Delft University of Technology

System identification beyond the Nyquist frequency

A kernel-regularized approach

van Haren, Max; Smith, Roy S.; Oomen, Tom

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

System identification beyond the Nyquist frequency: A kernel-regularized approach☆

Max van Haren [a],[*], Roy S. Smith [b], Tom Oomen [a],[c]

[a] *Department of Mechanical Engineering, Control Systems Technology Section, Eindhoven University of Technology, Groene Loper 5, Eindhoven, 5612 AE, The Netherlands*
[b] *Automatic Control Laboratory, ETH Zürich, Zürich, 8092, Switzerland*
[c] *Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, Delft, 2628 CN, The Netherlands*

ARTICLE INFO

ABSTRACT

Models that contain intersample behavior are important for control design of systems with slow-rate outputs. The aim of this paper is to develop a system identification technique for fast-rate models of systems where only slow-rate output measurements are available, e.g., vision-in-the-loop systems. In this paper, the intersample response is estimated by identifying fast-rate models through least-squares criteria, and the limitations of these models are determined. In addition, a method is developed that surpasses these limitations and is capable of estimating unique fast-rate models of arbitrary order by regularizing the least-squares estimate. The developed method utilizes fast-rate inputs and slow-rate output measurements and identifies fast-rate models accurately in a single identification experiment. Finally, both simulation and experimental validation on a prototype wafer stage demonstrate the effectiveness of the framework.

1. Introduction

Systems which have their outputs sampled at a reduced sampling rate compared to their inputs are becoming more prevalent, particularly in applications such as vision-in-the-loop systems (Fujimoto & Hori, 2004). Slow-rate output measurements generally constrain the identified models to reflect only slow system dynamics. However, fast-rate models are crucial for capturing higher-frequency behavior, especially beyond the Nyquist frequency. Therefore, fast-rate models of multirate systems are essential for intersample performance estimation (Chen & Francis, 1995; Sivashankar & Khargonekar, 1993) and sampled-data or multirate control (Bamieh, Pearson, Francis, & Tannenbaum, 1991; Li, Shah, & Chen, 2002; Oomen, van de Wal, & Bosgra, 2007).

System identification plays an essential role in understanding and modeling the dynamic behavior of systems for the use in control design. System identification finds broad application in areas such as digital twin creation (Cech, Beltman, & Ozols, 2022) and control design (Munnig Schmidt, Schitter, Rankers, & van Eijk, 2020). Two distinct approaches are non-parametric and parametric system identification. Non-parametric system identification techniques, such as Frequency Response Functions (FRFs) (Pintelon & Schoukens, 2012),

model systems without requiring a predefined structure. Unlike non-parametric approaches, parametric system identification generally involves a fixed model structure, e.g., using prediction error methods (Ljung, 1998; Söderström & Stoica, 1989). A key development in system identification involves the application of kernel methods (Ljung, Chen, & Mu, 2020). The unique properties of kernels offer a major advantage by enforcing high-level properties on the identified models, that in addition enable the identification of continuous-time models (Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). The high-level properties that are enforced by kernels include for example exponential decay and correlation of impulse response coefficients. For linear time-invariant equidistantly sampled systems, both parametric and non-parametric system identification techniques have been proven to work effectively.

Important developments have been made in identifying fast-rate models from slow-rate output measurements, primarily in terms of continuous-time and low-order parametric identification. First, continuous-time identification aims to identify a parametric continuous-time model using input–output data, as outlined in Unbehauen and Rao (1990). Typically, these methods require intersample assumptions on the input signals, e.g., zero-order hold or band-limited

signals (Ljung, 2009), and hence, do not utilize broadband fast-rate inputs. Second, several methods are developed for fast-rate parametric model identification using slow-rate output measurements, including methods for identification of low-order Finite Impulse-Response (FIR) (Ding & Chen, 2004) and output-error (Zhu, Telkamp, Wang, & Fu, 2009) models. Subspace (Li, Shah, & Chen, 2001) or hierarchical identification techniques (Ding, Ding, Liu, & Liu, 2011), which identify a lifted system representation, are also developed. Generally, these methods identify a fast-rate model using slow-rate output measurements by incorporating prior information through a certain model structure, which is limited to low-order models and therefore does not include non-parametric models.

A key development in non-parametric identification for fast-rate models using slow-rate output measurements incorporates prior information through a local smoothness assumption on the frequency response (van Haren, Mirkin, Blanken, & Oomen, 2023). The smoothness assumption allows one to disentangle aliased components in the frequency-domain. However, the local smoothness assumption is an approximation, since it assumes that the system can be approximated as a polynomial of a certain degree in the frequency domain, leading to a bias–variance trade-off (Pintelon & Schoukens, 2012, Section 7.2.6).

Although methods to identify fast-rate models using slow-rate output measurements have been developed for certain scenarios, current approaches are limited to specific cases, including low-order parametric modeling and frequency-domain approaches. The aim of this paper is to present a general framework to identify non-parametric fast-rate models using slow-rate output measurements, enabling more generic prior information. The developed method utilizes regularized estimators, which have been proven to work effectively in system identification for equidistantly sampled systems (Ljung et al., 2020). Specifically, system properties are enforced through the use of regularization or kernel techniques, expressed in either the frequency domain, similarly to Hallemans, Pintelon, Joukovsky, Peumans, and Lataire (2022) and Lataire and Chen (2016), or the time domain, that is illustrated in this paper. Compared to earlier regularized system identification approaches, the developed approach uniquely combines and utilizes slow-rate output measurements, fast-rate input measurements, and prior information of the system to identify fast-rate non-parametric models using slow-rate output measurements. The key contributions are the following.

C1 Extending well-established conditions for unique model identification to the identification of fast-rate FIR models using slow-rate output measurements, which are critical for practical applications (Section 3).

C2 Identification of arbitrary order fast-rate models using slow-rate output measurements, including non-parametric models, by incorporating prior information through kernel regularization (Section 4).

C3 Validation of the framework for identification of fast-rate models for systems with slow-rate output measurements using a simulation and an experimental setup consisting of a prototype wafer stage (Sections 5 and 6).

Notation. Fast-rate and slow-rate signals are denoted by subscript h and l, with sampling times and frequencies $T_h = \frac{1}{f_h}$ and $T_l = FT_h = \frac{1}{f_l}$, with downsampling factor $F \in \mathbb{N}_1$ and positive integers $\mathbb{N}$, discrete-time indices $n \in \mathbb{N}_{[0,N-1]}$ and $m \in \mathbb{N}_{[0,M-1]}$, with N, M being the number of samples. Fast-rate and slow-rate discrete Fourier transforms are denoted with

$$X_h(e^{j\omega_k T_h}) = \sum_{n=0}^{N-1} x_h(n)e^{-j\omega_k T_h n},$$
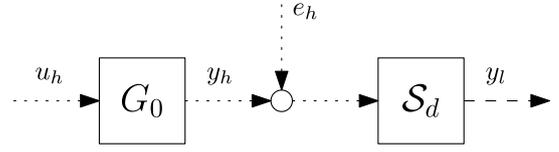$$X_l(e^{j\omega_k T_l}) = \sum_{m=0}^{M-1} x_l(m)e^{-j\omega_k T_l m}, \tag{1}$$

Fig. 1. Identification setup for systems with slow-rate output measurements.

with complex variable $j = \sqrt{-1}$, frequency grid $\omega_k = 2\pi \frac{k}{N} f_h = 2\pi \frac{k}{M} f_l$, and frequency bins $k \in \mathbb{N}$. The fast-rate forward shift operator is denoted by q, i.e., $q^{-1}x_h(n) = x_h(n-1)$. The Kronecker product is denoted with $\otimes$.

2. Problem formulation

In this section, the considered problem is presented. The identification setting and identifiability problem associated to fast-rate models using slow-rate output measurements are illustrated. Finally, the problem addressed in this paper is defined.

2.1. Identification setting

The goal is to identify fast-rate model G using slow-rate output measurements $y_l(m) = S_d y_h(n) = y_h(mF)$ and fast-rate inputs u_h, having sampling rates f_l and $f_h = Ff_l$ and known downsampler S_d. The open-loop identification setting is visualized in Fig. 1. The true fast-rate system G_0 is a linear time-invariant, single-input single-output system, and is described using the causal discrete-time state–space equations

$$G_0 : \begin{cases} x(n+1) & = Ax(n) + Bu_h(n), \\ y_h(n) & = Cx(n) + Du_h(n). \end{cases} \tag{2}$$

The output is disturbed with zero-mean, independent, and identically distributed noise e_h, with variance σ^2, and downsampled into

$$y_l(m) = S_d \left(y_h(n) + e_h(n) \right) = y_h(mF) + e_h(mF). \tag{3}$$

2.2. Identifiability of fast-rate models

In this section, identifiability of fast-rate models for systems with slow-rate outputs is investigated. The system G_0 is modeled using the FIR filter $G(q, \theta)$, which has fast-rate model output

$$\hat{y}_h(n) = G(q, \theta)u_h(n) = \sum_{i=0}^{P-1} \theta_i q^{-i} u_h(n), \tag{4}$$

with model order $P \in \mathbb{N}_{[1,N]}$, and FRF

$$G(e^{j\omega T_h}) = \sum_{i=0}^{N-1} \theta_i e^{-j\omega T_h i}, \tag{5}$$

which can be evaluated at any frequency $\omega \in \mathbb{R}$. Similarly to (3), the slow-rate model output is downsampled as $\hat{y}_l(m) = S_d \hat{y}_h(n) = \hat{y}_h(mF)$.

Identifiability of fast-rate models essentially involves determining whether the fast-rate input and the model structure allow for distinguishing between parameter values given the slow-rate output, which relates closely to the definition in Ljung and Glad (1994). Specifically, the fast-rate input should be sufficiently informative to distinguish between different parameter sets θ. Additionally, the model structure $G(q, \theta)$ should allow for distinguishing between different parameter sets θ, provided a sufficiently informative input. The identifiability of fast-rate models is defined in Definition 1.

Definition 1. The FIR coefficients θ of fast-rate model $G(q, \theta)$ in (4) are identifiable from slow-rate outputs if

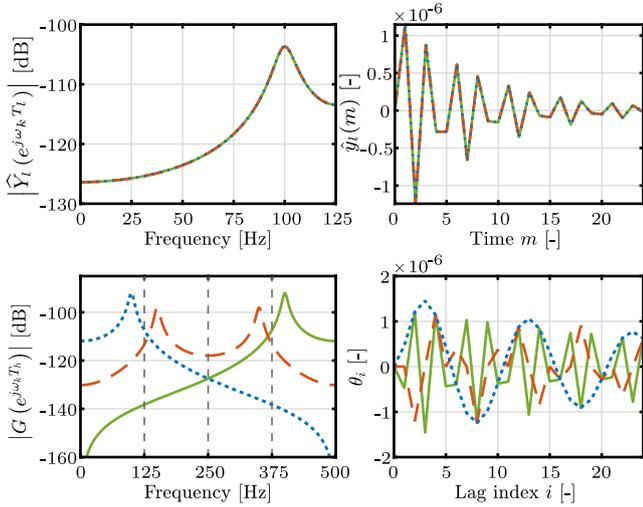$$S_d G(q, \theta)u_h(n) = S_d G(q, \theta^*)u_h(n), \quad \Rightarrow \theta = \theta^*. \tag{6}$$

Fig. 2. A non-identifiable model structure $G(q,\theta)$ for three specific impulse response coefficients θ (——), θ^* (— —), and θ' (····) (bottom right), shown by their equal downsampled output in time-domain $\hat{y}_l$ (top right) and frequency-domain $\hat{Y}_l\left(e^{jk\omega_k T_l}\right)$ (top left) for impulsive input u_h.

Fast-rate models are not always identifiable from slow-rate outputs, where several examples are illustrated in Example 1.

Example 1. Consider the model structure $G(q,\theta)$ (4) for three parameter values θ, θ^*, and θ' with order $P = 1000$. Let u_h be a fast-rate impulsive input with $N = P = 1000$ samples, being persistently exciting of order $P = N$ (Ljung, 1998, Lemma 13.1). After downsampling by $F = 4$, resulting in $M = 250$ samples in the slow-rate output y_l, the responses of the three models are identical, i.e.,

$$S_d G(q,\theta)u_h(n) = S_d G(q,\theta^*)u_h(n) = S_d G(q,\theta')u_h(n),$$

as illustrated in the top of Fig. 2. However, these are generated by different fast-rate models, as highlighted by the fast-rate FRFs and impulse response coefficients in the bottom of Fig. 2, indicating that the model is not identifiable according to Definition 1.

2.3. Problem definition and approach

The fast-rate model $G(q,\theta)$ from (4) is not always identifiable from slow-rate outputs given an arbitrary model order P and fast-rate inputs u_h, as illustrated by Example 1.

The problem considered in this paper is as follows. Given the fast-rate input signal u_h containing N samples and slow-rate output measurements y_l containing M samples, determine a fast-rate model $G(q,\theta)$ of arbitrary order $P \in \mathbb{N}_{[1,N]}$, by incorporating appropriate prior information of the system.

3. Identification of slow-rate systems

In this section, reduced-order FIR models are determined using a least-squares cost function, and several limitations thereof are shown, thereby constituting contribution C1. Specifically, these limitations include that the model order is restricted to the length of the data of the slow-rate output. Additionally, the input to the system cannot be zero-order hold, and hence, needs to be excited at the fast-rate.

First, the fast-rate model output in (4) is vectorized as

$$\hat{y}_h = \begin{bmatrix} \hat{y}_h(0) & \hat{y}_h(1) & \cdots & \hat{y}_h(N-1) \end{bmatrix}^\top. \tag{7}$$

Downsampling $\hat{y}_h$ in (7) with a factor F, which essentially involves selecting every F^{th} value of $\hat{y}_h$, results in the slow-rate model output $\hat{y}_l$, i.e.,

$$\hat{y}_l = \begin{bmatrix} \hat{y}_l(0) & \hat{y}_l(1) & \cdots & \hat{y}_l(M-1) \end{bmatrix}^\top = \Phi\theta \tag{8}$$

with parameter vector $\theta = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{P-1} \end{bmatrix}^\top$ and regressor matrix $\Phi \in \mathbb{R}^{M\times P}$, that is

$$\Phi = \begin{bmatrix} u_h(0) & 0 & \cdots & \cdots & 0 \\ u_h(F) & \cdots & u_h(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_h((M-1)F) & \cdots & \cdots & \cdots & u_h((M-1)F-P) \end{bmatrix}. \tag{9}$$

The goal is to identify θ given u_h and y_l. Similar to identifiability of fast-rate models in Definition 1, unique solutions for θ given u_h are defined as

$$S_d\Psi(\theta - \theta^*) = \Phi(\theta - \theta^*) = 0, \quad \Rightarrow \theta = \theta^*. \tag{10}$$

Unique fast-rate models are analyzed in Lemmas 1 and 2.

Lemma 1. *If $P \geq M$, there does not exist a unique fast-rate model $G(q,\theta)$ as defined in (10).*

Proof. Since matrix $\Phi \in \mathbb{R}^{M\times P}$ with $P > M$ has more columns than rows, the columns must be linearly dependent. This means there exists a non-zero vector $(\theta - \theta^*) \in \mathbb{R}^P$ such that $\Phi(\theta - \theta^*) = 0$, resulting in a non-unique model according to (10). $\square$

Lemma 1 illustrates that no more parameters θ can be identified than the M samples of the output y_l.

Lemma 2. *There does not exist a unique fast-rate model $G(q,\theta)$ with $P > 2$, given u_h and y_l, for zero-order hold input $u_h(mF + i) = u_h(mF) = u_l(m) \ \forall i \in \mathbb{N}_{[0,F-1]}$.*

Proof. The regressor matrix $\Phi \in \mathbb{R}^{M\times P}$ for zero-order hold inputs is equal to the first P columns of the matrix

$$\begin{bmatrix} u_l & \Phi_u \otimes 1_{1\times F} & 0_{M\times F-1} \end{bmatrix} = \begin{bmatrix} \Phi & | & \Phi^* \end{bmatrix} \in \mathbb{R}^{M\times N}, \tag{11}$$

where $\Phi^* \in \mathbb{R}^{M\times N-P}$ and

$$\Phi_u = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ u_l(0) & 0 & \cdots & \vdots \\ u_l(1) & u_l(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ u_l(M-2) & u_l(M-3) & \cdots & u_l(0) \end{bmatrix} \in \mathbb{R}^{M\times M-1}. \tag{12}$$

As a result of the repeating columns of Φ in (11), the matrix Φ has non-zero null-space. Therefore, there does not exist a unique fast-rate model $G(q,\theta)$ with $P > 2$ as defined in (10) for zero-order hold inputs. $\square$

Lemma 2 illustrates that zero-order hold inputs cannot be used to uniquely identify fast-rate models with $P > 2$.

The above conditions have to be taken into account for obtaining a unique estimate from a least-squares cost function. For estimating fast-rate models using slow-rate output measurements, the least-squares problem minimizes the squared error between measured and estimated slow-rate outputs (8) as

$$\min_\theta V(\theta) = \min_\theta \|y_l - \Phi\theta\|_2^2. \tag{13}$$

The least-squares problem (13) has minimizer

$$\theta = \left(\Phi^\top\Phi\right)^{-1}\Phi^\top y_l. \tag{14}$$

Lemma 1 illustrates that fast-rate FIR models cannot be uniquely determined for order $P > M$ using this generic least-squares cost function. Additionally, Lemma 2 shows that fast-rate models cannot be uniquely determined for zero-order hold inputs, and therefore, fast-rate inputs are required.

4. Kernel-regularized identification above the Nyquist frequency

In this section, accurate fast-rate models beyond the Nyquist frequency of slow-rate output measurements are determined for arbitrary orders $P \in \mathbb{N}_{[1,N]}$, including non-parametric models $P = N$, hence relaxing the limitation of $P < M$ and constituting contribution C2. This is realized through the use of regularization techniques that enforce certain properties of the estimated model.

First, the regularized estimation problem is posed. Second, several design aspects are illustrated, including regularization options and a procedure that summarizes the approach.

4.1. Regularized FIR estimation above the Nyquist frequency

Properties of the estimated FIR coefficients θ are enforced by regularizing the cost function in (13) with $\theta^\top K^{-1} \theta$, i.e.,

$$\min_\theta V(\theta) + \gamma \theta^\top K^{-1} \theta$$
$$= \min_\theta \|y_l - \Phi\theta\|_2^2 + \gamma \theta^\top K^{-1} \theta \tag{15}$$

with regularization parameter γ and regularization or kernel matrix K. The optimal solution to (15) results in regularized least-squares estimation of fast-rate FIR models using slow-rate output measurements, and is given by

$$\theta = K\Phi^\top \left(\Phi K \Phi^\top + \gamma I_M\right)^{-1} y_l, \tag{16}$$

with Φ from (8), and regularization matrix

$$K = \begin{bmatrix} k(0,0) & k(0,1) & \cdots & k(0, N-1) \\ k(1,0) & k(1,1) & \cdots & k(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ k(N-1,0) & k(N-1,1) & \cdots & k(N-1, N-1) \end{bmatrix}. \tag{17}$$

The entries $k(i,j)$ of regularization matrix K enforce certain high-level properties on the FIR coefficients, and are treated in more detail in Section 4.2. The fast-rate FIR model is uniquely determined as shown in Theorem 1, leading to the main result in this section.

Theorem 1. *The fast-rate FIR model using slow-sampled outputs of order $P \in \mathbb{N}_{[1,N]}$, when determined using the regularized solution in (16), is uniquely determined if $\gamma > 0$.*

Proof. $\Phi K \Phi^\top + \gamma I_M$ is invertible when $\gamma > 0$, since

$$x^\top \left(\Phi K \Phi^\top + \gamma I_M\right) x = x^\top \Phi K \Phi^\top x + \gamma x^\top x > 0,$$
$$\forall x \in \mathbb{R}^M \setminus \{0\}. \tag{18}$$

Therefore, $K\Phi^\top \left(\Phi K \Phi^\top + \gamma I_M\right)^{-1} y_l$ from (16) has a unique solution when $\gamma > 0$. □

In sharp contrast to the reduced-order least-squares estimator, the regularized least-squares estimator uniquely determines fast-rate FIR models of order $P \in \mathbb{N}_{[1,N]}$, independent of the order P or fast-rate inputs u_h. This is explained because the regularized least-squares estimator incorporates certain prior information about the fast-rate impulse response coefficients by using the regularization matrix P as the prior covariance matrix of θ (Pillonetto et al., 2014). An intuitive example illustrating that the regularized least-squares estimator is always capable of estimating a model of arbitrary order is given in Example 2.

Example 2. Let the length of the data in the slow-rate output be $M = 1$, i.e., $y_l \in \mathbb{R}$. If $\gamma > 0$, the model (16) is uniquely determined for order $P \in \mathbb{N}_{[1,N]}$ for any $N > 1$, according to Theorem 1. In sharp contrast, a fast-rate model determined with unregularized least-squares (14) can only be uniquely identified for $P = 1$.

Example 2 illustrates that the regularized least-squares estimator is capable of estimating a fast-rate model of arbitrary order even when the slow-rate output measurements are not informative. This is explained because the kernel-regularized estimator will always return a model of arbitrary order, independent of data length or persistence of excitation, by specifying prior information through the regularization matrix P.

Remark 1. Note that there is an equivalence between kernel regularized estimates and Gaussian process regression Pillonetto et al. (2014, Section 4.3). In addition, Gaussian process prediction, or reformulating (15) as a reproducing kernel Hilbert space function estimation problem (Pillonetto et al., 2014, Section 9), enables estimating the FIR model at arbitrary time indices.

4.2. Design aspects

In this section, several kernel functions are introduced and the developed approach is summarized in a procedure.

The selection of kernel functions $k(i,j)$ is essential, since they determine key regularization properties of the estimated fast-rate FIR model, including smoothness, decay, and resonant dynamics. Different choices of kernel functions impose different structural constraints, which influence a trade-off between model flexibility and model fit. The most generic kernel matrix is the identity matrix, i.e.,

$$k_T(i,j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \tag{19}$$

that is generally referred to as ridge regression or Tikhonov regularization (Tikhonov & Arsenin, 1977), that enables estimating $P > M$ FIR coefficients, in exchange for a certain amount of bias. Alternatively, kernels encoding properties that are reasonable to assume for impulse responses include the Diagonal-Correlated (DC) and stable-spline kernels. The DC kernel is given by

$$k_{DC}(i,j) = \lambda \alpha^{(i+j)/2} \beta^{|j-i|}, \tag{20}$$

where $0 \leq \alpha < 1$ accounts for exponential decay and $|\beta| < 1$ describes the correlation between neighboring impulse response coefficients. The stable-spline kernel is given by

$$k_{SS}(i,j) = \lambda \left(\frac{\alpha^{i+j+\max(i,j)}}{2} - \frac{\alpha^{3\max(i,j)}}{6} \right), \tag{21}$$

where α plays a similar role as for the DC kernel. Finally, the prior knowledge kernel, which encodes information about poles of a system, is given by Hallemans et al. (2022)

$$k_{pk,n}(i,j) = \alpha_n^{(i+j)/2} \left(\gamma_{1,n} \cos\left(\omega_n(i-j)\right) + \gamma_{2,n} \cos\left(\omega_n(i+j)\right) \right) \tag{22}$$

where α_n has a similar role as for the DC kernel (20), $\omega_n \in \mathbb{R}_{[0,2\pi)}$ is the normalized frequency of the pole, and

$$\gamma_{1,n} = \frac{\sigma_{1,n}^2 + \sigma_{2,n}^2}{2}, \quad \gamma_{2,n} = \frac{\sigma_{1,n}^2 - \sigma_{2,n}^2}{2}, \tag{23}$$

where $\sigma_{1,n}^2$ and $\sigma_{2,n}^2$ are the variances related to the response of the pole from the prior knowledge kernel (Hallemans et al., 2022). The dependence on n allows the prior knowledge kernel to encode information about multiple poles, which is achieved by summing the kernel functions over n. The prior knowledge kernel is especially effective in estimating systems with lightly-damped resonant dynamics, which can generally not be estimated accurately using kernels that assume a decaying impulse response such as the DC and stable-spline kernels.

Remark 2. Note that for $\gamma = 0$, the developed kernel-regularized method (16) recovers the least-squares FIR estimator (14) described in Section 3.

The kernel functions depend on the parameters λ, α, ρ, α_n, ω_n, and $\gamma_{i,n}$, which are typically referred to as hyperparameters. Selecting the hyperparameters is crucial, since they directly influence high-level properties of the fast-rate FIR model, such as the exponential decay α and the pole frequency ω_n.

The hyperparameters can be tuned in several ways. First, the hyperparameters can be optimized using the marginal likelihood (Pillonetto et al., 2014)

$$\hat{\eta} = \arg \min_{\eta} y_l^\top \left(\Phi K(\eta) \Phi^\top + \gamma I_M \right)^{-1} y_l + \log \det \left(\Phi K(\eta) \Phi^\top \right), \tag{24}$$

where the kernel matrix $K(\eta)$ from (17) is parameterized using the hyperparameters η. Additionally, the regularization parameter γ can be included in η and optimized simultaneously. Since only slow-rate output measurements are available, marginal-likelihood (24) optimizes the hyperparameters based on their effect on the slow-rate output measurements y_l. As a result, marginal likelihood optimization effectively selects the hyperparameters that balances slow-rate output data fit and model complexity (Rasmussen, 2006).

Alternatively, an additional data set can be used for cross-validation to optimize the hyperparameters by selecting them to generalize well to unseen data (Pillonetto et al., 2014), which is shown to be effective in practice.

The kernel functions encode prior information about the impulse response coefficients, and hence, introduce correlation between the coefficients that enable estimating the N coefficients from $M < N$ samples in y_l. The developed approach is summarized in Procedure 1, that links the main results in this paper.

Procedure 1 (*Identify fast-rate FIR model of arbitrary order using slow-rate output measurements and fast-rate inputs*).

1. *Construct fast-rate input u_h, e.g., random noise or a random binary sequence.*
2. *Apply input u_h to the system and record y_l.*
3. *Construct the regressor matrix Φ according to (8).*
4. *(Optional:) Tune the hyperparameters η and γ, e.g., based on marginal likelihood optimization (24).*
5. *Construct kernel matrix K using (17).*
6. *Calculate the P FIR coefficients θ via least-squares (14) and kernel-regularized (16) approaches.*

---

5. Simulation example

In this section, the developed method is validated using a Monte Carlo simulation, contributing to contribution C3.

5.1. Simulation setup

A Monte-Carlo simulation study is carried out for the mass–spring–damper system in Fig. 3, where the settings are listed in Table 1. The nominal system parameters are given by

$k_1 = 15$ N/m, $\quad k_2 = 100$ N/m, $\quad d_1 = 0.45$ N s/m,

$d_2 = 0.06$ N s/m, $\quad m_1 = 1$ kg, $\quad m_2 = 1$ kg.

During every Monte Carlo iteration, each parameter is perturbed randomly within 10% of its nominal value. The FRFs of the system during the Monte Carlo runs are seen in Fig. 4. Additional simulation settings are seen in Table 1.

Independent training and validation data sets are generated, where the excitation signals u_h are random phase multisines with the same flat amplitude spectrum but different phase realizations. The output of the training set is disturbed with additive white Gaussian noise e_h, where the signal-to-noise ratio, that is the ratio between the variance of y_h and e_h, is randomly chosen in $\mathbb{R}_{[40, 60]}$.

The following models are estimated using training data and compared on validation data in each Monte Carlo run.
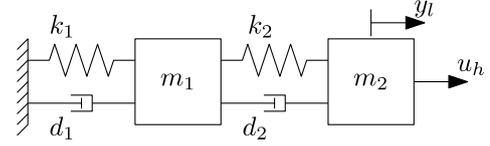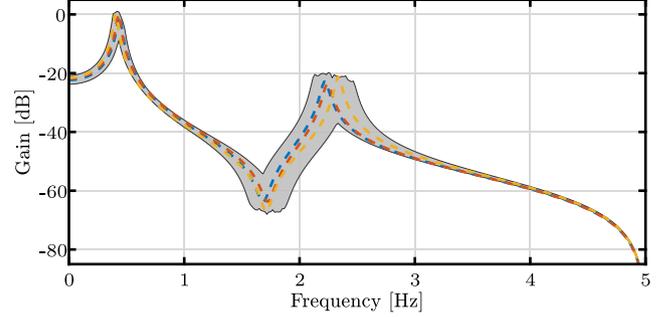

Fig. 3. Mass-spring–damper system considered.

Fig. 4. The FRFs during 100 Monte Carlo simulations are contained in (▢), where three FRFs are shown as (- -, - -, - -).

Table 1
Simulation settings.

Variable	Abbreviation	Value	Unit
Fast sampling time	T_h	0.1	s
Slow sampling time	T_l	0.3	s
Downsampling factor	F	3	–
Number of input samples	N	600	–
Number of output samples	M	200	–
Monte Carlo runs	–	100	–

G_{LS} Unregularized least-squares FIR from (14) of order $P \in \mathbb{N}_{[50,N]}$.

G_{DC} Kernel regularized FIR using (16) of order $P \in \mathbb{N}_{[50,N]}$, which uses the DC kernel k_{DC} (20).

G_{PK} Kernel regularized FIR using (16) of order $P \in \mathbb{N}_{[50,N]}$, which uses the kernel function with prior knowledge from (22), specifically

$$k_{PK}(i,j) = k_{DC}(i,j) + \sum_{n=1}^{2} k_{pk,n}(i,j). \tag{25}$$

The hyperparameters of the kernel regularized estimators are chosen as

$\lambda = 1$, $\qquad \omega_1 = 0.4 T_h \cdot 2\pi$, $\qquad \omega_2 = 2 T_h \cdot 2\pi$,

$\alpha = e^{-0.5 T_h}$, $\qquad \alpha_1 = \alpha_2 = e^{-0.5 T_h}$, $\qquad \gamma = 10^{-5}$,

$\beta = e^{-0.1 T_h}$, $\qquad \sigma_{1,n} = \sigma_{2,n} = 1$,

where $\gamma_{1,n}$ and $\gamma_{2,n}$ are determined using (23). Additionally, the regularization parameter γ and the hyperparameters ω_n and α_n are optimized using marginal likelihood optimization (24) with a gradient-based constrained optimizer to further enhance accuracy.

5.2. Simulation results

For each Monte Carlo run, the Goodness of Fit (GoF) of the fast-rate output y_h on the validation data is calculated,

$$\text{GoF} = 100 \left(1 - \frac{\sum_{n=0}^{N-1} \left(y_{h,v}(n) - \hat{y}_h(n) \right)^2}{\sum_{n=0}^{N-1} \left(y_{h,v}(n) - \bar{y}_{h,v} \right)^2} \right), \tag{26}$$

where $\hat{y}_h$ is estimated using the fast-rate FIR models, $y_{h,v}(n)$ are the validation values, and $\bar{y}_{h,v}$ is the mean value of $y_{h,v}(n)$. The mean plus and minus 2 times its sample standard deviation over the Monte Carlo runs are seen in Fig. 5. Additionally, the fast-rate FRFs have been evaluated for one Monte Carlo run, and are seen in Fig. 6. The following observations are made from the simulation results.
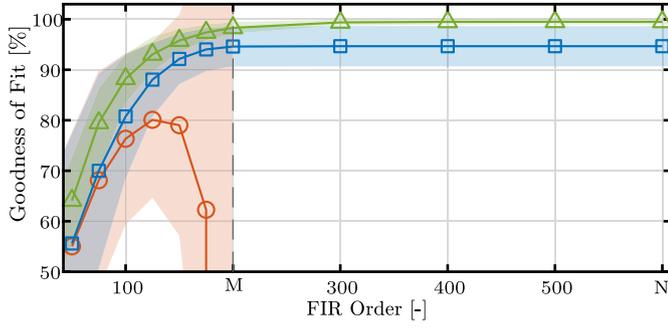
Fig. 5. Kernel regularized estimators using DC kernel (20) (──□──) or prior knowledge kernel (25) (──△──) are capable of achieving a higher GoF for all model orders compared to least-squares FIR models (──○──). In addition, the mean GoF ± 2 times its sample standard deviation (shaded areas) show that the prior knowledge kernel decreases variance.
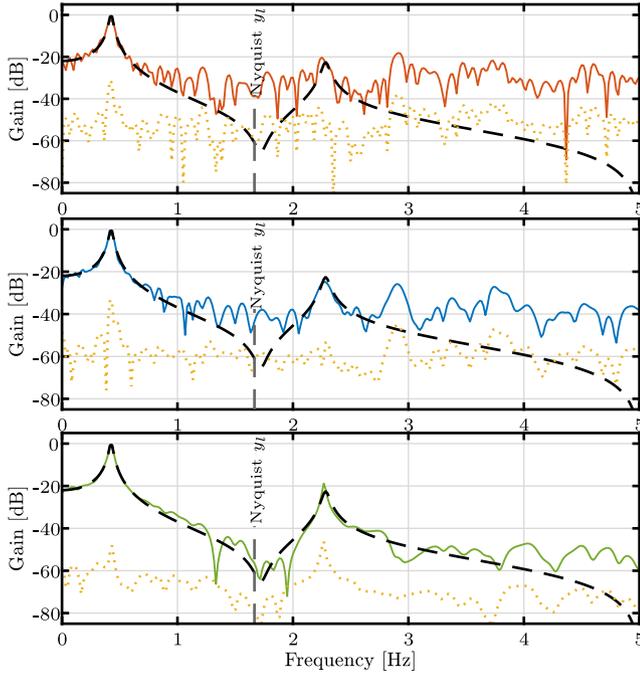

Fig. 6. The regularized FIR models with order $P = N = 600$ improve the identification of the true FRF (− −) using a DC kernel (──) (middle) and a prior knowledge kernel (──) (bottom). The least-squares FIR model with highest GoF having order $P = 150$ (──) (top) captures only the first resonance. The regularized FIR model with DC kernel partially captures the second resonance, while the prior knowledge kernel model captures both resonances. These differences are reflected in the scaled FRF error $0.1 \cdot \left| G_0 \left(e^{j\omega_k T_h} \right) - G \left(e^{j\omega_k T_h} \right) \right|$ (·····), which is lower for the regularized models.

- From the GoF in Fig. 5, the following is observed.

  – The least-squares FIR model is capable of estimating a model of order $P < M$, but not for $P > M$, supporting Lemma 1.

  – The kernel regularized FIR model with DC kernel accurately identifies models with low variance for all $P \in \mathbb{N}_{[50,N]}$, due to the additional prior that is encoded in the estimator.

  – The additional prior information about the system's poles introduced by the prior-knowledge kernel (25) enables a further improvement in GoF and a reduction in variance for $P > 200$, achieving a GoF of 99.50% compared to 94.67% for the DC kernel.

- From the FRFs in Fig. 6, the following is observed.

– The FRF of the least-squares FIR model with $P = 150$ in the top does not identify the FRF accurately, especially beyond the Nyquist frequency.

– The FRF of the regularized FIR models in the middle and bottom are significantly more accurate with lower FRF error, especially beyond the Nyquist frequency of the slow-rate output. In particular, the FIR model with prior knowledge kernel (bottom) accurately identifies the resonant behavior.

6. Experimental validation

In this section, the developed method is validated using an experimental prototype wafer stage with a slow-rate output, contributing to contribution C3.

6.1. Experimental setup

The experimental setup is the prototype wafer stage in Fig. 7 (Classens, van de Wijdeven, Heemels, & Oomen, 2023), which is used in semiconductor manufacturing and a prime example of mechatronic systems with slow-rate outputs.

The objective of the wafer stage is to accurately control the vertical displacement v of the point-of-interest on the wafer shown in Fig. 7(a), which is the point on the wafer during lithographic exposure. The chuck and wafer are levitated, and actuated by four Lorentz actuators. The vertical displacement of the chuck is measured by four linear encoders and used as inputs to the feedback controller as $y = \frac{1}{4} \left(y_1 + y_2 + y_3 + y_4 \right)$. Directly measuring the vertical displacement of the point-of-interest on the wafer v is not possible using linear encoders. The chuck of the wafer stage has internal lightly-damped structural modes. Consequently, measuring the vertical displacement on the outside of the chuck does not coincide with the vertical displacement at the wafer's point of interest (Heertjes et al., 2020; Oomen, Grassens, & Hendriks, 2015). Therefore, an external capacitive sensor, denoted by scanning sensor in Fig. 7(b), directly measures the vertical displacement of the point-of-interest. The external capacitive sensor is sampled at a reduced sampling rate compared to the actuators, resulting in a slow-rate output.

6.2. Identification setting

For this experimental validation, the goal is to identify the fast-rate equivalent model of the prototype wafer stage

$$G_z (q, \theta) = P_z \left(q, \theta \right) \left(I + C (q) P_y (q, \theta) \right)^{-1} : u_h \mapsto z_h, \tag{27}$$

using the fast-rate excitation u_h and slow-rate outputs z_l. The prototype wafer stage is operating in closed-loop control, where the control structure is seen in Fig. 8. For this experimental validation, the fast-rate excitation signal u_h is equally distributed over the four actuators, and is considered a disturbance to the plant. The excitation signal u_h is a random noise sequence with flat amplitude spectrum, resulting in a signal-to-noise ratio of approximately 45 dB. The scanning sensor is suspended above the wafer, and is positioned in the bottom left corner of the wafer, 110 mm in both directions from the center. The experimental settings are seen in Table 2. Similar to Section 5, FIR models are identified using unregularized and regularized least-squares cost functions, i.e., the following methods are compared.

G_{LS} Unregularized least-squares FIR from (14) of order $P \in \mathbb{N}_{[5,N]}$.

G_{DC} Kernel regularized FIR using (16) of order $P \in \mathbb{N}_{[5,N]}$, which uses the DC kernel k_{DC} (20).

G_{PK} Kernel regularized FIR using (16) of order $P \in \mathbb{N}_{[5,N]}$, which uses the kernel function with prior knowledge from (22), specifically

$$k_{PK}(i, j) = k_{DC}(i, j) + \sum_{n=1}^{3} k_{pk,n}(i, j). \tag{28}$$
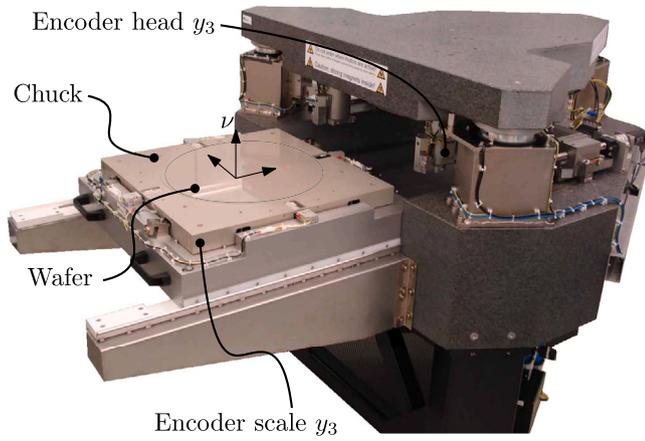
(a) Photograph of experimental setup.
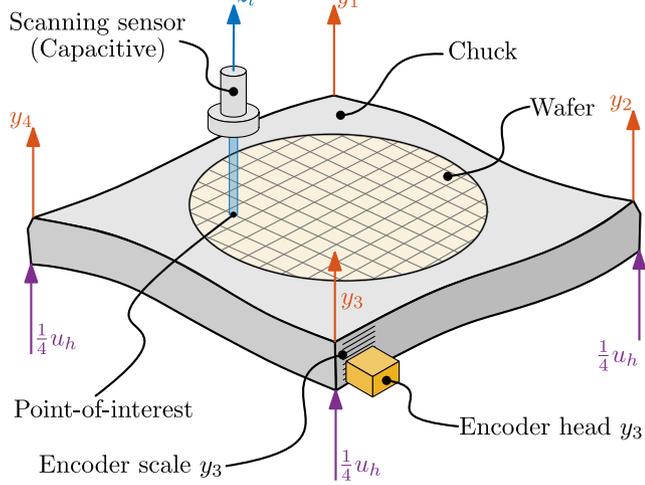
(b) Schematic overview of experimental setup.

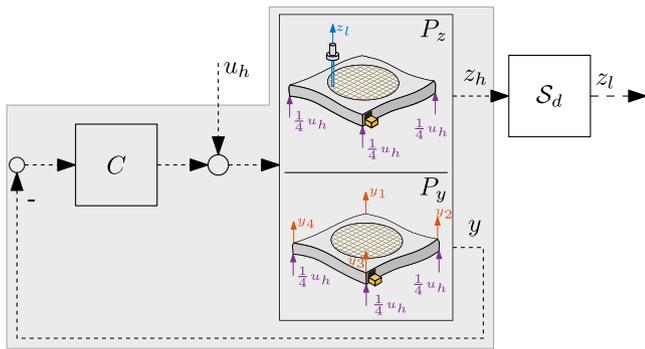**Fig. 7.** Prototype experimental wafer stage setup used in semiconductor manufacturing with a slow-rate output.

Fig. 8. Experimental feedback scheme used, where the equivalent system in (27) is to be identified.

Table 2
Experimental settings.

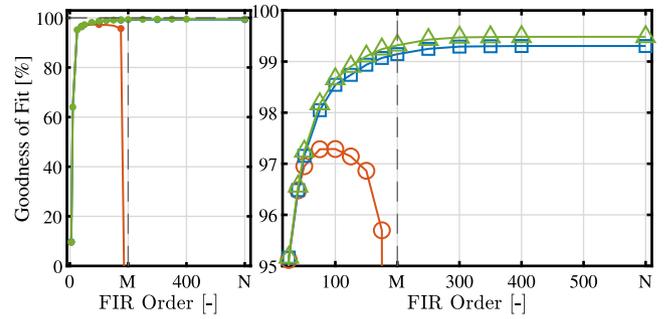| Variable	Abbr.	Value	Unit
Fast sampling time	T_h	0.5	ms
Slow sampling time	T_l	1.5	ms
Downsampling factor	F	3	–
Number of input samples	N	600	–
Number of output samples	M	200	–

Fig. 9. The GoF of the regularized FIR models with DC kernel (20) (-□-) and prior knowledge kernel (28) (-△-) are higher compared to the unregularized least-squares FIR model (-○-) for all model orders $P \in \mathbb{N}_{[5,N]}$ (left) and $P \in \mathbb{N}_{[25,N]}$ (right).

The hyperparameter of the kernels are initialized as

$$\lambda = 1, \qquad \alpha = \alpha_n = e^{-80T_h},$$
$$\omega_1 = 490T_h \cdot 2\pi, \quad \omega_2 = 510T_h \cdot 2\pi, \quad \omega_3 = 870T_h \cdot 2\pi,$$
$$\beta = e^{-20T_h}, \qquad \sigma_{1,n} = \sigma_{2,n} = 1, \qquad \gamma = 2 \cdot 10^{-1}$$

where $\gamma_{1,n}$ and $\gamma_{2,n}$ are determined using (23). The hyperparameters ω_n and α_n are refined using a gradient based constrained optimizer for the marginal likelihood (24).

For validation purposes, the fast-rate outputs z_h are recorded as well, and a validation data set with similar signal-to-noise ratio and the same amount of samples $N = 600$ is measured to compare the approaches. Finally, a significantly longer data set containing $N = 50000$ samples is measured, where z_h is recorded as well, to identify the validation benchmark FRF

$$G_0\left(e^{j\omega_k T_h}\right) : \frac{U_h\left(e^{j\omega_k T_h}\right)}{Z_h\left(e^{j\omega_k T_h}\right)}. \tag{29}$$

This validation FRF is identified with the local rational modeling method from McKelvey and Guérin (2012), using local rational degrees $R = 4$ and local window size $n_w = 150$.

6.3. Experimental results

The GoF in (26) for the validation data set is shown for various model orders in Fig. 9. The estimated FRFs are seen in Fig. 10, including an enlarged section of the resonant behavior around 500 and 870 Hz. The estimated phase is shown in Fig. 11. The following observations are made with respect to the experimental results.

• From the GoF in Fig. 9, the following is observed.

  – The GoF for the regularized FIR model estimates using either the DC or prior knowledge kernel is high for all model orders, close to 100% for $P \to N$.
  – The GoF for the least-squares FIR model estimates is significantly lower, and is not capable of estimating a model for $P > M$.
  – The regularized FIR model with prior knowledge kernel and $P = N$ achieves a GoF of 99.44%, and a root-mean-squared error of

$$\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left(y_{h,v}(n) - \hat{y}_h(n)\right)} = 1.65 \times 10^{-8} \text{ m},$$

  where the maximum absolute value and the RMS value of $y_{h,v}$ are $7.2 \cdot 10^{-7}$ m and $2.2 \cdot 10^{-7}$ m. In comparison, the least-squares FIR model with highest GoF 97.29% has a root-mean-squared error of 3.62×10^{-8} m, which is more than double.

• From the FRFs in Fig. 10, the following is observed.
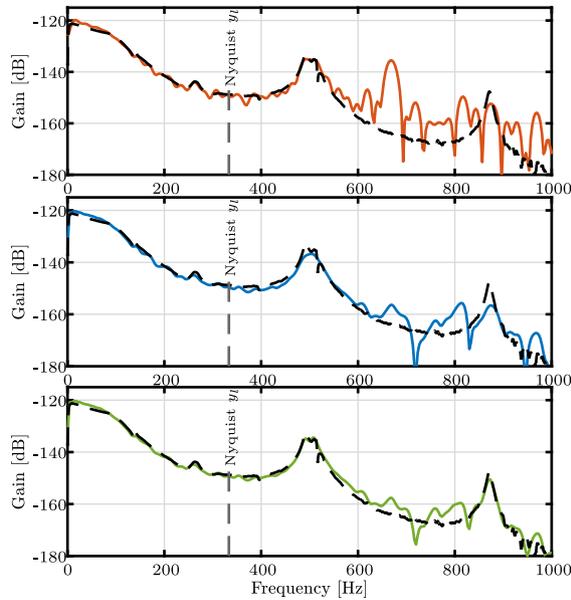
(a) The regularized FIR models with DC kernel (——) and prior knowledge kernel (——) improve the estimate of the validation FRF $G_0\left(e^{j\omega_k T_h}\right)$ (29) (- -) compared to the least-squares FIR model (——).
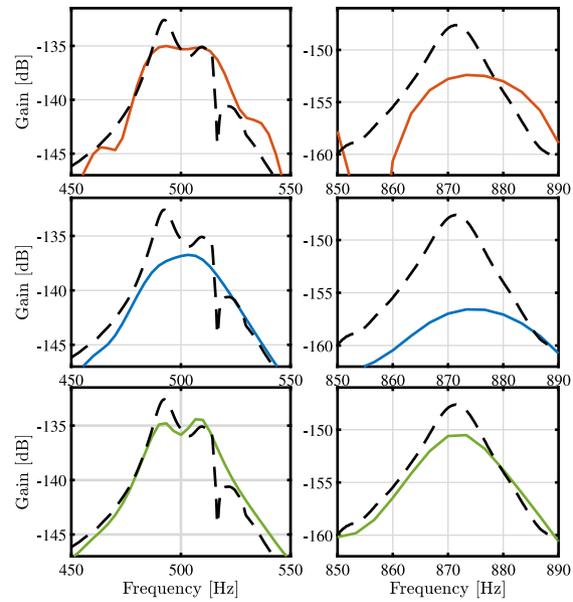
(b) Enlarged sections of the resonant behavior around 500 Hz (left) and 870 Hz (right) show that the FIR model with prior knowledge kernel accurately identifies the resonant behavior compared to the validation FRF $G_0\left(e^{j\omega_k T_h}\right)$ (29).

Fig. 10. The regularized FIR models with order $P = N = 600$ improve identification of the validation FRF $G_0\left(e^{j\omega_k T_h}\right)$ (29) for the prototype wafer stage (- -), using a DC kernel (——) (middle) and a prior knowledge kernel (——) (bottom), compared to the least-squares FIR model with highest GoF having order $P = 100 < M$ (——) (top).

- Both the regularized FIR models with DC kernel and prior knowledge kernel estimate the FRF accurately, without erratic variations, even beyond the Nyquist frequency of the slow-rate output.
- The resonant behavior at 500 and 870 Hz in Fig. 10(b) is estimated accurately by the regularized FIR model with prior knowledge kernel, which is explained by the additional prior information that is introduced.
- The least-squares FIR model inaccurately estimates the FRF beyond 550 Hz. Furthermore, it introduces several artificial resonances, especially around $f_l = 666\frac{2}{3}$ Hz, which is caused by aliasing of the gain around 0 Hz to this frequency.

• Fig. 11 shows that the phase of the wafer stage is estimated accurately for all methods up to 600 Hz. At higher frequencies, the regularized FIR model with prior-knowledge kernel provides the most accurate phase estimation.

7. Conclusions

The results in this paper determine conditions for unique identification of reduced-order FIR models and arbitrary order fast-rate models for systems with slow-rate output measurements, including non-parametric models, and additionally enables intersample response estimation. The method incorporates generic prior information through kernel regularization, that determines arbitrary order fast-rate models. This enables the usage of prior knowledge kernels that encode information about the poles of a system, resulting in a method which is highly suitable for lightly-damped systems.

The framework is validated using a Monte Carlo simulation example, which shows that the regularized FIR models have increased estimation quality and decreased variance. Furthermore, the framework is validated on a prototype wafer stage with a slow-rate output, showing that the regularized FIR models are capable of estimating the system, including lightly-damped resonant dynamics. The developed approach is a key enabler for control design of systems with slow-rate output measurements.
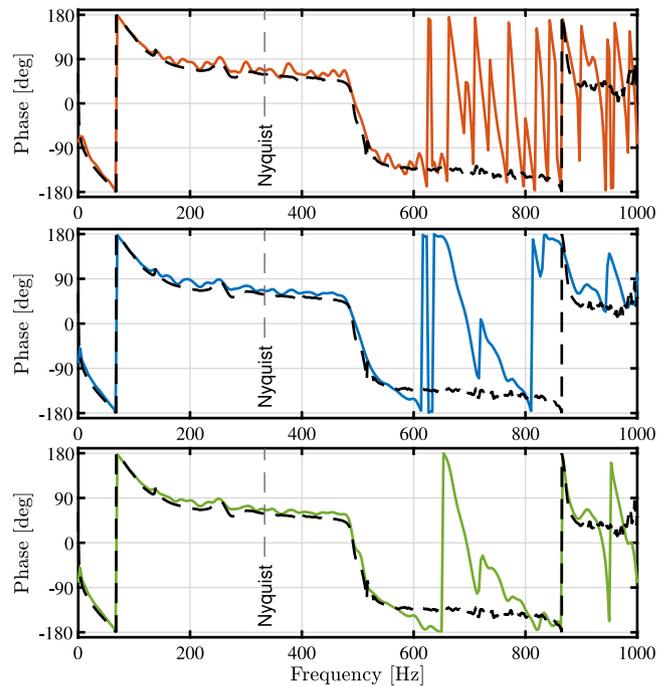

Fig. 11. Phase of prototype wafer stage (- -) estimated by unregularized FIR model (——), and regularized FIR models with DC kernel (——) and prior-knowledge kernel (——).

CRediT authorship contribution statement

Max van Haren: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Roy S. Smith:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Tom Oomen:** Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Max van Haren reports financial support was provided by Horizon Europe. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Leonid Mirkin for his help and fruitful discussions, and Koen Classens for gathering experimental data, which both led to the results in this paper.

References

Bamieh, B., Pearson, J., Francis, B. A., & Tannenbaum, A. (1991). A lifting technique for linear periodic systems with applications to sampled-data control. *Systems & Control Letters*, *17*(2), 79–88. http://dx.doi.org/10.1016/0167-6911(91)90033-B.

Cech, M., Beltman, A.-J., & Ozols, K. (2022). Digital twins and AI in smart motion control applications. In *2022 IEEE 27th international conference on emerging technologies and factory automation* (pp. 1–7). IEEE, http://dx.doi.org/10.1109/ETFA52439.2022.9921533.

Chen, T., & Francis, B. A. (1995). *Optimal sampled-data control systems*. Springer, http://dx.doi.org/10.1007/978-1-4471-3037-6.

Classens, K., van de Wijdeven, J., Heemels, M., & Oomen, T. (2023). From fault diagnosis and predictive maintenance to control reconfiguration. *Mikroniek*, (5), 5–12.

Ding, F., & Chen, T. (2004). Identification of dual-rate systems based on finite impulse response models. *International Journal of Adaptive Control and Signal Processing*, *18*(7), 589–598. http://dx.doi.org/10.1002/acs.820.

Ding, J., Ding, F., Liu, X. P., & Liu, G. (2011). Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data. *IEEE Transactions on Automatic Control*, *56*(11), 2677–2683. http://dx.doi.org/10.1109/TAC.2011.2158137.

Fujimoto, H., & Hori, Y. (2004). Visual servoing based on multirate control and dead-time compensation. *Journal of the Robotics Society of Japan*, *22*(6), 780–787. http://dx.doi.org/10.7210/jrsj.22.780.

Hallemans, N., Pintelon, R., Joukovsky, B., Peumans, D., & Lataire, J. (2022). FRF estimation using multiple kernel-based regularisation. *Automatica*, *136*, Article 110056. http://dx.doi.org/10.1016/j.automatica.2021.110056.

Heertjes, M., Butler, H., Dirkx, N., van der Meulen, S., Ahlawat, R., O'Brien, K., et al. (2020). Control of wafer scanners: Methods and developments. In *Proceedings of the IEEE American control conference* (pp. 3686–3703). http://dx.doi.org/10.23919/ACC45564.2020.9147464.

Lataire, J., & Chen, T. (2016). Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, *72*, 217–229. http://dx.doi.org/10.1016/j.automatica.2016.06.009.

Li, D., Shah, S. L., & Chen, T. (2001). Identification of fast-rate models from multirate data. *International Journal of Control*, *74*(7), 680–689. http://dx.doi.org/10.1080/00207170010018904.

Li, D., Shah, S. L., & Chen, T. (2002). Analysis of dual-rate inferential control systems. *Automatica*, *38*(6), 1053–1059. http://dx.doi.org/10.1016/S0005-1098(01)00295-3.

Ljung, L. (1998). *System identification: theory for the user*. Prentice Hall.

Ljung, L. (2009). Experiments with identification of continuous time models. In *Proceedings of the 15th IFAC symposium on system identification* (pp. 1175–1180). http://dx.doi.org/10.3182/20090706-3-FR-2004.00195.

Ljung, L., Chen, T., & Mu, B. (2020). A shift in paradigm for system identification. *International Journal of Control*, *93*(2), 173–180. http://dx.doi.org/10.1080/00207179.2019.1578407.

Ljung, L., & Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica*, *30*(2), 265–276. http://dx.doi.org/10.1016/0005-1098(94)90029-9.

McKelvey, T., & Guérin, G. (2012). Non-parametric frequency response estimation using a local rational model. In *Proceedings of the 16th IFAC symposium on system identification* (pp. 49–54). http://dx.doi.org/10.3182/20120711-3-BE-2027.00299.

Munnig Schmidt, R., Schitter, G., Rankers, A., & van Eijk, J. (2020). *The design of high performance mechatronics*. Delft University Press, http://dx.doi.org/10.3233/STAL9781643680514.

Oomen, T., Grassens, E., & Hendriks, F. (2015). Inferential motion control: Identification and robust control framework for positioning an unmeasurable point of interest. *IEEE Transactions on Control Systems Technology*, *23*(4), 1602–1610. http://dx.doi.org/10.1109/TCST.2014.2371830.

Oomen, T., van de Wal, M., & Bosgra, O. (2007). Design framework for high-performance optimal sampled-data control with application to a wafer stage. *International Journal of Control*, *80*(6), 919–934. http://dx.doi.org/10.1080/00207170701216329.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., & Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, *50*(3), 657–682. http://dx.doi.org/10.1016/j.automatica.2014.01.001.

Pintelon, R., & Schoukens, J. (2012). *System identification: a frequency domain approach*. John Wiley & Sons.

Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press.

Sivashankar, N., & Khargonekar, P. (1993). Robust stability and performance analysis of sampled-data systems. *IEEE Transactions on Automatic Control*, *38*(1), 58–69. http://dx.doi.org/10.1109/9.186312.

Söderström, T., & Stoica, P. (1989). *System identification*. Prentice Hall.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. John Wiley & Sons.

Unbehauen, H., & Rao, G. (1990). Continuous-time approaches to system identification—A survey. *Automatica*, *26*(1), 23–35. http://dx.doi.org/10.1016/0005-1098(90)90155-B.

van Haren, M., Mirkin, L., Blanken, L., & Oomen, T. (2023). Beyond Nyquist in frequency response function identification: Applied to slow-sampled systems. *IEEE Control Systems Letters*, *7*, 2131–2136. http://dx.doi.org/10.1109/LCSYS.2023.3284344.

Zhu, Y., Telkamp, H., Wang, J., & Fu, Q. (2009). System identification using slow and irregular output samples. *Journal of Process Control*, *19*(1), 58–67. http://dx.doi.org/10.1016/j.jprocont.2008.02.002.