Adaptive Time-Resolved Mass Spectrometry with Nanomechanical Resonant Sensors

Alper Demir

Koç University
Istanbul, Turkey
aldemir@ku.edu.tr

Abstract—Nanomechanical resonant sensors that are based on detecting and tracking the resonance frequency deviations due to events of interest are being advocated for a variety of applications. All sensor schemes currently in use are subject to a basic trade-off between accuracy and speed, while there is great interest in improving both in order to enable unprecedented and widespread applications. Based on a thorough understanding of the characteristics of current resonant sensor architectures, we propose adaptive and flexible sensor schemes. Unlike recently proposed time-resolved mechanical detection methods, the proposed schemes do not require ensemble averaging of the resonator response for many independent identical stimuli. Distinct one-time events can be detected in real-time with high time resolution with an accuracy that then improves considerably with elapsed time. While the proposed adaptive schemes also need to abide by the fundamental speed versus accuracy trade-off, we show that there is still “some room at the bottom” for improvement with sensor architecture innovations. Pareto optimal performance that reaches a bound that is imposed by the fundamental thermomechanical noise can be achieved.

Index Terms—nanomechanical resonant sensor, thermomechanical noise, detection noise, time-resolved mass spectrometry, real-time mass spectrometry, Kalman filter, adaptive filtering.

I. INTRODUCTION

NANOMECHANICAL resonant sensors offer unprecedented sensitivities and are used in a variety of applications, such as mass spectrometry [1]–[5] and atomic force microscopy (AFM) [6]–[10]. In mass spectrometry, the main sensing mechanism is based on tracking the resonance frequency deviations arising from the interactions of the nanomechanical resonator with the sample under study. Currently, three schemes are used for this purpose, namely the open-loop, feedback free (FF) approach, and closed-loop schemes based on frequency-locked loops (FLL) and self-sustaining oscillators (SSO). These resonator tracking architectures were analyzed and the accuracy versus speed trade-off characteristics were discussed in detail in [11], [12].

The speed-accuracy trade-off in the resonator tracking schemes arises from the characteristics of inherent noise sources affecting resonant sensors, i.e., thermomechanical and detection noise. In order to increase the accuracy (sensitivity) of the sensor, one has to simply reduce the effective bandwidth of the system for better noise filtering. However, bandwidth reduction results in a larger response time. All of the standard resonator tracking schemes mentioned above have to endure this basic trade-off [12]. On the other hand, noise reduction may also be achieved via ensemble averaging, as opposed to temporal filtering. In fact, recently proposed time-resolved mechanical motion detection techniques in atomic force microscopy [13]–[16] seem to be mostly based on some form of ensemble averaging of the resonator response for many identical stimuli. However, these techniques may be used only in cases where the events of interest have the form of multiple independent identical stimuli. They are not suitable for time-resolved detection of distinct one-time events that are typical in mass spectrometry applications.

The rigorous and thorough understanding that was developed in [11], [12] for the fundamentals and trade-offs of standard resonator tracking schemes paves the way to the conception of adaptively accurate resonator tracking schemes. An adaptive scheme, in principle, would provide very fast but somewhat coarse initial estimates of frequency shifts, that are then improved continually with elapsed time after an event of interest occurs. While such an adaptive scheme also needs to abide by the basic trade-off between noise filtering and speed, we show that there is still “some room at the bottom” [17] for improvement, unraveled by the detailed analysis we present in this paper. The performance boost here is due to sensor architecture innovations alone, and is in addition to the improvement one would get by using higher quality resonators and better electronic or optical hardware. We propose an adaptively accurate, time-resolved resonator tracking scheme that can detect distinct, one-time events either in real-time operation or offline processing. We show that this Pareto optimal (with respect to speed and accuracy) scheme achieves a performance bound that is imposed by the unavoidable,
fundamental thermomechanical noise of the resonator.

We first discuss adaptively accurate resonator tracking schemes in Section II with a simple and intuitive perspective. We then propose a scheme based on adaptive Kalman filtering in Section III and rigorously compare it with the three standard schemes in Section IV. Finally, we discuss practical implementation issues in Section V.

II. Resonator tracking with adaptive speed versus accuracy characteristics

The flexible speed versus accuracy characteristics of the FLL scheme discussed in [12] points to the possibility of constructing adaptively accurate resonator tracking schemes. Such an adaptive scheme would provide an ultra fast response immediately after an event of interest (causing a resonance frequency shift) with an estimate of the new value of the resonance frequency. It would then continuously update this estimate, increasing its accuracy as time progresses following the event of interest. In principle, such a scheme could be implemented with an FLL by adjusting the loop bandwidth (via the controller parameters) in an adaptive and automatic manner. This would involve real-time monitoring of the feedback error signal in the FLL, and increasing the loop bandwidth dramatically if the error exceeds a certain threshold, indicating that an event has occurred. The loop would then respond very rapidly and lock to the shifted resonance. Subsequently, the loop bandwidth needs to be gradually reduced for better noise filtering and hence increased accuracy of tracking. There are some challenges in implementing such a scheme though, such as the conception of an effective and robust technique for detecting events based on monitoring the error signal and ensuring the stability of the feedback loop with time-varying controller parameters. An engineering solution for these challenges may be found, though with some effort and increased system complexity.

In the simple FF scheme, shown in Fig. 1, the resonator is driven with a sinusoidal signal at a fixed frequency close to the resonance [12]. The resonator response at the same frequency is demodulated (down-converted to baseband using the drive signal) in order to compute the phase shift with respect to the drive. This phase shift can then be used to infer the difference between the resonance frequency and the drive frequency based on the resonator characteristics. The FF scheme does not offer a flexible speed versus accuracy characteristics as in the FLL, and is limited by the inherent resonator characteristics [12]. Thus, one may initially think that the FF scheme is not amenable to adaptation as described above. In fact, by examining the very simple model that was developed in [12] for the FF scheme, shown in Fig. 2, one can conceive a detection scheme as follows. We assume that the response speed of the FF scheme is limited by the resonator characteristics. That is, with a high quality factor resonator, the linewidth of the resonator is much smaller than the bandwidth of the low-pass filter $H_r(s)$ in the demodulator [18]. Then, the response of the FF scheme to an event that causes a sudden shift in the resonance frequency is simply determined by a very simple one-pole low-pass filter

$$H_r(s) = \frac{1}{1 + s\tau_r}, \quad (1)$$

with a bandwidth equal to the resonator linewidth, where $\tau_r$ is the inherent resonator time constant [12]. Thus, the step response of the FF scheme to an event that causes a sudden shift in the resonance frequency is simply given by

$$\Delta \omega(t) = \Delta \omega_e \left(1 - e^{-t/\tau_r}\right). \quad (2)$$

Straightforward determination of the unknown $\Delta \omega_e$ based on such a transient response requires at least $t \approx \tau_r$ for the exponential factor to decay, hence it is quite slow. However, prior knowledge of the form of the response in (2) and $\tau_r$, based on the simple but accurate model in Fig. 2, enables one to determine $\Delta \omega_e$ immediately after event occurrence. One would simply measure the response $\Delta \omega(t_m)$ at any $t = t_m > 0$ and evaluate

$$\Delta \omega_e = \frac{\Delta \omega(t_m)}{1 - e^{-t_m/\tau_r}}. \quad (3)$$

Fig. 1. Feedback-free (FF) resonator tracking scheme. LPF is a low-pass filter with transfer function $H_r(s)$. (Reproduced from [12, Fig. 4], Demir, Journal of Applied Physics 129, 044503 (2021), with the permission of AIP Publishing.)

Fig. 2. Phase-domain model for FF resonator tracking scheme. $\Delta \omega_e$ represents a sudden resonance frequency shift (root cause) due to an event of interest. The frequency shift measured by the FF resonator scheme is denoted by $\Delta \omega(t)$. $H_r(s)$ models the inherent resonator dynamics. $H_L(s)$ represents the phase noise in the resonator. $\tau_e$ is the resonator response time constant. (Reproduced from [12, Fig. 7], Demir, Journal of Applied Physics 129, 044503 (2021), with the permission of AIP Publishing.)
The fast free time-resolved electrostatic force microscopy (FFtrEFM) scheme [14] aiming to achieve better time resolution in AFM is in principle similar to the one described above.

The simplistic scheme represented by (3) disregards noise and assumes that the event time \( t = 0 \), in this case) is known. Thus, there are two problems that need to be addressed. First, a scheme which can detect events very soon after they occur needs to be devised. Second, the naive measurement and evaluation in (3) need to be replaced with an optimal estimation technique that takes both thermomechanical and detection noise into account. The second problem can be conveniently and properly addressed with the well established Kalman Filter [19], a least-squares based estimation technique for linear dynamical systems. The first problem can be addressed with the maximum-likelihood based technique described in [20], that augments a standard tracking-mode Kalman filter with a jump detection and estimation technique, turning it essentially into an adaptive filter. This adaptive Kalman filter is in principle similar to the adaptive FLL scheme we described above, in the sense that the system features an automatically adjusted bandwidth for an initial rapid response and a subsequent increase in accuracy with time. The adaptive Kalman filter is founded on rigorous and well established least-squares and maximum-likelihood based estimation formulations, whereas the adaptive FLL represents just a practical, engineering solution.

Kalman filtering, with respect to both effectiveness and complexity considerations, critically relies on the availability of an accurate and simple dynamical model of the system, as well as prior stochastic characterizations of the noise sources [19]. The model in Fig. 2 for the FF scheme, that was developed in [12], and the associated resonator and detection noise models are very accurate (especially for high \( Q \) resonators), and as simple as it gets. Thus, we are well poised to develop a robust, effective and simple resonator tracking technique based on an adaptive Kalman filter. Kalman filtering based techniques have been previously considered in the AFM literature [21]–[27]. However, they were based on more complicated dynamical system models as compared with the one in Fig. 2, resulting in an implementation with higher complexity. Furthermore, a rigorous analysis of the Kalman filtering based schemes with respect to speed versus accuracy characteristics, and performance comparisons with the traditional resonance frequency tracking architectures for mass spectrometry, seem to be lacking in the literature.

III. ADAPTIVE KALMAN FILTERING BASED TRACKING OF RESONATORS

A. Dynamical system model

The adaptive Kalman filtering (AKF) based resonator tracking scheme we propose is founded on the phase domain model in Fig. 2, which can be summarized as follows (assuming \( H_{th}(s) \) representing the inherent resonator response has a bandwidth that is much smaller than the one for \( H_{in}(s) \) representing the demodulator filter)

\[
\tau_r \frac{d}{dt} \Delta \theta(t) + \Delta \theta(t) = \tau_r \Delta \omega + \theta_{th}(t), \tag{4}
\]

\[
\Delta \theta_{th}(t) = \Delta \theta(t) + \theta_{th}(t),
\]

where \( \Delta \theta(t) \) is the inherent (internal) phase deviation of the resonator that is influenced by thermomechanical noise \( \theta_{th} \), and \( \Delta \theta_{th}(t) \) is the observed phase deviation that is corrupted by also detection noise \( \theta_d \). Noting \( \Delta \omega = \frac{\Delta \omega_{in}}{\tau_r} \) and \( \theta_{th} = \frac{\Delta \theta_{th}}{\tau_r} \), thus \( y = \frac{\Delta \theta_{th}}{\tau_r \Delta \omega} \), we divide both sides of (4) with \( \tau_r \Delta \omega \) to obtain

\[
\tau_r \frac{d}{dt} y(t) + y(t) = \theta_{th}(t) + \theta_{th}(t), \tag{5}
\]

\[
y_{th}(t) = \frac{\theta_{th}(t)}{\tau_r \Delta \omega}, \quad y_{th}(t) = \frac{\theta_{th}(t)}{\tau_r \Delta \omega}, \tag{6}
\]

where all \( y \) are fractional frequency quantities, \( y_{th} \) is the root cause due to an event of interest, \( y_{th} \) is the response of the resonator, the observable \( y_{th} \) includes also detection noise, and

\[
S_{dy}(\omega) = \frac{k_b T}{m Q \omega^3 A_{in}^2}, \quad S_{dy}(\omega) = \omega^2 S_{dy}(\omega), \tag{7}
\]

where \( k_b \) is Boltzmann’s constant, \( T \) is temperature, \( m \) is the mass of the resonator, \( \omega \) is the resonance frequency, \( Q \) is the quality factor, and \( A_{in} \) is the vibrational amplitude of the resonator motion. The dimensionless factor \( K_d \) quantifies detection noise background with respect to thermomechanical noise, typically satisfying \( 0 < K_d < 1 \) when thermomechanical noise is resolved [12].

We represent (5) in state-space form and append a new state equation that captures the dynamics of \( y_{th} \):

\[
\frac{d}{dt} y_{th}(t) = u_{th}(t),
\]

\[
\frac{d}{dt} y_{th}(t) = u_{th}(t) + \frac{y_{th}(t)}{\tau_r}, \tag{8}
\]

\[
y_{th}(t) = y_{th}(t) + y_{th}(t),
\]

where \( u_{th}(t) \) characterizes the event. For instance, \( u_{th}(t) = \Delta y \delta(t - t_{eo}) \), with \( \delta(t - t_{eo}) \) as a time-shifted Dirac delta function, represents an instantaneous resonance frequency jump at \( t = t_{eo} \) creating a step in \( y_{th}(t) \) and subsequently invoking a step response in \( y_{th}(t) \) and \( y_{th}(t) \). Thus,

\[
\frac{d}{dt} \begin{bmatrix} y_{th}(t) \\ y_{th}(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \frac{1}{\tau_r} & -\frac{1}{\tau_r} \end{bmatrix} \begin{bmatrix} y_{th}(t) \\ y_{th}(t) \end{bmatrix} + \begin{bmatrix} u_{th}(t) \\ 0 \end{bmatrix},
\]

\[
y_{th}(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} y_{th}(t) \\ y_{th}(t) \end{bmatrix} + y_{th}(t).
\]

The implementation of the proposed scheme is best accomplished in the sampled-data domain. In order to lead to a digital implementation, we discretize time in (9) with a time-step of \( h \) to obtain

\[
\begin{bmatrix} y_{th}[k + 1] \\ y_{th}[k + 1] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{\tau_r} & 1 - \frac{1}{\tau_r} \end{bmatrix} \begin{bmatrix} y_{th}[k] \\ y_{th}[k] \end{bmatrix} + \begin{bmatrix} u_{th}[k] \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_{th}[k] \end{bmatrix},
\]

\[
y_{th}[k] = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} y_{th}[k] \\ y_{th}[k] \end{bmatrix} + y_{th}[k], \tag{10}
\]
with

\[ y_n[k] \sim N(0, h S_{ym}), \]
\[ y_o[k] \sim N(0, BW_i S_{yo}), \]

(11)

where \( N(m, v) \) represents a Gaussian random variable with mean \( m \) and variance \( v \). \( u_o[k] = \Delta y \delta[k - k_o] \), with \( \delta[k - k_o] \) as a time-shifted Kronecker delta function, represents an instantaneous resonance frequency jump at \( k = k_o \). \( BW_i \) is the detection bandwidth (noise equivalent bandwidth\(^1\)), two-sided, expressed in Hz. \( H_c(s) \) in the IQ demodulator, typically set to be wider than the linewidth of the resonator, but limited by the resonance frequency \( \omega_r \). \( y_n[k] \) and \( y_o[k] \) are assumed to be uncorrelated with each other, and also in time, i.e., they are both discrete-time white noise processes. We define

\[ Q = \begin{bmatrix} 0 & 0 \\ 0 & h S_{ym} \end{bmatrix}, \quad R = [BW_i S_{yo}] \]

(12)

as covariance matrices. Kalman filter relies on the simple, discrete-time model of the resonator in (10) in order to make much faster (and optimal) predictions than otherwise allowed by the slow resonator response.

B. Tracking mode Kalman filter

The recursive prediction and update equations for the tracking mode Kalman filter [19] are as follows:

Predict

\[ \hat{y}[k|k-1] = F \hat{y}[k-1|k-1], \quad C[k|k-1] = FC[k-1|k-1]F^T + Q, \]

(13)

Observe

\[ \tilde{y}[k] = y_o[k] - H \hat{y}[k|k-1], \quad V[k] = H C[k|k-1] H^T + R, \]

(14)

Kalman gain

\[ K[k] = C[k|k-1] H^T V^{-1}, \]

(15)

Estimate

\[ \hat{y}[k|k] = \hat{y}[k|k-1] + K[k] \tilde{y}[k], \]

(16)

Estimate covariance

\[ C[k|k] = (I - K[k] H) C[k|k-1], \]

(17)

where

\[ y = \begin{bmatrix} y_n \\ y_o \end{bmatrix} \]

(18)

is the system state vector, \( \tilde{y} \) is the difference between the actual observation and the model based prediction for the observable resonator output \( y_o \), the \( [n|m] \) notation represents the estimate of a quantity \( \bullet \) at time \( n \) given the observations for up to time \( m \leq n \). \( K \) is the Kalman gain, and \( C \) is the \( 2 \times 2 \) covariance matrix for the state estimate. The first diagonal entry of \( C \) quantifies the accuracy of the estimate for the root cause \( y_o \) for the resonance frequency.

\(^1\)Defined as the cut-off frequency of an ideal low-pass filter which passes the same amount of noise power as the actual filter, equal to \( 2 \omega_{3dB} \) for a first-order filter where \( \omega_{3dB} \) is its (one-sided) 3-dB bandwidth. Hence, \( BW_i = \frac{\omega_{3dB}}{2 \pi} \), \( \omega_{3dB} = \frac{\omega_{BW}}{2} \), since \( BW_i \) is two-sided and expressed in Hz.

In tracking mode, \( y_o \) is time-invariant assuming no events occur, with \( u_o = 0 \). In this case, the Kalman filter recursively and continually updates its estimate for \( y_o \). The recursions for \( \hat{y} \) and the covariance matrix \( C \) converge to a fixed point. The steady-state error variance for the estimate \( \hat{y} \) at this fixed point is in fact zero, meaning that the estimate is perfect. However, this perfect estimate is, in principle, achieved only at convergence, i.e., given infinite time. Intuitively, one can steady-state Kalman filter (producing a perfect estimate) as a system with zero noise bandwidth, i.e., all of the noise is being filtered but at the same time the system has lost its ability to respond to any changes in the estimated quantity. From a practical point of view, the accuracy of the estimates produced by the Kalman filter needs to be assessed in a transient manner immediately following event occurrence. Before we can perform this characterization, we first need to introduce the event detection and estimation technique [20] that will augment the Kalman filter.

C. Event detection and Kalman filter adaptation

A sudden resonance frequency shift of an unknown size \( \Delta y \) at an unknown time \( k_o \), due to an event of interest, will manifest itself with \( u_o[k] = \Delta y \delta[k - k_o] \) in (10). The event detection technique proposed in [20] that we adopt is founded on a rigorous maximum-likelihood based estimation formulation, leading to an efficient implementation based on recursive prediction and update equations as in the tracking Kalman filter. We provide here an outline, and a somewhat imprecise but intuitive description for the technique, and refer the reader to [20] for the details.

The difference between the actual observation and the model based prediction for the observable resonator output \( y_o \) in (15), known as the residual or the innovation [19] in a Kalman filter, is continually monitored via recursive equations. At every point in a prescribed sliding time window, the residuals are projected onto sudden resonance frequency shift event signatures with sliding occurrence times within the window. The event time is determined based on a maximum-likelihood rule when the residual projections exceed a certain threshold. If an event is detected as such at a certain estimated time instant \( k_o \), a corresponding maximum-likelihood based estimate for the size \( \Delta y \) of the frequency shift is also computed. The accuracy of this estimate is quantified via the covariance (matrix) as in the tracking mode Kalman filter. Subsequent to event detection, the Kalman filter in tracking mode is reset by updating both the state estimate \( \hat{y} \) and the corresponding estimate covariance matrix \( C \). The frequency shift estimate \( \Delta \hat{y} \) at event detection is associated with a larger variance, since there is very little effective noise filtering. As a result, the Kalman gain \( K \) in (17) increases dramatically after the reset following event detection. This essentially corresponds to widening the bandwidth of the system, leading to a much faster response but also much less noise filtering. The Kalman filter then recursively and continually updates the estimate \( \Delta \hat{y} \), with the variance of this estimate and hence the Kalman gain continually decreasing in the meantime. Thus, the effective bandwidth of the Kalman filter is continually reduced after...
the reset, eventually ending up at the steady-state situation described before where the estimate becomes perfect with zero variance. In a practical setting, this detection scheme is to be used when multiple events occur with some time gap between them. In this case, the initial variance (quantifying accuracy) of the frequency shift estimate right after detection and how fast it decreases with time is the key.

The recursive equations for the detection of frequency shift events, and the maximum-likelihood estimators for the event time \( \hat{k}_o \) and the size \( \Delta \hat{y} \) of the frequency shift are as follows [20]:

\[
\text{for } k - M < m < k:
\begin{align*}
G[k; m] &= H \left[ \mathbf{F}^{k-m} - F \mathbf{L}[k-1; m] \right], \\
CI[k; m] &= G^T[k; m] V^{-1} G[k; m] + CI[k-1; m], \\
d[k; m] &= G^T[k; m] V^{-1} \tilde{y}[k] + d[k-1; m], \\
L[k; m] &= K[k] G[k; m] + F \mathbf{L}[k-1; m], \\
a[k; m] &= [1 \ 0] CI[k; m] [1 \ 0]^T \\
b[k; m] &= [1 \ 0] d[k; m] \\
I[k; m] &= \frac{b^2[k; m]}{a[k; m]},
\end{align*}
\]

(21)

at \( k \) :

\[
\begin{align*}
CI[k; k] &= H^T V^{-1} H, \\
d[k; k] &= H^T V^{-1} \tilde{y}[k], \\
L[k; k] &= K[k] H
\end{align*}
\]

The quantities above are recursively computed in a window of length \( M \) as above, where \( k \) is the current time and \( m \) indicates a time point in the window starting at \( k \) and going back to \( k - M \). When \( I[k; m] > \text{threshold} \), an event is detected and the estimates for the event time \( k_o \) and the size \( \Delta \hat{y} \) of the frequency shift are computed as follows. \( k_o \) is the value of \( m \) that maximizes \( I[k; m] \), and

\[
\Delta \hat{y}[k] = \frac{b[k; \hat{k}_o]}{a[k; \hat{k}_o]}.
\]

(22)

Here, \( k - \hat{k}_o \) is the detection delay time [20], and it can be controlled by adjusting the \text{threshold}, with lower values corresponding to shorter delays. However, making \text{threshold} too small may result in false alarms. Once an event is detected as such, the Kalman filter is reset by updating both the state estimate \( \tilde{y} \) and the corresponding estimate covariance matrix \( C \) as follows

\[
\begin{align*}
\tilde{y}[k|k]_{\text{new}} &= \tilde{y}[k|k]_{\text{old}} + \left\{ \mathbf{F}^{k-k_o} - \mathbf{L}[k; \hat{k}_o] \right\} \Delta \hat{y}[k] [1 \ 0]^T, \\
C[k|k]_{\text{new}} &= C[k|k]_{\text{old}} + \left\{ \mathbf{F}^{k-k_o} - \mathbf{L}[k; \hat{k}_o] \right\} C^{-1}[k; \hat{k}_o] + \left\{ \mathbf{F}^{k-k_o} - \mathbf{L}[k; \hat{k}_o] \right\}.
\end{align*}
\]

(23)

The variance of the initial estimate for \( \Delta \hat{y} \) is given by [20]:

\[
\text{var}(\Delta \hat{y}) = [1 \ 0] C^{-1}[k; \hat{k}_o] [1 \ 0]^T.
\]

(24)

D. Event frequency shift estimation accuracy

We next consider the transient behavior of the frequency shift estimate accuracy as a function of elapsed time following event detection. The analytical quantification of the estimate accuracy is rather challenging, due to the recursive nature of the predict and update equations of the Kalman filter as well as the equations in (21) for event detection and estimation [20]. However, the difficulty of the problem at hand is alleviated due to the very simple nature of the dynamical system modeled in (10) that we use. Still, derivation of a precise expression for the transient behavior of the estimate variance is not tractable. However, by employing reasonable and justifiable approximations, we were able to derive an approximate analytical expression that matches both the numerically computed results based on an evaluation of the recursive equations above, as well as the ones obtained from extensive Monte Carlo simulations. An analytical quantification of the accuracy versus speed characteristics of the Kalman filtering based technique is needed for one-to-one comparisons with the traditional resonator tracking schemes. By performing considerable and tedious algebra with the aid of a symbolic computer algebra package, we were able to derive the following expression for the variance of the frequency shift estimate \( \Delta \hat{y} \)

\[
\text{var}(\Delta \hat{y}) = \frac{Z + \sqrt{S_{\text{th}} t_e Z}}{2 t_e^2} (25)
\]

with

\[
Z = S_{\text{th}} t_e + 4 \mathbf{BW}_s S_{\text{th}} \tau_e^2 (26)
\]

where \( t_e \) is the elapsed time since the event. The small–\( t_e \) asymptote for this expression is given by

\[
\text{var}(\Delta \hat{y}) = 2 \mathbf{BW}_s S_{\text{th}} \left( \tau_e \right)^2 (27)
\]

and for the large–\( t_e \) asymptote, we have

\[
\text{var}(\Delta \hat{y}) = \frac{S_{\text{th}}}{\tau_e} (28)
\]

The two asymptotes intersect at the point

\[
t_e^* = 2 \mathbf{BW}_s \tau_e^2 S_{\text{th}} \frac{S_{\text{th}}}{S_{\text{th}}^2} = 2 \mathbf{BW}_s \tau_e^2 K_s^2 (29)
\]

For small values of \( t_e \), detection noise limits the accuracy of the estimate, since the bandwidth of the system is initially large. As \( t_e \) becomes larger, the effective bandwidth of the system is reduced continually, filtering more and more of detection noise. The estimate accuracy eventually becomes limited by the thermomechanical noise of the resonator.

IV. COMPARISON OF SPEED VERSUS ACCURACY CHARACTERISTICS

We quantify the quality of the estimate \( \Delta \hat{y} \), for a resonance frequency shift of \( \Delta y \), provided by various resonator tracking schemes in terms of their Mean Squared Error (MSE) (or with root mean squared error RMSE as its square root), defined by

\[
\text{MSE} = \mathbf{E} [ (\Delta \hat{y} - \Delta y)^2 ] = \text{var}(\Delta \hat{y} - \Delta y) + (\mathbf{E}[\Delta \hat{y} - \Delta y])^2
\]

\[
= \text{var}(\Delta \hat{y}) + \text{bias}^2 (30)
\]
where \( E[\cdot] \) denotes the probabilistic expectation operator. Thus, \( \text{MSE} \) is the sum of two terms, the variance and the square of the bias of the estimator. In general, both the variance and the bias are functions of the elapsed time \( t_e \) after an event.

All three of the standard FF, FLL and SSO schemes provide biased estimates for \( \Delta y \) [12]. The biases of these schemes can be directly determined using the frequency step response (FSTR) as follows

\[
\text{bias}(t_e) = \Delta y \left[ 1 - f_w(t_e) \right] \quad (31)
\]

where \( f_w(t_e) \) is normalized so that \( f_w(0) = 0 \) and \( f_w(t_e \to \infty) = 1 \) [12]. Thus, the biases of all three schemes are equal to the unknown frequency shift \( \Delta y \) immediately following the event. The estimators become (asymptotically) unbiased as \( t_e \to \infty \). In order to arrive at simple analytical expressions, we will approximate the FSTR of the three schemes as follows

\[
f_w^{\text{FF}}(t_e) = 1 - e^{-t_e/\tau_{\text{FF}}} \quad (32)
\]

\[
f_w^{\text{FLL}}(t_e) = 1 - e^{-t_e/\tau_{\text{FLL}}} \quad (32)
\]

\[
f_w^{\text{SSO}}(t_e) = 1 - e^{-t_e/\tau_{\text{SSO}}} \quad (32)
\]

where

\[
\tau_{\text{FLL}} = \frac{1}{\omega_{3dB}} = \frac{1}{2BW_{\text{FLL}}} = \frac{1}{2BW_{\text{SSO}}} \quad (33)
\]

is the time-constant for the effective low-pass filter (assumed first-order for the above) that models the detection/demodulator bandwidth for the FLL and SSO schemes, and \( BW_{\text{FLL}} = BW_{\text{SSO}} \) is its two-sided noise equivalent bandwidth expressed in Hz. For the FLL scheme, we assume that the loop bandwidth is chosen to be the same as the bandwidth of the filter in the demodulator [12].

The variances of the estimators for the FF, FLL and SSO schemes can be computed by simply integrating the spectral density of fractional frequency fluctuations, \( S_{\nu_0}(\omega) \), over the system bandwidth. We do this computation by multiplying \( S_{\nu_0}(\omega = 0) \) in [12, eqn. 72] with the corresponding noise equivalent bandwidth as follows

\[
\text{var}^{\text{FF}}(\Delta \hat{y}) = \frac{\omega_r}{4Q} \quad (34)
\]

\[
\text{var}^{\text{FLL}}(\Delta \hat{y}) = \frac{\omega_r}{4Q} (1 + K_d^2) \quad (34)
\]

\[
\text{var}^{\text{SSO}}(\Delta \hat{y}) = \frac{\omega_r}{4Q} (1 + K_d^2) \quad (34)
\]

where \( K_d \) is as given in (7), and \( \frac{\omega_r}{4Q} \) is the two-sided noise equivalent bandwidth (in Hz) for the resonator assuming a first-order Lorentzian line shape. The (steady-state) variances above are all independent of \( t_e \). The FF, FLL and SSO schemes are non-adaptive, with a fixed noise filtering bandwidth.

The AKF based scheme provides an unbiased estimator, with zero bias for all \( t_e \). On the other hand, due to its adaptive noise filtering bandwidth, the variance of the estimator, given in (25), is dependent on \( t_e \).

For FF, FLL, and SSO, MSE is finite at \( t_e = 0 \) and asymptotes to a nonzero value as \( t_e \to \infty \). MSE depends on \( \Delta y \), the frequency shift to be estimated, in fact dominated by it for small \( t_e \), but asymptotes to a value that is independent of it. For both FLL and SSO, estimate variance is proportional to the detection bandwidth \( BW_{\text{FLL}} = BW_{\text{SSO}} \). On the other hand, the amount of bias decreases as detection bandwidth is made larger. Thus, for a given \( t_e \), there is an optimal value for the detection bandwidth.

For AKF, MSE is infinite at \( t_e = 0 \) and asymptotes to zero at \( t_e \to \infty \). MSE is independent of \( \Delta y \) for all \( t_e \). However, it does not make sense to use this scheme at very small \( t_e \), where MSE is on the order of \( \Delta y \) (or larger) to be detected. Thus, even though MSE is independent of \( \Delta y \), \( \Delta y \) determines the smallest \( t_e \) at which it makes sense to use the AKF estimate. Using AKF at a very small \( t_e \) as such will make the reliable detection of event times impossible or non-robust. For small \( t_e \), MSE is proportional to \( BW_L \) for large \( t_e \), it is independent of it. Thus, \( BW_L \) can be made as small as the linewidth of the resonator, or a little larger, so that the inherent response of the resonator is not modified by the filter. Finally, it should be noted that MSE for \( t_e \to 0 \) and \( t_e \to \infty \) are obviously not meaningful from a practical point of view. We need to consider \( \tau_{\text{FLL}} < t_e < \tau_r \).

For FLL and SSO, one can choose the detection bandwidth in the range from the resonator linewidth, \( \omega_r/Q \), all the way to the resonance frequency \( \omega_r \). In fact, when the detection bandwidth is set to the linewidth of the resonator, FLL becomes equivalent to the FF scheme. For a given detection bandwidth, MSE for these schemes reaches the minimum asymptotic value, i.e., the variance in (34), when the bias becomes zero, which occurs approximately at

\[
t_e^m \approx 10 \tau_{\text{FLL}} = 10 \tau_{\text{SSO}} = \frac{5}{BW_{\text{FLL}}} = \frac{5}{BW_{\text{SSO}}} \quad (35)
\]

In order to compare AKF with FLL and SSO, we evaluate the frequency shift to be estimated, in fact dominated by it for small \( t_e \), but asymptotes to a value that is independent of it. For both FLL and SSO, estimate variance is proportional to the detection bandwidth \( BW_{\text{FLL}} = BW_{\text{SSO}} \). On the other hand, the amount of bias decreases as detection bandwidth is made larger. Thus, for a given \( t_e \), there is an optimal value for the detection bandwidth.

For AKF, MSE is infinite at \( t_e = 0 \) and asymptotes to zero at \( t_e \to \infty \). MSE is independent of \( \Delta y \) for all \( t_e \). However, it does not make sense to use this scheme at very small \( t_e \), where MSE is on the order of \( \Delta y \) (or larger) to be detected. Thus, even though MSE is independent of \( \Delta y \), \( \Delta y \) determines the smallest \( t_e \) at which it makes sense to use the AKF estimate. Using AKF at a very small \( t_e \) as such will make the reliable detection of event times impossible or non-robust. For small \( t_e \), MSE is proportional to \( BW_L \) for large \( t_e \), it is independent of it. Thus, \( BW_L \) can be made as small as the linewidth of the resonator, or a little larger, so that the inherent response of the resonator is not modified by the filter. Finally, it should be noted that MSE for \( t_e \to 0 \) and \( t_e \to \infty \) are obviously not meaningful from a practical point of view. We need to consider \( \tau_{\text{FLL}} < t_e < \tau_r \).

For FLL and SSO, one can choose the detection bandwidth in the range from the resonator linewidth, \( \omega_r/Q \), all the way to the resonance frequency \( \omega_r \). In fact, when the detection bandwidth is set to the linewidth of the resonator, FLL becomes equivalent to the FF scheme. For a given detection bandwidth, MSE for these schemes reaches the minimum asymptotic value, i.e., the variance in (34), when the bias becomes zero, which occurs approximately at

\[
t_e^m \approx 10 \tau_{\text{FLL}} = 10 \tau_{\text{SSO}} = \frac{5}{BW_{\text{FLL}}} = \frac{5}{BW_{\text{SSO}}} \quad (35)
\]

In order to compare AKF with FLL and SSO, we evaluate the \( t_e \) dependent MSE of AKF at the above \( t_e^m \). Here, we assume that \( t_e^m > t_e^s \), where \( t_e^s \) is given in (29). That is, we are assuming that \( t_e^m \) falls into the thermomechanical noise limited regime for AKF. The MSE for AKF can be found by evaluating (28) at \( t_e^m \) as follows

\[
\text{MSE}^{\text{AKF}}(t_e^m) = \frac{S_{\nu_0}BW_{\text{FLL}}}{5} = \frac{S_{\nu_0}BW_{\text{SSO}}}{5} \quad (36)
\]

Then,

\[
\frac{\text{MSE}^{\text{FLL}}(t_e^m)}{\text{MSE}^{\text{SSO}}(t_e^m)} = \frac{\text{MSE}^{\text{FLL}}(t_e^m)}{\text{MSE}^{\text{SSO}}(t_e^m)} = \frac{5}{5 (1 + K_d^2)} \quad (37)
\]

Thus, MSE for AKF is smaller than the minimum that can be achieved by fixed-bandwidth FLL and SSO at the same elapsed time after the event. The impact of detection noise on AKF can be minimized, and in fact, almost eliminated, since the detection bandwidth \( BW_L \) can be reduced all the way down to the resonator linewidth. On the other hand, for FLL and SSO, detection bandwidth \( BW_{\text{FLL}} \) and \( BW_{\text{SSO}} \) determine the response speed and can not be reduced with only noise filtering considerations. Furthermore, AKF, featuring an adaptive bandwidth, continually improves the frequency shift estimate quality beyond the MSE value at \( t_e^m \). On the other hand, MSE for FLL and SSO has a lower bound that is reached at \( t_e^m \) for a given bandwidth. A graphical illustration
of RMSE versus elapsed time following event detection, is provided in Fig. 3 for FF, fixed-bandwidth FLL, and AKF. With this illustration, one can observe that the AKF scheme conceptually behaves as an FLL with a time-varying, adaptive-bandwidth that is dramatically increased after event detection and gradually reduced thereafter. However, due to the optimality properties associated with AKF, it outperforms such a conceptual adaptive FLL as quantified by (37). We should also note that turning the adaptive FLL concept into a practical scheme would amount to a nontrivial engineering endeavor, a pointless exercise in our opinion given the optimality and generality of the AKF scheme.

We note that we used the standard definition of variance in (30), in contrast with the use of Allan variance in [12]. This is due to the fact that Allan variance is not appropriate for characterizing the accuracy of the frequency shift estimates provided by the AKF based scheme. As discussed above, AKF provides estimates with a time-varying variance, conceptually corresponding to a time-varying spectral density for fractional frequency fluctuations, i.e., $S_y(\omega)$ depends on $t_e$. This would make Allan variance in [12, eqn. 75] not only a function of the averaging time $\tau$, but also a function of the elapsed time $t_e$, rendering it confusing and uninterpretable. The factor in [12, eqn. 75] that multiplies $S_y(\omega)$ represents an averaging operation (over a duration of $\tau$), that is not an integral part of the sensing scheme but simply dictated by the definition of Allan variance. On the other hand, the time-varying variance for AKF (decreasing function of elapsed time $t_e$) represents an inherent noise averaging operation that is an integral part of the sensing scheme itself. For FF, FLL, and SSO, there is no inherent noise averaging, resulting in a constant standard variance. Thus, comparisons of FF, FLL, and SSO with AKF based on an error definition that mixes the two different kinds of averaging operation would become confusing and in fact incorrect.

V. PRACTICAL CONSIDERATIONS FOR ADAPTIVE KALMAN FILTER BASED RESONATOR TRACKING

The AKF based resonator tracking scheme can be implemented as an add-on to the FF scheme in Fig. 1. The frequency shift output of the FF scheme is simply scaled by the resonance frequency in order to obtain the fractional frequency shift, which would then serve as the observation $y_o$ in (5). Thus, AKF based resonator tracking amounts to processing the FF scheme output. This can be conveniently done in the sampled-data domain with a digital implementation. AKF can be implemented as an add-on software module in FPGA/DSP based lock-in setups, either for realtime/online operation or offline post-processing. The simple but accurate resonator model with very low complexity, shown in Fig. 2, is the key enabler for a robust AKF based resonator tracking technique that can operate in realtime. It should be noted that the simple resonator model requires the values of two resonator parameters, the resonance frequency $\omega$, and the quality factor $Q$, which can be measured in a pre-characterization step, and occasionally updated during AKF operation.

When AKF is implemented as an add-on at the output of the FF scheme as described above, the disadvantages of the open-loop FF scheme relating to resonance frequency drifts, usually due to slow thermal phenomena, are inherited [12]. However, this problem can be solved easily. The most straightforward and simplest approach would be to implement an add-on FLL with a loop bandwidth that is (much) smaller than the linewidth of the resonator, that is, a very slow FLL. The purpose of the FLL here is not to sense and track the resonance frequency shifts at fast time scales due to events of interest, but simply to keep the drive of the resonator close to the resonance over long time scales, compensating for resonance frequency drifts due to nonidealities as well as possibly accumulating frequency shifts due to events of interest. This FLL would operate on the (very) slow speed, (very) high accuracy extreme of the speed-accuracy trade-off characteristics. Since this FLL is very slow, the output of the demodulator will essentially reflect the open-loop, feedback-free resonator response at short time scales up to and exceeding the resonator time constant. The filter in the demodulator in this case has a bandwidth set to the linewidth of the resonator, but the loop bandwidth for the FLL can be set to a value that is (much) smaller than the linewidth by choosing the controller parameters accordingly as described in [12]. While the FLL is making slow corrections to the drive frequency to keep the resonator operating around its resonance, the AKF will detect, predict, estimate and track the faster variations in resonance frequency due to events of interest. Alternatively, instead of using an FLL in order to compensate for drifts and accumulated frequency shifts, one can directly use the output of the AKF itself in order to continually, and at a faster time scale, update the drive frequency for the resonator.

VI. CONCLUSIONS

We have shown that there is still room for improvement in nanomechanical resonant sensors with sensor architecture
innovations. Pareto optimality with respect to speed and accuracy can be achieved by using an adaptive Kalman filtering based scheme that is founded on well-established techniques in detection and estimation theory. In this paper, we have assumed an event where the resonance frequency experiences a sudden shift that persists, e.g., when a particle falls on and sticks to a NEMS resonator. However, the AKF technique presented can be easily modified to detect and track events where the resonance frequency experiences a sudden but transient shift: Any type of event with a resonance frequency shift signature that can be modeled as the impulse response of a simple linear system can be easily incorporated into the AKF scheme. In the case considered in this paper, events have a frequency shift signature that is modeled as the impulse response of an ideal integrator, i.e., a step change. Replacing the integrator with a simple first-order filter with a prescribed time constant may be used as the model for a transient event signature. Even if the time constant of the event is much smaller than the resonator response time, AKF will still be able to detect and extract the magnitude of the root cause of the transient event. Finally, the proposed AKF approach is certainly not restricted to nanomechanical resonant sensors, it can also be used for optical and microwave resonant sensors.

ACKNOWLEDGMENT

The author would like to thank Selim Hanay and Alper Erdogan for helpful discussions.

REFERENCES

[1] J. Chaste, A. Eichler, J. Moser, G. Ceballos, R. Rurila, and A. Bachtold, “A nanomechanical mass sensor with yoctogram resolution,” Nature Nanotechnology, vol. 7, no. 5, p. 301, 2012.
[2] A. K. Naik, M. S. Hanay, W. K. Hieberg, X. L. Feng, and M. L. Roukes, “Towards single-molecule nanomechanical mass spectrometry,” Nature Nanotechnology, vol. 4, no. 7, p. 445, 2009.
[3] M. S. Hanay, S. Kelber, A. K. Naik, D. Chi, S. Hentz, E. C. Bullard, E. Colinet, L. Duraffourg, and M. L. Roukes, “Single-protein nanomechanical mass spectrometry in real time,” Nature Nanotechnology, vol. 7, no. 9, pp. 602–608, 2012.
[4] S. Dominguez-Medina, S. Fostner, M. Defoort, M. Sansa, A.-K. Stark, M. A. Halim, E. Vernhes, M. Gely, G. Jourdan, T. Alava, P. Boulanger, C. Masselon, and S. Hentz, “Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators,” Science, vol. 362, no. 6417, pp. 918–922, 2018.
[5] A. H. Ghadimi, S. A. Fedorov, N. J. Engelsen, M. J. Bereyhi, R. Schilling, D. J. Wilson, and T. J. Kippenberg, “Elastic strain engineering for ultralow mechanical dissipation,” Science, vol. 360, no. 6390, pp. 764–768, 2018.
[6] T. Albrecht, P. Grüter, D. Horne, and D. Rugar, “Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity,” Journal of Applied Physics, vol. 69, no. 2, pp. 668–673, 1991.
[7] K. Kobayashi, H. Yamada, and K. Matsushige, “Frequency noise in frequency modulation atomic force microscopy,” Review of Scientific Instruments, vol. 80, no. 4, p. 043708, 2009.
[8] J. P. Adams, B. W. Erickson, J. Grossenbacher, J. Brugger, A. Niev-ergelt, and G. E. Fantner, “Harnessing the damping properties of materials for high-speed atomic force microscopy,” Nature Nanotechnology, vol. 11, no. 2, p. 147, 2016.
[9] M. Z. Baykara, “Noncontact atomic force microscopy for atomic-scale characterization of material surfaces,” in Surface Science Tools for Nanomaterials Characterization, Springer, 2015, pp. 273–316.
[10] F. J. Giessibl, “The qplus sensor, a powerful core for the atomic force microscope,” Review of Scientific Instruments, vol. 90, no. 1, p. 011101, 2019.
[11] A. Demir and M. S. Hanay, “Fundamental sensitivity limitations of nanomechanical resonant sensors due to thermomechanical noise,” IEEE Sensors Journal, vol. 20, no. 4, pp. 1947–1961, February 2020.
[12] A. Demir, “Understanding fundamental trade-offs in nanomechanical resonant sensors,” Journal of Applied Physics, vol. 129, no. 4, p. 044503, 2021.
[13] D. C. Coffey and D. S. Ginger, “Time-resolved electrostatic force microscopy of polymer solar cells,” Nature Materials, vol. 5, no. 9, pp. 735–740, 2006.
[14] R. Giridharagopal, G. E. Rayermann, G. Shao, D. T. Moore, O. G. Reid, A. F. Tillack, D. J. Masiello, and D. S. Ginger, “Submicrorsecond time resolution atomic force microscopy for probing nanoscale dynamics,” Nano Letters, vol. 12, no. 2, pp. 893–898, 2012.
[15] D. U. Karatay, J. S. Harrison, M. S. Glaz, R. Giridharagopal, and D. S. Ginger, “Fast time-resolved electrostatic force microscopy: Achieving sub-cycle time resolution,” Review of Scientific Instruments, vol. 87, no. 5, p. 053702, 2016.
[16] A. Mascaro, Y. Miyahara, T. Enright, O. E. Dagdeviren, and P. Grüter, “Review of time-resolved non-contact electrostatic force microscopy techniques with applications to ionic transport measurements,” Beilstein Journal of Nanotechnology, vol. 10, no. 1, pp. 617–633, 2019.
[17] R. R. Feynman, “Plenty of room at the bottom,” in AFIP Annual Meeting, 1959.
[18] Y. Sadeghi, A. Demir, L. G. Villanueva, H. Kühler, and S. Schmid, “Frequency fluctuations in nanomechanical silicon nitride string resonators,” Physical Review B, vol. 102, p. 214106, Dec 2020.
[19] T. Kailath, “An innovations approach to least-squares estimation—Part I: Linear filtering in additive white noise,” IEEE Transactions on Automatic Control, vol. 13, no. 6, pp. 646–655, 1968.
[20] A. Willisky and H. Jones, “A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems,” IEEE Transactions on Automatic Control, vol. 21, no. 1, pp. 108–112, 1976.
[21] 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 Transactions on Nanotechnology, vol. 13, no. 6, pp. 1257–1265, 2014.
[22] M. G. Ruppert, K. S. Karvinen, S. L. Wiggins, and S. O. Reza Moheimani, “A Kalman filter for amplitude estimation in high-speed dynamic mode atomic force microscopy,” IEEE Transactions on Control Systems Technology, vol. 24, no. 1, pp. 276–284, 2016.
[23] M. G. Ruppert, D. M. Harcombe, and S. O. R. Moheimani, “High-bandwidth demodulation in MF-AFM: A Kalman filtering approach,” IEEE/ASME Transactions on Mechatronics, vol. 21, no. 6, pp. 2705–2715, 2016.
[24] M. R. P. Ragazzon, M. G. Ruppert, D. M. Harcombe, A. J. Fleming, and J. T. Gravdahl, “Lyapunov estimator for high-speed demodulation in dynamic mode atomic force microscopy,” IEEE Transactions on Control Systems Technology, vol. 26, no. 2, pp. 765–772, 2018.
[25] D. R. Sahoo, A. Sebastian, and M. V. Salapaka, “Harnessing the transient signals in atomic force microscopy,” International Journal of Robust and Nonlinear Control, vol. 15, no. 16, pp. 805–820, 2005.
[26] ———, “Transient-signal-based sample-detection in atomic force microscopy,” Applied Physics Letters, vol. 83, no. 26, pp. 5521–5523, 2003.
[27] S. Salapaka and M. Salapaka, “Atomic force microscopy: Principles and systems viewpoint enabled methods,” in Feedback Control of MEMS to Atoms. Springer, 2012, pp. 109–139.

Alper Demir received a B.S. from Bilkent University, Ankara, Turkey, in 1991, and a Ph.D. from the University of California, Berkeley, CA, USA, in 1997. He was a member of technical staff at Bell Laboratories, Murray Hill, NJ, USA. He has been with Koç University, Istanbul, Turkey, since 2002. He received the 2002 Best of ICCAD Award, the 2003/2014 IEEE/ACM William J. McCalla ICCAD best paper awards, and the 2004 IEEE Guillemín-Cauer Award. He is an IEEE Fellow.