Dynamic decoupling in motion systems using a gradient approximation-based algorithm

Marcel Heertjes, Daan Hennekens, Arjan van Engelen, and Maarten Steinbuch

Abstract—To reduce cross coupling in the control of multi-input multi-output (MIMO) motion systems having diagonal dominance, a dynamic decoupling design is studied. The dynamic decoupling is used in parallel to a nominal (and static) decoupling with the aim to support a straightforward single-input single-output (SISO) control design. For a six degrees-of-freedom nano-accurate wafer stage, the off-diagonal entries of the dynamic decoupling matrix are given a fourth-order FIR filter structure as to counteract the coupled and dominant fourth-order plant behavior. The FIR filter coefficients are obtained using a gradient approximation-based algorithm which supports the means to obtain the optimized set of FIR coefficients through a limited set of perturbed-parameter experiments. In simulation, both performance as well as closed-loop stability are shown to be effectively dealt with.

I. INTRODUCTION

In nano-accurate motion systems, performance is key to the success of the control design. A clear example is given by the fast and nano-accurate motion control of the stages of wafer scanners used in the manufacturing of integrated circuits, see [12] and [15]. Being controlled in six degrees-of-freedom, a static input-output mapping is used to map both sensor information and actuator commands to a desired (orthogonal) basis about a reference point. A straightforward SISO control design is then used to achieve both control performance and robust stability. But as the coupled plant contains high-order dynamics related to flexible modes, a static decoupling generally sustains cross-talk among the individual servo axes. As a result, performance and stability of the feedback loops are compromised.

Improving upon the performance and stability properties of the feedback system without inducing unnecessary MIMO complexity in the control design, a dynamic decoupling is studied in which the SISO control design is kept valid. This preserves the effective usage of engineering expertise and system knowledge associated with such a design. Basically, we envision all MIMO plant characteristics which are generally less clear from an engineering standpoint to be properly addressed by the dynamic decoupling. Having inverted plant characteristics, the difficulty now lies in finding a sufficiently robust inverse as to improve upon servo performances.

To overcome this problem, a gradient approximation-based algorithm is considered. The idea relates to iterative feedback tuning [9] in which the optimization of controller parameters is done under feedback conditions, see also [10], and [13]. Given an objective function in terms of relevant servo signals, the set of controller parameters that minimize the objective function is found on the sole basis of perturbed-parameter experiments, see also [19], [20]. This renders the approach strictly data based. In terms of decoupling, other approaches are used in [8] and [16]. For a wafer stage application in particular, the considered approach is successfully applied to improve the feed-forward design [2], [14] and, for example, to solve for linear induction motor problems [21]. In this paper, a gradient approximation-based algorithm is used to find the coefficients of the finite impulse response (FIR) filters needed to construct the off-diagonal terms of a $6 \times 6$ dynamic decoupling matrix. The diagonal terms are excluded as to focus on the effects of coupling. The latter with the aim to provide the conditions that validate a SISO control approach. The filter order is kept limited because a major performance improvement is expected from the sole elimination of fourth-order coupled plant behavior [5]. In this sense the presented techniques are particularly suited but certainly not limited to motion control systems.

This paper has the following organization. In Section II, the dynamics and control of a wafer stage are addressed with specific attention to the current static decoupling. In Section III, the optimization approach is considered. Key is the derivation of the FIR filter coefficients used to dynamically decouple the system. In Section IV, improved decoupling is demonstrated through wafer stage simulations. A summary of the main conclusions is presented in Section V.

II. DYNAMICS AND CONTROL OF A WAFER STAGE

In the production of integrated circuits (ICs) wafer scanners are used to get a desired image onto a substrate via a lithographic process. An artist impression of a wafer scanner is given in Figure 1. Light from a laser passes a reticle which contains the image, through a lens which scales down the image, and onto a wafer which contains the substrate. Key to the lithographic process is the motion control of two sub-systems: the reticle stage supporting the reticle and the wafer stage supporting the wafer. As a benchmark, this paper merely considers the wafer stage.

The wafer stage consists of a long-stroke stage for course positioning (micrometer accuracy) and a short-stroke stage for fine positioning (nanometer accuracy). The short-stroke stage is controlled in six degrees-of-freedom according to the simplified motion control scheme of Figure 2. This scheme is centered about the MIMO wafer stage plant $P(s) \in \mathbb{C}^{6 \times 6}$ where $s$ represents the Laplace variable. The plant output $y = [x \ y \ z \ rx \ ry]^T$ is fed back and subtracted from the reference commands $r = [rx \ ry \ 0 \ rz \ 0 \ 0]^T$ to give the servo errors $e = [ex \ ey \ ez \ exr \ eyr]^T$ through $e = r - y$. Feedback control is reflected by the diagonal matrix $C_{fb} \in \mathbb{C}^{6 \times 6}$ having entries: $C_{fb,x}, C_{fb,y},$
in Laplace domain by
\[
P_d(s) = \begin{bmatrix}
\frac{1}{ms^2} & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{ms^2} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{J_x s^2} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{ms^2} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{J_y s^2} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{J_z s^2}
\end{bmatrix},
\]
with \( m = 22.57 \) kg the short-stroke wafer stage mass and \( J_x = 0.371 \) kgm\(^2\), \( J_y = 0.399 \) kgm\(^2\), and \( J_z = 0.745 \) kgm\(^2\), its principle moments of inertia, respectively. The matrix \( Q(r) \in \mathbb{R}^{6 \times 6} \) describes the relation between the center-of-gravity positions \( y_{cg} \) and the controller positions \( y \), or
\[
y = Q(r)y_{cg},
\]
with
\[
Q(r) = \begin{bmatrix}
1 & 0 & Q_{31}(r) & 0 & 0 & -Q_{61}(r) \\
0 & 1 & -Q_{32}(r) & 0 & Q_{61}(r) & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -Q_{31}(r) & Q_{32}(r) \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix},
\]
and \( Q_{31}(r) = r_y - y_{offset}, \ Q_{32}(r) = r_x - x_{offset}, \ Q_{61}(r) = r_z - z_{offset} \), and the offsets: \( x_{offset} = 10^{-4} \) m, \( y_{offset} = 4 \times 10^{-4} \) m, and \( z_{offset} = 2 \times 10^{-4} \) m.

Assume the aim of the decoupling \( D(r, s) \) in (1) is to create a SISO desired plant \( P_d(s) \), or
\[
P_d(s) := P(s)D(r, s),
\]
and moreover assume that the dynamic part \( D_d(s) \) in (1) is not needed to fully decouple the MIMO plant \( P(s) \), i.e., \( D_d(s) = 0 \) which is obviously not the case. It then follows that
\[
P_d(s) = P(s)P^{-1}_d(s)Q(r)^{-1}P_d(s) \implies P(s) = Q(r)P_d(s),
\]
i.e., \( P(s) \) is fully decoupled using the static decoupling in (2).

For the wafer stage application, the effect of the assumptions is shown in Fig.3 where both the measured plant characteristics \( P(j\omega) \) (solid,black) as well as the computed characteristics \( Q(r)P_d(j\omega) \) (dashed,gray) are shown for the vertical plane and in Bode magnitude representation. It is clear from the figure that the dynamic part of the decoupling \( D_d(j\omega) \) cannot be neglected; the off-diagonal measured frequency response functions do not match the computations. This is most clear in the high-frequency interval. As a result, \( D_d(j\omega) \neq 0 \). In discrete-time implementation, we therefore seek the dynamic decoupling part \( D_d(z) \) in z-domain such

---

**Fig. 1.** Artist impression of an industrial wafer scanner.

**Fig. 2.** Simplified short-stroke wafer stage control scheme.
III. A GRADIENT APPROXIMATION-BASED ALGORITHM

In a least-squares sense, optimal parameter values (in this case the FIR coefficients $\theta$) are derived under closed-loop conditions from a Newton-based update law, the latter being a function of the gradients of a representative objective function evaluated before and after known perturbations, see [14]. The procedure can be summarized as follows: define an objective function, derive the Newton-based update law, and derive the gradients – needed in the update law – from perturbed-parameter experiments.

The control objective is to improve upon the wafer stage servo performances during scanning, which is formalized by finding the set of FIR filter coefficients $\tilde{\theta}$ that minimizes the objective function $V(\theta, t)$, or

$$\tilde{\theta} := \arg \min_{\theta} V(\theta, t). \quad (12)$$

The objective function is chosen a positive semidefinite function of the servo errors resulting from time sampling in a relevant scanning interval. A realization is given by

$$V(\theta, t) = \tilde{e}^T(\theta, t)\tilde{e}(\theta, t), \quad \theta = [\theta_1 \cdots \theta_n]^T, \quad (13)$$

$$\tilde{e}(\theta, t) = \begin{bmatrix} e_x(\theta, t) \\ e_y(\theta, t) \\ e_{rx}(\theta, t) \\ e_{ry}(\theta, t) \\ e_{rz}(\theta, t) \\ e_{rxz}(\theta, t) \end{bmatrix},$$

where $\gamma_i > 0$, $i \in \{x, y, rz, z, rx, ry\}$, are scaling factors and $\theta \in \mathbb{R}^{n \times 1}$ represents the unknown FIR filter coefficients of the dynamic decoupling matrix in (9). $V(\theta, t)$ is assumed to be a twice continuously differentiable real valued and convex function in $\theta$ for which standard optimization tools yield the unknown FIR filter coefficients $\tilde{\theta}$.

Here consider a second-order Taylor-series expansion of $V$ about $\theta = 0$ (which reduces to the nominal closed-loop design and which is assumed in the neighborhood of $\theta$), or

$$V(\theta, t) \approx V(0, t) + \nabla V(0, t)^T\theta + \frac{1}{2}\theta^T\nabla^2 V(0, t)\theta, \quad (15)$$

from which it follows that

$$\begin{bmatrix} \frac{\partial V(\theta, t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial V(\theta, t)}{\partial \theta_n} \end{bmatrix} = \nabla V(0, t)^T + \theta^T\nabla^2 V(0, t), \quad (16)$$

where $\nabla V(0, t) \in \mathbb{R}^{n \times 1}$ represents the gradient matrix:

$$\nabla V(0, t) = \begin{bmatrix} \frac{\partial V(0, t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial V(0, t)}{\partial \theta_n} \end{bmatrix} = 2\begin{bmatrix} \frac{\partial \tilde{e}_x(0, t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \tilde{e}_x(0, t)}{\partial \theta_n} \end{bmatrix}, \quad (17)$$

for $k \in 0, 1, \ldots, m$ that satisfy (8); note that $n = 30(m+1)$ coefficients need to be optimized. For this purpose, we reside to a gradient approximation-based algorithm.
and $\nabla^2 V(0, t) \in \mathbb{R}^{n \times n}$ the Hessian matrix:

$$\nabla^2 V(0, t) = \begin{bmatrix}
\frac{\partial^2 V(0, t)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 V(0, t)}{\partial \theta_1 \partial \theta_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 V(0, t)}{\partial \theta_n \partial \theta_1} & \cdots & \frac{\partial^2 V(0, t)}{\partial \theta_n^2}
\end{bmatrix}$$

The motivation for applying (19), i.e., a single-trial optimization scheme, is clearly shown in terms of efficiency.

The validity of using the approximated Hessian in (24) relies on $e$ being affine in the FIR filter coefficients $\theta$. From Fig.2, it follows that

$$e(j\omega) \approx \begin{pmatrix} 1 + P(j\omega)D(r, j\omega)C_{fb}(j\omega) \end{pmatrix}^{-1} \ldots \begin{pmatrix} S \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} P \end{pmatrix} \begin{pmatrix} I \end{pmatrix}$$

(25)

Assume $\lim_{\omega \to 0} S(j\omega) = C_{fb}^{-1}(j\omega)D^{-1}(r, j\omega)P^{-1}(j\omega)\ldots
\begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} P \end{pmatrix} \begin{pmatrix} I \end{pmatrix} \begin{pmatrix} C_{fb}(j\omega) \end{pmatrix} \begin{pmatrix} r(j\omega) \end{pmatrix}$

(26)

Substitution of (8) gives:

$$e(j\omega) \approx C_{fb}^{-1}(j\omega)D^{-1}(r, j\omega)(P_{d}^{-1}(j\omega)Q(j\omega)^{-1} + D_d(j\omega)P_d^{-1}(j\omega))r(j\omega)$$

(27)

Because of diagonal dominance, see also Fig.3, hence $\|D^{-1}(r, j\omega)\| \approx \|D^{-1}(r)\|$, it follows that

$$e(j\omega) \approx C_{fb}^{-1}(j\omega)D_{s}^{-1}(r)(P_{d}^{-1}(j\omega)Q(r)^{-1} + D_d(j\omega)P_d^{-1}(j\omega))r(j\omega) - C_{fb}^{-1}(j\omega)C_{ff}(j\omega)r(j\omega)$$

(28)

Since $D_d(j\omega)$ is chosen affine in the FIR filter coefficients $\theta$, see (9), $e(\theta, j\omega)$ in the considered approximative sense is affine in $\theta$. Note that because of the low-frequency validity of Eqs. (25)-(28), a single trial optimization suffices in obtaining a static MIMO decoupling.

Remark 3.3: The validity of obtaining the optimized set of FIR filter coefficients $\theta$ in a single trial using (19) stems from the considered convexity properties. In the absence of convexity, one generally resides to the iterative scheme:

$$\theta_{k+1} = \theta_k - \lambda(\nabla^2 V(0, t))^{-1} \nabla V(0, t),$$

(29)

with $0 < \lambda \leq 1$ and $k$ the trial number. For three values $\lambda \in \{0.25, 0.5, 1\}$ Fig.4 shows the effect of convergence speed in minimizing the objective function $V(\theta, t)$ under $r_y(t)$ reference motion. The motivation for applying (19), i.e., a single-trial optimization scheme, is clearly shown in terms of efficiency.
IV. WAFER STAGE DYNAMIC DECOUPLING

The effect of optimization on improved wafer stage dynamic decoupling is shown in Fig.5. In Bode magnitude representation and for the MIMO vertical plant, it follows that the off-diagonal (and low-frequency) frequency response functions associate with significantly smaller amplitude levels after optimization (black) as compared to the levels prior to optimization (gray). Cross coupling between the vertical controller axes is significantly reduced (20 up to 40 dB).

For the entire plant $P(j\omega)$ this is made quantitatively through the relative gain array (RGA) number [18]:

$$\text{RGAnumber}(P(j\omega)) = \sum_{j=1}^{6} \sum_{i=1}^{6} \| (P(j\omega) \times P(j\omega)^{-T} \mathbf{I})_{ij} \|.$$

(30)

For each frequency point this is shown in frequency-domain representation in Fig.6. This figure confirms the significant improvements obtained in low-frequency interaction. Moreover, the improvements hardly relate to any increased high-frequency sensitivity. The latter is shown more clearly in Fig.7. For the MIMO sensitivity function $S(j\omega)$ evaluated at each frequency point separately this figure depicts the sensitivity’s maximum singular values $\sigma(S(j\omega))$. Due to low-frequency diagonal dominance (see also Fig.3) no low-frequency difference is visible prior to (gray) and after (black) optimization. More important, there is no significant high-frequency deterioration despite the clear presence of coupling in this interval.

In time-domain, the effect of the improved decoupling is shown in Fig.8. For a reference signal $r_y(t)$, it can be seen through simulation that the error responses prior to optimization (gray) are significantly larger than the corresponding responses after optimization (black). The improvements apply.
solely to the cross-talk axes and give a strong argument for SISO control.

The optimized FIR filter structure of (9) is shown in Fig.9. In Bode magnitude representation and for the vertical plane only, it can be seen that the optimized cross-talk frequency response functions include the effect of low-frequency zeros corresponding to the zeros encountered in the inverted plant model, see also Fig.5; an ideal decoupling would match the plant’s fourth-order behavior. Though the filter order suffices to capture the main characteristics associated with the plant’s fourth-order behavior. Though the filter order can easily be increased, its additional contribution to improved low-frequency decoupling is expected to be marginal, see [5]. More likely, the increased high-frequency sensitivity will act as a limit to keep the order small.

V. CONCLUSIONS

To improve upon the dynamic decoupling properties of a MIMO controlled and nano-accurate wafer stage motion system, data-based optimization is shown to be a powerful tool in achieving improved servo performances. The data-based approach is capable of dealing with model uncertainty and minimizes the effect on the performance variables at hand, i.e., the closed-loop servo error signals. Atop a nominal and static decoupling scheme, the optimization algorithm satisfies the linearity requirements needed to find an optimized set of FIR filter coefficients in a single optimization step. The FIR filters themselves provide the means to keep the number of perturbed-parameter experiments small. That is, one experiment per cross-axis FIR filter (regardless the filter order) giving a total of 30 perturbed-parameter experiments. In view of the plant’s main fourth-order characteristics, the filter order is kept limited to five coefficients per filter. Giving a low-frequency reduction of the cross-talk plant magnitude characteristics from 20 up to 40 dB, it is concluded that the system’s decoupling is efficiently and effectively improved giving access to improved robust stability and performance under SISO feedback control.

REFERENCES