High-Speed Tapping Mode AFM Utilizing Recovery of Tip-Sample Interaction

Jacques Noom, Carlas Smith, Gerard J. Verbiest, Allard J. Katan, Oleg Soloviev, and Michel Verhaegen

***Abstract*—We propose to use the State Estimation by Sum-of-Norms Regularisation (STATESON-)algorithm for recovering the tip-sample interaction in high-speed tapping mode atomic force microscopy (AFM). This approach enables accurate sample height estimation for each independent cantilever oscillation period, provided that the tip-sample interaction dominates the noise. The entire course of the cantilever deflection signal is compared to a modelled counterpart in subsequent convex minimisations, such that the sparse tip-sample interaction can be recovered. Afterwards, the sample height is determined using the minimum smoothed cantilever deflection per cantilever oscillation period. Results from simulation experiments are in favour of the proposed approach as it consistently reveals sharp edges in sample height, as opposed to both the conventional and a closely related existing approach. However, the non-processed cantilever deflection provided most accurate sample height estimation. It is recommended to implement the STATESON-algorithm in the form of a filter to use it in feedback control of the scanner and cantilever excitation.**

***Index Terms*—Atomic force microscopy, tip-sample interaction, state estimation.**

I. Introduction

THE Atomic Force Microscope (AFM), invented by Binnig and Quate in 1986 [1], is a versatile instrument which can be used for imaging and manipulation of biological samples. High-Speed tapping mode AFM (HS-AFM) facilitates videos of living biological samples at molecular scale [2], [3], [4], [5], enabling biologists to make new discoveries (e.g. [6]). A tiny cantilever with length of several micrometers oscillates above the sample, tapping the sample intermittently. Variations in cantilever oscillation amplitude enable controlling the scanner height. The sample height is determined directly from the control input. Both high scan speed and accuracy are required to produce a proper video with high frame rate. Considering that (living) biological samples are primarily imaged in liquid (e.g. [6], [7], [8]), a main limitation of conventional methods is the dependence on the Lock-In Amplifier (LIA) (or alternatively the phase-locked loop). Two drawbacks of the LIA are that (a) determination of the amplitude takes multiple cantilever oscillation periods, and (b) higher order cantilever dynamics containing important information of the tip-sample (t/s-) interaction are filtered out. This results in a shifted image with low sharpness at high scan speeds. Therefore, effort should be taken to process the available data more appropriately.

Sahoo and co-workers [9] used the entire course of the cantilever deflection signal to determine the presence of t/s-interaction at specific locations above the sample. The detection was based on the Willsky-Jones generalised likelihood ratio method [10], in which the impulsive disturbances are detected one by one in chronological order. An adaptive filtering scheme [10] provided estimates of the t/s-interaction. However, future estimates are corrupted by past (fixed) estimates and the procedure could only detect the presence of the sample, but not its height. In [11], the detection was extended for providing images by calculating the power of the innovation signal, which was presumably used as scaled sample height estimate.

More approaches for recovering the t/s-interaction in HS-AFM have been developed in [12], [13], [14], [15]. In contrast to the former two methods [12], [13], Karvinen et al. [14] employ the impulsive nature of the interaction, making the recovery less sensitive to noise. Although the time instances of the pulses are fixed in [14], the magnitudes of previously estimated pulses are more flexible than in [9], such that future estimates of the t/s-interaction are less influenced by previous estimates. Nevertheless, the assumptions in [14] involving fixed time instances of pulses is too restrictive. In [15], (semi-)periodicity of t/s-interaction is still assumed, requiring a cumbersome system augmentation involving *a priori* modelling of a chosen number of harmonics above the noise level.

In this manuscript, the complete t/s-interaction is estimated based on its sparsity by the State Estimation by Sum-of-Norms Regularisation (STATESON-)algorithm [16]. This pragmatic algorithm is a convex relaxation of the problem for recovering abrupt disturbances [16] in any linear dynamical system and therefore likely to be applicable to postprocess HS-AFM data for recovering the t/s-interaction. For this, the exclusive assumption

Manuscript received 24 April 2023; accepted 5 June 2023. Date of publication 9 June 2023; date of current version 20 June 2023. This work was supported in part by the ECSEL Joint Undertaking (JU) under Grant Agreement 826589, and in part by the European Union's Horizon 2020 Research and Innovation Programme and The Netherlands, Belgium, Germany, France, Italy, Austria, Hungary, Romania, Sweden and Israel. The review of this article was arranged by Associate Editor J. Lyding. *(Corresponding author: Jacques Noom.)*

Jacques Noom, Carlas Smith, and Michel Verhaegen are with the Delft Center for Systems and Control (DCSC), Delft University of Technology, 2628CD Delft, The Netherlands (e-mail: j.noom@tudelft.nl; c.s.smith@tudelft.nl; m.verhaegen@tudelft.nl).

Gerard J. Verbiest is with the Department of Precision and Microsystems Engineering, Delft University of Technology, 2628CD Delft, The Netherlands (e-mail: g.j.verbiest@tudelft.nl).

Allard J. Katan is with the Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Delft University of Technology, 2629HZ Delft, The Netherlands (e-mail: a.j.katan@tudelft.nl).

Oleg Soloviev is with the DCSC, Delft University of Technology, 2628CN Delft, The Netherlands, and also with the Flexible Optical BV, 2288GG Rijswijk, The Netherlands (e-mail: oleg.soloviev@gmail.com).

Digital Object Identifier 10.1109/TNANO.2023.3284654

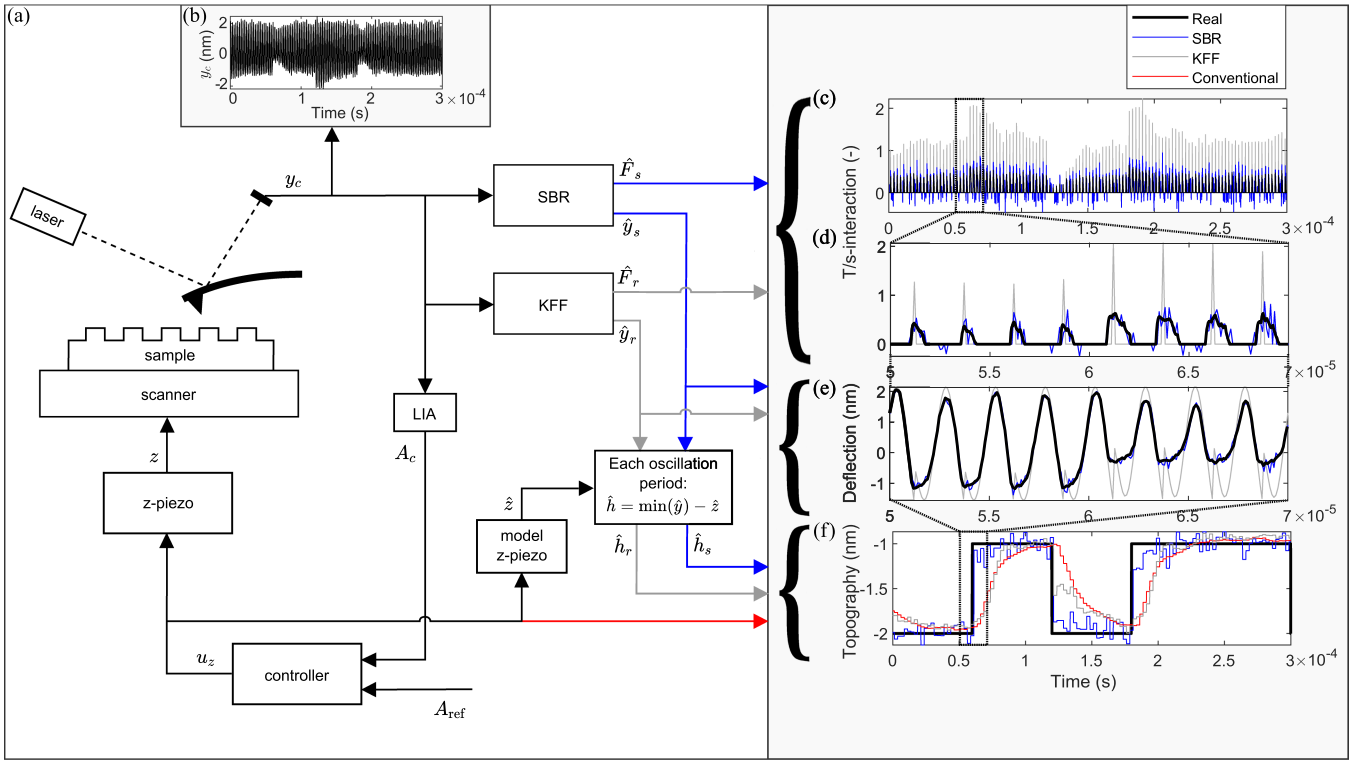

Fig. 1. (a) Illustration of AFM-operation with additional processing (the four blocks on the right-hand side). Directly measured output data y_c is captured and fed to SBR (see Section II-B) and KFF (from [14], summarised in Section II-A), which recover the tip-sample interaction and smooth (or filter) the output resulting in $\hat{F}$ and $\hat{y}$. The LIA extracts the oscillation amplitude A_c from y_c, which serves as input for the controller, which in turn determines the control input to the scanner u_z for adjusting the table height z. As the signal z cannot be measured, a model for the z-piezo is used to form an estimate $\hat{z}$, which can be used for the two new methods for estimating the sample height h with $\hat{h}$. (b)–(f) Simulation of the algorithms from measured cantilever deflection to estimates of the sample height. (b) The measured cantilever deflection. (c) The real t/s-interaction (black) with estimates using KFF (grey) and SBR (blue). (d) t/s-interactions and estimates enlarged. (e) The (smoothed) cantilever deflection corresponding to zoomed t/s-interactions. (f) The actual and estimated topographies in original time frame including the conventional estimate (red).

is made that the t/s-interaction is pulse-like. As opposed to [14], [15] where the idea was to use the estimated t/s-interaction as input for the controller, our final goal is to estimate the sample height as accurate as possible. After applying an equivalent extension for reconstructing the sample height to the method in [14], the results will be compared. In the analysis an additional approach has emerged which is also taken into account in the comparison.

The manuscript starts with presenting a state-space description [17] of the cantilever, which will be used in the evaluation of both the algorithm presented in [14] and the one presented in this article based on the STATESON-algorithm. After describing the two algorithms in Section II, the experimental design of the simulation study and corresponding results are presented in Sections III and IV, respectively. The results are discussed in Section V. Lastly, Section VI states the conclusions.

II. Methodology

The AFM is shown schematically in Fig. 1(a). The cantilever deflection is measured through a reflected laser beam on a photodetector. This signal y_c is in a conventional manner fed to a lock-in amplifier (LIA), after which its oscillation amplitude A_c is used for control u_z of the scanner height z. In this section the conversion from cantilever input u_c and output y_c to t/s-interaction $\hat{F}$ and estimated deflection $\hat{y}$ will be elaborated. This will (together with estimated table height $\hat{z}$) eventually lead to estimates of the sample height $\hat{h}$.

The cantilever dynamics can be approximated by the linear time-invariant (LTI-) system [17]

$$\begin{aligned} x(k+1) &= Ax(k) + B\bar{u}(k) + H\eta(k) \\ y_c(k) &= Cx(k) + \nu(k), \end{aligned} \tag{1}$$

with A, B, H and C the (known) system matrices; $y_c(k)$ the measured cantilever deflection at time step k; $\bar{u}(k) = u_c(k) + F(k)$ consists of the known input signal $u_c(k)$ (driving force) and the unknown t/s-interaction $F(k)$; $x(k)$ is the state; and $\eta(k)$ and $\nu(k)$ are the (approximately white) process and measurement noise with variances R_η and R_ν. With $u_c(k)$ and $y_c(k)$ measured for $k = \{1, 2, \ldots, N\}$ the first task is to recover $F(k)$ for $k = \{1, 2, \ldots, N-1\}$.

Due to sinusoidal actuation of the cantilever, the tip makes intermittently contact with the sample. This implies that $F(k)$ can be approximated with a sparse signal. The following two methods reconstruct $F(k)$ based on this assumption.

A. Kalman Filter Formulation

The method in Karvinen et al. [14] involves two assumptions in order to cast it in the so-called Kalman Filter Formulation (KFF). Firstly, the pulses occur regularly at specific time instances:

$$F(k) = \begin{cases} F(k) & \text{if } k = k_F T/T_s \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

with T the cantilever oscillation period in seconds, T_s the sampling time and $k_F \in \mathbb{N}_0$ is a non-negative integer. Secondly, the impulse response of system (1) becomes negligible after T_d seconds.

The KFF-approach is based on a new state-space representation with the state consisting of the magnitudes of the relevant past pulses:

$$\theta(k) = \begin{bmatrix} F(\ell_1(k)) & F(\ell_1(k) - \lfloor T/T_s \rceil) & \dots & F(\ell_2(k)) \end{bmatrix}^\top$$

and

$$\ell_1(k) = \left\lfloor \left\lfloor \frac{kT_s}{T} \right\rfloor \frac{T}{T_s} \right\rceil,$$
$$\ell_2(k) = \left\lfloor \left\lceil \frac{kT_s - T_d}{T} \right\rceil \frac{T}{T_s} \right\rceil.$$

with $\lfloor\bullet\rfloor$, $\lceil\bullet\rceil$, $\lfloor\bullet\rceil$ representing the floor, ceiling and nearest integer function. Note that the length of $\theta(k)$ can vary over time as T_d is not necessarily an integer multiple of T. The state-space equation becomes

$$\begin{aligned} \theta(k+1) &= \mathcal{A}(k)\theta(k) + \mu(k) \\ y_2(k) &= \mathcal{C}(k)\theta(k) + \gamma(k) \end{aligned} \tag{3}$$

with

$$\mathcal{C}(k) = \begin{bmatrix} \begin{cases} 0 & \text{if } k - \ell_1 = 0 \\ CA^{(k-\ell_1-1)}B & \text{if } k - \ell_1 > 0 \end{cases} \\ CA^{(k-\ell_1-1+T/T_s)}B \\ \vdots \\ CA^{(k-\ell_2-1)}B \end{bmatrix}^\top,$$

$$\mu(k) = \begin{cases} [\delta(k) \;\; 0 \;\; \dots \;\; 0]^\top & \text{if } k = (k_F T + \Delta)/T_s \\ [0 \;\; 0 \;\; \dots \;\; 0]^\top & \text{otherwise.} \end{cases}$$

The matrix $\mathcal{A}(k)$ is usually identity, except for the case when a new pulse appears in $\theta(k)$ (i.e. $\ell_1(k) \neq \ell_1(k-1)$) or if an old pulse disappears beyond the T_d-limit (i.e. $\ell_2(k) \neq \ell_2(k-1)$). See [14, Appendix A] for more details on updating $\mathcal{A}(k)$. The variable $\delta(k)$ is a white noise sequence with variance R_δ. On the contrary, $\gamma(k)$ is a result of both $\eta(\{1, 2, \dots, k\})$ and $\nu(k)$, from which the former introduces coloured noise according to (1). Yet, a Kalman Filter (which is designed for systems with white noise) is applied in [14] to estimate $\theta(k)$.

The Kalman Filter equations for system (3) read as:

$$\begin{aligned} K(k) = {} & P(k|k-1)\mathcal{C}(k)^\top \\ & \times \left(R_\gamma + \mathcal{C}(k)P(k|k-1)\mathcal{C}(k)^\top\right)^{-1} \end{aligned} \tag{4a}$$

$$\begin{aligned} \hat\theta(k|k) = {} & \hat\theta(k|k-1) \\ & + K(k)\left(y_2(k) - \mathcal{C}(k)\hat\theta(k|k-1)\right) \end{aligned} \tag{4b}$$

$$P(k|k) = \left(I - K(k)\mathcal{C}(k)\right)P(k|k-1) \tag{4c}$$

$$\hat\theta(k+1|k) = \mathcal{A}(k)\hat\theta(k|k) \tag{4d}$$

$$P(k+1|k) = \mathcal{A}(k)P(k|k)\mathcal{A}(k)^\top + Q(k) \tag{4e}$$

with R_γ the variance of sequence $\gamma(k)$. Furthermore, note that initial guesses $P(1|0)$ and $\hat\theta(1|0)$ are required for execution. Similarly to the matrix $\mathcal{A}$, the matrix $Q(k)$ usually consists of zeros, except for the case when a new pulse appears in $\theta(k)$, which results in the upper-left element of $Q(k)$ being equal to R_δ. The last entry of $\hat\theta(k|k)$ is expected to contain the most accurate estimate of $F(k)$ and is therefore used as final estimate $\hat F(k)$.

Choosing R_δ is nontrivial, for the definition of $\delta(k)$ in combination with updating rules of $\mathcal{A}(k)$ result in the nonzero elements in $F(k)$ being modelled as a Gaussian random walk (with mean zero), which is disputable. Besides, the choice of R_γ may be nontrivial due to the nonphysical definition of $\gamma(k)$, as opposed to $\eta(k)$ and $\nu(k)$.

The cantilever deflection can be reconstructed using

$$\begin{aligned} \hat x(k+1) &= A\hat x(k) + B\hat u(k) \\ \hat y_{\text{KFF}}(k) &= C\hat x(k) \end{aligned} \tag{5}$$

with $\hat u(k) = u_c(k) + \hat F(k)$.

B. The Proposed Approach: Sparsity-Based Reconstruction

The proposed approach entitled Sparsity-Based Reconstruction (SBR) employs the STATESON-algorithm presented in [16]. This algorithm is aimed at finding state estimates $\hat x(k)$ under additive pulse-like disturbances to the state equation. Sparsity of the t/s-interaction $F(k)$ is the only assumption. The algorithm starts with iterating the following two steps to find the time instances k at which a pulse occurred:

1) Minimise

$$\begin{aligned} &\left[\hat x_l(1), \hat F_l(k), \hat\eta_l(k)\right] = \\ &\underset{\substack{x(1), F(i), \eta(i) \\ 1 \le i \le N-1}}{\arg\min} \sum_{k=1}^{N} \left\| R_\nu^{-1/2}\left(y_c(k) - Cx(k)\right)\right\|_2^2 \\ &+ \lambda \sum_{k=1}^{N-1} \alpha_l(k) \left\| R_F^{-1/2} F(k)\right\|_1 \\ &+ \sum_{k=1}^{N-1} \left\| R_\eta^{-1/2}\eta(k)\right\|_2^2 \\ &\text{s.t. } x(k+1) = Ax(k) + Bu_c(k) + BF(k) + H\eta(k). \end{aligned} \tag{6}$$

2) Set

$$\alpha_{l+1}(k) = \left(\epsilon + \left\| R_F^{-1/2}\hat F_l(k)\right\|_1\right)^{-1} \tag{7}$$

where ϵ is a positive tuning parameter. Increase the iteration number $l = l + 1$ and return to step 1.

After convergence, a final estimate of the magnitude of the pulses can be found using

$$\begin{aligned}&[\hat{x}_{l+1}(1), \hat{F}_{l+1}(k), \hat{\eta}_{l+1}(k)] = \\ &\underset{\substack{x(1),F(i),\eta(i)\\ 1\le i\le N-1}}{\arg\min} \sum_{k=1}^{N} \left\| R_\nu^{-1/2} \left(y_c(k) - Cx(k) \right) \right\|_2^2 \\ &+ \sum_{k=1}^{N-1} \left\| R_\eta^{-1/2} \eta(k) \right\|_2^2 \\ &\text{s.t. } x(k+1) = Ax(k) + Bu_c(k) + BF(k) + H\eta(k) \\ &\quad F(k) = 0 \ \text{ if } k \notin \mathcal{T} \end{aligned} \tag{8}$$

with $\mathcal{T} = \{k | \hat{F}_l(k) \neq 0\}$.

The new parameters introduced in (6) have the following definitions. The quantity λ is a tuning parameter, $\alpha_l(k)$ is a variable weighting vector to enhance convergence and R_F is the estimated variance of $F(k|F(k) \neq 0)$. The 1-norm in (6) and (7) is used to enhance sparsity. Ohlsson et al. described a non-iterative procedure for finding an appropriate λ [16, eqs. (25)–(28)], though we found in the simulation experiments that additional tuning is useful.

Although the minimisation in (6) is convex due to the use of the 1-norm instead of the 0-norm, the disadvantage is estimates $\hat{F}_l(k)$ being biased towards zero. The concluding convex minimisation (8) corrects this bias using the time instances at which the estimates of $\hat{F}_l(k)$ in (6) of the latest iteration are nonzero[1]. Thus, the STATESON-algorithm utilizes only convex minimisations, which makes it computationally attractive.

Furthermore, optimisation problem (6) can be rewritten to a general 'lasso' form [18] by substituting $\eta(k)$ in the objective function and rearranging it to

$$\hat{\psi}_l = \underset{\psi}{\arg\min} \|\Phi^{-1}(\Gamma\psi - \varphi)\|_2^2 + \lambda \|\Xi_l \psi\|_1 \tag{9}$$

where

$$\Phi^{-1} = \begin{bmatrix} R_\nu^{-1/2} & 0 & 0 & \dots \\ 0 & R_\eta^{-1/2} & 0 & \dots \\ 0 & 0 & R_\nu^{-1/2} & \\ \vdots & \vdots & & \ddots \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} C & 0 & 0 & 0 & 0 & \dots \\ -A & -B & I & 0 & 0 & \dots \\ 0 & 0 & C & 0 & 0 & \\ 0 & 0 & -A & -B & I & \\ \vdots & \vdots & & & & \ddots \end{bmatrix},$$

$$\Xi_l = \begin{bmatrix} 0 & R_F^{-1/2}\alpha_l(1) & 0 & 0 & \dots \\ 0 & 0 & 0 & R_F^{-1/2}\alpha_l(2) & \\ \vdots & \vdots & & & \ddots \end{bmatrix},$$

$$\psi = \begin{bmatrix} x(1) \\ F(1) \\ x(2) \\ F(2) \\ \vdots \end{bmatrix}, \text{ and } \quad \varphi = \begin{bmatrix} y_c(1) \\ Bu_c(1) \\ y_c(2) \\ Bu_c(2) \\ \vdots \end{bmatrix}.$$

Optimisation problems in the 'lasso' form as in (9) can be solved efficiently, for instance using FISTA [19].

Likewise, optimisation problem (8) can be rewritten to an unconstrained least-squares problem

$$\hat{\psi}_{\text{LS}} = \underset{\psi_{\text{LS}}}{\arg\min} \|\Phi^{-1}(\Gamma_{\text{LS}}\psi_{\text{LS}} - \varphi)\|_2^2 \tag{10}$$

where Γ_{LS} and ψ_{LS} are constructed similarly to those in (9), though omitting the columns of Γ and rows of ψ corresponding to the zero estimated t/s-interactions $\hat{F}_l(k|\hat{F}_l(k) = 0)$. The solution to this problem can be found analytically [20].

The cantilever deflection can be reconstructed using

$$\hat{y}_{\text{SBR}}(k) = C\hat{x}(k). \tag{11}$$

Note that the smoothed states are directly available as outcome of the optimisation problem (10), as opposed to (5) in which the states need to be reconstructed using previous states and estimates of the t/s-interaction.

C. From t/s-Interaction to Sample Height

The idea in [14] was to use the estimated t/s-interaction as input for the controller. In this article, the goal is to estimate the sample height as accurate as possible at high scan speeds, rather than only the t/s-interaction. Moreover, the sample height is in our perspective the most appropriate input signal to the controller and the ultimate information to be obtained. Various models exist for converting the t/s-interaction to distance between tip and sample, such as the Hertzian, the Derjaguin-Müller-Toporov and the Johnson-Kendall-Roberts model [21]. Unfortunately, each model requires knowledge of several parameters of both tip and sample [21], [22], which are in addition likely to vary spatially for biological samples. Hence, those models will not be used.

To make the height reconstruction generally applicable, it is estimated for each oscillation period

$$\left(i - \frac{1}{2}\right) T \le kT_s < \left(i + \frac{1}{2}\right) T$$

using the minimum of the estimated cantilever deflection:

$$\hat{h}(i) = \min_k \left(\hat{y}(k)\right) - \hat{z}(i). \tag{12}$$

Naturally, the estimated table height $\hat{z}(i)$ is subtracted from the minimum cantilever deflection in order to arrive at the sample height. Accordingly, the estimated table height captures the coarse height profile and the minimum cantilever deflection term captures the finer variations in height.

[1] In practice, the set $\mathcal{T} = \{k|\hat{F}_l(k) > \varepsilon\}$ will be used, with ε a small number, so that very small values in $\hat{F}_l(k)$ are set to zero.

III. Experimental Design

The discussed methods will be tested in two simulation studies. A first simulation tests the methods for a one-dimensional scan with two steps of height 1 nm. The second simulation experiment comprises a scan of a 24-nm $\times$ 24-nm-sized grating consisting of discrete steps of height 1 nm. The sample is scanned with a cantilever tip radius of 1 nm, from left to right in 20 scan lines with scan speeds varying from 2 to 200 μm/s.

Signals A_c, u_z, z, $\hat{z}$ and $\hat{h}$ in Fig. 1(a) are sampled for each cantilever oscillation period $1/T = 400$ kHz, whereas the other signals are sampled with rate $1/T_s = 10$ MHz.

A piece-wise linear force-distance curve is used to simulate the tip-sample interaction. If the distance between tip and sample is smaller than 0 (*i.e.* the tip is intruding the sample), then the interaction is proportional to the depth of the tip in the sample with factor 1.6 nm^{-1}, otherwise there is no interaction.

For the cantilever immersed in liquid, a second-order model is used, identified from experimental data obtained from an SS-NEX Ando model AFM. The cantilever dynamics are identified by processing its thermal motion ($y_c(k)$ with input $u_c(k) = 0$) in a subspace identification algorithm [20], [23]. This yielded a resonance frequency of 552 kHz, quality factor 1.59 and steady-state gain 6.59 dB. The input of the cantilever $u_c(k)$ is a sinusoid with frequency $1/T = 400$ kHz such that the cantilever's free oscillation amplitude is ca. 2.15 nm. The thermal noise and measurement noise are inspected from measurement data and valued as $R_\eta = 0.0033$ and $R_\nu = 0.0151$.

The LIA is simulated using [24, eq. (2), (3), (8)]. The input $y_c(k)$ is mixed with reference signals $\cos(2\pi k T_s/T)$ and $\sin(2\pi k T_s/T)$. The resulting signals are fed to a fourth-order Butterworth low-pass filter. As the output of the LIA is sampled each oscillation period of the cantilever T, the Butterworth filter should filter out all frequencies higher than the Nyquist frequency $1/(2\,T)$ Hz and therefore, the cutoff frequency of the filter is chosen as $2/(5\,T)$ Hz.

The z-piezo is modelled as an all-pole second-order transfer function with resonance frequency $\omega_z = 40000 \times 2\pi$ and damping coefficient $\zeta_z = 0.5$. For simplicity, the steady-state gain of this model is unity, such that scaling of u_z in Fig. 1(a) is unnecessary before combining it with A_{ref}. A PI-controller regulating the table height is empirically designed with proportional and integral gains $k_p = 0.1$ and $k_i = 7.5 \times 10^4$. The setpoint oscillation amplitude of the cantilever is $A_{\text{ref}} = 1.6$ nm.

The parameters of the KFF-approach are chosen as follows. The variance $R_\delta = 30$: lower values result in a slower transient of the nonzero elements in $F(k)$ whereas higher values do not affect the solution significantly. The variance of $\gamma(k)$ is chosen to be $R_\gamma = 0.0012$ and the length of vector $\theta(k)$ is chosen as constant 5 (implying $T_d = 5T - T_s$). A smaller length of $\theta(k)$ results in a nonconverged estimate, whereas the choice of a larger vector does not affect the solution significantly. The initial conditions are chosen as $P(1|0) = I_5$ and $\hat{\theta}(1|0) = [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0]^\top$. Since T is in our case not necessarily an integer multiple of T_s, the pulse occurrences are chosen to be a fixed number of time steps before the input crosses 0 with positive derivative.

The SBR-approach is implemented with the following parameters. The variance of the t/s-interaction is $R_F = 10^{-2}$, tuning parameter $\lambda = 1$, the weighing factor $\alpha_1(k) = 1\ \forall k$ and the threshold value is $\varepsilon = 0.01$. In the simulation we restricted the algorithm to the first step (6) followed by (8) only. The overall problem is split into parts of 300 data points to prevent computational overload, and solved using FISTA [19], [25].

TABLE I
Residuals of Estimated Cantilever Deflection at Scan Speed 200 μm/s

Approximation method	Var($y_{\text{real}} - \hat{y}$)	Comp. time (s)
KFF ($\hat{y} = \hat{y}_{\text{KFF}}$)	0.0944	1.87
SBR ($\hat{y} = \hat{y}_{\text{SBR}}$)	0.0089	7.50
Raw measurements ($\hat{y} = y_c$)	0.0150	-

IV. Results

Fig. 1(b)–(f) presents the results of the first simulation experiment. From the noisy measured cantilever deflection in Fig. 1(b), the tip-sample interaction is reconstructed in Fig. 1(b) and (c). Fig. 1(c) reveals the difference between KFF and SBR. The KFF-approach is restricted to $\hat{F}(k)$ to be nonzero for one single value of k within one cantilever oscillation period. On the contrary, $\hat{F}(k)$ determined in SBR may be nonzero at more time instances. This results in Fig. 1(e) in better reconstruction of the cantilever deflection using SBR than using KFF, which is confirmed by the residuals in Table I.

Note that the residual of measured cantilever deflection Var($y_{\text{real}} - y_c$) (approximated by R_ν) has a magnitude in between that of SBR and KFF. This gives rise to the idea to apply (12) with $\hat{y}(k)$ replaced by the raw measurement data $y_c(k)$, meaning that in Fig. 1(a) the blocks "SBR" and "KFF" are bypassed to directly advance to the height estimation block representing (12). This additional approach will be evaluated in the second simulation experiment under the name Minimum of Raw Measurements (MRM).

Fig. 2 presents images resulting from the second simulation experiment. In a slow scan, all methods are able to recover the sample height accurately up to the effect of nonzero tip radius, with SBR and MRM slightly more noisy than the conventional method and KFF. For higher scan speeds, the proposed methods seem to increase only in noise, whereas the existing two methods show image artifacts due to slow response of LIA and controller. The computational time of the SBR method to evaluate all t/s-interactions in Fig. 2 with scan speed 200 μm/s (consisting of 48 000 data points), was 7.50 seconds using Matlab R2021a on an Intel i7-9750H CPU. This is slightly larger than that of KFF, but improves the performance significantly. Fig. 3 shows the sample height reconstruction performances for increasing scan speeds quantitatively.

V. Discussion

The results in Fig. 3 confirm the observation from Fig. 2 that the conventional method and KFF perform better for lower scan speeds. As expected, for higher scan speeds the SBR-method and MRM produce a lower mean-squared error (MSE). The KFF-approach has a performance curve similar to that of the conventional approach for the reason that it has large dependency on

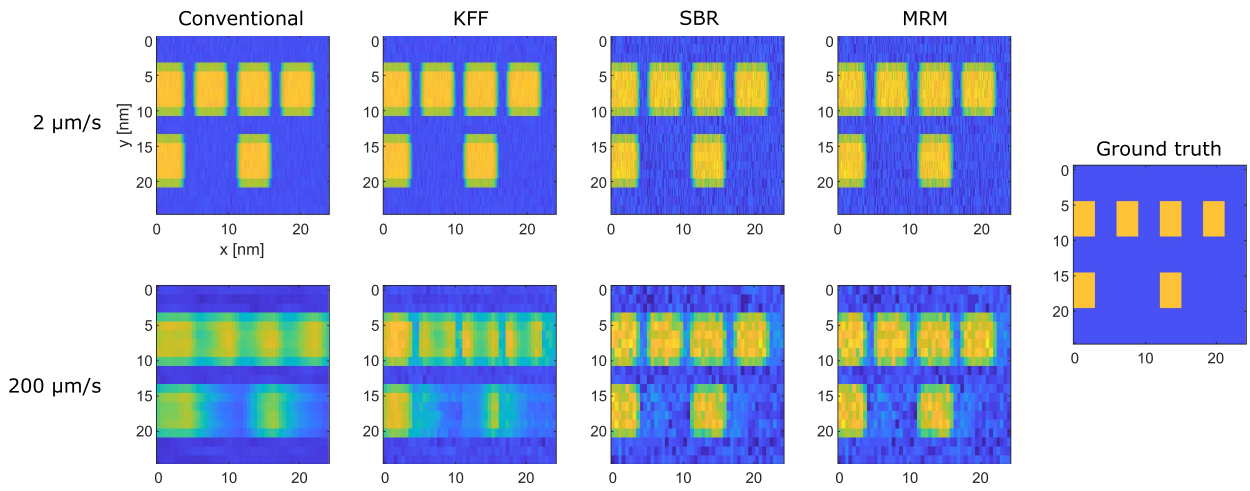

Fig. 2. Simulations of a 24 nm × 24 nm sample scan with steps in height of 1 nm, using a drive frequency of the cantilever of 400 kHz and 20 scan lines. The columns from left to right: the ground truth, the conventional approach ($-u_z$), KFF, SBR, and MRM. A slow scan of 2 μ m/s (top) and a fast scan of 200 μ m/s (bottom).

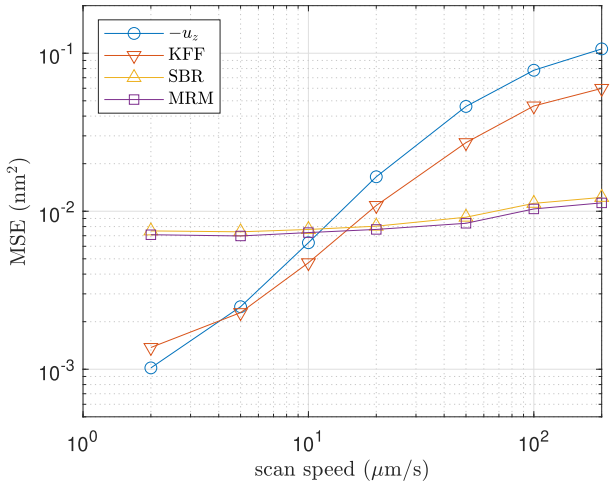

Fig. 3. Mean-Squared Errors (MSEs) of sample height reconstruction against scan speed. The conventional method (blue circle), KFF (red, down-pointing triangle), SBR (yellow, up-pointing triangle) and the MRM (purple, square).

TABLE II
SIGNIFICANCE BETWEEN SBR AND MRM BY MANN-WHITNEY-WILCOXON APPROXIMATED p-VALUES

Scan speed (μm/s)	p-value
2	0.0965
5	0.2108
10	0.4359
20	0.2437
50	0.1682
100	0.6114
200	0.2678

the scanner height, as can be seen in Fig. 1(e) and (f). It performs slightly better than the conventional approach for being able to reconstruct sharper edges in the sample, as seen in the fast scan in Fig. 2. The SBR-method and MRM also have a similar curve. Moreover, the Mann-Whitney-Wilcoxon approximated p-values between these methods in Table II (calculated over the squared errors in sample height estimation) show weak significance with $p > 0.05$, whereas other approaches differ with stronger significance with p-value < 0.05. Remarkably, MRM and SBR produce similar MSE for sample height reconstruction, whereas SBR has lower residual in cantilever deflection reconstruction as was shown in Table I. The reason for this contradiction could be that the minimum cantilever deflection typically occurs at time instances at which $\hat{F}(k)$ is nonzero. As subsequent elements in $\hat{F}(k)$ can be nonzero, it can fully compensate the measurement noise $\nu(k)$ at time instances around the minimum deflection, or even worse: it compensates the noise with errors. It is expected that the last-mentioned scenario occurred in this simulation. As the minimum estimated cantilever deflection determines the sample height, the image degrades in this scenario. This effect will be smaller for sparser t/s-interactions, which can be realised using a larger reference amplitude A_{ref} for gentler tapping or a higher quality factor of the cantilever.

VI. CONCLUSION

The proposed novel approach for sample height reconstruction for high-speed AFM demonstrates accurate recovery of tip-sample interaction and practically eliminates unwanted artefacts. Using the STATESON-algorithm, the SBR approach outperforms the most promising algorithm previously used in high-speed tapping mode AFM. Moreover, the STATESON-algorithm is a convex relaxation for recovering abrupt disturbances and not restricted to fixed time instances for identifying impulses. After recovering the tip-sample interaction, the sample height is determined using minimum smoothed cantilever deflection per cantilever oscillation period. Simulations show the ability to recover sharp edges in the sample, unlike the conventional approach.

In current simulation settings, the minimum of the raw cantilever deflection measurements provided unexpectedly a sample height reconstruction performance similar to that of SBR. Therefore we advise to use non-processed cantilever deflection for

estimating the sample height. Additionally, the SBR is pragmatic for providing insight in interactions during operation.

Future research should focus on implementing SBR as a filter, such that real-time sample height and interaction estimates form a reference for the scanner or can be used to control the cantilever excitation signal. Besides, it is recommended to find a procedure to tune the parameters involved in SBR automatically. This may for instance be done by optimising the image resolution using Fourier ring correlation [26]. The computational time may be further decreased depending on the computing platform and code optimisation, for instance by exploiting the sparsity and redundancy of the matrices involved in the optimisation problems.

REFERENCES

[1] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope," *Phys. Rev. Lett.*, vol. 56, no. 9, pp. 930–933, 1986.

[2] N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando, "Video imaging of walking myosin V by high-speed atomic force microscopy," *Nature*, vol. 468, no. 7320, pp. 72–76, 2010.

[3] T. Uchihashi, N. Kodera, and T. Ando, "Guide to video recording of structure dynamics and dynamic processes of proteins by high-speed atomic force microscopy," *Nature Protoc.*, vol. 7, no. 6, pp. 1193–1206, 2012.

[4] T. Ando, "High-speed atomic force microscopy coming of age," *Nanotechnology*, vol. 23, no. 6, 2012, Art. no. 062001.

[5] Y. F. Dufrêne et al., "Imaging modes of atomic force microscopy for application in molecular and cell biology," *Nature Nanotechnol.*, vol. 12, no. 4, pp. 295–307, 2017.

[6] J. M. Eeftens et al., "Condensin Smc2-Smc4 dimers are flexible and dynamic," *Cell Rep.*, vol. 14, no. 8, pp. 1813–1818, 2016.

[7] F. Jiao, K. S. Cannon, Y.-C. Lin, A. S. Gladfelter, and S. Scheuring, "The hierarchical assembly of septins revealed by high-speed AFM," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 5062.

[8] G. R. Heath et al., "Localization atomic force microscopy," *Nature*, vol. 594, no. 7863, pp. 385–390, 2021.

[9] D. R. Sahoo, A. Sebastian, and M. V. Salapaka, "Transient-signal-based sample-detection in atomic force microscopy," *Appl. Phys. Lett.*, vol. 83, no. 26, pp. 5521–5523, 2003.

[10] A. Willsky and H. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Trans. Autom. Control*, vol. 21, no. 1, pp. 108–112, Feb. 1976.

[11] D. R. Sahoo, P. Agarwal, and M. V. Salapaka, "Transient force atomic force microscopy: A new nano-interrogation method," in *Proc. IEEE Amer. Control Conf.*, 2007, pp. 2135–2140.

[12] G. Mohan, C. Lee, and S. Salapaka, "Control techniques for high-speed dynamic mode imaging in atomic force microscopes," in *Proc. IEEE 50th Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 651–656.

[13] P. Huang and S. B. Andersson, "High speed atomic force microscopy enabled by a sample profile estimator," *Appl. Phys. Lett.*, vol. 102, no. 21, 2013, Art. no. 213118.

[14] K. S. Karvinen, M. G. Ruppert, K. Mahata, and S. O. R. Moheimani, "Direct tip-sample force estimation for high-speed dynamic mode atomic force microscopy," *IEEE Trans. Nanotechnol.*, vol. 13, no. 6, pp. 1257–1265, Nov. 2014.

[15] A. Keyvani, G. van der Veen, M. S. Tamer, H. Sadeghian, H. Goosen, and F. van Keulen, "Real-time estimation of the tip-sample interactions in tapping mode atomic force microscopy with a regularized Kalman filter," *IEEE Trans. Nanotechnol.*, vol. 19, pp. 274–283, 2020.

[16] H. Ohlsson, F. Gustafsson, L. Ljung, and S. Boyd, "Smoothed state estimates under abrupt changes using sum-of-norms regularization," *Automatica*, vol. 48, no. 4, pp. 595–605, 2012.

[17] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems: Theory and Design*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 1997.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc.: Ser. B. (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.

[19] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[20] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[21] M. Kopycinska-Müller, R. H. Geiss, and D. C. Hurley, "Contact mechanics and tip shape in AFM-based nanomechanical measurements," *Ultramicroscopy*, vol. 106, no. 6, pp. 466–474, 2006.

[22] R. García and A. San Paulo, "Attractive and repulsive tip-sample interaction regimes in tapping-mode atomic force microscopy," *Phys. Rev. B*, vol. 60, no. 7, pp. 4961–4967, 1999.

[23] J. Noom, "Detection of tip-sample interaction in atomic force microscopy: Improving the image resolution," M.S. thesis, Delft Univ. Technol., Delft, The Netherlands, 2019.

[24] S. DeVore, A. Gauthier, J. Levy, and C. Singh, "Development and evaluation of a tutorial to improve students' understanding of a lock-in amplifier," *Phys. Rev. Phys. Educ. Res.*, vol. 12, no. 2, 2016, Art. no. 020127.

[25] T. Vu and P. State, "FISTA," 2016. [Online]. Available: https://github.com/tiepvupsu/FISTA

[26] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *J. Struct. Biol.*, vol. 183, no. 3, pp. 363–367, 2013.