

CONTROLLING A POSITION OF A MASS, ESPECIALLY IN A LITHOGRAPHIC APPARATUS

5 The invention relates to a controller arranged for controlling a position of a mass by providing the mass a mass acceleration by a control force depending on a desired mass acceleration. The invention is especially applicable in controlling a position of a substrate table or a mask table in a lithographic apparatus. Such a lithographic apparatus comprises:

- 10 - a radiation system for providing a projection beam of radiation;
 - a support structure for supporting patterning means, the patterning means serving to pattern the projection beam according to a desired pattern;
 - a substrate table for holding a substrate; and
 - a projection system for projecting the patterned beam onto a target portion of the
15 substrate.

 In such a lithographic apparatus, a substrate table supporting a substrate is moved in an XY operating region by actuators controlled by a controller. It has been observed that the controller error behaviour directly after accelerating to a constant velocity
20 depends on the position of the substrate table in the XY operating region. It also depends on the exact mass that is to be moved, which varies due to, for example, variations in the mass of the substrate. It is an object of the present invention to improve the controller in this respect.

 To that end the invention provides, in general terms, a controller as defined at the
25 outset wherein the controller is arranged to receive a feedback signal comprising status information of said mass, to calculate an estimated relation between the mass acceleration and said control force from said feedback signal and from said control force, and to use said estimated relation and said desired mass acceleration to determine said control force. Preferably, the status information comprises an indication of the
30 position, the speed and/or the acceleration of the mass. In a particular embodiment the feedback signal is a feedback acceleration signal of the mass (the acceleration signal being measured or calculated).

In such a controller, feedforward of the acceleration force adapts to position-dependent behaviour and mass variations, and the controller error becomes smaller and less dependent on the position of the table in its operating region.

5 In an embodiment, the estimated relation may be an estimated mass. This embodiment is applicable where the mass is operating as a rigid body.

In an embodiment, the controller may be arranged to calculate at least one of an estimated velocity coefficient, an estimated jerk coefficient, and an estimated snap coefficient from the feedback position signal and the control force, with the aim of either creating a better estimate of the mass, and/or to use at least one of the estimated
10 velocity coefficient, the estimated jerk coefficient, and the estimated snap coefficient to partly determine the control force. This may further improve the accuracy of the controller.

In another embodiment, the controller is arranged to calculate estimated filter coefficients of a general filter structure, such that the resulting filter describes the
15 relation between the acceleration of the mass and the applied control force. The controller is further arranged to use the estimated filter coefficients and the desired mass acceleration to partly determine the control force.

In a further embodiment, the invention relates to a lithographic projection apparatus comprising:

- 20 - a radiation system for providing a projection beam of radiation;
- a support structure for supporting patterning means, the patterning means serving to pattern the projection beam according to a desired pattern;
- a substrate table for holding a substrate; and
- a projection system for projecting the patterned beam onto a target portion of the
25 substrate,
- a controller as defined above, the mass being a movable object in the lithographic apparatus.

The invention also relates to a method of controlling a position of a mass by providing the mass a mass acceleration by a control force depending on a desired mass
30 acceleration characterised by receiving a feedback position signal indicating a position of the mass, calculating an estimated relation between the mass acceleration and the control force from the feedback position signal and from the control force and using the estimated relation and the desired mass acceleration to determine the control force.

In a further embodiment, the invention uses such a method in a device manufacturing method comprising:

- providing a substrate that is at least partially covered by a layer of radiation-sensitive material;
- 5 - providing a projection beam of radiation using a radiation system;
- using patterning means to endow the projection beam with a pattern in its cross-section; and
- projecting the patterned beam of radiation onto a target portion of the layer of radiation-sensitive material,
- 10 - controlling a position of the mass, the mass being at least one of the substrate table with the substrate and the support structure with the patterning means.

It is to be understood that the term "patterning means" as employed should be broadly interpreted as referring to means that can be used to endow an incoming
15 radiation beam with a patterned cross-section, corresponding to a pattern that is to be created in a target portion of the substrate; the term "light valve" can also be used in this context. Generally, the said pattern will correspond to a particular functional layer in a device being created in the target portion, such as an integrated circuit or other device (see below). Examples of such patterning means include:

- 20 - A mask. The concept of a mask is well known in lithography, and it includes mask types such as binary, alternating phase-shift, and attenuated phase-shift, as well as various hybrid mask types. Placement of such a mask in the radiation beam causes selective transmission (in the case of a transmissive mask) or reflection (in the case of a reflective mask) of the radiation impinging on the mask, according to the
25 pattern on the mask. In the case of a mask, the support structure will generally be a mask table, which ensures that the mask can be held at a desired position in the incoming radiation beam, and that it can be moved relative to the beam if so desired;
- A programmable mirror array. One example of such a device is a matrix-addressable surface having a viscoelastic control layer and a reflective surface. The
30 basic principle behind such an apparatus is that (for example) addressed areas of the reflective surface reflect incident light as diffracted light, whereas unaddressed areas reflect incident light as undiffracted light. Using an appropriate filter, the said undiffracted light can be filtered out of the reflected beam, leaving only the diffracted

light behind; in this manner, the beam becomes patterned according to the addressing pattern of the matrix-addressable surface. An alternative embodiment of a programmable mirror array employs a matrix arrangement of tiny mirrors, each of which can be individually tilted about an axis by applying a suitable localized electric field, or by employing piezoelectric actuation means. Once again, the mirrors are matrix-addressable, such that addressed mirrors will reflect an incoming radiation beam in a different direction to unaddressed mirrors; in this manner, the reflected beam is patterned according to the addressing pattern of the matrix-addressable mirrors. The required matrix addressing can be performed using suitable electronic means. In both of the situations described hereabove, the patterning means can comprise one or more programmable mirror arrays. More information on mirror arrays as here referred to can be gleaned, for example, from United States Patents US 5,296,891 and US 5,523,193, and PCT patent applications WO 98/38597 and WO 98/33096, which are incorporated herein by reference. In the case of a programmable mirror array, the said support structure may be embodied as a frame or table, for example, which may be fixed or movable as required; and

- A programmable LCD array. An example of such a construction is given in United States Patent US 5,229,872, which is incorporated herein by reference. As above, the support structure in this case may be embodied as a frame or table, for example, which may be fixed or movable as required.

For purposes of simplicity, the rest of this text may, at certain locations, specifically direct itself to examples involving a mask and mask table; however, the general principles discussed in such instances should be seen in the broader context of the patterning means as hereabove set forth.

Lithographic projection apparatus can be used, for example, in the manufacture of integrated circuits (ICs). In such a case, the patterning means may generate a circuit pattern corresponding to an individual layer of the IC, and this pattern can be imaged onto a target portion (e.g. comprising one or more dies) on a substrate (silicon wafer) that has been coated with a layer of radiation-sensitive material (resist). In general, a single wafer will contain a whole network of adjacent target portions that are successively irradiated via the projection system, one at a time. In current apparatus, employing patterning by a mask on a mask table, a distinction can be made between two different types of machine. In one type of lithographic projection apparatus, each

target portion is irradiated by exposing the entire mask pattern onto the target portion in one go; such an apparatus is commonly referred to as a wafer stepper or step-and-repeat apparatus. In an alternative apparatus — commonly referred to as a step-and-scan apparatus — each target portion is irradiated by progressively scanning the mask pattern under the projection beam in a given reference direction (the "scanning" direction) while synchronously scanning the substrate table parallel or anti-parallel to this direction; since, in general, the projection system will have a magnification factor M (generally < 1), the speed V at which the substrate table is scanned will be a factor M times that at which the mask table is scanned. More information with regard to lithographic devices as here described can be gleaned, for example, from US 6,046,792, incorporated herein by reference.

In a manufacturing process using a lithographic projection apparatus, a pattern (e.g. in a mask) is imaged onto a substrate that is at least partially covered by a layer of radiation-sensitive material (resist). Prior to this imaging step, the substrate may undergo various procedures, such as priming, resist coating and a soft bake. After exposure, the substrate may be subjected to other procedures, such as a post-exposure bake (PEB), development, a hard bake and measurement/inspection of the imaged features. This array of procedures is used as a basis to pattern an individual layer of a device, e.g. an IC. Such a patterned layer may then undergo various processes such as etching, ion-implantation (doping), metallization, oxidation, chemo-mechanical polishing, etc., all intended to finish off an individual layer. If several layers are required, then the whole procedure, or a variant thereof, will have to be repeated for each new layer. Eventually, an array of devices will be present on the substrate (wafer). These devices are then separated from one another by a technique such as dicing or sawing, whence the individual devices can be mounted on a carrier, connected to pins, etc. Further information regarding such processes can be obtained, for example, from the book "Microchip Fabrication: A Practical Guide to Semiconductor Processing", Third Edition, by Peter van Zant, McGraw Hill Publishing Co., 1997, ISBN 0-07-067250-4, incorporated herein by reference.

For the sake of simplicity, the projection system may hereinafter be referred to as the "lens"; however, this term should be broadly interpreted as encompassing various types of projection system, including refractive optics, reflective optics, and catadioptric systems, for example. The radiation system may also include components

operating according to any of these design types for directing, shaping or controlling the projection beam of radiation, and such components may also be referred to below, collectively or singularly, as a "lens". Further, the lithographic apparatus may be of a type having two or more substrate tables (and/or two or more mask tables). In such
5 "multiple stage" devices the additional tables may be used in parallel, or preparatory steps may be carried out on one or more tables while one or more other tables are being used for exposures. Dual stage lithographic apparatus are described, for example, in US 5,969,441 and WO 98/40791, both incorporated herein by reference.

The projection system as described above usually comprises one or more, for
10 instance six, projection devices, such as lenses and/or mirrors. The projection devices transmit the projection beam through the projection system and direct it to the target portion. In case the projection beam is EUV-radiation, mirrors should be used instead of lenses, in order to project the projection beam, since lenses are not translucent to EUV-radiation.

15 When an extreme ultraviolet projection beam is used for projecting relatively small patterns, the demands for the projection system concerning the accuracy are rather high. For instance, a mirror, that is positioned with a tilting error of 1 nm, can result in a projection error of approximately 4 nm on the wafer.

A projection system for projecting an extreme ultraviolet projection beam
20 comprises, for instance, 6 mirrors. Usually, one of the mirrors has a fixed spatial orientation, while the other five are mounted on a Lorentz actuated mount. These mounts can preferably adjust the orientation of the mirrors in 6 degrees of freedom (6-DoF-mounts) using 6 degrees Lorentz engines per mirror. The projection system further comprises sensors for measuring the spatial orientation of the projection
25 devices.

The projection system is mounted to the fixed world, for instance, a metro frame, using a 30 Hz mounting device. This is done to stabilize the projection beam and isolate it from vibrations and disturbances coming from the environment, such as adjacent systems. As a result of this mounting, unwanted disturbances above the 30 Hz
30 are almost completely filtered out. However, disturbances having a frequency of approximately 30 Hz, are not stopped by the mounting device and are even amplified.

Although specific reference may be made in this text to the use of the apparatus according to the invention in the manufacture of ICs, it should be explicitly understood

that such an apparatus has many other possible applications. For example, it may be employed in the manufacture of integrated optical systems, guidance and detection patterns for magnetic domain memories, liquid-crystal display panels, thin-film magnetic heads, etc. The skilled artisan will appreciate that, in the context of such
5 alternative applications, any use of the terms "reticle", "wafer" or "die" in this text should be considered as being replaced by the more general terms "mask", "substrate" and "target portion", respectively.

In the present document, the terms "radiation" and "beam" are used to encompass all types of electromagnetic radiation, including ultraviolet (UV) radiation (e.g. with a
10 wavelength of 365, 248, 193, 157 or 126 nm) and extreme ultra-violet (EUV) radiation (e.g. having a wavelength in the range 5-20 nm), as well as particle beams, such as ion beams or electron beams.

The invention will now be explained in connection with the accompanying drawings, which are only intended to show examples and not to limit the scope of
15 protection, and in which:

Figure 1 is a schematic general overview of a lithographic projection apparatus;
Figure 2 shows a control architecture according to the state of the art;
Figure 3 shows a basic scheme for on-line mass estimation in accordance with the
20 invention;
Figure 4 shows a circuit for generation of an estimation error;
Figure 5 shows several curves for mass estimation on a wafer stage. The curves include a set point acceleration curve, a mass estimation curve, a mass estimation curve using high-pass filters, and a mass estimation using an offset estimation;
25 Figure 6 shows an example of mass estimation and feedforward;
Figure 7 shows curves for the control error without and with mass estimation feedforward;
Figure 8 shows a circuit that can be used to estimate velocity, acceleration, jerk and snap feedforward;
30 Figure 9 shows curves for mass estimation when various other parameters are estimated;
Figure 10 shows curves for estimations of velocity, acceleration, jerk and snap components;

- Figure 11 shows a feed forward circuit for mass estimation and snap feedforward;
Figure 12a shows the controller error with and without snap feed forward, whereas
figure 12b shows a portion of figure 12a on an enlarged time scale;
Figure 13 shows mass estimation with and without snap feedforward;
- 5 Figure 14a shows the controller error with and without snap feedforward when
using 10 Hz high-pass filters for the estimation, whereas figure 14b shows a
portion of figure 14a on an enlarged time scale;
Figure 15 shows mass estimation using 10 Hz high-pass filters, without and with
snap feedforward;
- 10 Figure 16 shows mass estimation with 0.1 samples less delay in force path;
Figure 17 shows several curves for the exposed chuck during a negative move;
Figure 18 shows several curves for the exposed chuck during a positive move;
Figure 19 shows several curves for the measure chuck during a negative move;
Figure 20 shows several curves for the measure chuck during a positive move;
- 15 Figure 21 shows the effect of switching on the estimator only at maximum
acceleration;
Figure 22 shows an alternative least-square estimation for 1 parameter;
Figure 23 shows an alternative least-square estimation for 1 parameter with
forgetting factor;
- 20 Figure 24 shows a simplified offset estimation scheme;
Figure 25 shows an ARX filter structure for a simplified implementation to find the
optimal estimation mass;
Figure 26 shows a FIR filter architecture taking a forgetting factor into account;
Figure 27 shows the magnitude and phase of an estimated transfer function in an
alternative approach; the situation is shown for both a FIR filter and an ARX filter;
- 25 Figure 28 shows the feedforward for the FIR and ARX filters used in figure 27;
Figure 29 shows the estimated mass for the situation corresponding with figures 27
and 29;
Figures 30, 31 and 32 are showing curves which are similar as the curves in figures
30 27, 28 and 29, respectively, be it that the filters used are of a much lower order;
Figure 33 shows a circuit architecture with on-line feedforward estimation.

Figure 1 schematically depicts a lithographic projection apparatus 1 according to a particular embodiment of the invention.

The apparatus comprises:

- a radiation system Ex, IL, for supplying a projection beam PB of radiation (e.g. EUV radiation with a wavelength of 11-14 nm). In this particular case, the radiation system also comprises a radiation source LA;
- a first object table (mask table) MT provided with a mask holder for holding a mask MA (e.g. a reticle), and connected to first positioning means PM for accurately positioning the mask with respect to item PL;
- 10 - a second object table (substrate table) WT provided with a substrate holder for holding a substrate W (e.g. a resist-coated silicon wafer), and connected to second positioning means PW for accurately positioning the substrate with respect to item PL; and
- a projection system ("lens") PL for imaging an irradiated portion of the mask MA
- 15 onto a target portion C (e.g. comprising one or more dies) of the substrate W.

As here depicted, the apparatus is of a reflective type (i.e. has a reflective mask). However, in general, it may also be of a transmissive type, for example (with a transmissive mask). Alternatively, the apparatus may employ another kind of patterning means, such as a programmable mirror array of a type as referred to above.

- 20 The source LA (e.g. a laser-produced plasma or a discharge plasma EUV radiation source) produces a beam of radiation. This beam is fed into an illumination system (illuminator) IL, either directly or after having traversed conditioning means, such as a beam expander Ex, for example. The illuminator IL may comprise adjusting means AM for setting the outer and/or inner radial extent (commonly referred to as σ -
- 25 outer and σ -inner, respectively) of the intensity distribution in the beam. In addition, it will generally comprise various other components, such as an integrator IN and a condenser CO. In this way, the beam PB impinging on the mask MA has a desired uniformity and intensity distribution in its cross-section.

- 30 It should be noted with regard to Figure 1 that the source LA may be within the housing of the lithographic projection apparatus (as is often the case when the source LA is a mercury lamp, for example), but that it may also be remote from the lithographic projection apparatus, the radiation beam which it produces being led into

the apparatus (e.g. with the aid of suitable directing mirrors); this latter scenario is often the case when the source LA is an excimer laser. The current invention and claims encompass both of these scenarios.

The beam PB subsequently intercepts the mask MA, which is held on a mask
5 table MT. Having traversed the mask MA, the beam PB passes through the lens PL,
which focuses the beam PB onto a target portion C of the substrate W. With the aid of
the second positioning means PW (and interferometric measuring means IF), the
substrate table WT can be moved accurately, e.g. so as to position different target
portions C in the path of the beam PB. Similarly, the first positioning means PM can be
10 used to accurately position the mask MA with respect to the path of the beam PB, e.g.
after mechanical retrieval of the mask MA from a mask library, or during a scan. In
general, movement of the object tables MT, WT will be realized with the aid of a long-
stroke module (coarse positioning) and a short-stroke module (fine positioning), which
are not explicitly depicted in Figure 1. However, in the case of a wafer stepper (as
15 opposed to a step-and-scan apparatus) the mask table MT may just be connected to a
short stroke actuator, or may be fixed. Mask MA and substrate W may be aligned using
mask alignment marks M1, M2 and substrate alignment marks P1, P2.

The depicted apparatus can be used in two different modes:

1. In step mode, the mask table MT is kept essentially stationary, and an entire mask
20 image is projected in one go (i.e. a single "flash") onto a target portion C. The substrate
table WT is then shifted in the x and/or y directions so that a different target portion C
can be irradiated by the beam PB; and
2. In scan mode, essentially the same scenario applies, except that a given target
portion C is not exposed in a single "flash". Instead, the mask table MT is movable in a
25 given direction (the so-called "scan direction", e.g. the y direction) with a speed v , so
that the projection beam PB is caused to scan over a mask image; concurrently, the
substrate table WT is simultaneously moved in the same or opposite direction at a
speed $V = Mv$, in which M is the magnification of the lens PL (typically, $M = 1/4$ or
 $1/5$). In this manner, a relatively large target portion C can be exposed, without having
30 to compromise on resolution.

Below, the invention will be explained in detail.

1 INTRODUCTION

1.1 On-line mass estimation: why?

The accuracy of positioning the mask and wafer stages is largely dependent on the accuracy of the setpoint feedforward. Currently, only a setpoint acceleration

5 feedforward using the (calibrated) mass is used, as depicted in Figure 2. It is observed that Figure 2 is a simplified picture, without decoupling, etc. Moreover, it is observed that, below, the invention will be illustrated with reference to the wafer stage (i.e., the wafer table WT, together with a wafer W), but it should be understood that the invention is equally applicable with the mask stage.

10 Figure 2 shows a control architecture 3 for a wafer stage (WS) 12, as may be implemented by software in any type of computer arrangement (not shown) in the second positioning means PW (Figure 2). Such a computer arrangement may comprise a single computer or a plurality of computers acting in co-operation. However, alternatively, any suitable type of analog and/or digital circuits may be used as well, as
15 will be evident to persons skilled in the art. In that sense, Figure 2 (and the other architecture Figures described here) only show functional modules that can be implemented in many different ways. Apart from the wafer stage 12, all of the components shown in Figure 2, can be thought of as being included in the second positioning means PW.

20 The control architecture 3 comprises a comparator 16 receiving a position setpoint signal for the wafer stage 12 and an actual position signal from the wafer stage 12. The comparator 16 has its output connected to a PID control unit 2 that has its output connected to a first notch filter 4. The first notch filter 4 outputs a signal to a first adder unit 6 that has its output connected to a second notch filter 8. The first adder unit 6 also
25 receives a force setpoint signal and adds this to the output of the first notch filter 4. The force setpoint signal is delivered as output signal from a multiplier unit 14 that multiplies a received acceleration setpoint signal by a mass m_{ff} , i.e., the feedforward mass of the wafer stage 12.

The second notch filter 8 has its output connected to a second adder unit 10. The
30 second adder unit 10 can also receive additional feedforward signals. The second adder

unit 10 outputs the addition of its two input signals as a position control force to the wafer stage 12 to minimize the output of comparator 16 (the controller error).

It is observed that, for persons skilled in the art, the architecture of Figure 2 is straight forward to control both a desired position of the wafer stage 12 and a translation to a new position via a controlled acceleration in an XY operating region of the wafer stage 12. Especially, while the setpoint acceleration equals the second derivative of the setpoint position, the feedforward force generated by multiplier unit 14 results in a movement that already closely matches the setpoint position. The PID control unit 2 therefore only takes care of remaining deviations between the actual wafer stage path and that dictated by its setpoint position.

It has been observed that the controller error behaviour (output of comparator 16) is dependent on the position of the wafer stage 12: the error is larger in the corners of the XY operating region. Note that a decoupling (gainbalancing) matrix, that incorporates all gain effects like the mass, amplifier gain, motor constant (of an actuator, not shown, to translate the wafer stage 12), etc., is calibrated only in the centre of the operating region. The feedforward mass m_{ff} , as depicted in Figure 2, together with the accompanying delay correction, is also calibrated in the centre of the operating region. One possible explanation of the position-dependent behaviour lies in a changing 'gain' of the physical wafer stage 12, caused by one of the above-mentioned effects: mass, amplifier gain, motor force constant. Here, on-line mass estimation is proposed as a method to compensate for this effect. Besides position-dependence, other time-varying effects are countered as well, like ageing of components (amplifier, motor of actuator, etc.), or changed behaviour due to heating of components (amplifiers, motors). Also, variations in the mass of the substrates that are subsequently exposed are countered.

1.2 On-line mass estimation: basic principle

The basic idea of on-line mass estimation is to continuously estimate the mass of the wafer stage 12 from the wafer stage input force and the resulting position change, i.e., from the output signals from the second adder unit 10 and the wafer stage 12. The estimated mass is used to modify the feedforward coefficient m_{ff} as used in the multiplier 14. If the estimation is fast enough, position-dependent behaviour can be captured.

Note that not only the wafer stage mass is estimated but all aspects that influence the gain from the wafer stage input to its output are included in the estimation. Note also that, because only the feedforward is adjusted, and the controller gain is left unchanged, the instability risk is minimised.

- 5 This document describes the mass estimation design, complications and results.

2 ON-LINE MASS ESTIMATION

2.1 Architecture

The basic architecture of the on-line estimation according to the invention is shown in Figure 3. In Figure 3, the same reference numerals as in Figure 2 refer to the same components. In addition to Figure 2, a mass estimation unit 18 is used that receives the output signals from the second adder unit 10 and the wafer stage 12 as input signals. From those input signals, it calculates a mass estimation signal m_{est} that is output to the multiplier 14.

It is observed that, below, it will be assumed that the mass estimation unit 18 receives an input signal as to position x directly from the wafer stage 12. However, alternatively, mass estimation unit 18 may receive input signals from the output of other units, e.g., comparator 16 or PID unit 2, to calculate the mass estimation m_{est} .

Now, the calculations made by the 'mass estimation' unit 18 will be explained.

The main idea is to estimate the mass from the relation:

$$F = m \cdot a$$

The force F is generated by the controller, and is hence already available as output signal from second adder unit 10. The actual acceleration a , however, must be deducted from the actual position signal x received from the wafer stage 12, e.g., using a digital double differentiator 22 (Figure 4):

$$a = \frac{(1 - z^{-1})^2}{T_s^2} x$$

This double differentiator 22, however, suffers from 1 sample delay, which causes the force F to be one sample advanced with respect to acceleration a . For this reason, F must be delayed 1 sample as well. In addition, the force that is actuated will stay on an output of a Digital Analog Converter (not shown) for one sample, introducing another 0.5 sample delay. Further, the Input Output delay (calculation time of the motion controller computer) influences the time shift between the force and the acceleration. For this reason, the force must be delayed by a total of 2.35 samples (determined from

measurements on an actual system). To that end, a delay unit 20 is introduced, see Figure 4.

The least-squares method is used to estimate the mass. In general, the least-squares method estimates parameters of a model of which the output is linear in its parameters.

- 5 In this case, the model simply generates the estimated force $\hat{F}$, which equals the measured acceleration multiplied by the estimated mass $\hat{m}$:

$$\hat{F} = \hat{m} \cdot a$$

The estimated force $\hat{F}$ is subtracted from the actual force F to generate an estimation error e :

10
$$e = F - \hat{F} = F - \hat{m} \cdot a$$

This estimation error e is one of the variables used by the least-squares method to create its mass estimation $\hat{m}$ (see following paragraph). The resulting block diagram is shown in Figure 4 which shows the double differentiator 22 receiving the actual position signal from the wafer stage 12 and outputting acceleration a to a multiplier 24
15 that multiplies acceleration a by the estimated mass $\hat{m}$.

2.2 Least-squares estimation

In general, the least-squares method is a method to estimate parameters from input/output data. Here, only the recursive least-squares method is described because at each sample, a new estimation must be generated. This is in contrast with the situation
20 where all data is available beforehand, and an estimation has to be made only once.

A model output is generated that is the multiplication of a 'signal vector' ω and the previously estimated parameter vector $\hat{\theta}$:

$$\hat{y}(k) = \hat{\theta}^T(k-1)\omega(k).$$

The difference between the real output $y(k)$ and the estimated output $\hat{y}(k)$ serves as
25 estimation error: $e(k) = y(k) - \hat{y}(k)$. This estimation error $e(k)$ is used to update the estimated parameter vector:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \Gamma(k)\omega(k)e(k)$$

Here, the 'adaptation gain matrix' Γ is updated every sample:

$$\Gamma(k) = \frac{1}{\lambda} \left[\Gamma(k-1) - \frac{\Gamma(k-1)\omega(k)\omega^T(k)\Gamma(k-1)}{\lambda + \omega^T(k)\Gamma(k-1)\omega(k)} \right]$$

where: λ = "forgetting factor" (see further below)

- 5 In the case of the mass estimation as in the previous paragraph, all of the above equations reduce to scalar equations:

$$\begin{aligned} y(k) &= F(k) \\ \hat{y}(k) &= \hat{m}(k-1)a(k) \\ \omega(k) &= a(k) \\ \hat{\theta}(k) &= \hat{m}(k) \\ e(k) &= y(k) - \hat{y}(k) = F(k) - \hat{m}(k-1)a(k) \\ \Gamma(k) &= \frac{1}{\lambda} \left[\Gamma(k-1) - \frac{\Gamma(k-1)\omega(k)\omega^T(k)\Gamma(k-1)}{\lambda + \omega^T(k)\Gamma(k-1)\omega(k)} \right] \\ &= \frac{1}{\lambda} \left[\Gamma(k-1) - \frac{\Gamma^2(k-1)a^2(k)}{\lambda + \Gamma(k-1)a^2(k)} \right] \\ &= \frac{\Gamma(k-1)}{\lambda + a^2(k)\Gamma(k-1)} \end{aligned}$$

$$\hat{m}(k) = \hat{m}(k-1) + \Gamma(k)a(k)[F(k) - \hat{m}(k-1)a(k)]$$

Hence, in each sample, the following steps are taken:

1. Determine the current value of the signal vector $\omega(k)$. In the case of mass estimation, this equals the current value of the acceleration $a(k)$.
2. Determine the model output, based on the previous estimation of the parameter vector: $\hat{\theta}(k-1)$ and the current signal vector $\omega(k)$. In the case of the mass estimation, this equals the product of the previous mass estimate and the current value of the measured acceleration.
3. From the model output, and the actual output (in the case of the mass estimation: the current value of the force), the estimation error $e(k)$ can be calculated.

4. The adaptation gain matrix $\Gamma(k)$ is calculated using the above recursion equation. In the case of the mass estimation, where only one parameter is estimated, this is a scalar equation.

5. The parameter estimate is updated by using $\Gamma(k)$, $\omega(k)$ and $e(k)$. This produces a
5 new mass estimate.

The parameter λ denotes the 'forgetting factor'. If it is 1, the recursive least-squares method produces exactly the same output as the non-recursive version. This means that no parameter updates will be produced any more after a long period of time. To keep the method adapting the estimates, a value slightly below 1 should be chosen. In
10 practice, a value of 0.995 appears to be fine. A larger value slows down the estimation, a smaller value introduces noise in the estimation.

Note that in the case of estimating other parameters than the mass, only the definition of the signal vector and the parameter vector need to be changed.

Note also that when more parameters are estimated, the numerical complexity increases
15 quadratically because matrix calculations are involved.

2.3 Offset removal

A complication that developed early is the presence of offset on the control force. This is introduced by DACs, amplifiers, or even a tilted stone introducing a gravity component. While the acceleration is zero, still some force value is 'actuated'. If no
20 other forces are present, the estimator interprets this effect as an infinite mass because the control force results in zero acceleration.

One possibility to tackle this is to estimate the offset as a second parameter. For this, the controller can be provided with an offset estimator which estimates the offset such that the estimated offset can for example subsequently be subtracted from the total
25 mass estimation. However, this strategy does not work properly because due to the excitation format (the acceleration profile has relatively long periods of constant acceleration), the least-squares method cannot properly distinguish between offset and mass. In other words, the offset estimation actually results in disturbance of the mass estimation.

For this reason, a simpler solution is chosen: both the actuated force and the acceleration are filtered by a high-pass filter. For a start, the time constant is set at 1 Hz. Figure 5 shows the results of the mass estimation in the original situation, and the effects of either having an additional offset estimation or high-pass filters.

- 5 It can be seen that if the offset is not taken into account, the estimated mass at a position around $y=+150$ mm is approximately 22.62 kg, while the estimated mass at $y=-150$ mm is about 22.47 kg. This very large difference of 150 g is not visible any more if either the offset is estimated also, or the high-pass filters are used. However, the high-pass filter solution gets less disturbed in the 'jerk' phase of the setpoint, and is
10 therefore preferable.

2.4 Persistent excitation

- Figure 6 shows the mass and offset estimation as discussed in the previous paragraph. It can be seen that, for example during the acceleration phase from 2.1 to 2.2 sec, both the mass and the offset are adjusted. This is caused by the fact that in the constant-
15 acceleration region, the least-squares estimator cannot distinguish between an offset and a gain (both a higher offset and a higher mass could be the cause for a required higher force). This is a typical example of a too low 'persistent excitation', which more or less means that there is not enough frequency content in the signals to create a correct parameter update.

- 20 A similar problem exists during the constant-velocity phase, wherein the nominal controller force is zero, and so is the stage acceleration. In such a region, noise is the main contributor to the signal contents. The least-squares method reacts to such a condition by increasing its adaptation gain Γ . To avoid Γ to get out of bound, adaptation is switched off when the setpoint acceleration becomes zero. Another
25 alternative would be to limit the trace of Γ .

Note that the requirement on persistent excitation becomes more severe when the number of parameters increases (which, again, is clearly illustrated with the offset estimation example).

2.5 Results

The previous paragraphs already showed some mass-estimation results. However, the estimated mass was not yet made active in the feedforward path. This paragraph describes some results when the mass estimation is made active in the feedforward.

- 5 Figure 7 shows an example, obtained by repeatedly moving the wafer stage in Y direction from -150 to $+150$ mm and back. The plots start with two negative acceleration phases: one to decelerate from $+0.9$ m/s to zero velocity, one to accelerate to -0.9 m/s. The end of the plot shows the deceleration to zero again, followed by acceleration to positive velocity. This implies that the 'left' side of the plot lies around
10 $+150$ mm, while the right side of the plot is around -150 mm.

The top window of Figure 7 shows the servo error without mass-estimation feedforward. It can be seen that at the left side of the plot the peak error is about 62 nm, while at the right side the error is about 44 nm. Hence, the controller error is position-dependent, and is smaller at $Y=-150$ mm.

- 15 It can be seen that at the end of the 2nd negative acceleration phase (around $t=1.53$ sec), the servo error obtained with and without mass-estimation feedforward (middle window and upper window of Figure 7, respectively) is the same. Strikingly, at this point the mass estimation roughly equals the nominal value (bottom window of Figure 7). At the right side of the plot, the mass estimation increases during the two positive
20 acceleration phases. It can be seen that also the servo error increases to the original value of 65 nm.

- When further inspecting the plots of Figure 7, it becomes obvious that the servo error is smaller when the estimated mass is smaller, and hence a smaller acceleration feedforward is present. At the right side of the plot, it is advantageous to use the
25 nominal value, because the mass estimation produces a 20 g higher mass feedforward. In areas where the mass estimation is smaller than nominal, the servo error is smaller when using this estimated mass. The main conclusion here is that the usage of the nominal mass in the feedforward is not optimal: a slightly smaller value improves the servo error.

2.6 Improving mass estimation by estimating more parameters

The drift-like behaviour in mass estimation during the acceleration phase could be caused by the fact that other disturbances influence the mass estimation. Candidates are, for example, the absence of velocity feedforward in the control loop. Other candidates are jerk (derivative of acceleration) and snap (derivative of jerk) feedforward. Because no compensation exists for disturbances in velocity, jerk and snap, the estimator 'pushes' all these effects into the mass estimation.

To check whether this is actually the case, combinations of estimations were tested using the input/output traces of the previous paragraph:

1. Mass estimation only
2. Estimation of mass and velocity feedforward
3. Estimation of mass, velocity and jerk feedforward
4. Estimation of mass, velocity, jerk and snap feedforward

To be able to estimate more parameters, the signal vector and parameter vector need to be extended. For this purpose, jerk, snap and velocity must be created by gradual differentiation of the position of the wafer stage 12. Because each digital differentiation introduces 0.5 samples delay, the various signals must be delayed such that at the end they all have the same delay. The delay in the force signal must match this total delay. Figure 8 shows the creation of the signal vector and the place of the respective parameters d , m , e and g , where:

d = a velocity coefficient;

m = mass;

e = a jerk coefficient;

g = a snap coefficient.

- Figure 8 shows a first differentiator 28, a second differentiator 30, a third differentiator 32, and a fourth differentiator 34 connected in series. An actual position signal is received from the wafer stage 12 and input to the first differentiator 28. Thus, an actual velocity signal, an actual acceleration signal, an actual jerk signal and an actual snap signal, respectively, are present at the outputs of the first differentiator 28, the second

- differentiator 30, the third differentiator 32, and the fourth differentiator 34, respectively. The actual velocity signal is multiplied by estimated velocity coefficient $\hat{d}$ in multiplier 36 and then delayed by 1,5 time period by a delay unit 44. The actual acceleration signal is multiplied by estimated mass $\hat{m}$ in multiplier 24 and then delayed by one time period by a delay unit 46. The actual jerk signal is multiplied by estimated jerk coefficient $\hat{e}$ in multiplier 40 and delayed by 0,5 time period by a delay unit 50. The actual snap signal is multiplied by estimated snap coefficient $\hat{g}$ in multiplier 42. The outputs of the delay units 44, 46, 50 and of the multiplier 42 are shown to be added by adder units 52, 54, 56 to render an estimated force signal to subtraction unit 26.
- 10 Note that the estimated parameters need not all be used in the feedforward, but could have the only goal of making the mass estimation more stable.

- The fact that jerk and snap feedforwards could be required stems from the fact that the process is not only represented by a mass but also has higher-order dynamics. A first possibility is that the process is described by a mass plus one resonance frequency, yielding the following equation for a movement x of the wafer stage 12 as a reaction to a force F :

$$\frac{x}{F} = \frac{1}{ms^2} \frac{1}{\left(\frac{g}{m}\right)s^2 + \left(\frac{e}{m}\right)s + 1} = \frac{1}{gs^4 + es^3 + ms^2}$$

Adding a friction term yields:

$$\frac{x}{F} = \frac{1}{gs^4 + es^3 + ms^2 + ds}$$

- 20 The correct feedforward for such a process would look like:

$$F = (gs^4 + es^3 + ms^2 + ds)x_{SPG}$$

Where: x_{SPG} = the setpoint position as generated by the setpoint position generator.

In addition to the acceleration feedforward (ms^2), the velocity (ds), jerk (es^3) and snap (gs^4) feedforwards are clearly recognised.

- 25 Figure 9 shows the mass estimation result under the conditions mentioned above. When only the mass m is estimated, the typical rise in mass during the acceleration phase is observed. If the velocity coefficient d is adjusted as well, the mass estimation becomes

somewhat more stable. The mass estimation now drifts downwards during the acceleration. When, additionally, the jerk coefficient e is estimated, this result does not change significantly. However, when also the snap coefficient g is estimated, the mass estimation becomes the most stable.

- 5 Figure 10 shows the estimation of all four parameters. It can be seen that especially the snap adjusts stably to $2.7e-7 \text{ N s}^3/\text{m}$, but the other parameters are 'disturbed' during the jerk phase. Also, the jerk coefficient was expected to be zero because no time lag between the measured acceleration and force should be present any more (a jerk coefficient means that a constant force is required during the jerk phase, which is also
- 10 the case when the acceleration feedforward timing is not correct; hence the presence of jerk feedforward indicates a timing problem).

- Now, a test was performed, as follows. In three different X positions (-150 mm , 0 , $+150 \text{ mm}$), repeated Y-movements of $\pm 150 \text{ mm}$ were performed. During each acceleration/deceleration part, the estimated mass (when using the combined
- 15 velocity/mass/jerk/snap estimation as described above) is recorded by using the average of the last 100 points (20 msec) at the end of each acceleration/deceleration part (hence, before the jerk phase starts!). This yields an 'estimated mass' at six points within the wafer stage field. The estimated mass is summarised in the following tables. Note that the nominal mass, as calibrated in the feedforward calibration, equals 22.667 kg .

- 20 *Table 1: Estimated mass when velocity, acceleration, jerk & snap feedforwards are also estimated*

Y	X	-150 mm	0	+150 mm
-150 mm		22.662	22.654	22.657
+150 mm		22.682	22.675	22.675

Table 2: Estimated mass when only the mass is estimated

Y	X	-150 mm	0	+150 mm
-150 mm		22.674	22.667	22.666
+150 mm		22.696	22.688	22.688

In all X positions, the controller error varies between 40 and 60 nm when no feedforward adjustment is performed. The error is 60 nm in all positions when
 5 feedforward mass estimation is switched on. In Table 2 it can be seen that at the positions where the servo error is the same, the estimated mass matches the nominal mass. In the locations where the servo error is smaller when no feedforward adjustment takes place, the estimated mass is higher, and hence the original feedforward is actually smaller than required. Evidently, a slightly (20 g) too small acceleration feedforward by
 10 itself reduces the servo error.

2.7 Combination with snap feedforward

Estimating many parameters simultaneously may not be the perfect solution for all situations for various reasons:

1. Matrix calculations (Γ !) become complicated and use a lot of calculation time.
- 15 2. The excitation must be persistent enough, which differs with the signal type: the mass (=acceleration feedforward) should be estimated only when the acceleration is sufficiently large, therefore the estimation is switched off when the acceleration setpoint is smaller than some value. However, the jerk feedforward estimation requires a sufficient jerk, the snap estimation requires a sufficient snap, and the
 20 velocity feedforward estimation requires a sufficient velocity. Hence, the estimation of the different parameters should be switched on during different phases of the trajectory, which is impossible when using the least-squares method.
3. Not all parameters may be time-varying, focus should be placed on those parameters that change.

25 On the other hand, as observed in the previous paragraphs, the mass estimation appears to be disturbed by the fact that the process not only consists of a mass but also includes

higher-order dynamics (hence the estimation of the other 3 parameters). Assuming the snap feedforward takes away the higher-order dynamics from the mechanics, the mass estimator should now be connected to the control force minus the snap component, as shown in Figure 11 where the adder 10 receives a further input signal from a multiplier 58. The multiplier 58 multiplies a received snap setpoint signal by a feedforward snap coefficient g_{ff} . The mass estimation block 18 now receives a force signal from the output of Notch 2 block 8, i.e. excluding the snap feedforward component from multiplier 58. Because the snap feedforward compensates the higher dynamics of the wafer stage, the relation between stage acceleration and the output of Notch 2 better resembles a mass.

The following plots show preliminary results. Figure 12 shows the controller error without and with snap feedforward, while the mass estimation is switched on. This mass estimation is shown in Figure 13.

It can be seen that the mass estimation is not influenced by the presence of snap feedforward. The snap feedforward in this particular test by itself reduces the servo error about a factor of 2 (from 60 to 30 nm peak).

2.8 Again: offset removal

Inspecting the controller output, it was observed that the required output in standstill differed as much as 0.4 N in the extreme Y-positions. Note that the high-pass filters that should remove the offset were at a frequency of 1 Hz, while a complete move only takes about 0.32 sec. Hence, the high-pass filters may be set at a too low frequency. To test this, an experiment was performed by using a 10 Hz corner frequency for the high-pass filters, resulting in the controller errors of Figure 14 and mass estimation of Figure 15 (without and with snap feedforward). Note that also a velocity feedforward of 0.22 Ns/m was used. The plots show that at the same location, the estimator estimates a 20 g different mass, dependent on the movement direction (even when the sign of the acceleration is the same). This phenomenon is caused by nonlinear behaviour of the amplifiers.

2.9 Again: time shift between acceleration and force signals

Although the mass estimation has now become considerably faster, it can be seen that during the jerk phase quite a large disturbance remains visible. This appears to be caused by a remaining difference in timing between the generated acceleration and force signals. By decreasing the force delay from 2.35 to 2.25 samples, the mass estimation becomes more stable, as indicated in Figure 16.

2.10 Results with mass estimation and snap FF

With the knowledge gained in the previous paragraphs, a new test was performed using the following conditions:

snap feedforward gain	$3.4\text{e-}7 \text{ Ns}^3/\text{ms}$ (injection after notch2)
snap filter frequency & damping	700 Hz, $d=0.7$
snap delay correction	$400\text{e-}6 \text{ sec}$
Mass estimation high-pass filters	10 Hz
Mass estimation delay in force path	2.25 samples, force extracted before snap injection
Mass estimation forgetting factor	0.995
Velocity feedforward	0.22 Ns/m expose stage (WT), 0 Ns/m measure stage (MT)
Nominal mass feedforward as calibrated earlier	expose: 22.652 kg, measure: 22.601 kg

10

Combinations with and without snap feedforward and mass estimation were performed, both on the expose stage WT and the measure stage MT. It appears that in this case at the measure stage MT the nominal feedforward does not match very well. Each test was done 6 times: on 3 X-positions ($\sim 150 \text{ mm}$, 0 , $+150 \text{ mm}$), moves having both a positive and a negative direction were performed. For the centre position ($X=0$), the following four plots in figures 17, 18, 19, and 20, respectively, show the results, including the estimated mass. In each plot, also the peak servo error after the

15

- acceleration phase is indicated. In each of the Figures 17, 18, 19, and 20, the top left plot shows the original situation. The middle left plot shows the effect of mass estimation, with below it the estimated mass. The top right plot shows the result with snap feedforward only. The middle right plot shows the result with mass estimation and snap feedforward, with below it the estimated mass.

The peak controller errors for the expose stage WT are summarised in the following tables. These are ordered according to the position in the field where the measurement was done, and because both $Y=+150$ and $Y=-150$ are present in each test file, two values are present.

10 *Table 3: Peak controller error [nm], original condition*

Y	X	-150 mm	0	+150 mm
+150 mm		43.4	40.5	35.1
+150 mm		42.6	43.3	40.7
-150 mm		35.8	25.8	30.8
-150 mm		29.8	31.8	36.0

Table 4: Peak controller error [nm], mass estimation

Y	X	-150 mm	0	+150 mm
+150 mm		20.4	18.0	19.9
+150 mm		22.4	22.0	25.1
-150 mm		27.1	21.1	20.8
-150 mm		23.0	21.2	22.4

Table 5: Peak controller error [nm], snap feedforward

Y	X	-150 mm	0	+150 mm
+150 mm		23.0	19.3	13.0
+150 mm		20.2	23.4	12.7
-150 mm		8.7	13.7	12.0
-150 mm		11.1	10.0	12.2

Table 6: Peak controller error [nm], mass estimation and snap feedforward

Y	X	-150 mm	0	+150 mm
+150 mm		14.1	15.8	14.8
+150 mm		10.5	13.6	16.2
-150 mm		16.3	14.2	11.4
-150 mm		11.1	11.2	12.8

5 Regarding these new experiments, the following conclusions can be drawn:

1. For the measure stage MT, about 60 g mismatch exists between the feedforward mass and the estimated mass. The effect of the mass estimator is that the peak servo error decreases from more than 100 nm to about 35 nm.
2. At the expose stage WT, the mass estimation by itself also improves the controller error considerably (peak error decreases from 43 to 27 nm). The original error of 43 nm is relatively high due to non-exact mass calibration.
3. When snap feedforward is used, the mass estimator yields a slightly less varying mass than without snap feedforward.
4. When snap feedforward is used, the peak controller error in the wafer stage field becomes more constant. The peak error decreases from 23 nm to 16 nm. Note that the snap feedforward gain and timing was not tuned, and the estimated snap

indicates a smaller value than used in the machine. Some room for improvement appears to be present.

5 5. During the jerk phase, the mass estimator is rapidly varying. This can be improved by switching on the estimator only when the maximum acceleration is reached (up to now, it is active whenever the acceleration is nonzero). In that case, the estimator is very constant at the end of an acceleration phase, but has to settle more during the start of the acceleration phase. This is shown in Figure 21. The effect in the machine is tested in the following paragraph.

2.11 Again: results with mass estimation and snap FF

10 The same test was performed as in paragraph 1.10, with the only change that the mass estimation is only active when the setpoint acceleration has reached its maximum. Hence, no adjustment takes place any more during the jerk phase, for the reasons mentioned in paragraph 1.10. The results summarized for the expose stage WT are listed in the tables. Table 11 summarizes results for the four combinations of snap
15 feedforward and mass estimation. It can be seen that the combination of snap feedforward and mass estimation has performed slightly better than in the first test. Apparently, a constant feedforward mass during the end of the acceleration phase improves the maximum error. In the plots, note that when using mass estimation only, the servo error is always smaller when the estimated mass is smaller than the nominal
20 value, as observed earlier. This is no longer true if also snap feedforward is used.

Table 7: Peak controller error [nm], no mass estimation or snap feedforward

Y	X	-150 mm	0	+150 mm
+150 mm		28.0	23.9	17.8
+150 mm		30.6	18.1	19.7
-150 mm		13.8	19.3	17.7
-150 mm		16.1	20.3	13.3

Table 8: Peak controller error [nm], mass estimation

Y	X	-150 mm	0	+150 mm
+150 mm		24.1	20.4	23.4
+150 mm		25.9	25.4	24.0
-150 mm		27.9	25.6	23.0
-150 mm		25.7	24.3	24.7

Table 9: Peak controller error [nm], snap feedforward

Y	X	-150 mm	0	+150 mm
+150 mm		10.9	15.0	19.0
+150 mm		13.6	19.9	13.7
-150 mm		21.0	28.0	25.8
-150 mm		18.8	26.6	21.7

5 *Table 10: Peak controller error [nm], mass estimation and snap feedforward*

Y	X	-150 mm	0	+150 mm
+150 mm		11.4	14.1	11.8
+150 mm		9.9	13.9	9.7
-150 mm		13.1	10.4	10.8
-150 mm		10.9	10.9	11.1

Table 11: Peak position errors [nm], summarized

X	What? Y→	-150 mm	0	+150 mm
+150	Original	28.0	23.9	17.8
	Mass estimation	24.1	20.4	23.4
	Snap FF	10.9	15.0	19.0
	Mass est + snap FF	11.4	14.1	11.8
-150	Original	13.8	19.3	17.7
	Mass estimation	27.9	25.6	23.0
	Snap FF	21.0	28.0	25.8
	Mass est + snap FF	13.1	10.4	10.8

3 IMPLEMENTATION SIMPLIFICATION & ADDITIONS

An alternative implementation was developed for the case when only 1 parameter is estimated, as is true in the case of mass estimation.

3.1 Implementation simplification 1

- 5 In the case of mass estimation, the non-recursive least-squares method attempts to find the optimal estimated mass $\hat{m}$ in the equation:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \hat{m} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

$$A\hat{m} = F$$

- Here, a_i ($i = 1, 2, \dots, n$) is an acceleration sample (a_n is the most recent sample), while f_i ($i = 1, 2, \dots, n$) is a control force sample (f_n being the most recent one). This can be
10 written into:

$$A^T A \hat{m} = A^T F$$

$$\hat{m} \sum_{i=1}^n (a_i)^2 = \sum_{i=1}^n (a_i \cdot f_i)$$

$$\hat{m} = \frac{\sum_{i=1}^n (a_i \cdot f_i)}{\sum_{i=1}^n (a_i)^2}$$

Hence, the least-squares estimate can be found by the filter implementation shown in Figure 22.

- This form does not support using a forgetting factor λ yet. Implementing this factor
15 can be done as follows:

$$\hat{m} \sum_{i=1}^n \lambda^{n-i} (a_i)^2 = \sum_{i=1}^n \lambda^{n-i} (a_i \cdot f_i)$$

$$\hat{m} = \frac{\sum_{i=1}^n \lambda^{n-i} (a_i \cdot f_i)}{\sum_{i=1}^n \lambda^{n-i} (a_i)^2}$$

The matching filter implementation then looks like that in Figure 23. This alternative implementation has a simpler form than the original least-squares implementation, which involved two recursion equations.

3.2 Implementation simplification 2

- 5 In the case of offset estimation, the used model would look like:

$$F = \hat{m}a + \hat{d}$$

Where: $\hat{d}$ = estimated offset.

- 10 When the offset is estimated by itself (not simultaneous with the mass), effectively a signal vector is used which is a constant of 1. Using the structure of Figure 23, a_i is replaced by 1. The top filter is then fed with the control force, while the bottom filter is fed by an input of 1. It can be easily calculated that the bottom filter settles to a value of $\frac{1}{1-\lambda}$. This fixed value can then be used instead of the output of the bottom filter, yielding the structure of Figure 24.

4 ALTERNATIVE APPROACH: ON-LINE FEEDFORWARD ESTIMATION

4.1 Basic idea

In the first embodiment discussed above, the one and only estimated parameter that was
5 used in the feedforward, was the mass. This mass actually serves as the simplest form of 'inverse process dynamics': it is the inverse of the transfer from force to acceleration. The additional snap feedforward actually serves as a better 'inverse process dynamics' by including one (zero-damping) resonance.

An alternative feedforward would be a higher-order model of the inverse process. One
10 way to do this is to estimate a model of the process, inverse this model, and use this inverse model as a filter in the acceleration feedforward path. This method, however, has some severe drawbacks. First, the process usually has a high order of gain roll-off for higher frequencies, which translates into a strongly rising frequency characteristic of the process inverse. Furthermore, non-minimum phase zero's in the process transfer
15 function translate into unstable poles in the inverse. This especially poses a problem in the discrete domain, where non-minimum phase zero's are common.

A solution proposed here is to extend the mass estimator to an 'inverse process dynamics' estimator. The estimator then estimates the transfer function from measured acceleration to the applied force, rather than estimating the transfer function from the
20 applied force to the acceleration, and inverting this estimated transfer function. This way, a transfer function estimate of the inverse dynamics will result that is optimal in a least-squares sense. By choosing, for example, a FIR filter architecture, stability of the estimate is guaranteed.

Figure 33 shows the basic architecture for this alternative approach. The architecture is
25 similar to the one shown in Figure 3, and like reference numbers refer to the same components. However, the mass estimation unit 18 of Figure 3 has been generalised into a feedforward (FF) filter estimation unit 60. Moreover, the multiplier 14 of Figure 3 has been changed into a transfer function unit 62, arranged to apply a transfer function H_{ff} to the setpoint acceleration.

30 The difference between the architectures of Figure 3 and 33 is that no longer a mass is

estimated but the relation between acceleration and force. As may be evident to a person skilled in the art, if the wafer stage 12 (or any other mass to be controlled) performs as a "rigid body", then the architecture of Figure 30 reduces to the one of figure 3 since then the transfer function H_{ff} is the same as multiplying by mass m_{ff} . The difference between the two architectures is important when there are dynamics in the wafer stage 12.

The feedforward filter estimation unit 60 determines an estimated transfer function H_{est} from the measured acceleration to the applied force F . In the transfer function unit 62 this estimated transfer function H_{est} is used, together with the setpoint acceleration to provide an estimated input force. This estimated input force is deducted from the real input force and the difference is used by the least square's mechanism to produce a new estimated transfer function.

4.2 Estimating reverse transfer function: ARX structure

15

The most general structure of such an inverse process dynamics estimator that can be used is the ARX structure, as shown in Figure 25.

In general terms, the transfer function of this structure is:

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) - \dots - a_n y(k-n) + b_0 u(k) + b_1 u(k-1) + \dots + b_m u(k-m)$$

20 Hence, the signal vector $\omega(k)$ and parameter vector $\hat{\theta}(k)$ are defined by:

$$\omega(k) = [-y(k-1), -y(k-2), \dots, -y(k-n), u(k), u(k-1), \dots, u(k-m)]$$

$$\hat{\theta}(k) = [\hat{a}_1(k), \hat{a}_2(k), \dots, \hat{a}_n(k), \hat{b}_0(k), \hat{b}_1(k), \dots, \hat{b}_m(k)]$$

Here, the input u is formed by the measured acceleration. The output y then represents the estimated input force, which is compared with the real input force to create the estimation error.

25

4.3 Estimating reverse transfer function: FIR structure

Another architecture that can be used is a FIR filter. The advantage of the FIR filter is

that it cannot become unstable. The architecture is shown in Figure 26.

The FIR filter recursion equation is:

$$y(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_m u(k-m)$$

Consequently, the definition of the signal vector and parameter vector are:

$$\begin{aligned} \omega(k) &= [u(k), u(k-1), \dots, u(k-m)] \\ \hat{\theta}(k) &= [\hat{b}_0(k), \hat{b}_1(k), \dots, \hat{b}_m(k)] \end{aligned}$$

Again, the input $u(k)$ is formed by the measured acceleration and the filter output $y(k)$ represents the estimated input force, used to create the estimation error by subtracting it from the actual input force.

4.4 Test results

Figure 27 shows the estimated transfer function for a large number of estimated parameters for both the FIR and the ARX filter. The FIR filter has 20 FIR taps (21 parameters), whereas the ARX filter is of the 10th order (21 parameters). The resemblance between the FIR and ARX transfer functions is striking. Figure 28 shows the resulting feedforward force for both filters. The "overshoot" shows a striking resemblance with snap feedforward. Figure 29 shows the estimated mass, which equals the DC gain of each of the resulting filters. Figure 30 and Figure 31 show the results for a much lower filter order. Figure 30 shows estimated transfer functions for a FIR filter with 4 taps (5 parameters) and a 2nd order ARX filter (5 parameters), respectively. Figure 31 shows the feedforward force in both situations. The resulting feedforward is not very different from the one associated with Figures 27-29 but the resulting estimated mass, shown in Figure 32, now has become much less stable in the ARX case.

Claims

1. A controller arranged for controlling a position of a mass (12) by providing said mass a mass acceleration by a control force depending on a desired mass acceleration
5 characterised in that said controller is arranged to receive a feedback signal comprising status information of said mass (12), to calculate an estimated relation between the mass acceleration and said control force from said feedback signal and from said control force, and to use said estimated relation and said desired mass acceleration to determine said control force.
10
2. Controller according to claim 1, wherein the status information comprises an indication of the position, the speed and/or the acceleration of the mass.
3. Controller according to claim 1 or 2, wherein the feedback signal is a feedback
15 position signal indicating a position of said mass.
4. Controller according to one of the preceding claims, wherein said controller is arranged to determine said estimated relation by means of a least-squares method.
- 20 5. Controller according to any of the preceding claims, wherein said controller is arranged to remove offset on said control force.
6. Controller according to claim 5, wherein said controller comprises a high-pass filter for removing the offset on said control force.
25
7. Controller according to one of the preceding claims, wherein said estimated relation is an estimated mass ($\hat{m}$).
8. Controller according to claim 7, wherein said controller is arranged to calculate
30 said estimated mass by:

$$\hat{m} = \frac{\sum_{i=1}^n \lambda^{n-i} (a_i \cdot f_i)}{\sum_{i=1}^n \lambda^{n-i} (a_i)^2}$$

where:

$\hat{m}$ = estimated mass

a_i ($i = 1, 2, 3, 4, \dots, n$) is an acceleration sample.

5 f_i ($i = 1, 2, 3, 4, \dots, n$) is a control force sample.

λ is a forgetting factor.

9. Controller according to any of the claims 7 and 8, wherein said controller is
 10 arranged to calculate at least one of an estimated velocity coefficient ($\hat{d}$), an estimated
 jerk coefficient ($\hat{e}$), and an estimated snap coefficient ($\hat{g}$) from said feedback position
 signal and possibly to use said at least one of said estimated velocity coefficient ($\hat{d}$),
 said estimated jerk coefficient ($\hat{e}$), and said estimated snap coefficient ($\hat{g}$) to
 determine said control force.

15

10. A lithographic projection apparatus comprising:
 - a radiation system for providing a projection beam of radiation;
 - a support structure for supporting patterning means, the patterning means serving
 to pattern the projection beam according to a desired pattern;
 20 - a substrate table for holding a substrate; and
 - a projection system for projecting the patterned beam onto a target portion of the
 substrate,
 - a controller according to any of the preceding claims, said mass being a movable
 object in said lithographic projection apparatus.

25

11. Lithographic projection apparatus according to claim 10, wherein said movable
 object is at least one of said support structure with patterning means and said substrate
 table with a substrate.

12. Method of controlling a position of a mass (12) by providing said mass a mass acceleration by a control force depending on a desired mass acceleration characterised by receiving a feedback position signal indicating a position of said mass (12),
- 5 calculating an estimated relation between the mass acceleration and said control force from said feedback position signal and from said control force and using said estimated relation and said desired mass acceleration to determine said control force.
13. Method according to claim 12, wherein said estimated relation is an estimated
- 10 mass ($\hat{m}$).
14. Method according to claim 12 or 13 in a device manufacturing method comprising:
- providing a substrate that is supported by a substrate table and, that is at least
 - 15 partially covered by a layer of radiation-sensitive material;
 - providing a projection beam of radiation using a radiation system;
 - using patterning means supported by a support structure to endow the projection beam with a pattern in its cross-section; and
 - projecting the patterned beam of radiation onto a target portion of the layer of
 - 20 radiation-sensitive material,
 - controlling said position of said mass, said mass being at least one of said substrate table with said substrate and said support structure with said patterning means.