathcal{C}_{lp}(s)\Pi_{i=1}^{5}\mathcal{C}_{n,i}(s). \tag{22}$$

The nonlinear PID-based controller uses hybrid integrator-gain systems [21], *i.e.*, nonlinear integrators that use logic to decide between integrator mode or gain mode. A hybrid integrator-gain system $\mathcal{H}$ (abbreviated as HIGS) is defined in time-domain by

$$\mathcal{H}(e, u, \dot{e}) := \begin{cases} \dot{x} &= \omega_h e, \quad \text{if } (e, u, \dot{e}) \in \mathcal{F}_{\omega_h} \\ x &= k_h e, \quad \text{if } (e, u, \dot{e}) \in \mathcal{F}_{k_h} \\ u &= x, \end{cases} \tag{23}$$

with state x, input e with corresponding time derivative $\dot{e}$, control output u, parameters $\omega_h \in [0, \infty)$ and $k_h \in (0, \infty)$ representing the integrator frequency and gain value, respectively, and with $\mathcal{F}_{\omega_h}$ and $\mathcal{F}_{k_h}$ denoting the regions where either integrator or gain mode is active; the initial condition is assumed to be zero, *i.e.*, $x(0) = 0$. $\mathcal{H}$ in (23) is input affine and sign-equivalent, *i.e.*, its input and output e, u always have equal sign. As such, $\mathcal{H}$ can give a clear phase advantage over any linear integrator [52], see for example [21] for further details.

The nonlinear PID-based controller is cast in a similar structure as the linear PID-based controller (22) where (for brevity) the notch filters $\mathcal{C}_{n,i}$ are chosen identical as in (21). The nonlinear PID-controller $\mathcal{C}_{pid}^{\mathcal{H}}$ is represented in time-domain by the block diagram of Fig. 25. Note that the structure of the linear PID controller $\mathcal{C}_{pid}$ along with its parameters is mostly preserved. Different, however, is the integrator branch of $\mathcal{C}_{pid}^{\mathcal{H}}$ (further denoted by $\mathcal{C}_i^{\mathcal{H}}$) that uses a HIGS-filter $\mathcal{H}$ from (23) in series connection with a simple integrator and a matched parallel gain. Note that with respect to the simple integrator in $\mathcal{C}_{pid}$, $\mathcal{C}_i^{\mathcal{H}}$ introduces an extra state.

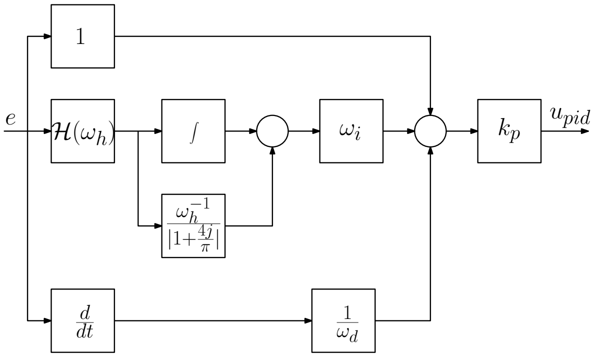

Fig. 25: Block diagram of the HIGS-based PID controller $\mathcal{C}_{pid}^{\mathcal{H}}$ in time-domain.

The choice for the specific structure of the nonlinear integrator $\mathcal{C}_i^{\mathcal{H}}$ is motivated by describing function analysis. The resulting frequency response is shown in the Bode diagram of Fig. 26. In the figure, it can be seen that an

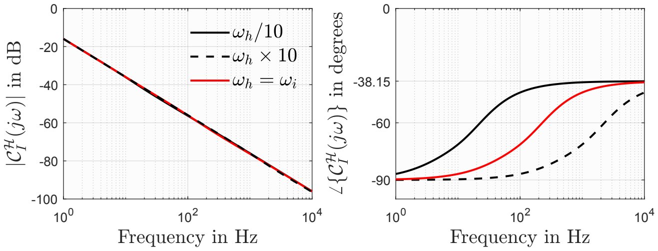

Fig. 26: Bode diagram of the HIGS-based integrator $\mathcal{C}_i^{\mathcal{H}}$ via describing function analysis.

integrator is obtained with the characteristic 20dB/decade amplitude decay similar to a linear integrator, but with a phase lag that can reach ≈ 38.15. It is important to observe that parameter ω_h (with $k_h = 1$) can influence the phase lag reduction independent from amplitude, hence extra freedom in dealing with inherent design limitations; see for example [23] for the describing function of $\mathcal{H}$, which will further be denoted by $\mathcal{D}$. For the HIGS low-pass filter $\mathcal{C}^{\mathcal{H}}_{lp}$ that has the same controller parameters as its linear counterpart $\mathcal{C}_{lp}$ in (21) but with two extra parameters k_h, ω_h (and similarly one extra state), see [39].

Regarding tuning rules, the explicit choice is made of keeping all shared parameters between $\mathcal{C}_{fb}$ and $\mathcal{C}^{\mathcal{H}}_{fb}$ identical. The only choice left then is with the HIGS parameter ω_h of both HIGS elements; note that $k_h = 1$ can be normalized without loss of generality, because $\mathcal{H}$ in (23) is independent of input amplitude. A balanced choice for ω_h in $\mathcal{C}^{\mathcal{H}}_{pid}$ is $\omega_h = \omega_i/8$. For the lowpass filter $\mathcal{C}^{\mathcal{H}}_{lp}$, ω_h is chosen as $\omega_h = \omega_{lp}/3$. With these controller components and tuning rules, the nonlinear PID-based controller becomes

$$\mathcal{C}^{\mathcal{H}}_{fb}(p) = \mathcal{C}^{\mathcal{H}}_{pid}(p)\mathcal{C}^{\mathcal{H}}_{lp}(p)\Pi^{5}_{i=1}\mathcal{C}_{n,i}(p), \tag{24}$$

with p the continuous time-differential operator. A frequency-domain representation of (24) does only exist in an approximate sense, for example through the describing function:

$$\mathcal{D}^{\mathcal{H}}_{fb}(j\omega) \approx \mathcal{D}^{\mathcal{H}}_{pid}(j\omega)\mathcal{D}^{\mathcal{H}}_{lp}(j\omega)\Pi^{5}_{i=1}\mathcal{C}_{n,i}(j\omega), \tag{25}$$

with $\mathcal{D}^{\mathcal{H}}_{fb}$ the describing function of $\mathcal{C}^{\mathcal{H}}_{fb}$, and $\mathcal{D}^{\mathcal{H}}_{pid}, \mathcal{D}^{\mathcal{H}}_{lp}$ the describing functions of $\mathcal{C}_{pid}, \mathcal{C}_{lp}$, respectively. Note that in (25), in general $\mathcal{D}^{\mathcal{H}}_{fb}(j\omega) \neq \mathcal{D}^{\mathcal{H}}_{pid}(j\omega)\mathcal{D}^{\mathcal{H}}_{lp}(j\omega)\Pi^{5}_{i=1}\mathcal{C}_{n,i}(j\omega)$.

To study the possibilities of nonlinear feedback control in dealing with inherent design limitations, consider the simplified wafer stage control system in which the wafer stage plant is assumed to be a mass system, *i.e.*, a double integrator plant of the form $\mathcal{P} \approx 1/(ms^2)$, with $m = 20$ kg. For a desired bandwidth of $\omega_b = 2\pi \times 300$ rad/s and using the earlier discussed tuning rules, the openloop as well as closed loop frequency response functions using either (22) or (25) are shown in Fig. 27. In the Nyquist plots (left), it can be seen that the robustness margins obtained with the linear design (through $\mathcal{C}_{fb}\mathcal{P}$) are all improved by the nonlinear design when viewed through $\mathcal{D}^{\mathcal{H}}_{fb}\mathcal{P}$, *i.e.*, when based on describing functions. More specifically, the modulus margin improves from 4 to 1.5 dB, the phase margin from 38 to 49 degrees, and the gain margin from 19 to 35 dB. In terms of the closed-loop sensitivity functions $\mathcal{S}$ and $\mathcal{S}^{\mathcal{H}}$ (right) the waterbed effect that holds true for the linear design seems to be outperformed in case of the nonlinear design.

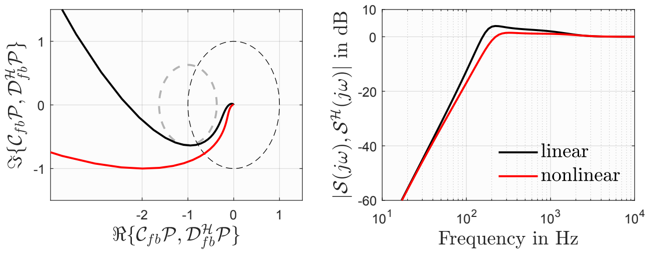

Fig. 27: Nyquist (left) and Bode (right) magnitude of both the linear design with $\mathcal{C}_{fb}\mathcal{P}$ and the HIGS-based design $\mathcal{D}^{\mathcal{H}}_{fb}\mathcal{P}$.

Regarding time-domain behavior consider Fig. 28 which shows the closed-loop responses to (a) a step input (left part), and (b) a sinusoidal input at 200 Hz (right part) either using $\mathcal{C}_{fb}$ from (22) or the true nonlinear controller $\mathcal{C}^{\mathcal{H}}_{fb}$ from (24). In the left plot, it is clear that both overshoot and settling times improve, overshoot is reduced from 82% to 22% while the settling time, *i.e.*, the time needed to reach 5% of the final value, reduces from 6.7 milliseconds to 3.9 milliseconds. The rise time, *i.e.*, the time needed for the response to rise from 10% to 90% of the final value, does increase from 0.4 milliseconds to 0.8 milliseconds. This is inevitable for any HIGS-based control design that bounds its integrator output. In the right plot, the response to a sinusoidal input at 200 Hz clearly illustrates the favorable steady-state properties obtained with the nonlinear design. However, the reduction from 2.6 μm to 1.8 μm, *i.e.*, a 3.2 dB difference, does not match exactly with the sensitivity approximation from Fig. 27. Namely, $|\mathcal{S}(j400\pi)| = 3.6$ dB and $|\mathcal{S}^{\mathcal{H}}(j400\pi)| = -1.4$ dB yields a difference of 5 dB, which gives a 1.8 dB overestimate.

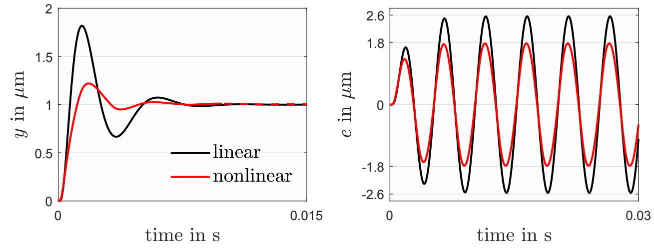

Fig. 28: Time series plot of the closed-loop design to step input (left) and sinusoidal input at 200 Hz (right).

E. Research directions

Challenges for stage control design, as already advocated in the previous sections, can be directed to dealing with (a) stage flexibilities under increasingly aggressive motion profiles, and (b) design limitations from the linear feedback control design that limit increase of control bandwidth.

For the considered feedforward control methods, *i.e.*, acceleration-snap feedforward control and compliance compensation, one can think of several developments. Firstly, dealing with position-dependent control aspects as a result of the quasi-static deformation varying over the wafer scanning position. For compliance compensation, this variation can be stored in look-up tables for a grid of distinct wafer positions. The quasi-static contributions follow from the 'frozen' dynamics, which are assumed to be LTI at these locations. During scanning, position-dependent compliance compensation can then be obtained by interpolating among these values as a function of the setpoints in the x- and y-direction. This disregards transients, *i.e.*, time-varying aspects in the dynamic response during scanning [35], [37].

To deal with these transients, appropriate LPV modelling and control is advocated in [38]. Furthermore, the estimation of quasi-static deformation, which is often obtained from finite element (FE) models, may benefit from recent developments in multisine identification [57], and should at least address the effect of machine-to-machine variation, which is generally not captured in the FE model. In this regard, interesting developments are to be expected from local parametric methods [45]. Secondly, an interesting aspect of stage control development lies in the inferential nature of the scanning process. We cannot measure directly under the lens where the exposure process takes place. Using inferential control methods [66], which are often based on observer design and state estimation [63], one expects to be able to control the area of exposure on the wafer located under the lens. Lastly, developments in (machine) learning control may pave the way to dealing with setpoint variation occurring in wafer stage systems. See for example [55] regarding norm-based ILC approaches with basis functions.

For the considered feedback control methods and in view of inherent design limitations several challenges can be identified. For example, the need for rigorous robust control design that takes both frequency-domain as well as time-domain performance measures into account [23], easy and non-conservative frequency-domain tools for closed-loop stability analysis [42], and formal proofs and insights on the possibilities for nonlinear control design to exceed beyond the possibilities of a linear design. Also, the relation between frequency-domain control design of the quasi-linear system and the corresponding time-domain performance of the hybrid system is still an open issue.

VI. STAGES: PART TWO - CONTROL OF THERMAL-INDUCED DEFORMATION

Both reticle and wafer heat up and locally deform due to the exposure light generated by the light source, recall Section III. This causes overlay and focus errors at wafer level. These problems are referred to as reticle heating and wafer heating, respectively. In this section, problem formulations, solution directions, and control challenges for both heating phenomena will be discussed.

A. Reticle heating problem

Exposing the reticle to ultraviolet light causes reticle heating and leads to deformations in the reticle image field. In turn, these deformations result in distortions of the projected image on the wafer. The most dominant distortion shape is magnification, yet for increasing source power higher order deformations are becoming more and more dominant.

The heating in the reticle stage system is described by the thermo-mechanical dynamics

$$\mathcal{P} := \begin{cases} \dot{T} &= AT + B_Q Q \\ z &= C_z([T^0,\ T^1,\ \ldots,\ T^n])T. \end{cases} \tag{26}$$

These dynamics consist of (a) an LTI thermal system, governed by the differential equation of the temperature field T subject to the heat load Q, and (b) a thermo-mechanical mapping C_z from the reticle temperature field to the image distortions z. The mapping can assumed to be static, though it is nonlinear in T due to the nonlinear expansion behavior of the reticle material. Specifically, this nonlinearity is formulated as a polynomial function in T that approximately describes the expansion coefficient of the reticle material. The (uncorrected) overlay error is a function of the image distortions and is obtained as $OVL = f(z)$ where $f(z) = \max(|z|)$.

Without any corrective measures, the contribution of reticle heating to the overlay error is unacceptably large. An example of the maximum uncorrected overlay error contribution as function of time is shown in Fig. 29 by the blue curve. Typically, this curve is characterized by a sawtooth profile, as a direct consequence of the intermittent heating and cooling of the reticle whenever a wafer is exposed. On top of the sawtooth profile, a global increase of the overlay error over the duration of the so-called wafer lot (multiple wafers being batch-wise exposed) is observed, as a result of the gradual average temperature rise of the reticle. The main goal of reticle heating error correction is to minimize the overlay contribution.

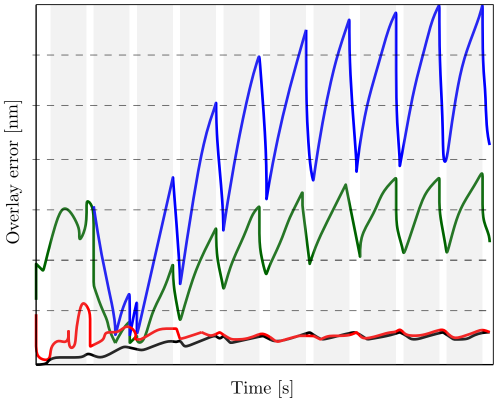

Fig. 29: Overlay error contributions as a function of time (blue: no correction; black: ideal correction; green: TIS-based correction; red: observer-based correction).

B. Reticle heating error correction

This section discusses the main principles of reticle heating error correction (RHEC) [8]. Both a traditional static error correction mechanism and a more advanced observer-based approach are discussed.

Minimizing the overlay contribution is achieved by performing corrective actions with both the reticle and the wafer stage. The set of possible corrective actions that can be taken is limited. The correctable portion of the image distortions is described by $\tilde{z} = \Phi_z\theta$ where $\Phi_z \in \mathbb{R}^{n_z \times n_\theta}$ is a set of basic n_θ correctable shapes and $\theta \in \mathbb{R}^{n_\theta}$ are the associated coefficients. The corrected overlay error is then expressed as $OVL_c = f(z - \tilde{z})$. Supposing that the true image distortions z are known exactly, the coefficients θ are obtained by regression as $\theta = (\Phi_z^T\Phi_z)^{-1}\Phi_z^T z$. The resulting corrected overlay error represents the smallest overlay error that can be achieved given the limited set of correctable shapes in Φ_z. This overlay error is shown by the black line in Fig. 29.

In practice, however, the image distortions z are unknown as these cannot be measured directly. The key problem in RHEC is therefore to accurately estimate the wafer image distortions from available measurement data, and determine the correction coefficients θ. These measurements are available in the form of the readings of the displacements of a set of so-called transmission image sensor (abbreviated by TIS) markers, located just outside the reticle image field. The measurements are described by the output equation

$$y_{TIS} = C_{TIS}([T^0,\ T^1,\ \ldots,\ T^n])T, \tag{27}$$

where C_{TIS} is a nonlinear mapping matrix, similar to C_z in (26). The TIS markers can only be read at a specific instance in between wafer exposures, *i.e.*, during the indents in the sawtooth profile.

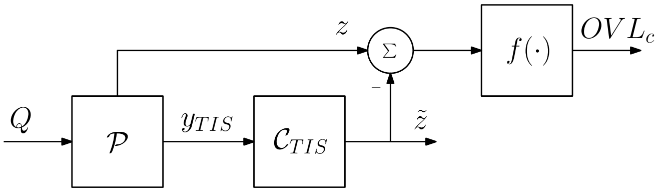

Fig. 30: TIS-based reticle heating error correction.

In traditional RHEC, a static TIS-based correction is performed just before the exposure of each wafer. The method is schematically depicted in Fig. 30. Herein, $\mathcal{C}_{TIS}$ represents the correction mechanism:

$$\mathcal{C}_{TIS}: \quad \tilde{z} = \Phi_z \theta, \quad \theta = (\Phi_y^T \Phi_y)^{-1} \Phi_y^T y_{TIS}, \tag{28}$$

where $\Phi_y \in \mathbb{R}^{n_y \times n_\theta}$ represents the set of correctable shapes evaluated at the marker locations. Although the TIS-based RHEC is relatively simple and effective, it has the disadvantage that the short time-scale dynamic behavior (the sawtooth) cannot be corrected. Moreover, the limited number of TIS markers and their placement outside of the image field result in a limited estimation capacity of the in-field distortions. An example of the achievable overlay error using TIS-based RHEC is shown by the green line in Fig. 29. Clearly, this approach reduces the error, though the limitations are immediately observed from the sawtooth-shaped error profile.

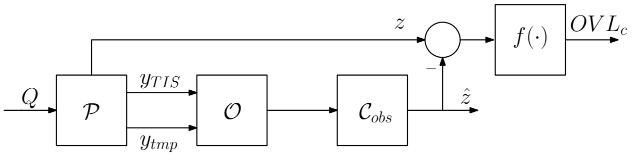

Fig. 31: Observer-based reticle heating error correction.

For EUV machines requiring higher source powers and more stringent performance requirements, the traditional static RHEC approach is expected to no longer suffice, paving the way to more advanced techniques. Such an advanced approach is observer-based RHEC [33], [9], [36]. In this method, a dynamic thermo-mechanical model of (26) runs in parallel to the actual system, producing real-time predictions of the temperature field, deformations, and the resulting image distortions. The model-based prediction errors are corrected for using the TIS marker readings, and additionally real-time temperature measurements $y_{tmp} = C_{tmp}T$. The observer-based RHEC scheme is depicted in Fig. 31. Herein, $\mathcal{O}$ represents the observer, $\hat{z}$ the estimated image distortions and the correction mechanism $\mathcal{C}_{obs}$ is described by

$$\mathcal{C}_{obs}: \quad \tilde{z} = \Phi_z \theta, \quad \theta = (\Phi_z^T \Phi_z)^{-1} \Phi_z^T \hat{z}. \tag{29}$$

Observer-based RHEC has several advantages over its TIS-based counterpart. First of all, the use of a dynamic model enables computation of the distortions not only at the instances of the TIS marker readings, but at each sample during the exposure cycle. Second, the model allows for more accurate estimation of the distortions inside the image field, despite the markers being located outside the field. Third, combining a model with measurement data allows for suppression of measurement noise, *e.g.*, by using Kalman filtering techniques [32], [31]. An example of the achievable overlay error using observer-based RHEC is shown in red in Fig. 29. The overlay error is reduced substantially compared to TIS-based correction and converges toward the best achievable performance during the duration of the wafer lot.

C. Outlook

Observer-based RHEC may be promising for the minimization of reticle heating induced overlay errors. Nevertheless, the fact that the performance of such a method hinges on the availability of a high quality model also introduces many modeling challenges, both from a first principles and an experimental perspective. The basis for the model is typically obtained by first principles modeling. To arrive at models that can be implemented in real-time control schemes, model-order reduction techniques that can effectively reduce the thermo-mechanical model as a whole, including possible nonlinear thermo-mechanical relationships, are indispensable. Still, since a model is never perfect, additional experimental modeling techniques, *e.g.*, system identification and model calibration, are required. Specifically, such identification and calibration techniques must allow for accurate identification of the, possibly nonlinear, standard plant, *i.e.*, the total system including the relations from uncontrollable inputs to unmeasurable outputs. Such calibration methods also require model order reduction techniques that preserve the original interpretation of physical parameters in the reduced order model [44]. To name but of a few of the main challenges in this direction.

D. Wafer heating problem

The net effect of the thermal loads causes a local heating and expansion of the wafer, also called wafer heating. A schematic representation of the wafer and its environment in an EUV system illustrating wafer heating is shown in Fig. 20, where the exposure light (in purple) hits the wafer. If no special measures are taken, the thermal loads will cause an unacceptably large wafer deformation in the slit at the edge of the wafer, see Fig. 32. That is, raw overlay error that occurs before any metrology corrections. To deal with wafer heating, a two-step solution strategy is adopted. First, the heating load due to the exposure light can be actively counteracted by means of a cooling hood (CH) [56]. The cooling hood basically consists of two cold bodies hovering above the wafer. Via pressurized gas between the wafer and the cooling hood to increase the thermal contact, both cold bodies extract heat from the wafer, as close as possible to the slit. The active control strategy for the cooling hood is that the devices before and after the slit (in y-direction) each provide half of the cooling power to compensate for the heating power of the expose load. Moreover, by switching the cooling power simultaneously with the expose light an active control strategy is obtained that renders a heat-neutral situation. However, by the non-ideal location and geometry of the cooling devices, active control via the cooling hood does not work perfectly. Nevertheless, the cooling hood will reduce the raw wafer displacement in the slit down by a factor of ≈10, see Fig. 33. Second, in conjunction with the cooling hood, a form of error correction may be used, specifically an advanced thermo-mechanical model of the wafer and the wafer clamp [19] for real-time-with-preview prediction of the raw wafer displacements in the slit. Subsequently, the correctable part of the predicted wafer displacements can be fed forward to the setpoints for the stages, hence the technique is called wafer heating feedforward correction (WHFFC). Successful WHFFC would reduce the overlay error by an additional factor of ≈6, see Fig. 34.

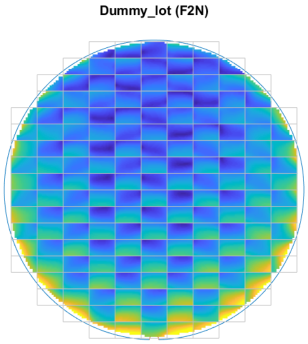

Fig. 32: Overlay error without cooling hood and without metrology corrections.

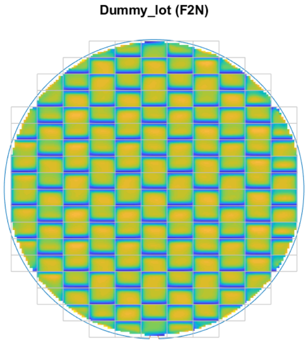

Fig. 33: Overlay error with cooling hood and without metrology corrections.

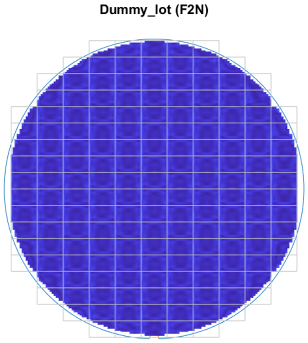

Fig. 34: Overlay error with cooling hood and with metrology corrections.

E. Wafer heating feedforward correction

This section discusses the global idea of wafer heating feedforward correction (or further abbreviated as WHFFC). The basic idea of WHFFC can be explained as follows, see Fig. 35. Let us consider the plant $\mathcal{P}$, which represents the actual hardware of the wafer and its environment, *i.e.*, the mechanics, which is mechanically actuated/sensed by the force actuator/position sensor and thermally actuated/sensed by the cooling hood/temperature sensor. This system is typically controlled using a position feedback controller $\mathcal{C}_{fb}$ and subject to a position setpoint r. In the context of this paper, the plant is subject to the actual disturbances (EUV and IR) that are lumped in thermal power t. The variables of interest, mainly the wafer displacements (or deformations) in the slit, are represented by the performance variables, which refer to overlay metric o. To compensate for the actual heating power, the cooling hood requires a high-quality estimate $\hat{t}$ of the actual heating power, which is provided by the thermo-mechanical model $\hat{\mathcal{P}}$. Subsequently, the cooling hood controller $\mathcal{C}_{ch}$ is providing steering such

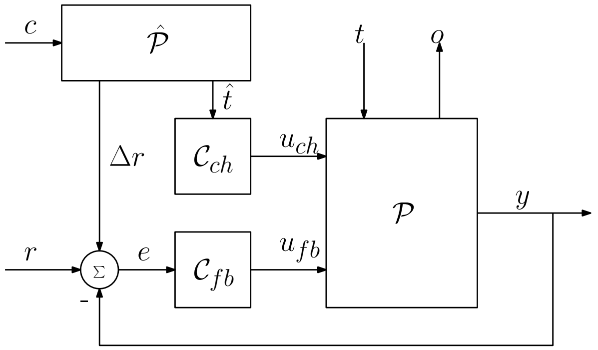

Fig. 35: Illustration of wafer heating feedforward correction (WHFFC).

that the devices before and after the slit (in y-direction) each provide half of the cooling power needed to compensate for the estimated heating power. To arrive at the final performance requirement, the remaining wafer deformation is predicted by the thermo-mechanical model $\hat{\mathcal{P}}$. Subsequently, this information is transferred into a suitable modification Δr of the original setpoint r of the stages to compensate for this deformation. The benefit of this error correction mechanism is that it is able to work with different so-called use cases c, *i.e.*, it is able to cope with different parameters and conditions set by a customer, like specific recipe parameters (field size, EUV dose, routing), specific types of reticles and wafers, and system-specific parameters, *e.g.*, effective EUV source power [24].

The thermo-mechanical model needed for the wafer heating error correction, in its simplest form, can be represented by the following linear state-space description:

$$\begin{cases} \dot{T} & = AT + Bw \\ z & = CT \\ z_{slit} & = Sz. \end{cases} \tag{30}$$

It combines (a) a dynamic differential equation in a set of local temperatures T in the system and (b) a static linear relation that maps the local temperatures T into the local displacements z on the wafer. Moreover, because for overlay the momentary wafer displacements locally in the slit are relevant, in (c) the relevant selection of the wafer displacements is made via a time-dependent (or x/y-position-dependent) selection matrix S. The first equation in (30) is also referred to as the transient (dynamic) thermal model and is composed of thermal system properties. The thermal model parameters appear explicitly in the model that is used for WHFFC via the matrices A and B in (a). Matrix C in the second equation in (30) is called the thermo-mechanical matrix or deformation matrix that is a function of mechanical system properties. The C-matrix is determined by the mechanical model parameters. The relation between a change of the thermal state variables in T and the displacements in z is assumed to occur instantaneously in (30), which is justified for the problem at hand. Due to this, the thermal calculation in (a) and the displacement calculation in (b) and (c) can be separated. The number of state variables in the thermo-mechanical model used for WHFFC is in the order of $20 \cdot 10^3 - 40 \cdot 10^3$, depending on the model resolution.

F. Outlook

Assuming certain physical effects are captured perfectly in a parametric model in a qualitative sense, it is still likely that the numerical values of the parameters do not perfectly match the physical ones. Examples of such imperfect parameter values playing a role in the model for WHFFC are both Young's modulus (E) and Poisson's ratio (ν) of materials, see also [20]. The innovation mechanism requires an adaptation mechanism for the model parameters, based on calibration experiments. The specific adaptation is commanded by the outcome of an (optimization) algorithm that is fed by measured overlay data and use case information. The challenge is found in the effectiveness of the innovation mechanism in terms of achievable accuracy and adaptation speed.

For the thermo-mechanical model for WHFFC to be of high-quality, *i.e.*, for the model to have a sufficiently small mismatch with reality, the resolution of the model must be sufficiently high, *i.e.*, the finite element discretization error must be sufficiently small, which generally implies a large number of state variables in the thermo-mechanical model. Furthermore, the dominant disturbance input is a moving thermal load, which complicates the application of traditional model reduction techniques [62]. The challenge is found in the need to obtain a thermo-mechanical model that simultaneously unifies real-time execution and spatial high-resolution.

VII. OUTLOOK

The control of wafer scanners is expected to continuously benefit from theory and methods originating from the field of systems & control. This involves aspects like linear and nonlinear theory, continuous and digital control, SISO and MIMO control, Kalman filtering, adaptive control, system identification, stochastic control, control of distributed parameter systems, as well as learning control [41]. It also involves many application domains, traditionally from branches of physics like classical mechanics, thermodynamics, electromagnetism and electronics, and optics. Also, other domains like mathematics and in particular statistics play an important role in the sense of statistical process control that only is expected to increase via holistic lithography [25]. Next-generation wafer scanners may also benefit from advances in other fields, for example the field of artificial intelligence (AI), and in particular machine learning and deep neural networks. With the risk of being incomplete, the following directions for control may prove instrumental:

- Plant identification: Identification challenges are found throughout almost all modules of the wafer scanner and range from the identification of highly nonlinear and time-varying plants in the source, identification of complex optics that contain many subsystems, accurate identification of the (possibly nonlinear) standard plant in reticle and wafer heating problems including the relations from inputs that cannot be actuated to unmeasurable outputs, to the identification of the time-varying quasi-static deformation occurring in fast-motion stage systems when scanning along the wafer. Apart from identification in steady-state, identification during transient behavior becomes increasingly important as to save measurement time or to identify systems that never reach a steady state.
- Large-scale optimization: Where in the past wafer scanner modules were largely isolated and given individual sub-specifications to meet overall system performance, next-generation scanners are expected to become increasingly connected either within sub-modules as well as among submodules. This poses many challenges regarding large-scale optimization and control design. In several cases, the connections are made by introducing

for example disturbance feedforward control as to avoid any impact on the closed-loop stability properties of the connected module [6], [27]. In other cases, the connection involves two-way interaction inducing MIMO closed-loop stability aspects that may significantly increase the complexity of the control design [43]. The increasing difficulty (or often impossibility) to develop a sufficiently detailed system-of-systems (SoS) model has motivated some researchers to pose system-of-systems control as an open problem [30]. A solution to this problem is expected to contribute significantly to the control of next-generation wafer scanners.

- Nonlinear control design: Though for valid reasons of predictability, maintenance, and ease of design many wafer scanner components are desirably designed as linear as possible [50], increasingly more nonlinear components enter the machine as to meet increased performance specifications. Examples are given by reluctance and piezoelectric actuators, Peltier elements, but also spindles having friction and being introduced to reduce cost, or inherently nonlinear systems like the laser in the source, the expansion behavior of the reticle material, or the cables and hoses needed to transport (cooling) fluids and electrical power to the moving stages. Apart from the need for nonlinear control design to deal with nonlinear plant behavior, nonlinear control is expected to provide an increasingly-relevant alternative to deal with performance limitations occurring in the linear control design [52], *i.e.*, the waterbed effect, Bode's gain-phase relationship, and transient effects like overshoot and settling times, recall Section IV. Herein formalizing performance limitations of the nonlinear control design [28], especially in the frequency domain, remains an open issue.
- Parameter and disturbance estimation: Be it energy and timing control in the source laser, deformation control of frames and connections in the optics that are performance-relevant but typically too small to measure, (quasi-static) deformation control of motion stages regarding point-of-interest, or thermal deformation control, developments in estimation and model reduction techniques are increasingly required in the control of wafer scanners. Especially, regarding techniques that preserve the original interpretation of the physical parameters in the (reduced) model [44]. Furthermore, the model should capture nonlinear and time-varying aspects, should be real-time executable and, in case of thermal deformation control, should enable a sufficiently high spatial resolution. These properties are not yet found in current reconstruction and (state) estimation techniques.

REFERENCES

[1] R.S. Abhari, B. Rollinger, A.Z. Giovannini, O. Morris, I. Henderson, S.S. Ellwi. (2012) Laser-produced plasma light source for extreme-ultraviolet lithography applications, Journal of Micro/Nanolithography, MEMS, and MOEMS, 11(2), pp. 021114-1 - 021114-13.

[2] P. Alagna, O. Zurita, V. Timoshkov, P. Wong, G. Rechtsteiner, J. Baselmans, J. Mailfert. (2015) Optimum ArFi laser bandwidth for 10nm node logic imaging performance. In Proceedings of the SPIE Optical Microlithography XXVIII, San Jose, CA, 942609.

[3] ASML. (2017) A backgrounder on extreme ultraviolet (EUV) lithography, ASML website https://medium.com/@ASMLcompany/a-backgrounder-on-extreme-ultraviolet-euv-lithography-a5fccb8e99f4.

[4] V. Bakshi (editor). (2018) EUV Lithography, Second Edition. SPIE Press, ISBN 9781510616783.

[5] M.A. Beijen. (2018) Disturbance feedforward control for vibration isolation systems : analysis, design, and implementation, PhD thesis, TU Eindhoven, The Netherlands.

[6] M.A. Beijen, M.F. Heertjes, J. van Dijk, W.B.J. Hakvoort. (2018) Self-tuning MIMO disturbance feedforward control for active hard-mounted vibration isolators, Control Engineering Practice, 72, pp. 90-103.

[7] M.A. Beijen, R. Voorhoeve, M.F. Heertjes, T.A.E. Oomen. (2018) Experimental estimation of transmissibility matrices for industrial multi-axis vibration isolation systems. Mechanical Systems and Signal Processing, 107, pp. 469-483.

[8] C. Bikcora. (2013). Modeling and prediction of thermally induced imaging errors in deep- and extreme-ultraviolet lithography. Master Thesis, Eindhoven University of Technology, Eindhoven.

[9] C. Bikcora, S. Weiland, W.M.J. Coene. (2014) Thermal deformation prediction in reticles for extreme ultraviolet lithography based on a measurement-dependent low-order model, IEEE Transactions on Semiconductor Manufacturing, 27(1), pp. 104-117.

[10] G.M. Blumenstock, C. Meinert, N.R. Farrar, A. Yen. (2014) Evolution of light source technology to support immersion and EUV lithography. In Proceedings of the SPIE, 5645, pp. 188-195.

[11] F.A.J. Boeren, D.J.H. Bruijnen, N. van Dijk, T.A.E. Oomen. (2014) Joint input shaping and feedforward for point-to-point motion: automated tuning for an industrial nanopositioning system, Transactions on Control Systems Technology, 24(6), pp. 572-581.

[12] M. Boerlage, R. Tousain, M. Steinbuch. (2004) Jerk derivative feedforward control for motion systems, In Proceedings of the American Control Conference, Boston, MA, pp. 4843-4848.

[13] D.A. Bristow, M. Tharayil, A.G. Alleyne. (2006) A survey of iterative learning control. IEEE Control Systems Magazine, 26(3), pp. 96-114.

[14] T. Brunner, D. Corliss, S. Butt, T. Wiltshire, C.P. Ausschnitt, M. Smith. (2006) Laser bandwidth and other sources of focus blur in lithography. Journal of Microlithography, Microfabrication, and Microsystems, 5(4), pp. 043003-1 - 043003-7.

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

[16] H. Butler. (2018) Support structure, method and lithographic apparatus, Patent, EP2018/057887.

[17] T. Cacouris, R. Rao, R. Rokitski, R. Jiang, J. Melchior, K. O'Brien. (2012) Advanced light source technologies that enable high-volume manufacturing of DUV lithography extensions, Advanced Lithography, DOI 10.1117/12.917827.

[18] B.H. Chang, Y. Hori. (2003) Trajectory Design considering Derivative of Jerk for Head-positioning of Disk Drive System with Mechanical Vibration. In Proceedings of the American Control Conference, Denver, CO, pp. 4335-4340.

[19] J. Chang, C.J. Martin, R.L. Engelstad, E.G. Lovell. (2002). Thermomechanical global response of the EUVL wafer during exposure, In Proceedings of the SPIE, 4688, pp. 755-766.

[20] J. Chang, R.L. Engelstad, E.G. Lovell. (2003). Comparison of EUV and optical device wafer heating, In Proceedings of the SPIE, 5037, pp. 690-699.

[21] D.A. Deenen, M.F. Heertjes, W.P.M.H. Heemels, and H. Nijmeijer. (2017) Hybrid integrator design for enhanced tracking in motion control. In Proceedings of the American Control Conference, Seattle, WA, pp. 2863-2868.

[22] W.J. Dunstan, R. Jacques, R. Rafac, R. Rao, F. Trintchouk. (2006) Active spectral control of DUV light sources for OPE minimization. In Proceedings of the SPIE Optical Microlithography XIX, San Jose, CA, 61542J.

[23] S.J.A.M. van den Eijnden, M.F. Heertjes, H. Nijmeijer. (2019) Robust stability and nonlinear loop-shaping design for hybrid integrator-gain-based control systems. In Proceedings of the American Control Conference, Philadelphia, PA, pp. 3063-3068.

[24] I. Fomenkov. (2017) EUV Lithography: progress in LPP source power

scaling and availability, In Proceedings of the International Workshop on EUV Lithography, Berkeley, CA, presentation.

[25] C. Fouquet. (2014) Holistic Lithography, Investor day, ASML small talk 2014, London, UK.

[26] R.H. French *et al.*. (2005) Imaging of 32-nm 1:1 lines and spaces using 193-nm immersion interference lithography with second-generation immersion fluids to achieve a numerical aperture of 1.5 and a k_1 of 0.25. Journal of Microlithography Microfabrication and Microsystems, 4(3), pp. 031103-1-14.

[27] M.F. Heertjes, B. Temizer, M. Schneiders. (2013). Self-tuning in master–slave synchronization of high-precision stage systems. Control Engineering Practice, 21(12), pp. 1706–1715.

[28] B.G.B. Hunnekens, N. van de Wouw, H. Nijmeijer. (2012) Variable gain motion control for transient performance improvement. In Proc. of the American Control Conference, Montréal, Canada, pp. 2467-2472.

[29] J. Holterman, T.J.A. de Vries. (2004) Active damping within an advanced microlithography system using piezoelectric smart disc, Elsevier Mechatronics, 14, pp. 15-34.

[30] M. Jamshidi. (2009) Control of system of systems, In Proceedings of the International Conference of Industrial Informatics, Cardiff, Wales, UK, pp. 1-16.

[31] S. Julier, J. Uhlmann (1997). A New Extension of the Kalman Filter to Nonlinear Systems. In Proceedings of the SPIE, 3068, pp 182-1932.

[32] R. Kalman (1960). A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME – Journal of Basic Engineering. Nr 82 (Series D), pp 35-45.

[33] N. Kant, M.J.H. Luttikhof. (2017) Method of reducing effects of reticle heating and/or cooling in a lithographic process. Patent, WO 2017/050523 A1.

[34] D. Karnopp, M.J. Crosby, R.A. Harward. (1974) Vibration control using semi-active force generators, Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control, 96, pp. 619–626.

[35] Y. Kasemsinsup, M.F. Heertjes, H. Butler, S. Weiland. (2016) Exact plant inversion of flexible motion systems with a time-varying state-to-output map. In Proceedings of the European Control Conference, Aalborg, Denmark, pp. 2483-2488.

[36] Y.H. Kim *et al.*. (2017). Reticle heating feed-forward control (RHC2) on NXT:1980Di immersion scanner for enhanced on-product overlay. In Proceedings of the SPIE Advanced Lithography, San Jose, CA, 10147, pp. 17-1-17-6.

[37] N. Kontaras, M.F. Heertjes, H.J. Zwart. (2016) Continuous compliance compensation of position-dependent flexible structures. IFAC-PapersOnLine, 49(13), pp. 76-81.

[38] N. Kontaras, M.F. Heertjes, H.J. Zwart, M. Steinbuch. (2018) Resonant-dynamics LTV feedforward for flexible motion systems, In Proceedings of the 2018 American Control Conference, Milwaukee, WI, USA, pp. 6012-6017.

[39] S.J.A.M. van den Eijnden, Y. Knops, M.F. Heertjes. (2019) A hybrid integrator-gain based low-pass filter for nonlinear motion control. In Proceedings of the Conference on Control Technology and Applications (CCTA), Copenhagen, Denmark, pp. 1108-1113.

[40] H.J. Levinson. (2010) Principles of Lithography; third edition. SPIE press, Bellingham, Washington, USA.

[41] W.S. Levine (editor). (2010) The control handbook: 2nd edition. CRC Press - IEEE press, Boca Raton, Florida, USA, ISBN 9781420073669.

[42] S.J.L.M. van Loon, K.G.J. Gruntjens, M.F. Heertjes, N. van de Wouw, W.P.M.H. Heemels. (2017) Frequency-domain tools for stability analysis of reset control systems. Automatica, 82, pp. 101-108.

[43] V.A. Looijen, M.F. Heertjes. (2018) Robust synchronization of motion in wafer scanners using particle swarm optimization. In Proceedings of the Conference on Control Technology and Applications, Copenhagen, Denmark, pp. 1102-1107.

[44] D. Lou, S. Weiland. (2018) Parametric model order reduction for large-scale and complex thermal systems. 2018 European Control Conference (ECC), Limassol, pp. 2593-2598.

[45] R. van der Maas, A. van der Maas, R. Voorhoeve, T.A.E. Oomen. (2017) Accurate FRF identification of LPV systems: nD-LPM with application to a medical x-ray system, IEEE Transactions on Control Systems Technology, 25(5), pp. 1724-1735.

[46] F. Matrichard, M. Evans. (2015) Review: tilt-free low-noise seismometry. Bulletin of the Seismological Society of America, 105(2A), pp. 497-510.

[47] H. Megens. (2007) An introduction to photolithography: overlay, ASML Images, Fall Edition 2007, Veldhoven, The Netherlands.

[48] T. Mizuno, T. Tourniya, M. Takasaki. (2003) Vibration isolation system using negative stiffness, JSME International Journal, Series C, Mechanical systems, machine elements and manufacturing, 46(3), pp. 807-812.

[49] G.E. Moore. (1965) Cramming more components on integrated circuits, Electronics, 38(8), pp. 1-3.

[50] R. Munnig-Schmidt, G. Schitter, J. van Eijk. (2011) The design of high performance mechatronics. Delft University Press, Delft, The Netherlands.

[51] G. Van der Poel. (2010) An exploration of active hard mount vibration isolation for precision equipment, Ph.D. thesis, University of Twente, Enschede, The Netherlands.

[52] C. Prieur, I. Queinnec, S. Tarbouriech, L. Zaccarian. (2018) Analysis and synthesis of reset control systems. In Foundations and Trends in Systems and Control, 6(2-3), pp. 117-338.

[53] A. Preumont. (2011) Vibration Control of Active Structures - Third Edition. Springer-Verlag, Berlin Heidelberg.

[54] D.J. Riggs, R.R. Bitmead. (2009) Rejection of aliased disturbances in a pulsed light source. In Proceedings of the Conference on Decision and Control, Shanghai, China, pp. 8148-8153.

[55] R. de Rozario, R. Pelzer, S. Koekebakker, T. Oomen. (2018) Accommodating trial-varying tasks in iterative learning control for LPV systems, applied to printer sheet positioning, In Proceedings of the American Control Conference, Milwaukee, WI, pp. 5213-5218.

[56] J. van Schoot *et al.*. (2017) High-NA EUV lithography - enabling Moore's law in the next decade, In Proceedings of the International Workshop on EUV Lithography, Berkeley, CA, presentation.

[57] J. Schoukens, K. Godfrey, M. Schoukens. (2018). Nonparametric data-driven modeling of linear systems: estimating the frequency response and impulse response function. IEEE Control Systems Magazine, 38(4), pp. 49-88.

[58] N. Singer, W. Singhose, W. Seering. (1999) Comparison of filtering methods for reducing residual vibration. European Journal of Control, 5, pp. 208-218.

[59] M. Steinbuch, M. Norg. (1998) Advanced motion control: an industrial perspective, European Journal of Control, 4, pp. 278-293.

[60] Y.R. Teo, A.J. Fleming. (2015) Optimal integral force feedback for active vibration control, Journal of Sound and Vibration, 356, pp. 20–33.

[61] H.L. Trentelman, A.A. Stoorvogel, M.L.J. Hautus. (2001) Control theory for linear systems, Springer, London.

[62] D.W.M. Veldman, R.H.B. Fey, H.J. Zwart, M.M.J. van de Wal, J.D.B.J. van den Boom, H. Nijmeijer. (2018) Semi-analytic approximation of the temperature field resulting from moving heat loads, International Journal of Heat and Mass Transfer, 122, pp. 128-137.

[63] K.W. Verkerk, J. Achterberg, C.M.M. van Lierop, S. Weiland. (2012) Modal observer design for a flexible motion system with state dependent sensor positions. In Proceedings of the Conference on Decision and Control, Maui, Hawaii, pp. 4498-4504.

[64] B. Voigtländer, P. Coenen, V. Cherepanov, P. Borgens, T. Duden, F.S. Tautz. (2017) Low vibration laboratory with a single-stage vibration isolation for microscopy applications, Review of Scientific Instruments, 88(2), pp. 023703-1-023703-7.

[65] M.J. Vervoordeldonk, M.C.J. Baggen. (2009) Position control system, a lithographic apparatus and a method for controlling a position of a movable object, Patent, US 20090268185-A1.

[66] R. Voorhoeve, N. Dirkx, T. Melief, W.H.T.M. Aangenent, T.A.E. Oomen. (2016) Estimating structural deformations for inferential control: a disturbance observer approach, IFAC-PapersOnLine, 49(21), pp. 642-648.

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