Understanding Fundamental Tradeoffs in Nanomechanical Resonant Sensors

Alper Demir

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

arXiv:2010.04447v1 [physics.app-ph] 9 Oct 2020

***Abstract*—Nanomechanical resonators are used as high performance detectors in a variety of applications such as mass spectrometry and atomic force microscopy. Initial emphasis in nanomechanical resonant sensor research was on increasing the sensitivity to the level of a single molecule, atom and beyond. On the other hand, there are applications where the speed of detection is crucial, prompting recent works that emphasize sensing schemes with improved time resolution. We first develop a general modeling framework encompassing all resonator tracking schemes currently in use, by extending recent previous work. We then explore the fundamental trade-offs between accuracy and speed in three resonant sensor architectures, namely the feedback-free open-loop approach, positive-feedback based self-sustaining oscillator, and negative-feedback based frequency-locked loop scheme. We comparatively analyze them in a unified manner, clarify some misconceptions that seem to exist in the literature, and unravel their speed versus accuracy characteristics.**

***Index Terms*—nano-mechanical resonant sensor, phase-locked loop, frequency-locked loop, NEMS oscillator, thermo-mechanical noise, detection noise, Allan deviation.**

I. Introduction

Nanomechanical resonators are used as the core components in various sensing, detection, spectroscopy and microscopy techniques, *e.g.,* in mass spectrometry [1]–[5] and atomic force microscopy [6]–[10], Almost all of these techniques are based on monitoring the resonance frequency shifts arising from the interaction of the nanomechanical resonator with its environment and the sample under study.

There are three prevalent schemes for measuring and tracking the resonance frequency of a nanomechanical resonator. In the simplest feedback-free (FF) approach [6], [11], shown in Fig. 1, the resonator is driven with a sinusoidal signal at a constant amplitude, and a constant frequency close to the resonance frequency. The vibratory motion of the resonator arising from this excitation is transduced and demodulated in order to measure its amplitude and phase difference with respect to the drive. Any event of interest will result in a deviation in the measured response amplitude and the phase difference, which can be used in conjunction with the resonator characteristics in order to compute the induced resonance frequency shift and other quantities such as mass changes that relate to the cause. This feedback-free scheme has two major disadvantages. The resonance frequency drifts over time, albeit at a slow rate, due to phenomena other than the events of interest. If the resonance frequency drifts too far away from the drive frequency, accurate measurement and computation of the frequency shift becomes impossible due to degraded, flat resonator response. Even if this drift problem is solved, the feedback-free scheme has an inflexible trade-off characteristics between accuracy and speed that is essentially dictated by the resonator characteristics, to be detailed later.

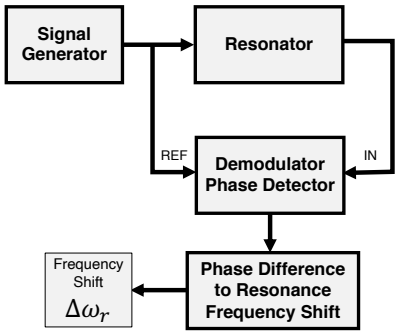


Fig. 1. Feedback-free resonator tracking (figure based on [12, Fig. 1])

A *frequency-locked loop* (FLL) [2], [3], [12]–[14] based resonator tracking scheme, shown Fig. 2, addresses the drift problem of the feedback-free approach, by continually updating the frequency of the drive signal generator with a (negative) feedback control loop so that the resonator is always driven at resonance. This solves the drift problem in the sense that the resonator always operates in the region of its characteristics that allows accurate frequency shift discrimination. However, resonance frequency drifts due to other nonideal phenomena may still interfere with the detection and tracking of the events of interest that also result in a frequency shift. Fortunately, in most applications, drift phenomena occur at a much slower time scale, allowing the discrimination of the phenomena of interest. The closed-loop FLL scheme has another major advantage, offering flexibility in response speed. It is well established that resonators with higher quality factors offer better sensitivity [12]. However, a high quality factor also results in a slow mechanical response dictating a lower detection speed. On the other hand, the speed with which an FLL responds to frequency shift events can be tailored based on the requirements of the application. Importantly,

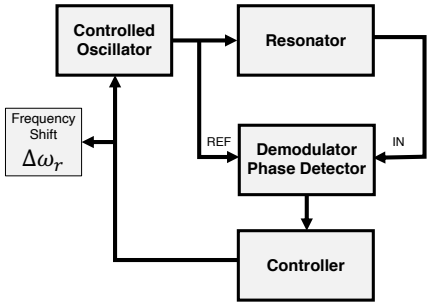


Fig. 2. Frequency-locked loop (figure based on [12, Fig. 1])

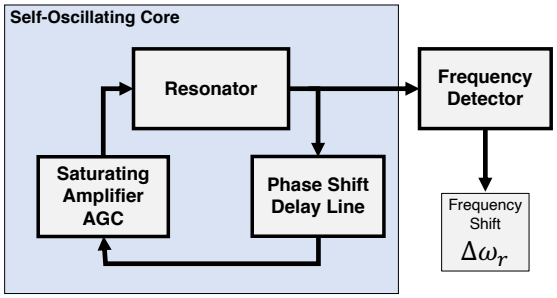


Fig. 3. Self-sustained oscillator (figure based on [12, Fig. 1])

this can be done in such a way so that the FLL response time is much shorter than, and in fact independent of, the mechanical response time of the resonator. Unfortunately, this added benefit does not come for free. As we show in detail later, some of the sensitivity improvement obtained by using a resonator with a higher quality factor has to be forfeited. On the other hand, the FLL dynamics can be designed to have a slower response when compared with the mechanical response of the resonator, for an extra sensitivity boost. This flexible speed-accuracy tradeoff characteristics offered by an FLL is most useful, but not very well recognized and deeply understood in the literature. Finally, the FLL scheme has a built-in utility for the measurement of the resonance frequency via the feedback control signal that sets the frequency of the drive signal generator.

Another closed-loop method in use for nanomechanical resonant sensors is the so-called *self-sustaining oscillator* (SSO) scheme [6], [15], [16], shown in Fig. 3, where the resonator is used as the frequency selective element in a classical positive-feedback based oscillator architecture with amplification and time delay (phase shift) in the loop. In the oscillator core, there is no external signal generator. Instead, the periodic oscillation signal (with a frequency close to the resonance frequency) is generated by the marginal instability of the positive-feedback loop itself. Oscillators of this sort require a mechanism for amplitude stabilization, which is usually provided by a nonlinearity. For nanomechanical oscillators, this nonlinearity can be implemented in the amplifier placed in the loop, via gain saturation and automatic gain control (AGC), as in the FM-AFM scheme [6], [15] that is widely used in atomic force microscopy. In this case, the resonator operates in the linear regime. In other self-oscillating schemes, the resonator is driven harder to operate in the nonlinear Duffing regime [17]–[20]. Unlike in the case of FLL, the self-oscillating scheme needs to be augmented with a frequency measurement method such as a frequency counter, an FM detector based on Hilbert transforms or narrowband IQ demodulators, or a *phase-locked loop* (PLL) [6], [15]. A PLL that locks an independent, controllable signal generator to the oscillation signal generated by the self-oscillating core has some advantages as discussed in [15]. We note the distinction between the PLL here and the FLL discussed above, which are both negative-feedback based control systems [12]. The FLL locks the frequency of the signal generator directly to the resonator, whereas the PLL locks both the frequency and phase of the signal generator to the oscillation signal produced by the positive-feedback based self-oscillating core. In the FLL, the resonator is driven with an external signal generator, whereas in the self-oscillating core, it is driven with an amplified and phase shifted version of the signal itself generates via positive-feedback. A hybrid scheme that combines an FLL with a self-oscillation loop was proposed in [15], where the resonator is driven by a signal that is obtained by combining its own amplified response with the output of the signal generator in the PLL. The self-oscillating and the hybrid schemes described above offer similar advantages over the feedback-free approach. The resonator always operates at or near resonance in the linear regime, or at other desired operating points in the Duffing regime [17]–[20]. This solves the drift problem. The response time of a SSO to an event of interest is also much shorter than, and in fact can be made independent of, the mechanical response time of the resonator [6], as in the case of an FLL, which we derive in a rigorous manner in this paper. On the other hand, the SSO scheme suffers from amplification of detection noise circulating in the positive feedback loop, resulting in an inferior accuracy (sensitivity) as compared with an FLL, which we quantify precisely. Furthermore, the SSO scheme also offers a flexible speed-accuracy trade-off characteristics, but not as versatile as the one in an FLL.

In Section II, we develop a unifying theory and models for resonators, basic configurations in resonant sensors and noise sources. The general modeling framework we develop in this paper encompasses all kinds of resonator tracking schemes, and extends our recent previous work in [12] that considered only the FLL. We describe the three standard resonator tracking schemes, FF, FLL and SSO in Section III, and perform a comparative characterization in Section IV.

II. Theory

A. Resonator model

We consider a resonator that is modeled as a damped harmonic oscillator as follows [21]

$$\frac{d^2}{dt^2}\,x(t) + \frac{\omega_\mathrm{r}}{Q}\,\frac{d}{dt}\,x(t) + \omega_\mathrm{r}^2\,x(t) = \frac{F(t)}{m} \qquad (1)$$

where x is the displacement, m is the mass, $F(t)$ represents a force excitation, ω_r is the resonance frequency, and Q is the

quality factor. We express

$$x(t) = \mathcal{R}\{s(t)\, e^{j\omega_o t}\} = \tfrac{1}{2}\left[s(t)\, e^{j\omega_o t} + s^*(t)\, e^{-j\omega_o t}\right] \quad (2)$$

where $\cdot^*$ denotes complex-conjugate, $\mathcal{R}\{\cdot\}$ is the real part operator. With a reasonably high Q, $s(t)$ is the slowly varying complex amplitude of $x(t)$. The carrier frequency ω_o, while not necessarily equal to the resonance frequency ω_r, is chosen such that $|\omega_r - \omega_o| < \frac{\omega_r}{Q} \ll \omega_r$. We substitute (2) into (1) to obtain

$$\frac{1}{2}\left[e^{j\omega_o t}\, \mathbf{S}(t) + e^{-j\omega_o t}\, \mathbf{S}^*(t)\right] = \frac{F(t)}{m} \quad (3)$$

$$\mathbf{S}(t) = \frac{d^2}{dt^2}s + \left(\frac{\omega_r}{Q} + 2j\omega_o\right)\frac{d}{dt}s + \left(\omega_r^2 - \omega_o^2 + j\frac{\omega_r\omega_o}{Q}\right)s$$

where the t dependence of $s(t)$ was omitted for notational simplicity. We assume that $s(t)$ is slowly varying with a maximum bandwidth of $\frac{w_r}{Q}$. Hence, the t derivative of $s(t)$ results in a maximum amplification of $\frac{w_r}{Q}$. With $|\omega_r - \omega_o| < \frac{\omega_r}{Q} \ll \omega_r$, the terms above have the following amplification factor orders:

$$\underbrace{\frac{d^2}{dt^2}}_{\frac{w_r^2}{Q^2}} s + \underbrace{\frac{\omega_r}{Q}\frac{d}{dt}}_{\frac{w_r^2}{Q^2}} s + \underbrace{2j\omega_o\frac{d}{dt}}_{\frac{w_r^2}{Q}} s + \underbrace{\left(\omega_r^2 - \omega_o^2 + j\frac{\omega_r\omega_o}{Q}\right)}_{\frac{w_r^2}{Q}} s$$

The first two terms above are Q times smaller than the last two terms, and hence we neglect them and also use $\omega_r^2 - \omega_o^2 \approx 2\omega_o(\omega_r - \omega_o)$ to obtain

$$\mathbf{S}(t) \approx 2j\omega_o\frac{d}{dt}s(t) + \left(2\omega_o(\omega_r - \omega_o) + j\frac{\omega_r\omega_o}{Q}\right)s(t) \quad (4)$$

$s(t)$ (and its derivative with respect to t) changes very little over the cycle time of the resonator, *i.e.,*

$$\mathbf{S}(t') \approx \mathbf{S}(t) \quad \text{for} \quad t \le t' \le t + \tfrac{2\pi}{\omega_o} \quad (5)$$

We then multiply both sides of (3) with $\frac{\omega_o}{\pi}e^{-j\omega_o t}$ and integrate over a time interval of $\frac{2\pi}{\omega_o}$ to obtain

$$\begin{aligned}\frac{\omega_o}{2\pi}\int_t^{t+\frac{2\pi}{\omega_o}} \left[\mathbf{S}(t') + e^{-j2\omega_o t'}\, \mathbf{S}^*(t')\right]dt' \approx \\ \mathbf{S}(t) \approx \frac{1}{m}\frac{\omega_o}{\pi}\int_t^{t+\frac{2\pi}{\omega_o}} e^{-j\omega_o t'} F(t')dt'\end{aligned} \quad (6)$$

We combine the above with (4)

$$\begin{aligned}2j\omega_o\frac{d}{dt}s(t) + \left(2\omega_o(\omega_r - \omega_o) + j\frac{\omega_r\omega_o}{Q}\right)s(t) \approx \\ \frac{1}{m}\frac{\omega_o}{\pi}\int_t^{t+\frac{2\pi}{\omega_o}} e^{-j\omega_o t'} F(t')dt'\end{aligned} \quad (7)$$

multiply both sides with $-j\frac{Q}{\omega_r\omega_o}$

$$\begin{aligned}\frac{2Q}{\omega_r}\frac{d}{dt}s(t) + \left(1 + j\frac{2Q}{\omega_r}(\omega_o - \omega_r)\right)s(t) \approx \\ \frac{-j}{m}\frac{Q}{\omega_r}\frac{1}{\pi}\int_t^{t+\frac{2\pi}{\omega_o}} e^{-j\omega_o t'} F(t')dt'\end{aligned} \quad (8)$$

and define $\tau_r = \frac{2Q}{\omega_r}$ as the intrinsic resonator time-constant to obtain

$$\begin{aligned}\tau_r\frac{d}{dt}s(t) + [1 + j\tau_r(\omega_o - \omega_r)]s(t) \approx \\ \frac{-j}{m}\frac{Q}{\omega_r}\frac{1}{\pi}\int_t^{t+\frac{2\pi}{\omega_o}} e^{-j\omega_o t'} F(t')dt'\end{aligned} \quad (9)$$

B. Feedback-free driven resonator step response

We choose the external force drive as

$$F(t) = \mathcal{R}\{f(t)\, e^{j\omega_o t}\} = \tfrac{1}{2}\left[f(t)\, e^{j\omega_o t} + f^*(t)\, e^{-j\omega_o t}\right] \quad (10)$$

where $f(t)$ is a slowly varying complex amplitude, similar to $s(t)$ in (2). We substitute the above in (9), assuming

$$f(t') \approx f(t) \quad \text{for} \quad t \le t' \le t + \tfrac{2\pi}{\omega_o} \quad (11)$$

and obtain

$$\tau_r\frac{d}{dt}s(t) + [1 + j\tau_r(\omega_o - \omega_r)]s(t) = \frac{-j}{m}\frac{Q}{\omega_r\omega_o}f(t) \quad (12)$$

The above serves as a compact model for evaluating the resonator response to various kinds of excitations. We first choose $f(t) = A_f e^{j\theta_f}$, where the amplitude A_f and the phase θ_f are time-invariant, which implies a sinusoidal drive $F(t) = A_f\cos(\omega_o t + \theta_f)$. The steady-state response of the resonator to this excitation can be computed by setting the t derivative in (12) to zero as follows

$$s(t) = \frac{-j}{m}\frac{Q}{\omega_r\omega_o}\frac{A_f e^{j\theta_f}}{[1 + j\tau_r(\omega_o - \omega_r)]} = A_r e^{j\theta_r} \quad (13)$$

where

$$\begin{aligned}A_r = \quad & \frac{1}{m}\frac{Q}{\omega_r\omega_o}\frac{A_f}{\sqrt{1 + \tau_r^2(\omega_o - \omega_r)^2}} \\ \theta_r = \quad & \theta_f - \tfrac{\pi}{2} - \operatorname{atan}\left(\tau_r(\omega_o - \omega_r)\right)\end{aligned} \quad (14)$$

implying a sinusoidal response $x(t) = A_r\cos(\omega_o t + \theta_r)$.

We next consider an excitation as follows

$$f(t) = \begin{cases} A_f\, e^{j\theta_f} & t < 0 \\ A_f\, e^{j(\Delta\omega_e t + \theta_f)} & t \ge 0 \end{cases} \quad (15)$$

where the drive frequency is stepped at $t = 0$ by $\Delta\omega_e$, with $\Delta\omega_e < \frac{\omega_r}{Q}$. We evaluate the transient response of the resonator due to the above excitation for $t \ge 0$, by setting the response

$$s(t) = A_r(t)e^{j(\Delta\omega_e t + \theta_r(t))} \quad (16)$$

with the initial condition

$$s(0) = A_r(0)e^{j\theta_r(0)} = \frac{-j}{m}\frac{Q}{\omega_r\omega_o}\frac{A_f e^{j\theta_f}}{[1 + j\tau_r(\omega_o - \omega_r)]} \quad (17)$$

with $A_r(0)$ and $\theta_r(0)$ as in (14). We substitute (15) and (16) in (12) for $t \ge 0$, and after some manipulations, we obtain

$$\begin{aligned}\tau_r\frac{d}{dt}\left[A_r(t)e^{j\theta_r(t)}\right] + [1 + j\tau_r(\omega_o + \Delta\omega_e - \omega_r)]A_r(t)e^{j\theta_r(t)} \\ = \frac{-j}{m}\frac{Q}{\omega_r\omega_o}A_f e^{j\theta_f}\end{aligned} \quad (18)$$

We observe that this excitation scenario also captures the case where the drive frequency is fixed, but the resonance frequency suddenly changes, *i.e.,* Δw increase in the drive frequency may

equivalently be intrerpreted as a Δw decrease in the resonance frequency. The solution of the above differential equation with the initial condition in (17) is given by (19). This response has a decay time constant τ_{r}, but also exhibits oscillatory behavior with beat frequency $\omega_{\mathsf{r}} - \omega_{\mathsf{o}} - \Delta\omega_{\mathsf{e}}$ [6].

$$A_{\mathsf{r}}(t)e^{j\theta_{\mathsf{r}}(t)} = \frac{-j\,Q}{m\omega_{\mathsf{r}}\,\omega_{\mathsf{o}}}\left[\frac{A_{\mathsf{f}}e^{j\theta_{\mathsf{f}}}}{1+j\tau_{\mathsf{r}}(\omega_{\mathsf{o}}-\omega_{\mathsf{r}})}e^{-t/\tau_{\mathsf{r}}}e^{j(\omega_{\mathsf{r}}-\omega_{\mathsf{o}}-\Delta\omega_{\mathsf{e}})t} + \frac{A_{\mathsf{f}}e^{j\theta_{\mathsf{f}}}}{1+j\tau_{\mathsf{r}}(\omega_{\mathsf{o}}+\Delta\omega_{\mathsf{e}}-\omega_{\mathsf{r}})}\Big(1-e^{-t/\tau_{\mathsf{r}}}e^{j(\omega_{\mathsf{r}}-\omega_{\mathsf{o}}-\Delta\omega_{\mathsf{e}})t}\Big)\right] \quad (19)$$

C. Feedback-free resonator phase response

Due to its sharpness around the resonance frequency, the phase response is preferred for measuring resonance frequency shifts. To first order, the amplitude of the resonator response around the resonance frequency can be approximated as constant with $A_{\mathsf{r}}(t) = A_{\mathsf{rss}}$. We set $\omega_{\mathsf{r}} = \omega_{\mathsf{o}}$ and

$$A_{\mathsf{r}}(t) = A_{\mathsf{rss}} = \frac{Q\,A_{\mathsf{f}}}{m\,\omega_{\mathsf{r}}^2} \quad \text{for all} \quad t \quad (20)$$

in (18) for

$$\tau_{\mathsf{r}}\frac{d}{dt}\Big[e^{j\theta_{\mathsf{r}}(t)}\Big] + (1+j\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}})e^{j\theta_{\mathsf{r}}(t)} = -je^{j\theta_{\mathsf{f}}} \quad (21)$$

with the initial condition $\theta_{\mathsf{r}}(0) = \theta_{\mathsf{f}} - \frac{\pi}{2}$. We set $\theta_{\mathsf{r}}(t) = \theta_{\mathsf{r}}(0) + \Delta\theta_{\mathsf{r}}(t)$, with $\Delta\theta_{\mathsf{r}}(t)$ as the phase shift due to the resonance frequency shift and subsititute into (21) for

$$\tau_{\mathsf{r}}\frac{d}{dt}\Big[e^{j\Delta\theta_{\mathsf{r}}(t)}\Big] + (1+j\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}})e^{j\Delta\theta_{\mathsf{r}}(t)} = 1 \quad (22)$$

Assuming a small phase shift, we use $e^{j\Delta\theta_{\mathsf{r}}(t)} \approx 1 + j\Delta\theta_{\mathsf{r}}(t)$ in the above to obtain

$$j\tau_{\mathsf{r}}\frac{d}{dt}\Delta\theta_{\mathsf{r}}(t) + j\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}} + j\Delta\theta_{\mathsf{r}}(t) - \tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}}\Delta\theta_{\mathsf{r}}(t) = 0 \quad (23)$$

Since both $\Delta\omega_{\mathsf{e}}$ and $\Delta\theta_{\mathsf{r}}(t)$ are assumed to be small, we neglect the second-order last term in the above for

$$\tau_{\mathsf{r}}\frac{d}{dt}\Delta\theta_{\mathsf{r}}(t) + \Delta\theta_{\mathsf{r}}(t) = -\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}} \quad (24)$$

Thus, the phase response of the resonator to sudden resonance frequency shifts of size $\Delta\omega_{\mathsf{e}}$, *i.e.,* its *frequency step response* (FSTR), can be obtained by simply applying a *phase step* of size $\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}}$ at the input of a one-pole low-pass filter based resonator model with transfer function

$$H_{\mathsf{R}}(\mathsf{s}) = \frac{1}{1+\mathsf{s}\,\tau_{\mathsf{r}}} \quad (25)$$

A resonator model as above was also derived both in [12] and [13] and using two different techniques. We believe that the derivation we have presented above, a third alternative, is most general and systematic. The solution of (24), with initial condition $\Delta\theta_{\mathsf{r}}(0) = 0$, is

$$\Delta\theta_{\mathsf{r}}(t) = -\tau_{\mathsf{r}}\Delta\omega_{\mathsf{e}}\Big(1-e^{-t/\tau_{\mathsf{r}}}\Big) \quad (26)$$

D. Positive-feedback based oscillator model

We consider a self-oscillating core, where an amplified and phase shifted version of the resonator response is used to drive itself. Based on the Barkhausen criterion, positive-feedback based oscillators need to have an amplitude stabilization mechanism, that ensures a consistent response at a desired level. This is provided by a nonlinearity, *e.g.,* the Duffing nonlinearity of the resonator, a saturating amplifier, or an AGC loop. Furthermore, in order to ensure proper positive-feedback and self-sustaining oscillations, a phase condition needs to be fulfilled around the loop, which may be achieved with a phase shifter, a delay line, or a time-derivative in some cases.

We consider a self-oscillating core where the resonator operates in the linear regime, and amplitude stabilization is provided by a saturating amplifier. We set the excitation $F(t)$ in (9) as follows

$$F(t) = e^{j\frac{\pi}{2}}\,h(x(t)) \quad (27)$$

where $h(\cdot)$ above is a nonlinear odd-function, linear for small values of its argument and with a saturating characteristics for large input, *e.g.,* a $\tanh$ like function, or a polynomial with only first and third order terms. The $\frac{\pi}{2}$ phase shift ensures proper positive feedback and satisfaction of the Barkhausen criterion at the resonance frequency, which may also be achieved with a time derivative operating either on the argument or the return value of $h(\cdot)$.

We assume that the solution of the self-oscillating core is sinusoidal or narrowband with the excitation as in (27), without higher-order harmonic components. For the feedback-free driven case, this assumption is justified based on the linearity of the resonator and the system and the fact that the drive itself is nearly sinuosidal and narrowband. However, in the self-oscillating case, the system is nonlinear due to the nonlinear function $h(\cdot)$ in (27), even though the resonator itself is still operating in the linear regime. That is, even if the argument $x(t)$ of $h(\cdot)$ is narrowband, $h(x(t))$ will have higher-order harmonic components. However, these will be filtered by the resonator, especially well, when Q is large.

We consider the integral in the right-hand-side of (9) with $F(t)$ as in (27) and with $x(t)$ in (2)

$$\int_{t}^{t+\frac{2\pi}{\omega_{\mathsf{o}}}} e^{-j\omega_{\mathsf{o}}t'}\,h\Big(\tfrac{1}{2}\Big[s(t')\,e^{j\omega_{\mathsf{o}}t'} + s^*(t')\,e^{-j\omega_{\mathsf{o}}t'}\Big]\Big)dt' \quad (28)$$

In the above operation, a nearly sinusoidal, narrowband signal with carrier frequency ω_{o} is passed through a nonlinear function $h(\cdot)$. The outcome will have narrowband components that are centered around the harmonics of ω_{o}. The integral above extracts the narrowband signal component around the fundamental ω_{o}, assuming that $s(t)$ is slowly varying as before. When $h(\cdot)$ is an odd, memoryless nonlinear function as described above, it can be shown that this component can be expressed as

$$h_{\mathsf{D}}(|s(t)|)\,s(t) \quad (29)$$

where $h_{\mathsf{D}}(\cdot)$ is a nonlinear function, related to but distinct from $h(\cdot)$, known as a *describing function* [22]. $h_{\mathsf{D}}(|s(t)|)$ essentially represents an amplitude-dependent gain. When $h(\cdot)$ is a $\tanh$ like function, the corresponding $h_{\mathsf{D}}(\cdot)$ has a flat characteristics in the region where $h(\cdot)$ is linear, and decays to small values when $h(\cdot)$ saturates. Hence, (9) can be simplified as follows

$$\begin{aligned}\tau_{\mathsf{r}}\frac{d}{dt}s(t)+[1+j\tau_{\mathsf{r}}(\omega_{\mathsf{o}}-\omega_{\mathsf{r}})]s(t)\\ =\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}}\,\omega_{\mathsf{o}}}h_{\mathsf{D}}(|s(t)|)\,s(t)\end{aligned}\tag{30}$$

We assume a steady-state, sinusoidal solution to the above in the form $s(t) = A_{\mathsf{r}}e^{j\theta_{\mathsf{r}}}$, with time-invariant A_{r} and θ_{r}, and substitute in (30) to obtain

$$[1+j\tau_{\mathsf{r}}(\omega_{\mathsf{o}}-\omega_{\mathsf{r}})]A_{\mathsf{r}}e^{j\theta_{\mathsf{r}}}=\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}}\,\omega_{\mathsf{o}}}h_{\mathsf{D}}(A_{\mathsf{r}})\,A_{\mathsf{r}}e^{j\theta_{\mathsf{r}}}\tag{31}$$

We observe that the above equation does not impose any restriction on the phase θ_{r}. It is well-known that autonomous oscillators have a *free* phase, determined by the initial conditions, that may freely diffuse due to noise [23]. Without noise in the system, θ_{r} takes an arbitary but time-invariant value. We eliminate the terms in (31) that involve θ_{r}

$$[1+j\tau_{\mathsf{r}}(\omega_{\mathsf{o}}-\omega_{\mathsf{r}})]A_{\mathsf{r}}=\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}}\,\omega_{\mathsf{o}}}h_{\mathsf{D}}(A_{\mathsf{r}})\,A_{\mathsf{r}}\tag{32}$$

A_{r} is a real-valued solution of the above equation. The trivial solution $A_{\mathsf{r}} = 0$ is in fact unstable due to positive feedback and net gain in the loop. Any disturbance or nonzero initial condition will kick the solution away from $A_{\mathsf{r}} = 0$. For (32) to have a positive and real-valued nontrivial solution A_{r}, both of the following conditions, obtained from the real and imaginary parts of (32), need to be satisfied

$$\begin{aligned}\omega_{\mathsf{o}}&=\omega_{\mathsf{r}}\\ \frac{1}{m}\frac{Q}{\omega_{\mathsf{r}}\,\omega_{\mathsf{o}}}h_{\mathsf{D}}(A_{\mathsf{r}})&=1\end{aligned}\tag{33}$$

Thus, the oscillation (carrier) frequency is in fact equal to the resonance frequency ω_{r}, ensured by the $\frac{\pi}{2}$ phase shift introduced into the positive-feedback loop. A_{r} is the unique solution of the nonlinear equation in (33).

E. Self-oscillating core step response

We evaluate the transient response of the self-oscillating core to a sudden change in the resonance frequency, from $\omega_{\mathsf{r}1}$ to $\omega_{\mathsf{r}2}$ at $t = 0$. We use the model in (30) for $t < 0$, by choosing $\omega_{\mathsf{o}} = \omega_{\mathsf{r}} = \omega_{\mathsf{r}1}$. The steady-state solution for $t < 0$ is then given by

$$s(t) = A_{\mathsf{r}1}\,e^{j\theta_{\mathsf{r}1}} \quad \text{for} \quad t<0\tag{34}$$

with arbitrary $\theta_{\mathsf{r}1}$ determined by initial conditions at $t = -\infty$, and $A_{\mathsf{r}1}$ as the solution of

$$\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}^2}h_{\mathsf{D}}(A_{\mathsf{r}1})=1\tag{35}$$

and a carrier $e^{j\omega_{\mathsf{r}1}t}$. We use the model in (30) for $t \geq 0$, by choosing $\omega_{\mathsf{o}} = \omega_{\mathsf{r}1}$ and $\omega_{\mathsf{r}} = \omega_{\mathsf{r}2}$

$$\begin{aligned}\tau_{\mathsf{r}}\frac{d}{dt}s(t)+[1+j\tau_{\mathsf{r}}(\omega_{\mathsf{r}1}-\omega_{\mathsf{r}2})]s(t)\\ =\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(|s(t)|)\,s(t)\end{aligned}\tag{36}$$

with the initial condition in (34). Assuming a solution in the form $s(t) = A_{\mathsf{r}}(t)\,e^{j\theta_{\mathsf{r}}(t)}$ in the above, we obtain

$$\begin{aligned}&\tau_{\mathsf{r}}\left[\frac{d}{dt}A_{\mathsf{r}}(t)\right]e^{j\theta_{\mathsf{r}}(t)}\\ &+\left[1+j\tau_{\mathsf{r}}\left(\frac{d}{dt}\theta_{\mathsf{r}}(t)+\omega_{\mathsf{r}1}-\omega_{\mathsf{r}2}\right)\right]A_{\mathsf{r}}(t)e^{j\theta_{\mathsf{r}}(t)}\\ &=\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}}(t))\,A_{\mathsf{r}}(t)e^{j\theta_{\mathsf{r}}(t)}\end{aligned}\tag{37}$$

All terms above have a common factor $e^{j\theta_{\mathsf{r}}(t)}$ that can be eliminated to obtain

$$\begin{aligned}\tau_{\mathsf{r}}\frac{d}{dt}A_{\mathsf{r}}(t)+\left[1+j\tau_{\mathsf{r}}\left(\frac{d}{dt}\theta_{\mathsf{r}}(t)+\omega_{\mathsf{r}1}-\omega_{\mathsf{r}2}\right)\right]A_{\mathsf{r}}(t)\\ =\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}}(t))\,A_{\mathsf{r}}(t)\end{aligned}\tag{38}$$

We note that $A_{\mathsf{r}}(t)$ is a positive and real-valued quantity, the real and imaginary parts of the above equation yields

$$\begin{aligned}\frac{d}{dt}\theta_{\mathsf{r}}(t)&=\omega_{\mathsf{r}2}-\omega_{\mathsf{r}1}\\ \tau_{\mathsf{r}}\frac{d}{dt}A_{\mathsf{r}}(t)+A_{\mathsf{r}}(t)&=\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}}(t))\,A_{\mathsf{r}}(t)\end{aligned}\tag{39}$$

with the initial conditions $\theta_{\mathsf{r}}(0) = \theta_{\mathsf{r}1}$ and $A_{\mathsf{r}}(0) = A_{\mathsf{r}1}$. The solution for the phase is given by

$$\theta_{\mathsf{r}}(t) = (\omega_{\mathsf{r}2}-\omega_{\mathsf{r}1})\,t+\theta_{\mathsf{r}1} = \Delta\omega_{\mathsf{e}}\,t+\theta_{\mathsf{r}1}\tag{40}$$

The solution for the amplitude $A_{\mathsf{r}}(t)$ can not be obtained analytically, since the governing differential equation is nonlinear. However, it can be approximated as follows. $A_{\mathsf{r}2}$, the new steady-state amplitude at the new resonance frequency $\omega_{\mathsf{r}2}$, is the solution of

$$\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}2})=1\tag{41}$$

Let

$$A_{\mathsf{r}}(t) = A_{\mathsf{r}2}+\Delta A_{\mathsf{r}}(t)\tag{42}$$

and assuming $A_{\mathsf{r}2}-A_{\mathsf{r}1}$ and $\Delta A_{\mathsf{r}}(t)$ are small, use the following approximation

$$\begin{aligned}&\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}}(t))\approx\\ &\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}h_{\mathsf{D}}(A_{\mathsf{r}2})+\underbrace{\frac{1}{m}\frac{Q}{\omega_{\mathsf{r}1}\,\omega_{\mathsf{r}2}}\frac{d}{d\bullet}h_{\mathsf{D}}(\bullet)\Big|_{\bullet=A_{\mathsf{r}2}}}_{\gamma}\Delta A_{\mathsf{r}}(t)\\ &\approx 1+\gamma\,\Delta A_{\mathsf{r}}(t)\end{aligned}$$

in (39) for

$$\begin{aligned}\tau_{\mathsf{r}}\frac{d}{dt}\Delta A_{\mathsf{r}}(t)+A_{\mathsf{r}2}+\Delta A_{\mathsf{r}}(t)=\\ [1+\gamma\,\Delta A_{\mathsf{r}}(t)][A_{\mathsf{r}2}+\Delta A_{\mathsf{r}}(t)]\end{aligned}\tag{43}$$

Ignoring the second-order term in ΔA_{r}, we obtain

$$\tau_{\mathsf{r}} \frac{d}{dt} \Delta A_{\mathsf{r}}(t) = \gamma\, A_{\mathsf{r2}}\, \Delta A_{\mathsf{r}}(t) \tag{44}$$

where $\gamma\, A_{\mathsf{r2}}$ is a dimensionless, negative valued quantity. Based on the above equation, and given that $A_{\mathsf{r}}(0) = A_{\mathsf{r1}}$, we obtain

$$A_{\mathsf{r}}(t) = A_{\mathsf{r2}} + (A_{\mathsf{r1}} - A_{\mathsf{r2}})\, e^{\gamma\, A_{\mathsf{r2}}\, \frac{t}{\tau_{\mathsf{r}}}} \tag{45}$$

F. Thermomechanical noise in the resonator model

The *thermomechanical noise* of the resonator can be modeled as a white noise source at the drive input $F(t)$ in (1), with a (two-sided) spectral density

$$\mathcal{S}_{\mathsf{thm}}(\omega) = 2\, m\, \frac{\omega_{\mathsf{r}}}{Q}\, k_{\mathsf{B}}\, T \tag{46}$$

where k_{B} is Boltzmann's constant, and T is temperature [24]. In order to incorporate this noise source into the model that was developed in Section II-A, we use a complex amplitude representation [25] as follows

$$\begin{aligned} N_{\mathsf{thm}}(t) &= \mathcal{R}\big\{ n_{\mathsf{thm}}(t)\, e^{j\omega_{\mathsf{o}} t} \big\} \\ &= \tfrac{1}{2} \big[n_{\mathsf{thm}}(t)\, e^{j\omega_{\mathsf{o}} t} + n^{*}_{\mathsf{thm}}(t)\, e^{-j\omega_{\mathsf{o}} t} \big] \end{aligned} \tag{47}$$

with

$$n_{\mathsf{thm}}(t) = n_{\mathsf{thm}\mathcal{R}}(t) + j\, n_{\mathsf{thm}\mathcal{I}}(t) \tag{48}$$

where the real and imaginary parts $n_{\mathsf{thm}\mathcal{R}}(t)$ and $n_{\mathsf{thm}\mathcal{I}}(t)$ are both white noise sources, uncorrelated with each other, and with twice the spectral density [25] in (46):

$$\mathcal{S}_{\mathsf{thm}\mathcal{R}}(\omega) = \mathcal{S}_{\mathsf{thm}\mathcal{I}}(\omega) = 4\, m\, \frac{\omega_{\mathsf{r}}}{Q}\, k_{\mathsf{B}}\, T \tag{49}$$

The resonator model with thermomechanical noise, obtained from (12), is then given by

$$\begin{aligned} \tau_{\mathsf{r}} \frac{d}{dt} s(t) + [1 + j\tau_{\mathsf{r}}(\omega_{\mathsf{o}} - \omega_{\mathsf{r}})] s(t) = \\ \frac{-j}{m} \frac{Q}{\omega_{\mathsf{r}}\, \omega_{\mathsf{o}}} [f(t) + n_{\mathsf{thm}}(t)] \end{aligned} \tag{50}$$

The noise source can also be incorporated into the phase domain model developed in Section II-C by modifying (22) (with the frequency shift $\Delta\omega_{\mathsf{e}}$ set to zero) as follows

$$\tau_{\mathsf{r}} \frac{d}{dt} \Big[e^{j\Delta\theta_{\mathsf{r}}(t)} \Big] + e^{j\Delta\theta_{\mathsf{r}}(t)} = 1 + \tfrac{1}{A_{\mathsf{f}}} n_{\mathsf{thm}}(t) \tag{51}$$

We note here that any phase shift (possibly time-varying) applied to $n_{\mathsf{thm}}(t)$ does not change the stochastic properties of its real and imaginary parts, *i.e.,* , $n_{\mathsf{thm}}(t)$ and $e^{j\theta(t)} n_{\mathsf{thm}}(t)$ are stochastically equivalent. Thus, we ignore any such phase shifts that are applied to $n_{\mathsf{thm}}(t)$. Subsequently, the noise term in the above dictates a noise term in (24) as follows

$$\begin{aligned} \tau_{\mathsf{r}} \frac{d}{dt} \Delta\theta_{\mathsf{r}}(t) + \Delta\theta_{\mathsf{r}}(t) &= \tfrac{1}{A_{\mathsf{f}}} n_{\mathsf{thm}\mathcal{I}}(t) \\ &= \underbrace{\tfrac{1}{A_{\mathsf{rss}}} \tfrac{Q}{m\, \omega_r^2}\, n_{\mathsf{thm}\mathcal{I}}(t)}_{\theta_{\mathsf{th}}(t)} \end{aligned} \tag{52}$$

where we used (20). Thus, the phase deviation $\Delta\theta_{\mathsf{r}}(t)$ due to thermomechanical noise of the resonator can be modeled with a white noise process $\theta_{\mathsf{th}}(t)$, at the input of the one-pole filter represented by the transfer function in (25). $\theta_{\mathsf{th}}(t)$ is given by

$$\theta_{\mathsf{th}}(t) = \frac{1}{A_{\mathsf{rss}}} \frac{Q}{m\, \omega_r^2}\, n_{\mathsf{thm}\mathcal{I}}(t) \tag{53}$$

G. Detection noise

The thermomechanical noise of the resonator is the most fundamental, unavoidable noise source. However, there are other noise sources that need to be taken into account, for instance, the ones that are introduced during the transduction of the mechanical displacement of the resonator into an optical and/or electrical signal and in the subsequent amplification in the electrical domain. Regardless of the actual physical source, we lump all of these noise sources into one, represent them as an additive, *detection* noise source at the output of the resonator (added to $s(t)$, at the input of any electrical amplifier), and define it as follows

$$n_{\mathsf{d}}(t) = \mathcal{K}_{\mathsf{d}}\, \frac{Q}{m\, \omega_{\mathsf{r}}^2}\, n_{\mathsf{thmD}}(t) = n_{\mathsf{d}\mathcal{R}}(t) + j\, n_{\mathsf{d}\mathcal{I}}(t) \tag{54}$$

representing a baseband equivalent, complex-valued white noise source, as in (48). In (54), the dimensionless factor $\mathcal{K}_{\mathsf{d}}$ determines whether the thermomechanical noise is resolved above the detection noise background ($\mathcal{K}_{\mathsf{d}} < 1$) or buried in it ($\mathcal{K}_{\mathsf{d}} > 1$), as implied by (50). The value of $\mathcal{K}_{\mathsf{d}}$ can be easily determined experimentally when $\mathcal{K}_{\mathsf{d}} < 1$, by simply measuring the spectrum of the transduced resonator displacement signal in the electrical domain in an open-loop configuration with no excitation on the resonator [11]. The (square root) of the ratio of the spectrum level of the (white) background to the level of the peak at the resonance frequency yields $\mathcal{K}_{\mathsf{d}}$ [11, Fig. 1(b)]. $n_{\mathsf{thmD}}(t)$ in (54) is a noise source that is uncorrelated with $n_{\mathsf{thm}}(t)$ but has the same stochastic characteristics, *i.e.,* spectral density.

The phase deviation $\theta_{\mathsf{d}}(t)$ of a driven resonator due to detection noise, in an open-loop configuration, can be modeled by adding detection noise to the resonator response as follows

$$\begin{aligned} A_{\mathsf{rss}} + n_{\mathsf{d}\mathcal{R}}(t) + j\, n_{\mathsf{d}\mathcal{I}}(t) =& [A_{\mathsf{rss}} + \Delta A_{\mathsf{r}}(t)]\, e^{j\theta_{\mathsf{d}}(t)} \\ \approx& A_{\mathsf{rss}} + \Delta A_{\mathsf{r}}(t) + j\, A_{\mathsf{rss}}\, \theta_{\mathsf{d}}(t) \end{aligned}$$

where we used $e^{j\theta_{\mathsf{d}}(t)} \approx 1 + \theta_{\mathsf{d}}(t)$ and ignored the second-order noise term $\Delta A_{\mathsf{r}}(t)\, \theta_{\mathsf{d}}(t)$. We identify

$$\theta_{\mathsf{d}}(t) = \frac{n_{\mathsf{d}\mathcal{I}}(t)}{A_{\mathsf{rss}}} = \frac{\mathcal{K}_{\mathsf{d}}}{A_{\mathsf{rss}}}\, \frac{Q}{m\, \omega_{\mathsf{r}}^2}\, n_{\mathsf{thmD}\mathcal{I}}(t) \tag{55}$$

The white noise process $\theta_{\mathsf{d}}(t)$ above, representing detection noise, is to be added to the *output* of the one-pole filter in (25), whereas $\theta_{\mathsf{th}}(t)$ in (53), representing thermomechanical noise, is introduced at its *input*.

H. Noise in the self-oscillating core

The thermomechanical noise of the resonator and detection noise are introduced into the model of the self-oscillating core in (30)

$$\begin{aligned} &\tau_{\mathsf{r}} \frac{d}{dt} s(t) + [1 + j\tau_{\mathsf{r}}(\omega_{\mathsf{o}} - \omega_{\mathsf{r}})] s(t) \\ &\quad = \frac{1}{m} \frac{Q}{\omega_{\mathsf{r}}\, \omega_{\mathsf{o}}} \Big[h_{\mathsf{D}}(|s(t)|)\, [s(t) + n_{\mathsf{d}}(t)] + n_{\mathsf{thm}}(t) \Big] \end{aligned} \tag{56}$$

as in (50). Above, we did not include $n_{\mathsf{d}}(t)$ in the argument of $h_{\mathsf{D}}(\cdot)$, ignoring the second-order, detection noise effect in determining the amplitude-dependent gain value for the

saturating amplifier. $n_{\mathsf{d}}(t)$ in (56) represents both the noise introduced during the transduction as well as the noise of the amplifier (in an input-referred manner). $n_{\mathsf{d}}(t)$ is amplified along with the signal component $s(t)$ and fed back to drive the resonator. As before, the phase shift or time delay introduced in the self-oscillating core does not change the stochastic properties of $n_{\mathsf{d}}(t)$.

In order to evaluate the impact of noise on the amplitude and phase of the signal generated by the self-oscillating core, we modify (38) as follows, with $\omega_{\mathsf{r}} = \omega_{\mathsf{r1}} = \omega_{\mathsf{r2}}$,

$$\begin{aligned} &\tau_{\mathsf{r}} \frac{d}{dt} A_{\mathsf{r}}(t) + \left[1 + j\tau_{\mathsf{r}} \frac{d}{dt}\theta_{\mathsf{r}}(t)\right] A_{\mathsf{r}}(t) \\ &\quad = \frac{1}{m}\frac{Q}{\omega_r^2}\left[h_{\mathsf{D}}(A_{\mathsf{r}}(t))\,[A_{\mathsf{r}}(t) + n_{\mathsf{d}}(t)] + n_{\mathsf{thm}}(t)\right] \end{aligned} \tag{57}$$

by ignoring any phase shifts that are applied to $n_{\mathsf{thm}}(t)$ and $n_{\mathsf{d}}(t)$ in deriving (57) from (56). Seperating the real/imaginary parts of the above equation results in

$$\begin{aligned} &\tau_{\mathsf{r}} A_{\mathsf{r}}(t)\, \frac{d}{dt}\theta_{\mathsf{r}}(t) = \\ &\quad \frac{1}{m}\frac{Q}{\omega_r^2}[h_{\mathsf{D}}(A_{\mathsf{r}}(t))\, n_{\mathsf{d}\mathcal{I}}(t) + n_{\mathsf{thm}\mathcal{I}}(t)] \\ &\tau_{\mathsf{r}} \frac{d}{dt} A_{\mathsf{r}}(t) + A_{\mathsf{r}}(t) = \\ &\quad \frac{1}{m}\frac{Q}{\omega_r^2}\left[h_{\mathsf{D}}(A_{\mathsf{r}}(t))\,[A_{\mathsf{r}}(t) + n_{\mathsf{d}\mathcal{R}}(t)] + n_{\mathsf{thm}\mathcal{R}}(t)\right] \end{aligned} \tag{58}$$

The above equations can be simplified, nothing that there is a one way coupling between them (from $A_{\mathsf{r}}(t)$ to $\theta_{\mathsf{r}}(t)$, but not in the reverse direction) and by expressing $\theta_{\mathsf{r}}(t) = \theta_{\mathsf{rss}} + \Delta\theta_{\mathsf{r}}(t)$ and $A_{\mathsf{r}}(t) = A_{\mathsf{rss}} + \Delta A_{\mathsf{r}}(t)$ with the steady-state, noiseless A_{rss} as the solution of

$$\frac{1}{m}\frac{Q}{\omega_r^2} h_{\mathsf{D}}(A_{\mathsf{rss}}) = 1 \tag{59}$$

θ_{rss} is an arbitrary but constant phase determined by the initial conditions. We proceed as in (42)-(44) to obtain

$$\begin{aligned} &\tau_{\mathsf{r}}[A_{\mathsf{rss}} + \Delta A_{\mathsf{r}}(t)]\, \frac{d}{dt}\Delta\theta_{\mathsf{r}}(t) = \\ &\quad \frac{1}{m}\frac{Q}{\omega_r^2}[h_{\mathsf{D}}(A_{\mathsf{rss}})\, n_{\mathsf{d}\mathcal{I}}(t) + \ n_{\mathsf{thm}\mathcal{I}}(t)] \\ &\tau_{\mathsf{r}} \frac{d}{dt}\Delta A_{\mathsf{r}}(t) - \gamma\, A_{\mathsf{rss}}\, \Delta A_{\mathsf{r}}(t) = \\ &\quad \frac{1}{m}\frac{Q}{\omega_r^2}[h_{\mathsf{D}}(A_{\mathsf{rss}})\, n_{\mathsf{d}\mathcal{R}}(t) + \ n_{\mathsf{thm}\mathcal{R}}(t)] \end{aligned} \tag{60}$$

The amplitude deviation $\Delta A_{\mathsf{r}}(t)$ and the phase deviation $\Delta\theta_{\mathsf{r}}(t)$ are due to noise. Due to amplitude stabilization in the self-oscillating core, $\Delta A_{\mathsf{r}}(t)$ stays small compared with A_{rss}. Based on the second equation above, $\Delta A_{\mathsf{r}}(t)$ is essentially the output of a one-pole (with time constant $\frac{-\tau_r}{\gamma A_{\mathsf{rss}}}$) low-pass filter, with the input as a white noise process. We thus approximate the first equation above using $A_{\mathsf{rss}} + \Delta A_{\mathsf{r}}(t) \approx A_{\mathsf{rss}}$, and obtain

$$\begin{aligned} \frac{d}{dt}\Delta\theta_{\mathsf{r}}(t) &= \frac{1}{\tau_{\mathsf{r}}}\frac{1}{A_{\mathsf{rss}}}\frac{Q}{m\,\omega_r^2}[h_{\mathsf{D}}(A_{\mathsf{rss}})\, n_{\mathsf{d}\mathcal{I}}(t) + \ n_{\mathsf{thm}\mathcal{I}}(t)] \\ &= \frac{1}{\tau_{\mathsf{r}}}\left[\frac{Q}{m\,\omega_r^2} h_{\mathsf{D}}(A_{\mathsf{rss}})\, \theta_{\mathsf{d}}(t) + \theta_{\mathsf{th}}(t)\right] \end{aligned} \tag{61}$$

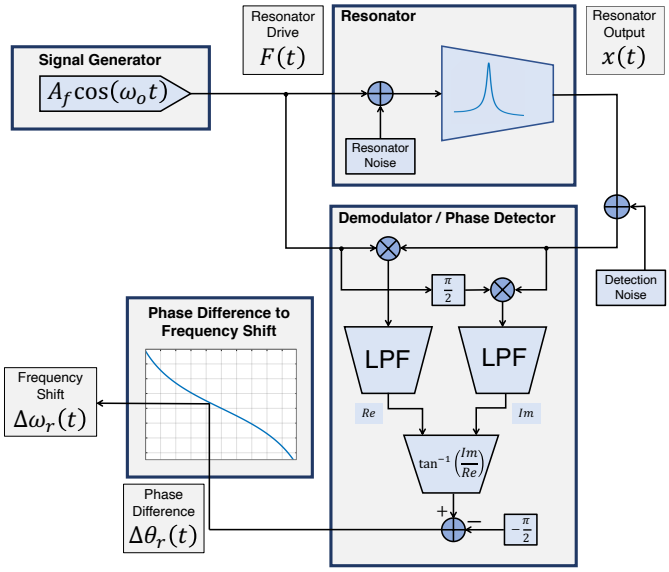


Fig. 4. Feedback-free (FF) scheme (figure based on [12, Fig. 2])

The above equation implies that the phase deviation $\Delta\theta_{\mathsf{r}}(t)$ is the (time) integral of a white noise process, thus in the form of a Brownian motion (random walk). The fact that the phase noise of autonomous oscillators has a Browinan motion nature is well established [23].

III. Resonator tracking schemes

A. Feedback-free (FF) resonator tracking

In the feedback-free scheme, shown in Fig. 4, the signal that drives the resonator has a fixed frequency. At steady-state, the drive signal and the resonator output are at the same (drive) frequency, but there is a (constant) phase difference between them, equal to $-\frac{\pi}{2}$ if the drive frequency is equal to the resonance frequency. Any resonance frequency shift is detected and inferred by continually monitoring this phase difference [11], and mapping it to a frequency shift based on the (inverse) of the phase response of the resonator given by

$$\theta_{\mathsf{r}}(t) = -\tfrac{\pi}{2} - \operatorname{atan}\left(\tau_{\mathsf{r}}\, \Delta\omega_{\mathsf{r}}(t)\right) \approx -\tfrac{\pi}{2} - \tau_{\mathsf{r}}\, \Delta\omega_{\mathsf{r}}(t) \tag{62}$$

As above, the phase response can be approximated as a linear function around the resonance frequency. Thus, the map from the phase difference to the frequency shift takes the following simple form

$$\Delta\omega_{\mathsf{r}}(t) = -\frac{\Delta\theta_{\mathsf{r}}(t)}{\tau_{\mathsf{r}}} \tag{63}$$

We recall that the phase/frequency shift response above exhibits a transient following a resonance frequency shift causing event, as in (26), arising from the inherent dynamics of the resonator. That is, the *full* impact of the event can not be instantaneously observed in the phase response of the resonator. This observation is in contrast with an apparent misconception in the nanomechanical sensors field that the phase/frequency response of a resonator has an instantaneous nature, whereas the amplitude response exhibits a transient dictated by its inherent dynamics. In fact, both the amplitude and phase exhibit inherent transient characteristics, as shown

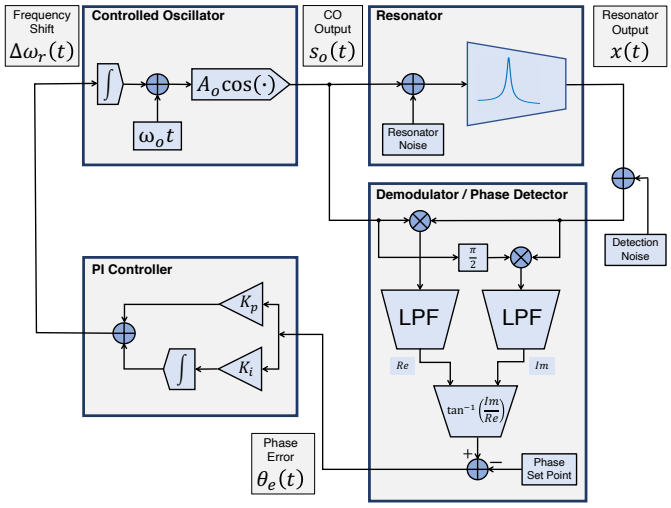


Fig. 5. Frequency-locked loop (FLL) (figure based on [12, Fig. 2])

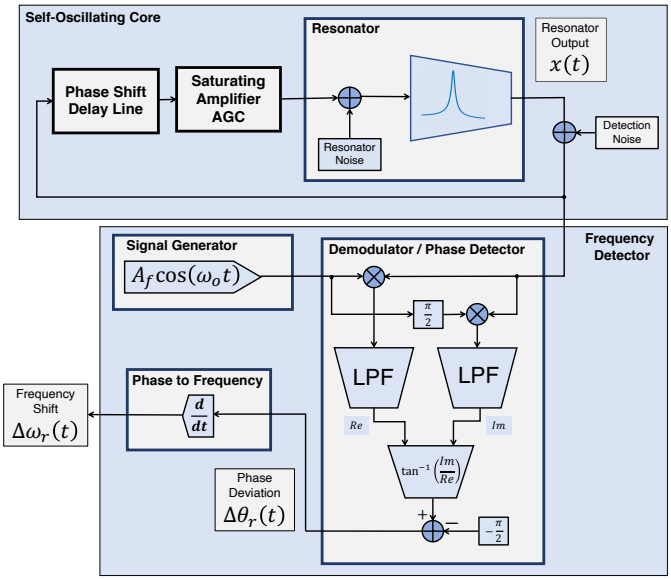


Fig. 6. Self-sustained oscillator (SSO) (figure based on [12, Fig. 2])

in (19) and (26). Eqn. (26) was verified experimentally with the measurement results reported in [11, Apdx A]. We will delve deeper into this issue later in our treatment, when we compare the feedback-free approach with the other closed-loop schemes.

The phase difference between two narrowband (nearly sinusoidal) signals can be extracted using a narrowband IQ demodulator as shown in Fig. 4. Alternatively, one can use a Hilbert transform based scheme [25], [26]. Since the two schemes are equivalent for narrowband signals, the IQ demodulator is preferred due to its less complex nature when compared with the Hilbert transform.

The IQ demodulator uses low-pass filters (LPFs) in order to filter out the frequency components, that are at twice the resonance frequency, produced by the multipliers. The bandwidths of these filters need to be chosen to be less than (twice) the resonance frequency. This imposes a further limitation on the bandwidth (speed) of the feedback-free resonator tracking scheme, in addition to the inherent low-pass filtering represented by (25) due to the resonator dynamics. For high quality factor resonators, the bandwidth limitation imposed by the resonator characteristics is the limiting factor.

The bandwidth of the LPFs in the demodulator and inherent resonator dynamics not only influence the transient response of the feedback-free tracking scheme to frequency shifts caused by events of interest, but also determine the noise performance and hence the sensitivity together with the sources of noise.

B. Frequency-locked loop (FLL) based resonator tracking

In the closed-loop FLL scheme, shown in Fig. 5, the frequency of the drive signal is continually updated with a negative-feedback loop in such a way so that the resonator is always operating at its resonance. This is accomplished by feeding the phase difference (minus the phase set point $-\frac{\pi}{2}$) between the drive signal and the resonator response as an error signal into a PI controller. The output of the controller is used to set the frequency of the signal generator, *i.e.,* a controlled oscillator (CO). Thus, any resonance frequency shift will result in a nonzero phase error, and a subsequent update in the controller output, which can be directly used to continually monitor the frequency of the CO and hence the resonance frequency. In the FLL scheme, there is no need for an explicit map from phase error to frequency shift. This is inherently accomplished by the FLL dynamics, without relying on information regarding the phase response characteristics of the resonator.

The phase error signal in the FLL scheme is produced in exactly the same manner as the phase difference in the feedback-free aproach. Thus, one may conclude that the speed of the FLL response to events of interest will be also limited by the inherent resonator dynamics. However, the phase difference signals are used in a different manner in the two schemes. In the feedback-free scheme, the phase difference is directly mapped to the frequency shift, whereas in the FLL, it is further processed by the controller and also shaped by the closed-loop dynamics. The response characteristics of the FLL may be tailored so that it is much faster, in fact, not limited by the resonator dynamics. We present a precise, quantitative analysis of the FLL later in our treatment. Qualitatively, the controller in the FLL can be designed so that the negative-feedback either speeds up or, if necessary, slows down the FLL with respect to the inherent resonator response, albeit with implications for noise filtering bandwidth. This allows one to trade speed for accuracy or vice versa.

C. Self-sustained oscillator (SSO) based resonator tracking

One embodiment of the self-sustained oscillator based resonator tracking scheme is shown in Fig. 6, composed of a self-oscillating core based on positive-feedback and a separate frequency detector. In Fig. 6, the frequency detector is realized using an IQ demodulator that includes a signal generator at a fixed frequency, which basically measures the total phase difference between the nearly sinusoidal signal generated by the self-oscillating core, called the oscillator output from now

on, and the sinusoidal signal output of the signal generator. If the oscillator frequency is exactly equal to the signal generator frequency, then the phase difference is constant. If the two frequencies are different, then the phase difference is a linear function of time with the slope as the frequency difference. Thus, in order to compute the frequency difference, the phase difference is differentiated *w.r.t.* to time as in Fig. 6. We note that the use of the demodulator in Fig. 6 is quite different when compared with its use, in Figs. 4 and 5, in the feedback-free and the FLL schemes, where the signal generator that drives the resonator also drives one of the inputs of the demodulator. On the other hand, in the frequency detector in Fig. 6, one demodulator input is set to the self-generated oscillator output, and the other one is a fixed frequency signal unrelated to the resonator drive or response. A resonance frequency shift in the feedback-free scheme results in a (steady-state) *time-invariant phase difference*, whereas for the self-sustained oscillator scheme, the resulting *phase difference changes linearly with time*. As a result, the mapping of the phase difference to a frequency shift is done via (63) in the feedback-free scheme, whereas with a time derivative in the frequency detector in Fig. 6. There seems to be a confusion regarding this issue in the literature. For instance, authors in [26] seem to be using a feedback-free scheme, but employ a time derivative in order to map the phase information into a frequency shift.

In Section II-E, we derived the (resonance) frequency step response of the self-oscillating core, and obtained (40), reproduced here for convenience:

$$\theta_{\mathsf{r}}(t) = (\omega_{\mathsf{r}2} - \omega_{\mathsf{r}1})\, t + \theta_{\mathsf{r}1} = \Delta\omega_{\mathsf{e}}\, t + \theta_{\mathsf{r}1}$$

The above phase response was derived for a sudden resonance frequency shift of size $\Delta\omega_{\mathsf{e}}$ that occurs at $t = 0$. Thus, the time-derivative of the phase deviation above produces an *instantaneous* frequency shift from 0 to $\Delta\omega_{\mathsf{e}}$ at $t = 0$. The fact that a self-oscillating core produces an instantaneous response to a resonance frequency shift was stated and exploited in inventing the FM-AFM scheme in [6], however, a theoretical derivation was not provided. The theory and analysis we have presented in Section II-E provides the rigorous theoretical foundation for this observation. Our theory in Section II, covering all three resonator tracking schemes in a unified manner, clarifies the misconceptions and confusions in the field regarding the transient and instantaneous nature of amplitude/phase/frequency step responses of resonators in various configurations. We note that the phase step response in all of the schemes is a *continuous function of time*, does not exhibit any unphysical jumps. However, the *continuous phase response* maps to a *continuous frequency step response* in the feedback-free case, whereas to a *discontinuous frequency step response* in the self-oscillating scheme. Even though the inherent frequency step response of a self-oscillating core is indeed instantaneous, any practical frequency detection scheme will come with a bandwidth and speed limitation in measuring this instantaneous frequency response [6], [27]. For instance, the frequency detection scheme in Fig. 6 based on an IQ demodulator uses LPFs with a bandwidth less than (twice) the resonance frequency. Thus, the frequency shift output of the frequency detector will be smoothed with a time-constant

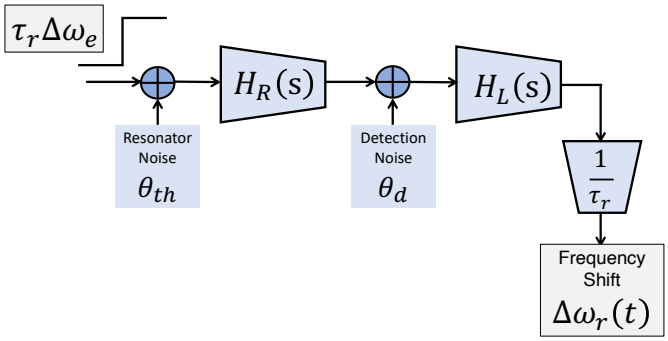


Fig. 7. Phase-domain model for FF scheme

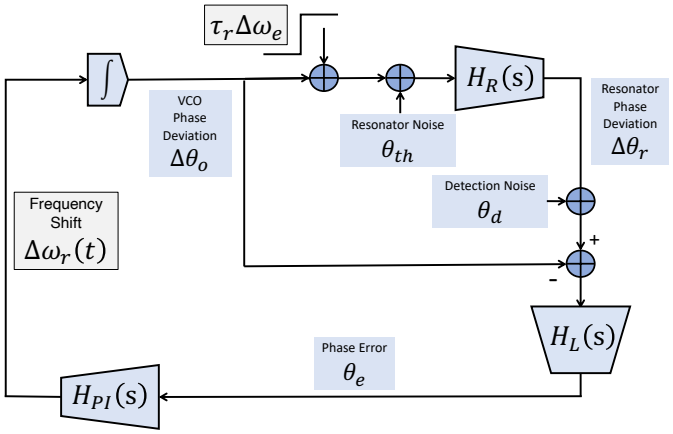


Fig. 8. Phase-domain model for FLL (figure based on [12, Fig. 3])

that is dictated by the bandwidths of the LPFs, even though the inherent frequency step response of the self-oscillating core is sharp. Other frequency detection schemes also have bandwidth limitations. For instance, a PLL which locks the phase and frequency of a signal generator to the oscillator output has a certain loop bandwidth that dictates its response speed.

IV. Comparative Characterization of Resonator Tracking Schemes

We next perform a comparative analysis of the three resonator tracking schemes under consideration, using the theory developed in Section II. There are two performance criteria we consider, simply put, *speed* and *accuracy*. We quantify speed in terms of how quickly the system responds to resonance frequency jumps, *i.e.,* with *Frequency Step Response* (FSTR). Accuracy, or sensitivity, is assessed based on various characterizations of the frequency fluctuations due to noise in the frequency shift output of the system in tracking mode.

In order to simplify performance characterization and enable comparative analysis, we distill and capture the results of the theory developed in Section II by creating simplified phase/frequency domain models for all of the tracking schemes, shown as block diagrams in Figs. 7, 8 and 9. In these diagrams, $\Delta\omega_{\mathsf{e}}$ represents a resonance frequency step change due to an event of interest. All functions of s represent Laplace domain transfer functions. $H_{\mathsf{R}}(\mathsf{s})$, given in (25), models the inherent resonator dynamics. $H_{\mathsf{L}}(\mathsf{s})$ represents

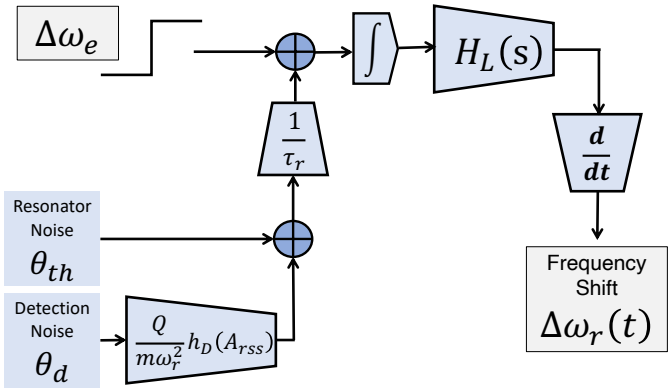


Fig. 9. Phase-domain model for SSO

the low-pass filtering (with a bandwidth less then (twice) the resonance frequency and a unity gain for DC inputs) in the IQ demodulator. $H_{\mathsf{PI}}(s)$ is for the controller

$$H_{\mathsf{PI}}(\mathsf{s}) = K_{\mathsf{p}} + \frac{K_{\mathsf{i}}}{\mathsf{s}}. \tag{64}$$

We consider a *proportional-integral* (PI) controller in this work, as it is very commonly used. Our treatment in the paper, however, can be applied to other controllers in a straightforward manner. The (white) noise sources representing the thermomechanical noise of the resonator, θ_{th}, and the detection noise, θ_{d}, are characterized as given in (53) and (55).

A. Frequency step response: Speed

Frequency step response can be easily characterized by computing the transfer function from the frequency step input $\Delta\omega_{\mathsf{e}}$ to the frequency shift output $\Delta\omega_{\mathsf{r}}$, using the block diagrams in Figs. 7, 8 and 9:

$$\begin{aligned} H^{\mathsf{FF}}_{\mathsf{FST}}(\mathsf{s}) &= H_{\mathsf{R}}(\mathsf{s})\, H_{\mathsf{L}}(\mathsf{s}) \\ H^{\mathsf{FLL}}_{\mathsf{FST}}(\mathsf{s}) &= \frac{(\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})}{\mathsf{s}^2 + \frac{\mathsf{s}}{\tau_{\mathsf{r}}} + (\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})} \\ H^{\mathsf{SSO}}_{\mathsf{FST}}(\mathsf{s}) &= H_{\mathsf{L}}(\mathsf{s}) \end{aligned} \tag{65}$$

The step response is then simply given by

$$f_{\mathsf{str}}(t) = \mathcal{L}^{-1}\left\{\frac{H_{\mathsf{FST}}(\mathsf{s})}{\mathsf{s}}\right\} \tag{66}$$

where $\mathcal{L}^{-1}\{\bullet\}$ denotes the inverse Laplace transform, with $f_{\mathsf{str}}(0) = 0$ and $f_{\mathsf{str}}(t \to \infty) = 1$ since $H_{\mathsf{FST}}(\mathsf{s} \to 0) = 1$.

In (65), $H^{\mathsf{FF}}_{\mathsf{FST}}(\mathsf{s})$ and $H^{\mathsf{SSO}}_{\mathsf{FST}}(\mathsf{s})$ are simple, but $H^{\mathsf{FLL}}_{\mathsf{FST}}(\mathsf{s})$ needs some clarification. $H^{\mathsf{FLL}}_{\mathsf{FST}}(\mathsf{s})$ was obtained by writing and solving an equation [12] based on the block diagram in Fig. 8. We choose K_{p} and K_{i}, as suggested in [12], [13],

$$K_{\mathsf{p}} = \omega_{\mathsf{FLL}} \quad , \quad K_{\mathsf{i}} = \frac{\omega_{\mathsf{FLL}}}{\tau_{\mathsf{r}}} \tag{67}$$

where ω_{FLL} is the desired FLL loop bandwidth. The above choice yields [12], [13]

$$H^{\mathsf{FLL}}_{\mathsf{FST}}(\mathsf{s}) = \frac{H_{\mathsf{L}}(\mathsf{s})}{H_{\mathsf{L}}(\mathsf{s}) + \frac{\mathsf{s}}{\omega_{\mathsf{FLL}}}} \tag{68}$$

The simplified transfer function above, with a low-pass characteristics, is independent of, and not limited by, the resonator dynamics, as a direct consequence of the choice in (67). Thus, while the open-loop FF approach suffers from a bandwidth limitation imposed by the inherent resonator dynamics, both the FLL and SSO schemes do not. This is enabled by feedback (negative-feedback for FLL, and positive-feedback for SSO), which in a sense speeds up the response of the system, albeit at the expense of worse noise performance, as we discuss later. In the case of FLL, this happens through a pole-zero cancellation [12], [13] in the transfer function.

For a detailed characterization of speed, one can simply evaluate (65) and (66) and compute the settling time of FSTR. Roughly, we can conclude that the response time constant of both the FLL and the SSO schemes can be as small as $\frac{1}{\omega_{\mathsf{r}}}$ limited by the required low-pass filtering in the demodulator/phase detector, whereas for the FF scheme, it is at least $\frac{Q}{\omega_{\mathsf{r}}}$ limited by the inherent resonator dynamics. One can, of course, for all schemes, make the response slower in order to obtain better steady-state tracking noise performance, *i.e.,* for more accuracy, as we discuss below.

B. Spectral density of frequency fluctuations: Accuracy

In order to characterize the fluctuations due to noise in the frequency shift output of the system, we need to compute transfer functions from the two noise sources θ_{th} and θ_{d} to the output $\Delta\omega_{\mathsf{r}}$. This can be done easily using the block diagrams in Figs. 7, 8 and 9:

$$\begin{aligned} H^{\mathsf{FF}}_{\theta_{\mathsf{th}}}(\mathsf{s}) &= \frac{1}{\tau_{\mathsf{r}}}\, H_{\mathsf{R}}(\mathsf{s})\, H_{\mathsf{L}}(\mathsf{s}) \\ H^{\mathsf{FF}}_{\theta_{\mathsf{d}}}(\mathsf{s}) &= \frac{1}{\tau_{\mathsf{r}}}\, H_{\mathsf{L}}(\mathsf{s}) \\ H^{\mathsf{FLL}}_{\theta_{\mathsf{th}}}(\mathsf{s}) &= \frac{1}{\tau_{\mathsf{r}}} \frac{(\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})}{\mathsf{s}^2 + \frac{\mathsf{s}}{\tau_{\mathsf{r}}} + (\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})} \\ H^{\mathsf{FLL}}_{\theta_{\mathsf{d}}}(\mathsf{s}) &= \frac{1}{\tau_{\mathsf{r}}} \frac{1}{H_{\mathsf{R}}(\mathsf{s})} \frac{(\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})}{\mathsf{s}^2 + \frac{\mathsf{s}}{\tau_{\mathsf{r}}} + (\mathsf{s}K_{\mathsf{p}} + K_{\mathsf{i}})H_{\mathsf{L}}(\mathsf{s})} \\ H^{\mathsf{SSO}}_{\theta_{\mathsf{th}}}(\mathsf{s}) &= \frac{1}{\tau_{\mathsf{r}}}\, H_{\mathsf{L}}(\mathsf{s}) \\ H^{\mathsf{SSO}}_{\theta_{\mathsf{d}}}(\mathsf{s}) &= \frac{Q}{m\,\omega_r^2} h_{\mathsf{D}}(A_{\mathsf{rss}})\, \frac{1}{\tau_{\mathsf{r}}}\, H_{\mathsf{L}}(\mathsf{s}) \end{aligned} \tag{69}$$

The transfer functions above are used in computing the spectral density of the fractional frequency fluctuations, *i.e.,* $y_{\mathsf{r}} = \frac{\Delta\omega_t}{\omega_{\mathsf{r}}}$, as follows

$$\mathcal{S}_{y_{\mathsf{r}}}(\omega) = \frac{|H_{\theta_{\mathsf{th}}}(j\omega)|^2\, \mathcal{S}_{\theta_{\mathsf{th}}}(\omega) + |H_{\theta_{\mathsf{d}}}(j\omega)|^2\, \mathcal{S}_{\theta_{\mathsf{d}}}(\omega)}{\omega_{\mathsf{r}}^2} \tag{70}$$

where $\mathcal{S}_{\theta_{\mathsf{th}}}(j\omega)$ and $\mathcal{S}_{\theta_{\mathsf{d}}}(j\omega)$ are the (white) spectral densities of θ_{th} and θ_{d}, given by

$$\begin{aligned} \mathcal{S}_{\theta_{\mathsf{th}}}(\omega) &= \left[\frac{1}{A_{\mathsf{rss}}} \frac{Q}{m\,\omega_r^2}\right]^2 4\,m\,\frac{\omega_{\mathsf{r}}}{Q}\,k_{\mathsf{B}}\,T = \frac{4\,Q\,k_{\mathsf{B}}\,T}{m\,\omega_{\mathsf{r}}^3\,A_{\mathsf{rss}}^2} \\ \mathcal{S}_{\theta_{\mathsf{d}}}(\omega) &= \mathcal{K}_{\mathsf{d}}^2\, \mathcal{S}_{\theta_{\mathsf{th}}}(\omega) \end{aligned} \tag{71}$$

The transfer functions in (69) all have low-pass characteristics. The spectral density of fractional frequency fluctuations, $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$, is thus in the form of low-pass filtered white noise. The value of the spectral density at zero (low) frequency, $\mathcal{S}_{y_{\mathsf{r}}}(\omega = 0)$, is of interest, since it determines the accuracy

of the resonator tracking scheme at steady-state, after the step response has settled following a resonance frequency shift due to an event of interest. $\mathcal{S}_{y_\mathsf{r}}(\omega=0)$ is a measure of accuracy/sensitivity that is independent of speed/bandwidth, as opposed to, for instance, mean-squared fluctuations that can be computed by integrating $\mathcal{S}_{y_\mathsf{r}}(\omega)$ over the system bandwidth. While the bandwidth/speed of a flexible resonator tracking scheme may be adjusted for speed versus accuracy trade-off, $\mathcal{S}_{y_\mathsf{r}}(\omega=0)$ quantifies a fundamental sensitivity limitation. We evaluate $\mathcal{S}_{y_\mathsf{r}}(0)$ using (69), (70) and (71) to obtain

$$\begin{aligned}
\mathcal{S}_{y_\mathsf{r}}^{\mathsf{FF}}(0) &= \frac{k_\mathsf{B}\,T}{m\,Q\,\omega_\mathsf{r}^3\,A_\mathsf{rss}^2}\left(1+\mathcal{K}_\mathsf{d}^2\right)\\
\mathcal{S}_{y_\mathsf{r}}^{\mathsf{FLL}}(0) &= \frac{k_\mathsf{B}\,T}{m\,Q\,\omega_\mathsf{r}^3\,A_\mathsf{rss}^2}\left(1+\mathcal{K}_\mathsf{d}^2\right)\\
\mathcal{S}_{y_\mathsf{r}}^{\mathsf{SSO}}(0) &= \frac{k_\mathsf{B}\,T}{m\,Q\,\omega_\mathsf{r}^3\,A_\mathsf{rss}^2}\left[1+\left(\mathcal{K}_\mathsf{d}\,\frac{Q}{m\,\omega_r^2}h_\mathsf{D}(A_\mathsf{rss})\right)^2\right]
\end{aligned} \tag{72}$$

We observe that the FF and FLL schemes have the same performance, whereas for SSO, the performance is degraded due to the amplification of detection noise through the amplifier in the positive feedback loop as well as the resonator itself. The FF and FLL schemes do not suffer from amplified detection noise, since the resonator is directly driven with a clean signal generator instead of an amplified version of the signal transduced from its own output. If thermomechanical noise is very well resolved, that is, if detection noise is negligible in comparison, with $\mathcal{K}_\mathsf{d}\ll 1$, then all three schemes have the same performance as characterized by (72).

C. Speed versus Accuracy

There is a direct trade-off between accuracy and speed in resonator tracking systems. Any mechanism that speeds up the response is bound to be more susceptible to noise. Conversely, more noise filtering for improved accuracy, in some form or another, will result in a slower response. The accuracy versus speed trade-off characteristics of resonator tracking schemes are not always well understood and articulated in the literature. We next compare the three schemes under consideration with respect to their speed-accuracy trade-off characteristics. We first consider the case where thermomechanical noise is well resolved, *i.e.,* detection noise is negligible in comparison. We then consider detection noise.

For high-Q resonators, the response speed of the FF scheme is determined by the inherent resonator dynamics, as discussed in Section IV-A. The response time constant is equal to the resonator time constant $\frac{2Q}{\omega_\mathsf{r}}$. In accordance, the thermomechanical noise of the resonator is filtered by the resonator transfer function, with a bandwidth of $\frac{\omega_\mathsf{r}}{2Q}$ as in (69). As a result, when Q is increased, thermomechanical noise bandwidth decreases in proportion, and at the same time, so does the noise spectral density within that bandwidth, as given in (72). Thus, one obtains a *quadratic reduction* in the noise power that is passed through. However, this comes at the expense of increased response time that is proportional to Q. Thus, in the FF scheme, increasing Q results in a quadratic improvement in accuracy at the expense of a linear degradation in speed, assuming that thermomechanical noise is dominant. The bandwidth of the low-pass filter $H_\mathsf{L}(s)$ in the IQ demodulator is typically larger than $\frac{\omega_\mathsf{r}}{2Q}$, with a maximum value of ω_r. One may set the bandwidth of $H_\mathsf{L}(s)$ to a value smaller than $\frac{\omega_\mathsf{r}}{2Q}$. This would result in further filtering for the thermomechanical noise of the resonator, as in (69), in addition to the one provided by the inherent resonator dynamics, but at the same time further slowing down of the response.

The FLL scheme features a flexible speed-accuracy trade-off characteristics. The flexibility is brought about by feedback and the adjustable controller parameters. The transfer function $H_{\mathsf{FST}}^{\mathsf{FLL}}(\mathsf{s})$ in (65) that dictates the frequency step response and $H_{\theta_\mathsf{th}}^{\mathsf{FLL}}(\mathsf{s})$ in (69) that filters thermomechanical noise are the same. The bandwidth and characteristics of this filter that has a low-pass characteristics may be adjusted via the controller parameters K_p and K_i, which were chosen in (67) so that the prescribed bandwidth is independent of the resonator characteristics, in particular, its Q. As a result, when Q is increased, the FLL scheme does not suffer from a speed degradation. At the same time, thermomechanical noise spectral density decreases since it is inversely proportional to Q. When compared with the FF scheme, the flexibility of the FLL scheme enables one to give up the quadratic improvement in accuracy for only a linear improvement but with no degradation in speed. One can also design the controller in such a way so that one achieves better than linear improvement in accuracy with some degradation in speed. In fact, one may choose to endure worse than linear degradation in speed for a better than quadratic improvement in accuracy. In this case, the bandwidth of the FLL would be designed to be even smaller than $\frac{\omega_\mathsf{r}}{2Q}$ that is imposed by the inherent resonator characteristics. The bandwidth adjustment for an FLL can be done in a manner that is independent of the resonator characteristics.

For the SSO scheme, the transfer function $H_{\mathsf{FST}}^{\mathsf{SSO}}(\mathsf{s})$ in (65) that dictates the frequency step response and $H_{\theta_\mathsf{th}}^{\mathsf{SSO}}(\mathsf{s})$ in (69) that filters thermomechanical noise are the same, and determined by the low-pass filter $H_\mathsf{L}(s)$ in the IQ demodulator (frequency detector). The SSO scheme also features a flexible speed-accuracy characteristics. The bandwidth of $H_\mathsf{L}(s)$ may be adjusted, with a maximum value of ω_r, with the possibility of being smaller than $\frac{\omega_\mathsf{r}}{2Q}$. While speed is proportional to this bandwidth, accuracy is inversely proportional. Thus, the speed-accuracy trade-off characteristics of the SSO scheme is similar to the one offered by the FLL, in the sense that bandwidth can be adjusted, independent of the resonator characteristics such as Q. This was the main motivation in the invention of the FM-AFM technique in [6], which is an SSO based scheme. However, the FLL scheme is more versatile. The loop transfer function that determines the frequency step response and thermomechanical noise filtering, given in (68), can be chosen to have a bandwidth that is less than the bandwidth of $H_\mathsf{L}(s)$ in the demodulator.

Detection noise in the FF scheme is simply filtered by the low-pass filter $H_\mathsf{L}(s)$ in the IQ demodulator. In the case of FLL, nature of detection noise filtering is dependent on controller parameters. With controller parameters as in (67), detection noise transfer function of an FLL takes the form

$$H_{\theta_\mathsf{d}}^{\mathsf{FLL}}(\mathsf{s}) = \frac{1+\mathsf{s}\,\tau_\mathsf{r}}{\tau_\mathsf{r}}\,\frac{H_\mathsf{L}(\mathsf{s})}{H_\mathsf{L}(\mathsf{s})+\frac{\mathsf{s}}{\omega_{\mathsf{FLL}}}} \tag{73}$$

When $\omega_{\mathsf{FLL}}\,\tau_{\mathsf{r}} \gg 1$, *i.e.,* the loop bandwidth is larger than the resonator linewidth [11], this transfer function satisfies

$$\begin{aligned} H^{\mathsf{FLL}}_{\theta_{\mathsf{d}}}(0) &= \frac{1}{\tau_{\mathsf{r}}} \\ H^{\mathsf{FLL}}_{\theta_{\mathsf{d}}}(j\omega) &\approx \omega_{\mathsf{FLL}}\, H_{\mathsf{L}}(j\omega) \quad \text{for} \quad \omega > \omega_{\mathsf{FLL}} \end{aligned} \tag{74}$$

Thus, the above implies that FLL loop dynamics in fact amplifies (with respect to the FF scheme, see (69)) detection noise at frequencies above the loop bandwidth ω_{FLL}, with an amplification factor of $\omega_{\mathsf{FLL}}\,\tau_{\mathsf{r}} = 2Q\,\frac{\omega_{\mathsf{FLL}}}{\omega_{\mathsf{r}}} = \frac{\omega_{\mathsf{FLL}}}{\Gamma}$, where $\Gamma = \frac{\omega_{\mathsf{r}}}{2Q}$ is the (one-sided) linewidth of the resonator. The ramifications of this detection noise amplification predicted by our theory was experimentally observed in the measurements reported in [11, Fig. 5] for a closed-loop FLL scheme. On the other hand, when the loop bandwidth ω_{FLL} is chosen to be *smaller* than the resonator linewidth Γ, detection noise is in fact *attenuated* (with respect to the FF scheme). This prediction of our theory was also experimentally verified in [11, Fig. 5].

The detection noise amplification in the FLL scheme can be alleviated to a degree by reducing the bandwidth of $H_{\mathsf{L}}(s)$ down to ω_{FLL}. Furthermore, noise amplification above the loop bandwidth has minimal impact on the accuracy performance of the FLL scheme as we show later. On the other hand, the SSO scheme suffers from detection noise amplification all the way down to zero frequency, in the amount $\frac{Q}{m\,\omega_r^2}\,h_{\mathsf{D}}(A_{\mathsf{rss}})$. This will result in an accuracy performance degradation that can not be circumvented even when one is willing to trade-off speed performance for it.

D. Figure-of-Merit: Mean-Squared Error

Following our somewhat qualitative treatment of speed versus accuracy trade-offs, we next perform a quantitative evaluation of the three sensor schemes and compare them in detail. As a figure-of-merit that quantifies speed and accuracy in a unified manner, we use *Root Mean Squared Error* (RMSE). RMSE combines *frequency step response* (FSTR) for speed, with *Allan Deviation* (AD) for accuracy. FSTR was defined in Section IV-A.

AD [28]–[32] is the standard measure of frequency stability [33], established in the frequency standards community but also widely used for resonant sensors. *Allan Variance* can be computed with

$$\sigma_y^2(\tau) = \frac{4}{\pi\tau^2}\int_{-\infty}^{+\infty} \frac{\left[\sin\left(\frac{\omega\,\tau}{2}\right)\right]^4}{\omega^2}\,\mathcal{S}_{y_{\mathsf{r}}}(\omega)\,d\omega \tag{75}$$

where $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$ is the spectrum of fractional frequency fluctuations given in (70), and τ is the *averaging time* for the fluctuations. AD $\sigma_y(\tau)$ is uniquely determined by $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$, as dictated by (75). For all resonator tracking schemes, $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$ has the form of low-pass filtered white noise (Lorentzian spectrum), assuming detection noise is negligible. In this case, AD first increases as a function of averaging time τ, reaches its peak value σ_y^{max} at around the response time constant (inverse of the bandwidth of $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$), and decreases as τ is increased further [12]. For τ larger than the response time constant, $\sigma_y^2(\tau)$ is determined by $\mathcal{S}_{y_{\mathsf{r}}}(0)$, simply given by

$$\sigma_y^2(\tau) = \frac{\mathcal{S}_{y_{\mathsf{r}}}(0)}{\tau} \tag{76}$$

The above expression is valid for averaging time τ that is *larger* than the time-constant of the (low-pass) noise filtering mechanisms that reduce $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$ as compared with $\mathcal{S}_{y_{\mathsf{r}}}(0)$ for ω larger than the filtering bandwidth. Since this time-constant (inverse of filtering bandwidth) also determines the response time of the resonant sensor to resonance frequency jump events, the relevant time scales of practical interest are in fact represented by (76). This justifies our use of $\mathcal{S}_{y_{\mathsf{r}}}(0)$ in Section IV-B as an ultimate measure of accuracy that is independent of speed. If τ is increased even further, $\sigma_y(\tau)$ starts increasing again due to thermal drift and other slow nonideal drift phenomena [11]. The behavior of $\sigma_y(\tau)$ for the three time scales is illustrated in Fig. 10 (upper right graph). For these illustrations, we assumed that the detection noise is negligible, and thermomechanical noise is dominant.

The case when detection noise is not negligible (while thermomechanical noise is still well resolved above the detection noise floor) and the detection bandwidth (of the filter $H_{\mathsf{L}}(\mathsf{s})$) is larger than the resonator linewidth was experimentally investigated in [11]. In this scenario, $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$, the spectrum of fractional frequency fluctuations in (70), exhibits low-pass filtered white noise behavior at two different time scales for the FF scheme. That is, $\mathcal{S}_{y_{\mathsf{r}}}(\omega)$ is the sum of two Lorentzian spectra, one for thermomechanical noise with a corner frequency equal to the resonator linewidth, and the other for detection noise with a corner frequency equal to the detection bandwidth. In this case, $\sigma_y(\tau)$ exhibits two peaks and a valley between them, followed by an eventual increase due to drift. The first peak (at smaller τ) occurs at the detection filter time constant, whereas the second peak occurs at the resonator response time. Behavior in $\sigma_y(\tau)$ as such for the FF scheme predicted by our theory was shown in [11] to agree very well with experimental results. The depth of the valley between the two peaks is determined by the detection noise level with respect to thermomechanical noise [11].

The problem with AD is that it misleadingly suggests better accuracy performance for τ smaller than the response time of the resonant sensor. In fact, it is not meaningful to use AD in this region for accuracy characterization. AD is appropriate for characterizing random fluctuations in frequency, as well as deterministic frequency drifts, but not for deterministic frequency *offsets*. In fact, AD simply evaluates to zero when there is a deterministic frequency offset, *i.e.,* a bias error. Thus, the bias error needs to be characterized separately. This can be readily done via FSTR $f_{\mathsf{str}}(\tau)$, defined by (66), normalized so that $f_{\mathsf{str}}(0) = 0$ and $f_{\mathsf{str}}(\tau \to \infty) = 1$. The frequency estimate provided by any resonant sensor scheme has a bias

$$\textbf{bias}(\tau) = \Delta y\,[1 - f_{\mathsf{str}}(\tau)] \tag{77}$$

in addition to the random fluctuations characterized by AD, where Δy is the root cause amount of fractional frequency

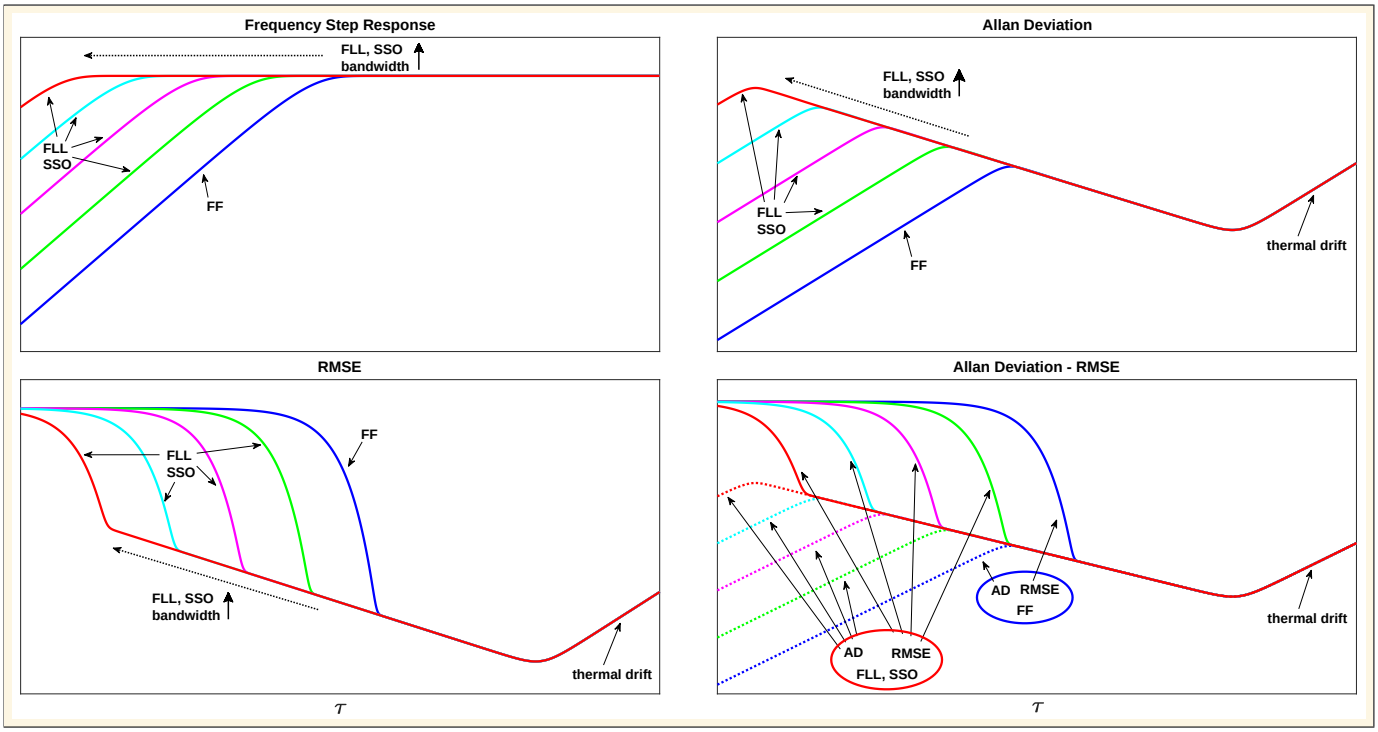


Fig. 10. Frequency Step Response, Allan Deviation and Root Mean Squared Error

shift. *Root Mean Squared Error* (RMSE) which combines the two is defined as follows

$$\mathsf{RMSE}(\tau) = \sqrt{\sigma_y^2(\tau) + \mathbf{bias}^2(\tau)} \tag{78}$$

For τ less then the response time, RMSE is dominated by the bias term. As τ is increased, $f_{\mathsf{str}}(\tau)$ increases, and hence $[1 - f_{\mathsf{str}}(\tau)]$ decreases. The bias becomes small and then negligible when τ exceeds the response time, where RMSE becomes equal to $\sigma_y(\tau)$.

Fig. 10 graphically illustrates the behaviors of FSTR, AD and RMSE, for the FF, FLL and SSO schemes. We recall that, for these illustrations, we assumed that the impact of detection noise is negligible as compared with the thermomechanical noise of the resonator, and hence the thermomechanical noise is well resolved above the detection noise floor. FLL and SSO have the same performance as derived before. Fig. 10 illustrates and compares the performance of FF (with bandwidth fixed at the resonator linewidth) with FLL and SSO as the bandwidth for these schemes is increased. As seen in these illustrations, using AD alone for performance characterization misleadingly suggests that the FF scheme has better performance for small values of τ. However, RMSE correctly indicates better performance for FLL and SSO schemes for time scales below the response time of FF. On the other hand, all schemes have the same performance at long time scales beyond the response time of FF.

We note that another figure-of-merit called FSTDEV (Frequency STep DEViation) that also combines FSTR with AD was proposed in [12]. Upon further and careful evaluation, we concluded that FSTDEV has some deficiencies in appropriately characterizing performance for resonant sensor schemes with varying bandwidths. On the other hand, RMSE is a universal, well established error measure for estimators that takes into account both random deviations due to noise and biases, *i.e.,* deterministic offsets.

V. Conclusions

We have presented an in-depth theory and comparative analyses for the main resonant sensor schemes that are currently in use, and unraveled their speed versus accuracy trade-off characteristics. We have developed accurate but simple and analytically tractable models, and attained a deep understanding of various issues in designing and optimizing resonant sensors. The predictions based on the theory presented in this paper for various scenarios as system parameters are varied, for both the FF and FLL schemes, were shown to agree very well with experimental results [11]. Even though we have focused on nanomechanical resonant sensors in this paper, our treatment and the proposed techniques are generally applicable to all kinds of resonant sensors in other domains, *e.g.,* optics or microwaves, where the sensing scheme is based on tracking the resonance frequency of a resonator.

VI. Acknowledgments

The author would like to thank Selim Hanay, Matt Matheny, Selim Olcum, Pedram Sadeghi, Silvan Schmid, Guillermo Villanueva and the attendees of NMC 2019 in Lausanne for many helpful discussions on nanomechanical resonant sensors.

References

[1] J. Chaste, A. Eichler, J. Moser, G. Ceballos, R. Rurali, 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. Hiebert, 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. Grütter, 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. D. Adams, B. W. Erickson, J. Grossenbacher, J. Brugger, A. Nievergelt, 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] P. Sadeghi, A. Demir, L. G. Villanueva, H. Kähler, and S. Schmid, "Frequency fluctuations in nanomechanical silicon nitride string resonators," *arXiv:2009.12331 preprint*, September 2020.

[12] 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.

[13] S. Olcum, N. Cermak, S. C. Wasserman, and S. R. Manalis, "High-speed multiple-mode mass-sensing resolves dynamic nanoscale mass distributions," *Nature Communications*, vol. 6, p. 7070, 2015.

[14] S. K. Roy, V. T. K. Sauer, J. N. Westwood-Bachman, A. Venkatasubramanian, and W. K. Hiebert, "Improving mechanical sensor performance through larger damping," *Science*, vol. 360, no. 6394, 2018.

[15] U. Dürig, H. R. Steinauer, and N. Blanc, "Dynamic force microscopy by means of the phase-controlled oscillator method," *Journal of Applied Physics*, vol. 82, no. 8, pp. 3641–3651, 1997.

[16] X. Feng, C. White, A. Hajimiri, and M. L. Roukes, "A self-sustaining ultrahigh-frequency nanoelectromechanical oscillator," *Nature Nanotechnology*, vol. 3, no. 6, pp. 342–346, 2008.

[17] E. Kenig, M. C. Cross, L. G. Villanueva, R. B. Karabalin, M. H. Matheny, R. Lifshitz, and M. L. Roukes, "Optimal operating points of oscillators using nonlinear resonators," *Physical Review E*, vol. 86, no. 5, p. 056207, 2012.

[18] L. G. Villanueva, E. Kenig, R. B. Karabalin, M. H. Matheny, R. Lifshitz, M. C. Cross, and M. L. Roukes, "Surpassing fundamental limits of oscillators using nonlinear resonators," *Physical Review Letters*, vol. 110, no. 17, p. 177208, 2013.

[19] E. Kenig, M. C. Cross, J. Moehlis, and K. Wiesenfeld, "Phase noise of oscillators with unsaturated amplifiers," *Physical Review E*, vol. 88, no. 6, p. 062922, 2013.

[20] A. Demir and M. S. Hanay, "Numerical analysis of multi-domain systems: Coupled nonlinear PDEs & DAEs with noise," *IEEE Transactions on Computer-Aided Design of Integrated Circuits & Systems*, vol. 37, no. 7, pp. 1445–1458, July 2018.

[21] B. Hauer, C. Doolin, K. Beach, and J. Davis, "A general procedure for thermomechanical calibration of nano/micro-mechanical resonators," *Annals of Physics*, vol. 339, pp. 181–207, 2013.

[22] W. E. Vander Velde, *Multiple-input describing functions and nonlinear system design*. McGraw-Hill, New York, 1968.

[23] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterisation," *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655–674, May 2000.

[24] K. L. Ekinci, Y. T. Yang, and M. L. Roukes, "Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems," *Journal of Applied Physics*, vol. 95, no. 5, pp. 2682–2689, 2004.

[25] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*. Springer Science & Business Media, 1999.

[26] R. Giridharagopal, G. E. Rayermann, G. Shao, D. T. Moore, O. G. Reid, A. F. Tillack, D. J. Masiello, and D. S. Ginger, "Submicrosecond time resolution atomic force microscopy for probing nanoscale dynamics," *Nano Letters*, vol. 12, no. 2, pp. 893–898, 2012.

[27] A. Mascaro, Y. Miyahara, T. Enright, O. E. Dagdeviren, and P. Grütter, "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.

[28] D. W. Allan, "Statistics of atomic frequency standards," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 221–230, Feb 1966.

[29] V. F. Kroupa, *Frequency stability: Fundamentals and measurement*. IEEE, 1983.

[30] E. Rubiola, *Phase noise and frequency stability in oscillators*. Cambridge University Press, 2009.

[31] A. Makdissi, F. Vernotte, and E. D. Clercq, "Stability variances: a filter approach," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 5, pp. 1011–1028, May 2010.

[32] Wikipedia, "Allan Variance – Wikipedia, The Free Encyclopedia," 2019.

[33] "IEEE standard definitions of physical quantities for fundamental frequency and time metrology—random instabilities," *IEEE Std 1139-2008*, Feb 2009.



Alper Demir received the B.S. degree from Bilkent University, Ankara, Turkey, in 1991, and the M.S. and Ph.D. degrees from the University of California, Berkeley, CA, USA, in 1994 and 1997. He was with Motorola, Austin, TX, USA, Cadence Design Systems, San Jose, CA, USA, Bell Laboratories, Murray Hill, NJ, USA, MIT, Cambridge, MA, USA, and University of California at Berkeley. He has been a Faculty Member with Koç University, Istanbul, Turkey, since 2002. Dr. Demir was a recipient of the 2002 Best of ICCAD Award, the 2003/2014 IEEE/ACM William J. McCalla ICCAD best paper awards, and the 2004 IEEE Guillemin-Cauer Award. He was named an IEEE Fellow in 2012.