CHALMERS
UNIVERSITY OF TECHNOLOGY

Autoregressive Stochastic Clock Jitter Compensation in Analog-to-Digital Converters

Downloaded from: https://research.chalmers.se, 2026-11-16 19:46 UTC

Citation for the original published paper (version of record):
Gerosa, D., Hou, R., Björk, V. et al (2026). Autoregressive Stochastic Clock Jitter Compensation in Analog-to-Digital Converters. IEEE Transactions on Signal Processing, 74: 1818-1829.
http://dx.doi.org/10.1109/TSP.2026.3686550

N.B. When citing this work, cite the original published paper.

© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

Autoregressive Stochastic Clock Jitter Compensation in Analog-to-Digital Converters

Daniele Gerosa, Rui Hou, Vimar Björk, Ulf Gustavsson, and Thomas Eriksson, *Member, IEEE*

***Abstract*—This paper addresses the mathematical modeling and compensation of stochastic discrete-time clock jitter in analog-to-digital converters (ADCs). We model the stochastic clock jitter as a first-order autoregressive (AR(1)) process, and we propose two novel, computationally efficient, pilot-assisted dejittering algorithms for baseband signals: one based on solving a sequence of weighted least-squares problems, and another that exploits the correlated jitter structure via a Kalman filter-based routine. We also propose a conditional maximum-likelihood estimator for the autoregressive parameters, enabling near-optimal Kalman-filter performance even when such parameters vary over time. We further provide a mathematical analysis of the induced linearization errors, and we complement the theory with synthetic simulations to evaluate the proposed techniques across different scenarios. The proposed techniques are shown to yield a $1-15$ dB improvement in signal-to-noise-and-distortion ratio (SINADR) and $0.02-1.6$ dB in symbol error vector magnitude (EVM), depending on impairment severity and pilot density. The Kalman smoother generally provides superior performance by leveraging additional temporal information.**

***Index Terms*—Analog-to-digital converters, stochastic jitter, autoregressive process, weighted least-squares, Kalman smoother, linearization.**

I. Introduction

ANALOG-TO-DIGITAL converters are hardware devices that convert analog signals into digital signals. They are never ideal, and they typically introduce various imperfections into the sampled signals. These imperfections arise from non-ideal hardware and may appear as direct-current (DC) voltage offset bias, quantization noise, aliasing, nonlinear distortion due to clipping, and irregularities in the sampling instants (jitter) [41]. This paper focuses on the latter imperfection, which arises from random phase fluctuations in the sampling clock generated by the clock source and its conditioning/distribution circuitry.

From a theoretical perspective, sampling instants that deviate from the ideal multiples of the sampling interval do not necessarily preclude the reconstruction of bandlimited analog signals, provided that the average sampling rate is large enough. This is established by various extensions of the classical interpolation results by Shannon, Nyquist, Whittaker and Kotelnikov [23]. The caveat is that the timing deviations must be known. In practical applications, these timing deviations are not available a priori and must therefore be inferred from measurements. Considerable effort has been devoted to the modeling, estimation and compensation of sampling-clock jitter in ADCs. Several works use out-of-band pilot (or training) tones injection, as in [37], [33], [40], [14], where the baseband payload is typically filtered to isolate the pilot tone. The jitter realizations are estimated from this tone and then used to compensate (i.e., dejitter) the payload. A time-domain technique that time-multiplexes pilot samples to unknown data samples is described in [38] in the context of time-interleaved (ti) ADCs. Other statistical signal processing methods proposed in the literature aim to directly estimate the payload without first necessarily estimating the jitter realizations. Minimum Mean Squared Error (MMSE) estimators are proposed in [42], [31], [39], [17], under different payload models.

In parallel with algorithmic methods, the literature has also devoted substantial attention to how sampling clock jitter should be modeled statistically. While static (i.e., constant in time) jitter is mostly relevant in the field of ti ADCs [9], stochastic jitter is often modeled as white noise [42], [31], [39], [17], [22], [2] or as colored process with a pronounced $1/f^2$ spectral decay component [40], depending on whether the jitter is caused by an internal or an external clock reference [44]. Clocks with external references can also add an additional low-offset $1/f^3$ region in their spectrum, which is often neglected for simplicity (e.g., by retaining only the $1/f^2$ component) [44], [8]. These spectral features are commonly observed in oscillators [44].

To the best of our knowledge, autoregressive (AR) processes were first employed in [6] to model sampling jitter in sigma-delta modulators. Wiener-process models are also common in phase-noise studies [12], [28]. In this work, we model sampling-clock jitter as a first-order autoregressive process (AR(1)), which is mathematically tractable and can reproduce the relevant spectral characteristics with an appropriate choice of parameters. More details are provided in Section II.

Received 6 August 2025; revised 13 February 2026 and 30 March 2026; accepted 17 April 2026. Date of publication 22 April 2026; date of current version 21 May 2026. The associate editor coordinating the review of this article and approving it for publication was Jiafeng Xie. *(Corresponding author: Daniele Gerosa.)*

Daniele Gerosa and Thomas Eriksson are with the Communication, Antennas and Optical Networks, Electrical Engineering Department, Chalmers University of Technology, 41296 Göteborg, Sweden (e-mail: daniele.gerosa@chalmers.se; thomase@chalmers.se).

Rui Hou is with Ericsson AB, 16483 Stockholm, Sweden.

Vimar Björk is with Ericsson AB, 41756 Göteborg, Sweden.

Ulf Gustavsson is with Ericsson AB, 41756 Göteborg, Sweden, and also with the Communication, Antennas and Optical Networks, Electrical Engineering Department, Chalmers University of Technology, 41296 Göteborg, Sweden.

Digital Object Identifier 10.1109/TSP.2026.3686550

© 2026 The Authors. This work is licensed under a Creative Commons AttributionNonCommercialNoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

We exclude effects such as nonlinearities, phase noise, and imperfect bandlimiting. Such effects will also degrade the signal quality. However, most of them are relatively orthogonal to the jitter, and can be treated separately using independent techniques[1]. In the present work, we assume that the residual distortion from these other impairments can be modeled as additive and incorporated into the thermal noise term (see e.g. [24]).

The main theoretical contributions of this paper are summarized as follows:

- ➢ Following [6], we model the stochastic clock jitter as an autoregressive process of order 1 with model parameter close to 1.
- ➢ We propose and compare two novel jitter tracking and compensation algorithms based on pilot samples. The first uses Kalman filtering and smoothing to exploit the jitter statistical structure. The second uses an optimally weighted least-squares approach and is more "model-agnostic" (Section III). Depending on the scenario, both techniques are shown to yield a $1-15$ dB improvement in Signal-to-Noise-and-Distortion Ratio (SINADR) and $0.02-1.6$ dB in symbol error vector magnitude (EVM) over the uncompensated case, assuming sufficiently high symbol or sample pilot density.
- ➢ We propose a Maximum Likelihood Estimation technique for autoregressive parameter estimation within the Kalman smoother routine.

Furthermore, the analysis also includes:

- ➢ rigorous justifications for the approximations arising in Taylor expansions and an outline of the regime in which these approximations are valid (Section II-C);
- ➢ synthetic simulations to evaluate the techniques presented (Section IV) and to compare them with existing methods.

A. Notation and Symbols

ADCs act on continuous-time signals by discretizing and digitizing them, so that the output of an ADC is a discrete sequence of values $\{x(t_n)\}_{n\in\mathbb{N}}$ when the input is the analog signal $x(t)$. If $x(t)$ is a stochastic process, we can model the ADC output as a sequence of random variables. We will denote sequences (finite or infinite) by bold lowercase letters $\mathbf{x} := (x_1, x_2, \dots)$. If $\mathbf{x}$ and $\mathbf{y}$ are two such sequences, $\mathbf{x} \odot \mathbf{y}$ indicates their Hadamard componentwise product. Bold uppercase letters $\mathbf{A}, \mathbf{B}, \dots$ will represent matrices; the symbol "$\approx$" stands for "approximately equal to" while the symbol "$\ll$" means "much less than". We will use $\widehat{\cdot}$ for estimates of quantities, and $\widetilde{\cdot}$ for noisy measurements of quantities.

If X is a complex-valued random variable defined on the probability space (Ω, Σ, P), then $\mathbb{E}[X]$ denotes its expected value, $\mathbb{E}[X] = \int_\Omega \mathfrak{Re}[X]\, dP + i \int_\Omega \mathfrak{Im}[X]\, dP$ and $\mathrm{var}(X)$ denotes its variance, $\sigma_X^2 = \mathrm{var}(X) = \mathbb{E}[|X - \mathbb{E}[X]|^2]$.

For a set E, χ_E indicates its characteristic function and $|E|$ its cardinality.

Derivatives: By $x'_n = x'(nT_s)$ we mean the first-order time derivative of the realization of the process $x(t)$ evaluated at times nT_s; for a bandlimited Gaussian process this operation is well-defined [3] and is equivalent to the n-th sample of the discrete derivative operator D applied to the (infinite) sequence of samples, i.e.

$$\frac{d}{dt}x(t)\big|_{t=nT_s} = x'(nT_s) = (D\mathbf{x})_n.$$

This derivative operator D acts as time domain convolution

$$(D\mathbf{x})_n = \frac{1}{T_s}\sum_k h_k x_{n-k}$$

with the ideal non-causal bandlimited derivative filter given by $h_k = (-1)^k/k$ for $k \neq 0$ and $h_0 = 0$; its frequency response is $H_D(\Omega) = i\Omega$. These expressions are commonly found and used in the relevant literature, cf. [34], [9], [25].

Oftentimes throughout the text, we will encounter continuous-time processes y (Wiener, Ornstein-Uhlenbeck, white noise) whose derivative may not exist in the standard pointwise sense. However, upon bandlimiting and sampling, their discrete $D\mathbf{y}$ counterparts are well-defined. To keep the notation as light as possible, we will write, with a slight abuse of notation, y'_n in lieu of $(D\mathbf{y})_n$.

II. Signal and System Model

We consider baseband analog signals that are modeled as continuous-time bandlimited complex-valued stationary Gaussian processes $x(t, \omega)$ so that, for each time instant $t \in T \subseteq \mathbb{R}$, the random variable $(\omega \mapsto x(t, \omega))$ is Gaussian with 0 mean and variance σ_x^2. We also assume that x has flat power spectral density $\mathcal{S}_x(f) = \sigma_x^2/(2W)\chi_{\{|f|\leq W\}}(f)$. This standard model, commonly used to describe signals in communication applications (cf. [29]), possesses analytically well-behaved sample paths: by Theorem 11 in [3] the maps $t \mapsto x(\omega, t)$ are indeed holomorphic for almost every $\omega \in \Omega$. Therefore the "n-th derivatives objects" $x^{(n)}$ are well-defined for all $n \in \mathbb{N}$.

A. Stochastic Jitter Description and Problem Statement

An ideal ADC samples signals at uniformly spaced instants. This means that $t_{n+1} - t_n$ is constant for all $n \geq 0$. In real-world applications, however, imperfections in the local oscillator (LO) circuitry introduce jitter, causing the actual sampling instants to deviate from the ideal in a non-uniform way: i.e., the actual sampling takes places at times $\tilde{t}_n = nT_s + \xi_n$ where T_s is the ADC sampling interval and ξ_n is the time-dependent (stochastic) jitter. The literature presents different models for the (discrete) process $\{\xi_n\}_{n\geq 1}$: in [31] $\xi_n \sim \mathcal{U}([-a, a])$ and independent while in [42], [39], [2] the ξ_n are modeled as i.i.d. Gaussian with 0 mean and variance σ^2. [40] describes ξ_n as a "slowly varying" Gaussian process. To the best of our knowledge, only [6] models the time domain clock jitter as an autoregressive process, while the Wiener process is more broadly used e.g. in [28], [12], [20]. Several sources characterize the jitter process in the frequency domain by deriving formulas for its power spectral density. These expressions turn out to coincide with or be very close to the power spectral

[1]As an example, this is the case in many Ericsson products.

densities of AR(1) processes; compare, for example, [6, eq. 9] with (2), and [8, eq. 4] with (4).

Therefore, in the present work we model the jitter process as a discrete autoregressive process of order 1 AR(1) with evolution relation given by

$$\xi_n = \varphi \xi_{n-1} + \epsilon_n, \tag{1}$$

where $\varphi \approx 1$ yet $\varphi < 1$, $\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\xi_0 \sim \mathcal{N}(0, \sigma_\epsilon^2/(1-\varphi^2))$ (stationary initialization). It can be shown [19] that the power spectral density of this process is

$$\mathcal{S}_\xi(e^{i\omega}) = \frac{\sigma_\epsilon^2}{1+\varphi^2-2\varphi\cos(\omega)},\ \omega \in [-\pi, \pi], \tag{2}$$

from which it is straightforward to observe that much of the process power is concentrated near $\omega = 0$ (or 0 Hz): indeed

$$\mathcal{S}_\xi(e^{i\omega}) = \frac{\sigma_\epsilon^2}{(1-\varphi)^2 + 2\varphi(1-\cos(\omega))} \approx \frac{\sigma_\epsilon^2}{\omega^2} \tag{3}$$

where the last approximation in (3) holds as long as $\varphi \approx 1$, $\omega \approx 0$ but $|\omega| > |1-\varphi|$. Moreover notice that $\mathcal{S}_\xi(e^{i0}) = \sigma_\epsilon^2/(1-\varphi)^2$. These observations show that the discrete process (1) can accurately model the jitter spectral decay as experimentally measured [44]. In addition, we remark that the Ornstein-Uhlenbeck process, which can be seen as the continuous counterpart of (1), has the Lorentzian profile

$$\mathcal{S}_{\mathrm{OU}}(\omega) = \frac{\sigma^2}{\gamma^2 + \omega^2} \tag{4}$$

as spectral density, from which it is again visible the $1/\omega^2$ decay, now for all ω "large enough".

If the jitter is not too large, with respect to the sampling interval, its effect on the discretized signal is often [40], [39], [31], [10] approximated via a first-order Taylor expansion:

$$x(nT_s + \xi_n) \approx x(nT_s) + \xi_n x'(nT_s). \tag{5}$$

As we will show in Proposition II.1, the higher order terms in the Taylor expansion can be considered in some sense negligible.

From here on, we will also set

$$y_n := x(nT_s) + \xi_n x'(nT_s) + w_n \tag{6}$$

as our model under investigation, with w_n a sequence of independent zero-mean Gaussian random variables with variance σ_w^2 modeling additive white noise in measurements. It has been observed [40] that ξ_n may be weakly correlated with the w_n, but in this work, we will consider the two processes as independent.

The problem we address in this work is to estimate the process ξ_n given the measurements y_n $\forall n$ together with pilot samples x_n for $n \in I \subseteq \mathbb{N}$ and, as a byproduct, to derive a dejittered estimation of x_n.

B. How Much Jitter is "Too Much" Jitter?

One way to quantify how much jitter can be tolerated by the system is by setting a target SINADR level. Since the disturbance is modeled as additive (cf. (5)), we can write the SINADR for our signal and jitter model (excluding quantization noise)

$$\begin{aligned}\mathrm{SINADR}_{\mathrm{dB}} &= 10\log_{10}\left[\frac{\mathrm{var}(x_n)}{\mathrm{var}(w_n)+\mathrm{var}(\xi_n x_n')}\right] \\ &= 10\log_{10}\left[\frac{\mathrm{var}(x_n)}{\mathrm{var}(w_n)+\mathrm{var}(\xi_n)\mathrm{var}(x_n')}\right] \\ &= 10\log_{10}\left[\frac{3(1-\varphi^2)\sigma_x^2}{3(1-\varphi^2)\sigma_w^2 + 4\pi^2 W^2 \sigma_\epsilon^2 \sigma_x^2}\right],\end{aligned} \tag{7}$$

where we used $\sigma_\xi^2 = \sigma_\epsilon^2/(1-\varphi^2)$, $\mathbb{E}[\xi_n] = \mathbb{E}[x_n'] = 0$ and the fact that ξ_n and x_n' are independent. By solving eq. (7) for σ_ϵ^2 it is easy to see how much jitter could be tolerated (for a desired SINADR level). At the same time, for the first-order Taylor expansion (5) to hold, the jitter ξ_n cannot be too large; therefore one reasonable assumption is to make the "small jitter" hypothesis (SJH), in line with the literature surveyed so far,

$$\sigma_\xi / T_s \ll 1. \tag{SJH}$$

We will refer to the left-hand side of the latter as *jitter percentage*. Notice that when $\sigma_w = 0$, (7) reduces to

$$-20\log_{10}\left[2\pi W \sigma_\xi/\sqrt{3}\right] \approx -20\log_{10}\left[2\pi W \sigma_\xi\right] + 4.77\mathrm{dB},$$

which is the standard jitter SNR formula found in classical literature (see e.g. [35]), applied to bandlimited signals.

The interaction between the two additive distortions in (6) will also play a role in our analysis. For convenience we introduce a Noise-to-Distortion Ratio (NDR), defined as the ratio between the measurement white noise power and the jitter-induced distortion power. Although NDR is not a standard performance metric in the ADC literature, it provides a concise, dimensionless parameter that captures the relative strength of the two additive impairments in our model and helps organize the numerical results. In particular, we set

$$\begin{aligned}\mathrm{NDR}_{\mathrm{dB}} &:= 10\log_{10}\left(\frac{\mathrm{var}(w_n)}{\mathrm{var}(\xi_n x_n')}\right) \\ &= 10\log_{10}\left[\frac{3\sigma_w^2(1-\varphi^2)}{4\pi^2 W^2 \sigma_x^2 \sigma_\epsilon^2}\right].\end{aligned} \tag{8}$$

Since the observation model (6) contains two additive impairments, their relative strength affects the behavior of the proposed estimators. For this reason, we will report and interpret our experimental results also as a function of $\mathrm{NDR}_{\mathrm{dB}}$. Note that when $\mathrm{NDR}_{\mathrm{dB}} = 0$, the two additive terms in (6) have the same variance.

C. Approximation Errors Analysis

There are two main approximations used in our analysis, whose errors need to be rigorously analyzed. In the first-order expansion in (5) under (SJH) the neglected higher-order terms are, in some sense, negligible, as the following Proposition clarifies:

Proposition II.1: Let $\mathbf{x} = \{x_n\}_{n\in\mathbb{N}}$ be a discrete zero-mean, wide-sense stationary (WSS) bandlimited Gaussian process, and $\boldsymbol{\xi} = \{\xi_n\}_{n\in\mathbb{N}}$ a discrete autoregressive process of order 1

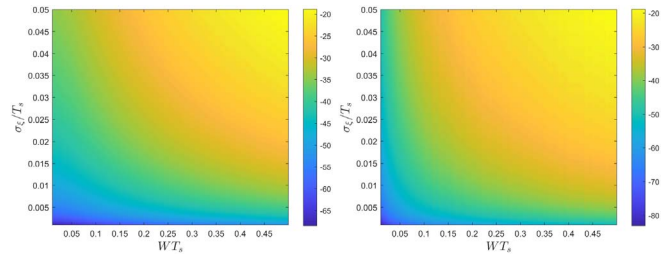


Fig. 1. Level curves of the term multiplying $\sigma^2_{x'}$ in (12); $\varphi = 0.95$ left, $\varphi = 0.999$ right, in dB. The double integral in (12) was numerically evaluated via nested `integral` in MATLAB and its accuracy was double-checked via a separate Monte Carlo where instances of the processes $\mathbf{x}$, $\boldsymbol{\xi}$ etc. were generated and the left-hand side of (12) estimated.

with variance $\sigma^2_{\xi_n} \equiv \sigma^2_\xi$, independent of $\mathbf{x}$. Assume that (SJH) holds; then we have

$$\operatorname{var}(\xi_n^2 x_n'') \ll \operatorname{var}(\xi_n x_n') \quad \forall\, n. \tag{9}$$

The second source of error comes from the derivative of y in (6), that will replace the unknown x_n' in our algorithms:

$$y_n' = x_n' + u_n' + w_n', \tag{10}$$

where we set $\mathbf{u} = \boldsymbol{\xi} \odot D\mathbf{x}$. The term w_n' is violet noise bandlimited to $1/2T_s$, whose variance equals $\pi^2\sigma_w^2/3T_s^2$. For a fixed NDR level r in dB, the white noise variance satisfies $\sigma_w^2 = 10^{r/10}\operatorname{var}(\xi x')$. Therefore

$$\sigma^2_{w'} = \frac{\pi^2 10^{r/10}\sigma_\xi^2}{3T_s^2}\sigma^2_{x'} \approx 3.29\frac{\sigma_\xi^2}{T_s^2}10^{r/10}\sigma^2_{x'} \tag{11}$$

and thus $\sigma^2_{w'} \ll \sigma^2_{x'}$ provided that (SJH) holds and the white noise is not too strong. The term $D(\boldsymbol{\xi} \odot \mathbf{x}')_n$ has a more convoluted structure, but its impact can be explicitly assessed, as shown in Proposition II.2, the proof of which can be found in the appendix. The power relation between u_n' and x_n' can be evaluated explicitly by means of equation (12). Fig. 1 illustrates this ratio for a few values of φ, σ_ξ/T_s and WT_s.

Proposition II.2: Let $\mathbf{x} = \{x_n\}_{n\in\mathbb{N}}$ be a discrete zero-mean, wide-sense stationary bandlimited Gaussian process (sampled every T_s seconds), with derivative $\mathbf{x}' = \{x_n'\}_{n\in\mathbb{N}}$, and $\boldsymbol{\xi} = \{\xi_n\}_{n\in\mathbb{N}}$ a discrete autoregressive process of order 1 with parameter $\varphi \in (0,1)$ and variance $\sigma^2_{\xi_n} \equiv \sigma^2_\xi$, independent of $\mathbf{x}$. Let D denote the ideal time-domain derivative operator. Then

$$\operatorname{var}(D(\boldsymbol{\xi} \odot \mathbf{x}')_n) = \frac{\sigma_\xi^2}{T_s^2}\frac{3(1-\varphi^2)}{32\pi^4(WT_s)^3} I \sigma^2_{x'} \tag{12}$$

where

$$I = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\omega^2((\omega-\nu)_{2\pi})^2}{1 - 2\varphi\cos(\nu) + \varphi^2} \chi_{\{|(\omega-\nu)_{2\pi}| \le 2\pi WT_s\}}\, d\omega\, d\nu$$

is a Poisson kernel-type double integral.

Remark II.1: A stochastic version of the Peano remainder theorem for Taylor expansions holds, cf. Lemma 3 in [43]. Further results comparing (conditional) variances of the objects $x(nT_s + \xi_n, \omega)$ and $x(nT_s, \omega) + x'(nT_s, \omega)\xi_n$ would require some care. First of all, the measurability/randomness of the maps $\omega \mapsto x(nT_s + \xi_n(\omega), \omega)$ or $(\omega_1, \omega_2) \mapsto x(nT_s + \xi_n(\omega_2), \omega_1)$ must be carefully assessed, depending on whether the jitter process is assumed to live in the same probability space of the process x or not. A standard sufficient condition for the measurability of $x(t,\omega)$ on the whole product space $\mathbb{R}\times\Omega$ is the left- or right-continuity in t to hold *for all* $\omega \in \Omega$ (see Lemma 6.4.6 in [5]), which might not be the case for our signal model. In that case, one should replace x with an indistinguishable copy obtained by redefining it on a P-null set so that $t \mapsto x(t,\omega)$ is left-/right-continuous for every ω.

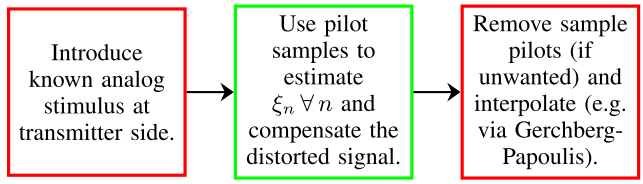


Fig. 2. Schematic three-step process.

III. Pilot-Sample-Based Jitter Compensation

This section contains the main contributions of this paper. We present two jitter tracking techniques able to mitigate the negative impact of this disturbance. Both algorithms assume knowledge of pilot samples at known time index positions, meaning that the true sample values at $n \in I \subsetneq \mathbb{N}$ will be assumed to be known a priori. The output stream of samples will be jittered, including the samples at pilot positions. A noisy estimate of the jitter process sample values can be obtained by rearranging equations (5) and (6) at pilot positions:

$$\xi_n \approx \widetilde{\xi}_n = (y_n - x_n)/x_n', \quad n \in I. \tag{13}$$

Once the jitter process is estimated using one of the methods detailed in this section, the jitter distortion in (6) must be subtracted, the pilots removed, and the signal interpolated using e.g. the Gerchberg-Papoulis algorithm [26], [13] to "fill the gaps" left after the removal of pilot samples. Depending on the density of sample pilots, the ADC sampling frequency must be high enough to ensure the uniqueness of the bandlimited reconstruction. The workflow is summarized by the scheme in Fig. 2; the focus of this paper is within the green box.

In contrast to pilot tone-based techniques [40], no knowledge of the signal's derivative is assumed by our techniques, i.e. x_n' is *not* assumed to be known a priori at pilot positions or elsewhere.

Moreover, the knowledge of sparse signal's sample values x_n at pilot positions is insufficient to compute its derivative at those positions. In our algorithms, we will use y_n' as a proxy instead, as justified by the analysis in Section II-C. For illustration, Fig. 3 shows a tracking example: despite the large amount of missing data, both methods recover the overall behavior of the underlying autoregressive process.

A. Robust Blockwise Polynomial Interpolation

In this section, the true latent jitter process is treated as blockwise deterministic, yet unknown, and slowly varying. Under the model (5), noisy estimates of the jitter at pilot positions can be

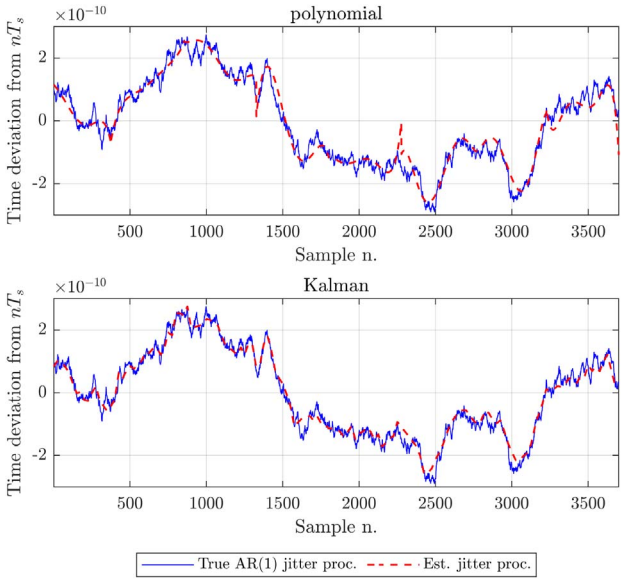


Fig. 3. Jitter trackers comparison example, Algorithm 1 (left) vs Algorithm 2 (right).

obtained from (13) as $\widetilde{\xi}_n = \xi_n + w_n/x'_n$. The additive noise is conditionally Gaussian but heteroscedastic because its variance is time-dependent; moreover, the unknown values x'_n are to be replaced with their estimations y'_n. We set $v_n = w_n/y'_n$. The idea is to perform an optimal short-time (i.e., blockwise) polynomial interpolation of the jitter process at the observation points $\widetilde{\xi}_n$, $n \in I$ for a given I.

Let N be the total number of samples observed. Since we want the pilot samples to be uniformly spaced, we fix a value $K \in \mathbb{N}$ (the number of data samples between pilot samples) and set

$$I = \{n(K+1)+1 : n = 0, \dots, \lfloor (N-1)/(K+1) \rfloor\};$$

we can write $I = \bigcup_{l=1}^{L} I_l$, where each I_l is a (sorted) set of strictly increasing integer indices such that the last element of I_l coincides with the first element of I_{l+1}. Moreover, we require that each I_l contains the same number C of elements, except possibly for I_L. We indicate indices belonging to I_l with $i_c^{(l)} \in I_l$. As noted earlier in this section, we assume that there exist L vectors $\overline{\boldsymbol{\beta}}^{(l)}$ of polynomial coefficients such that $\mathbf{M}\overline{\boldsymbol{\beta}}^{(l)} = \boldsymbol{\xi}^{(l)}$, where $\mathbf{M}$ is the rectangular measurements matrix

$$\mathbf{M} = \begin{bmatrix} 1 & i_1^{(1)}T_s & \dots & (i_1^{(1)}T_s)^d \\ 1 & i_2^{(1)}T_s & \dots & (i_2^{(1)}T_s)^d \\ \vdots & \vdots & \ddots & \vdots \\ 1 & i_C^{(1)}T_s & \dots & (i_C^{(1)}T_s)^d \end{bmatrix} \tag{14}$$

and $\boldsymbol{\xi}^{(l)} = (\xi_{i_1^{(l)}}, \dots, \xi_{i_C^{(l)}})^t$. Our cost function is the conditional mean-square error

$$\mathbb{E}\left[\left\| \mathbf{M}\mathbf{S}\widetilde{\boldsymbol{\xi}}^{(l)} - \boldsymbol{\xi}^{(l)} \right\|_2^2 \middle| \mathbf{y}' \right], \tag{15}$$

that we seek to minimize as a function of $\mathbf{S}$. We do not use ordinary least-squares because the measurement noise $\widetilde{\boldsymbol{\xi}}^{(l)}$ has time-varying variance. Equation (15) can be written as standard bias-variance decomposition (by expanding the square)

$$\left\| (\mathbf{I} - \mathbf{M}\mathbf{S})\boldsymbol{\xi}^{(l)} \right\|_2^2 + \mathbb{E}\left[\left\| \mathbf{M}\mathbf{S}\mathbf{v}^{(l)} \right\|_2^2 \middle| \mathbf{y}' \right], \tag{16}$$

and by imposing the (necessary) unbiasedness condition $\mathbf{S}\mathbf{M} = \mathbf{I}$ (cf. (3.28) in [32]) we find that the second term in (16) is minimized if

$$\widehat{\mathbf{S}}_l = (\mathbf{M}^t \mathbf{W}_l \mathbf{M})^{-1} \mathbf{M}^t \mathbf{W}_l \tag{17}$$

where $\mathbb{E}[\mathbf{v}^{(l)}(\mathbf{v}^{(l)})^t]^{-1} = \mathbf{W}_l \propto \operatorname{diag}\{(y'_{i_c^{(l)}})^2\}_{c=1}^{C}$, while the bias term vanishes. Notice that $\widehat{\mathbf{S}}_l\mathbf{M} = \mathbf{I}$; in addition we recall that (17) solves the weighted least squares problem for heteroscedastic noise [1].

The sought polynomial coefficients are thus computed via $\widehat{\boldsymbol{\beta}}^{(l)} = \widehat{\mathbf{S}}_l \widetilde{\boldsymbol{\xi}}^{(l)}$, leading to an explicit formula for the interpolating / denoising polynomial functions:

$$p_l(t) := \widehat{\beta}_1^{(l)} + \widehat{\beta}_2^{(l)} t + \dots + \widehat{\beta}_{d+1}^{(l)} t^d.$$

The estimated jitter process at all time instants will hence be

$$\widehat{\xi}_n = \begin{cases} p_1(nT_s), & 1 \le n \le i_C^{(1)} \\ p_l((n - i_1^{(l)} + 1)T_s), & i_1^{(l)} < n \le i_C^{(l)},\ l = 2, \dots, L. \end{cases} \tag{18}$$

The dejittered signal is then estimated as:

$$\widehat{x}_n = y_n - \widehat{\xi}_n y'_n. \tag{19}$$

We summarize the algorithm in the following flowchart:

B. Kalman Smoother

In this section, we apply Kalman filtering and smoothing to our jitter tracking problem, which are standard techniques that provide optimal recursive estimates for systems described using the state-space model formalism. The state-space model for the linearized jitter tracking problem under investigation is

$$\begin{cases} \xi_n = \varphi \xi_{n-1} + \epsilon_n \\ y_n = x_n + \xi_n x'_n + w_n, \end{cases} \tag{20}$$

from which it should be obvious to the reader how Kalman filtering and smoothing can estimate the process ξ_n provided that sufficiently accurate values of the parameter φ, σ_ϵ^2 and white noise variance σ_w^2 are available.

Our implementation of the Kalman filter and smoother routine is described in Algorithm 2. It follows the standard textbook routine (see, e.g., [36]), but it includes a slight modification to handle missing data at non-pilot positions: the measurement covariances R_n are set to $= +\infty$, making the Kalman gain null. This correctly encodes the fact that the autoregressive state is observed only at the pilot indices (i.e., measurement updates are performed only on pilots). Equivalently, between

two consecutive pilot updates separated by K data samples, the AR(1) dynamics imply the K-step relation

$$\xi_n = \varphi^{K+1}\xi_{n-K-1} + \nu_n, \quad \nu_n \sim \mathcal{N}\left(0,\ \sigma_\epsilon^2 \frac{1-\varphi^{2(K+1)}}{1-\varphi^2}\right), \tag{21}$$

and the smoother interpolates the intermediate states using (20). Notice that we replace the observation "matrices" x'_n with their estimates y'_n. Algorithm 2 clarifies our implementation.

1) AR(1) process and white noise parameters estimation via conditional Maximum Likelihood: To achieve optimal performance, the Kalman smoother requires exact knowledge of the three parameters φ, σ_ϵ, and σ_w. One way to obtain these parameter values is to estimate them during a preliminary calibration phase and then set them as constants. However, this solution would not account for potential time variations in the parameters. We propose in this section a conditional maximum likelihood estimation technique for φ, σ_ϵ, and σ_w that can be repeated efficiently for each block of N samples. A numerical evaluation is deferred to Section IV.

At pilot positions $n \in I$ we can rearrange the second equation in (20) and define

$$m_n := \frac{y_n - x_n}{x'_n} = \xi_n + \frac{w_n}{x'_n},$$

which is conditionally Gaussian given the realized value of x'_n. In practice, x'_n is not available and we use the plug-in approximation $x'_n \approx y'_n$; accordingly, all conditional statements below are understood with respect to the realized values of y'_n.

Assume that pilots are uniformly spaced with K data samples between two pilots, i.e., consecutive pilots are separated by $K+1$ time indices. Then the conditional covariance structure of $\{m_n\}_{n\in I}$ can be written as

$$\begin{aligned}\gamma_m(n;d) &:= \mathbb{E}\left[m_{n+d(K+1)}\, m_n \,\middle|\, \mathbf{x}'\right] \\ &= \frac{\sigma_\epsilon^2}{1-\varphi^2}\,\varphi^{|d|(K+1)} + \frac{\sigma_w^2}{(x'_n)^2}\,\chi_{\{0\}}(d),\end{aligned} \tag{22}$$

where $d \in \mathbb{Z}$ is the pilot lag (so that $n + d(K+1) \in I$).

If $N \in \mathbb{N}$ samples are collected by the ADC, the covariance in (22) can be written as an $M \times M$ symmetric real matrix with $M = \lceil N/(K+1) \rceil$, that we decompose as $\mathbf{\Sigma} + \mathbf{D}$, where $\mathbf{\Sigma}_{ij} = \sigma_\epsilon^2 \varphi^{(K+1)|i-j|}/(1-\varphi^2)$ has a Toeplitz Kac-Murdoch-Szegő structure and $\mathbf{D} = \mathrm{diag}(\sigma^2_{w_{i_n}})$. As already noted, $\mathbf{m}$ is conditionally Gaussian, and thus its conditional negative log-likelihood function ℓ given the measurements $\widetilde{\mathbf{m}}$ is

$$\begin{aligned}\ell(\boldsymbol{\theta}; \widetilde{\mathbf{m}}) = &\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(\det(\mathbf{\Sigma}+\mathbf{D})) \\ &+ \frac{1}{2}\widetilde{\mathbf{m}}^t(\mathbf{\Sigma}+\mathbf{D})^{-1}\widetilde{\mathbf{m}}\end{aligned} \tag{23}$$

with $\boldsymbol{\theta} = (\sigma_\epsilon, \varphi, \sigma_w)$. It can be shown that

$$\det(\mathbf{\Sigma}) = \left(\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)^M (1-\varphi^{2(K+1)})^{M-1} \neq 0; \tag{24}$$

additionally $\mathbf{\Sigma}^{-1}$ is tridiagonal and has the closed-form expression

$$\begin{aligned}(\mathbf{\Sigma}^{-1})_{ij} = &\frac{(1-\varphi^2)}{\sigma_\epsilon^2(1-\varphi^{2(K+1)})} \\ &\times \begin{cases} 1 & \text{if } i=j=1 \text{ or } i=j=M, \\ 1+\varphi^{2(K+1)} & \text{if } i=j,\ i\notin\{1,M\}, \\ -\varphi^{K+1} & \text{if } |i-j|=1, \\ 0 & \text{else.}\end{cases}\end{aligned} \tag{25}$$

Therefore, if we write $(\mathbf{\Sigma}+\mathbf{D}) = \mathbf{\Sigma}(\mathbf{I} + \mathbf{\Sigma}^{-1}\mathbf{D})$, the complexity of calculating $\det(\mathbf{\Sigma}+\mathbf{D})$ is $O(M)$ because $(\mathbf{I} + \mathbf{\Sigma}^{-1}\mathbf{D})$ is tridiagonal and $\det(\mathbf{\Sigma})$ is known from (24).

Similarly, we can use the Woodbury matrix identity to deal with the inversion of $\mathbf{\Sigma}+\mathbf{D}$:

$$(\mathbf{\Sigma}+\mathbf{D})^{-1} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}(\mathbf{D}^{-1} + \mathbf{\Sigma}^{-1})^{-1}\mathbf{\Sigma}^{-1},$$

from which we can deduce that the complexity of evaluating $\widetilde{\mathbf{m}}^H(\mathbf{\Sigma}+\mathbf{D})^{-1}\widetilde{\mathbf{m}}$ is also $O(M)$, since $(\mathbf{D}^{-1} + \mathbf{\Sigma}^{-1})$ is tridiagonal. Hence, since the evaluation of (23) can be performed very efficiently by exploiting the problem structure and using standard matrix theory tricks, we can, in turn, efficiently minimize (23) to obtain the maximum likelihood estimate of $\boldsymbol{\theta}$ for use in the Kalman filter routine.

C. Complexity Analysis

The Kalman smoother processes each sample within the stream; a single iteration in both forward and backward passes has constant cost, thus the total cost for N samples is $O(N)$, no matter what the pilot density $p = (L(C-1)+1)/N$ is.

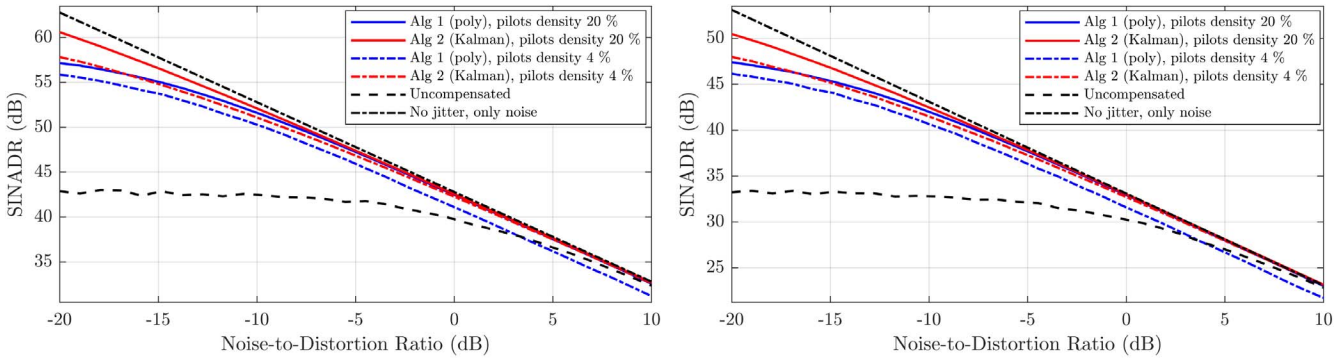
In the polynomial technique, the matrix to be inverted in (17) is $(d+1)\times(d+1)$; the full cost of (17), including matrix multiplications, is thus $O(C(d+1)^2 + (d+1)^3)$. Since this is repeated L times, the total cost for calculating the polynomial coefficients will be $O(L(C(d+1)^2 + (d+1)^3))$; the polynomial evaluations cost $O(N(d+1))$, bringing the total complexity to $O(N(d+1) + L(C(d+1)^2 + (d+1)^3))$. Rewriting the latter as function of the pilot density provides a complexity of

$$\begin{aligned}O\Bigg(N\left((d+1) + p(d+1)^2\frac{C+d+1}{C-1}\right) \\ + (d+1)^2\frac{C+d+1}{C-1}\Bigg).\end{aligned}$$

IV. Simulation Results

A. Baseline Evaluation

Throughout most of this section, the test signal x is a bandlimited Gaussian process (cf. Section II) with a cutoff frequency of 40 MHz, sampled at 100 MS/s and normalized so that $\sigma_x^2 = 1$. The discrete-time waveform was generated by applying the desired brickwall filter to a sequence u of $N = 2^{18}$ independent real-valued Gaussian random variables. The samples were taken at the uniform time instants nT_s, and the jittered samples



(a) $\sigma_\xi/T_s = 5 \cdot 10^{-3}$, $\varphi = 0.999$ (b) $\sigma_\xi/T_s = 1.5 \cdot 10^{-2}$, $\varphi = 0.999$

Fig. 4. Varying NDR analysis for two different jitter levels. The white noise power is progressively increased. The "No jitter, only noise" line shows σ_x^2/σ_w^2 in dB.

$x(nT_s + \xi_n)$ were obtained by evaluating the underlying bandlimited continuous-time waveform via a truncated version of the Whittaker–Shannon interpolation formula,

$$x(nT_s + \xi_n) = \sum_{m=-\infty}^{\infty} x(mT_s)\,\mathrm{sinc}\left(\frac{nT_s + \xi_n - mT_s}{T_s}\right). \tag{26}$$

We designated a subset of sample indices $I \subset \{0, \ldots, N-1\}$ as *pilot positions* and assumed that the corresponding nominal (unjittered) samples $\{x(nT_s)\}_{n\in I}$ were available as references; the effects of pulse shaping and matched filtering were therefore neglected in this section. In a first set of simulations, we tested the relationship between background white noise and jitter, and their effects on the proposed algorithms. In Fig. 4, we fix the amount of jitter at two levels. We swept over 30 different white noise power levels (and ran 5 experiments each) so that the NDR ranges from -20 to 10 dB, and we assessed the overall SINADR (cf. formula (7)). Two different pilot-sample densities are considered as well, while keeping fixed both the block length $C = 500$ and the polynomial degree $d = 4$ fixed for Algorithm 1. The Kalman smoother has oracle knowledge of σ_w^2, σ_ϵ^2 and $\varphi = 0.999$. As expected, the Kalman smoother displays superior performance throughout the entire NDR range; improvements after compensation up to $1-2$ dB, after which the jitter disturbance is essentially drowned in noise and thus no longer distinguishable.

Another batch of simulations is shown in Fig. 5 where the NDR is kept constant, while the density of the (uniformly spaced) pilot samples is gradually increased from 1% to 20%.

Both techniques benefit from the higher pilot density, with a notable jump in performance for Algorithm 1 when going from 1 to 2%; at the same time, the Kalman smoother performs well even with pilot density as low as 1%.

In Fig. 6 we show the algorithms' performance for increasing jitter levels at constant noise power; the latter is chosen so that the NDR is -10 dB at the lowest jitter level. The compensated signals achieve $6-15$ dB improvements in SINADR up to ≈4% jitter, beyond which the polynomial method overtakes the Kalman smoother and continues to deliver consistent gains up to 10% of jitter.

Algorithm 1: Robust Polynomial Interpolation.

Require: Observations y_n, y'_n, pilot indices I and x_n for $n \in I$; number of samples N, measurements matrix $\mathbf{M}$ (14).

1: **for** $l = 1$ to L **do**
2: Form measurements vector $\widetilde{\boldsymbol{\xi}}^{(l)}$ according to (13)
3: Minimize (15) in $\mathbf{S}$: $\widehat{\mathbf{S}}_l \leftarrow (\mathbf{M}^t\mathbf{W}_l\mathbf{M})^{-1}\mathbf{M}^t\mathbf{W}_l$
4: Interpolating polynomial coefficients $\widehat{\boldsymbol{\beta}}^{(l)} \leftarrow \widehat{\mathbf{S}}_l\widetilde{\boldsymbol{\xi}}^{(l)}$
5: Interpolated values at non pilots pos. $\widehat{\boldsymbol{\xi}}^{(l)}$ (18)
6: **end for**

7: **Return** de-jittered signal $\widehat{x}_n = y_n - \widehat{\xi}_n y'_n$ (19) $\forall\, n$.

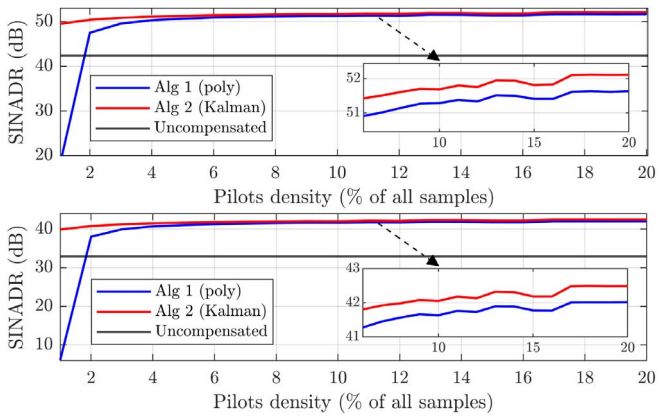


Fig. 5. From top to bottom: $\sigma_\xi/T_s = 5 \cdot 10^{-3},\ 1.5 \cdot 10^{-2}$ respectively. NDR is fixed to $= -10$ dB.

This numerical observation highlights one limitation of a model-based approach, such as the Kalman filter. When the standard deviation of the jitter process is sufficiently large so that the first-order Taylor expansion becomes increasingly unreliable, the measurement model in (20) becomes invalid. Although higher-order Taylor terms may be lumped into w_n, the latter would become colored and both signal- and jitter-dependent. Moreover errors due to model mismatch

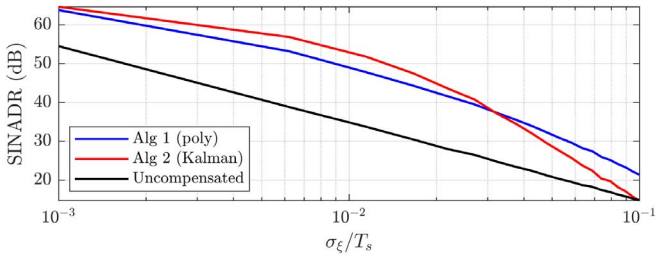


Fig. 6. Fixed noise power (-10 dB NDR at lowest jitter level, -42 dB NDR at 4%), increasing jitter.

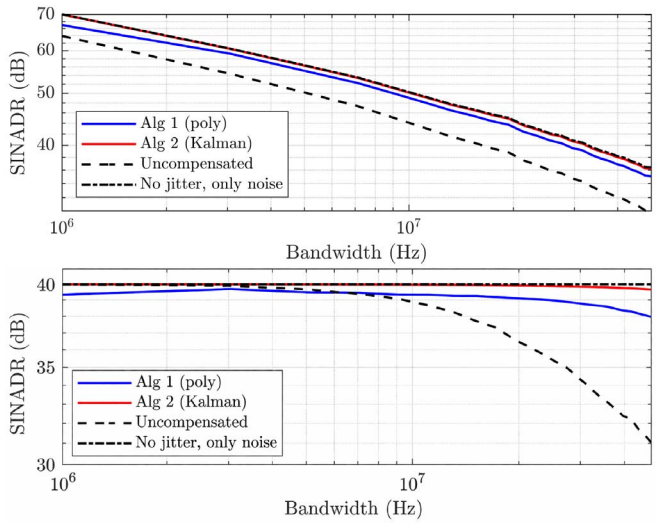


Fig. 7. Algorithm performances vs signal bandwidth. Fixed NDR $= -5$ dB (top) and fixed SNR $= 40$ dB (bottom). $\varphi = 0.999$, $\sigma_\xi/T_s = 1.5 \cdot 10^{-2}$.

introduced in the Kalman recursion at certain time instants propagate, causing additional inaccuracies. In contrast, the polynomial interpolation technique, being block-based, does not propagate errors and is more robust to misspecified noise variance, especially if the polynomial degree is low.

We note that the parameters d and C, together with the pilot density, interact in the polynomial technique. In these experiments, we did not attempt to identify their optimal combination; instead, we chose sub-optimal values empirically to avoid both over-fitting and under-fitting. The problem of optimally interpolating Wiener-type processes with polynomial functions has already been studied in, e.g., [4]. Thus, a more rigorous investigation of this interplay is left for future work.

A final numerical simulation is shown in Fig. 7, where we swept over different signal bandwidths to show the SINADR of the compensated signals, in a manner similar to Fig. 4. As the bandwidth increases, we kept either the NDR (left) or the SNR (right) constant.

B. Robustness of the Kalman Smoother to Parameter Inaccuracies

We numerically tested the parameters estimation technique described in Section III-B1, and the results are summarized in Fig. 8. The plot shows a negligible performance loss when

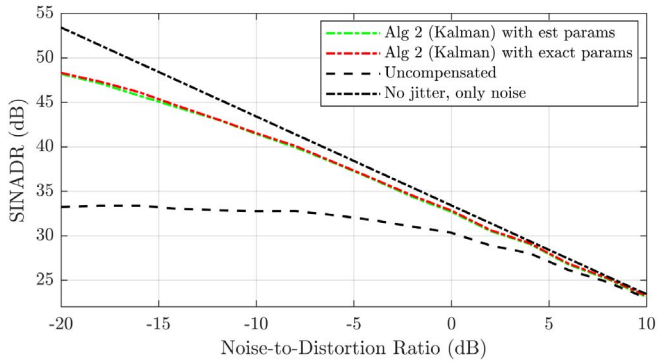


Fig. 8. Kalman smoother SINADR performance with noise and process parameters estimated from data. $\varphi = 0.999$ and $\sigma_\xi/T_s = 1.5 \cdot 10^{-2}$.

the parameters obtained by minimizing (23) are used in Algorithm 2. Due to the non-convexity of (23) with respect to $\boldsymbol{\theta}$, a careful optimization strategy is required. In our Monte Carlo simulations, we used the built-in MATLAB `fmincon` solver coupled with iterative randomization of the starting point with 1000 different initializations and some crude constraints on the parameter values. The test signal and the parameters values were kept fixed throughout the simulation, while the white-noise and jitter realizations were regenerated at each run.

C. Radar Use Case Example: Suppression of Jitter From Nearly Co-Channel Interferer

The numerical examples displayed in the previous section consider jitter suppression as an in-band distortion; the power spectral densities of compensated and uncompensated signals are mostly similar because the term $\xi x'$ has smaller variance than the signal itself if (SJH) holds. This can be rigorously proved by modifying the proof of Proposition II.2. The fact that the term $\xi x'$ produces mainly in-band distortions is not entirely obvious, because its power spectral density is not compactly supported. However, its out-of-band contributions are negligible. This can be seen by studying the integral function

$$I(\nu) = \int_{-\overline{\omega}}^{\overline{\omega}} \frac{\omega^2}{1 + \varphi^2 - 2\varphi\cos(\nu - \omega)}\, d\omega, \tag{27}$$

which exhibits two peaks at $|\nu| \approx \overline{\omega}$ and decays rapidly once $|\nu| > \overline{\omega}$.

To better illustrate the spectral impact of the proposed methods, we outline a different (and somewhat idealized) scenario motivated by radar applications. We assume a bistatic pair consisting of an illuminator of opportunity (IO) at known location and an antenna receiver with two channels, one for surveillance (SC) and one for reference (RC). The RC acquires high-fidelity copies of the IO waveform, while the SC primarily captures targets' echoes. However, the SC is also corrupted by direct-path interference (DPI) which is usually much stronger than echoes and can thus completely overshadow them.

The signal at the ADC output can be written as

$$z_n = x_n + \xi_n x'_n + \zeta_n + \xi_n \zeta'_n + w_n. \tag{28}$$

Algorithm 2: (Inexact) Kalman Filter and Smoother Flow for (20).

Require: Observations y_n, y'_n, pilot indices I and x_n for $n \in I$; number of samples N; AR(1) parameter φ; AR process variance $Q_n = \sigma_\epsilon^2$; measurement noise variance $R_n = \sigma_w^2$ for $n \in I$, $R_n = +\infty$ elsewhere; initial state $\widehat{\xi}_{0|0}$ and covariance $P_{0|0}$.

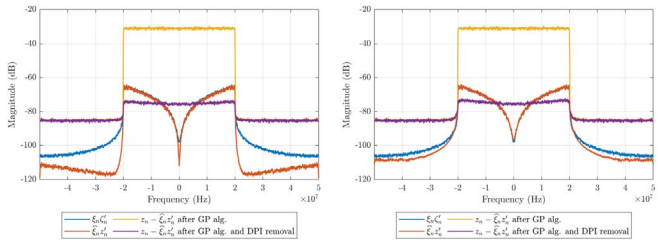
Forward Pass
for $n = 1$ to N **do**
  if $n > 1$ **then**
    Prediction:
    $\widehat{\xi}_{n|n-1} \leftarrow \varphi \widehat{\xi}_{n-1|n-1}$
    $P_{n|n-1} \leftarrow \varphi^2 P_{n-1|n-1} + Q_n$
  else
    Set $\widehat{\xi}_{n|n-1} \leftarrow \widehat{\xi}_{0|0}$
    Set $P_{n|n-1} \leftarrow P_{0|0}$
  end if
  Measurement Update:
  Define $H_n \leftarrow y'_n$
  Innovation: $\widetilde{y}_n \leftarrow y_n - x_n - H_n \widehat{\xi}_{n|n-1}$
  Innovation covariance: $S_n \leftarrow |H_n|^2 P_{n|n-1} + R_n$
  Kalman gain: $K_n \leftarrow P_{n|n-1} \overline{H_n} / S_n$
  Update state estimate: $\widehat{\xi}_{n|n} \leftarrow \widehat{\xi}_{n|n-1} + K_n \widetilde{y}_n$
  Update error covariance: $P_{n|n} \leftarrow P_{n|n-1} - K_n H_n P_{n|n-1}$
end for

Backward Pass (Kalman Smoother)
Let $\widehat{\xi}_{N|N}$ be the final filter estimate from above
$\widehat{\xi}^{\text{smooth}}_{N|N} \leftarrow \widehat{\xi}_{N|N}$
for $n = N - 1$ down to 1 **do**
  Compute smoother gain: $C_n \leftarrow \varphi P_{n|n} / P_{n+1|n}$
  Smoothed state: $\widehat{\xi}^{\text{smooth}}_{n|N} \leftarrow \widehat{\xi}_{n|n} + C_n (\widehat{\xi}^{\text{smooth}}_{n+1|N} - \widehat{\xi}_{n+1|n})$.
end for

Return de-jittered signal (cf. (19)): $\widehat{x}_n \leftarrow y_n - \widehat{\xi}^{\text{smooth}}_{n|N} y'_n$

Here, ζ_n is the contribution from the DPI, while x_n is the wanted signal. The term $\zeta_n + x_n$ is assumed to be known if $n \in I$. By assumption $\sigma_\zeta^2 \gg \sigma_x^2$ and $\sigma_{\xi\zeta'}^2 > \sigma_x^2$. This means that even the interferer-induced jitter term dominates the signal of interest.

We consider a situation in which the term ζ_n can be canceled with negligible impact on the other terms (at least if the target is moving and there is a nontrivial Doppler component in x_n), for example by projecting it away using subspace methods in combination with the RC signal knowledge, see, e.g. [7]. After dejittering z_n using the methods presented in the previous section and optimally interpolating the "holes" left after removing the pilot samples using the Gerchberg-Papoulis algorithm, we then project away ζ_n, and what is left is the desired echo signal. Fig. 9 shows examples of power spectral densities of the different compensation stages described in this section.



(a) Kalman (b) Polynomial

Fig. 9. Power spectral density illustrations at different compensation stages.

D. Relation to Existing Techniques

In a significant portion of the existing literature, clock jitter is modeled as Gaussian and spectrally white, and the mitigation schemes proposed there are designed accordingly. In that context, the notion of "tracking" is vague. Adapting those techniques to the correlated-jitter scenario is non-obvious, thus extensive work is needed and deemed beyond the scope of the present work. The work closest in spirit to ours is [40], where clock jitter is modeled with a $1/f^2$ spectral decay and the jitter process is tracked by placing a sideband pilot tone that is first downconverted to DC and then lowpass-filtered. The resulting phase oscillations are extracted by minimizing an expected value, which results in an infinite-impulse-response (IIR)-type recursion. We therefore benchmark our routines with the technique in [40].

We nonetheless stress that our algorithms and [40] remain structurally different: our proposed methods exploit known pilot samples intertwined with the payload, whereas [40] separates pilots and payload in frequency via a pilot tone. It is not entirely clear how to construct a scenario in which all relevant quantities (bandwidth usage, pilot-symbol density, pulse-shaping filters, etc.) can be consistently mapped across the three techniques; nevertheless, it is still possible to draw qualitative conclusions, which we do in this section.

We generated $N_{\text{sym}} = 43691$ Gaussian data symbols s_i and we pulse-shaped them into a real-valued 40 MHz-wide (from -20 to 20 MHz) waveform using a windowed sinc interpolation, and the resulting analog waveform $s(t)$ was sampled at 120 Msps at the ADC side. The symbols at uniformly spaced indices $I \subseteq \{1, \ldots, N_{\text{sym}}\}$ were designated as pilots; those at I_d as data. Since the oversampling factor is an integer, there is no intersymbol interference (ISI) at the symbol-synchronous sampling instants, and therefore knowledge of sample values at those instants can be inferred from pilot symbol values.

The analog waveform $s(t)$ serves as a common test signal for all methods so that there are no differences in digital signal processing operations that could lead to different values of the baseline $\text{EVM}_{\text{uncomp}}$ (30). For [40], the pilot tone is injected outside the payload band (hence using a larger effective bandwidth) and it is removed prior to symbol recovery, so that the

TABLE I
COMPARATIVE SUMMARY OF THE TWO PROPOSED METHODS

	Polynomial Interpolation 1	**Kalman Smoother 2**
High-level complexity (cf. Section III-C)	$O\left(N\left((d+1)+p(d+1)^2\frac{C+d+1}{C-1}\right)+(d+1)^2\frac{C+d+1}{C-1}\right)$	$O(N)$
Robustness to low pilot density (cf. Figure 5)	3–4% density required for optimal performance	As low as 1% density is enough for optimal performance
Robustness to model mismatch (cf. Figure 6)	Less sensitive to measurement model mismatch	More sensitive to measurement model mismatch
Robustness to parameter inaccuracies (cf. Section III-B1)	Model agnostic	Parameters can be accurately estimated

baseline $\text{EVM}_{\text{uncomp}}$ is evaluated on the payload component in I_d only.

The optimal forgetting factor in the infinite impulse response-type recursion [40] was computed using oracle knowledge of the jitter process realization, not feasible in practice, while the parameters φ, σ_ϵ and σ_w were estimated using the Maximum Likelihood routine described in section III-B1. In the polynomial method, we have not optimized the degree and the batch lengths.

After inserting jitter and white noise, we performed jitter estimation and compensation. We then downsampled and applied the matched filter, along with the optimal rescaling factor, to obtain estimates $\{\widehat{s}_i\}_{i\in I_d}$ of the data symbol values. The performances are here assessed via the Δ Error Vector Magnitude (EVM) defined as follows:

$$\Delta\text{EVM}_{\text{dB}} := 20\log_{10}\left(\frac{\text{EVM}_{\text{uncomp}}}{\text{EVM}_{\text{comp}}}\right) \tag{29}$$

and

$$\text{EVM}_{\text{comp}} = \sqrt{\frac{\frac{1}{|I_d|}\sum_{i\in I_d}|\widehat{s}_i - s_i|^2}{\frac{1}{|I_d|}\sum_{i\in I_d}|s_i|^2}} \tag{30}$$

and equivalently for $\text{EVM}_{\text{uncomp}}$, replacing the $\widehat{s}_i$ with their uncompensated counterparts.

The outcome of this qualitative analysis is shown in Fig. 10. We plot the $\Delta\text{EVM}_{\text{dB}}$ on the y axis whereas on the x axis we report two overlaid quantities, $p = |I|/N_{\text{sym}}$ (pilot symbol density) for our algorithms and % p_t of the unit total power budget allocated to the pilot tone for [40]. By construction $\text{EVM}_{\text{uncomp}}$ is constant across methods and across all density/power levels. Note that while $p = 0 \Leftrightarrow p_t = 0$ (no pilots), the relationship between the two is not necessarily linear.

Algorithms 1 and 2 display a qualitative behavior consistent with that seen in Fig. 5: at low pilot density, Algorithm 1 struggles to deliver a positive gain, and both curves somewhat flatten beyond a certain threshold. The method in [40] performs better overall; however, the three methods are not directly comparable. Moreover, [40] is favored by its use of additional bandwidth and oracle (genie) knowledge of the jitter process realization; at higher power levels [40] is expected to display a performance loss due to low SNR [14].

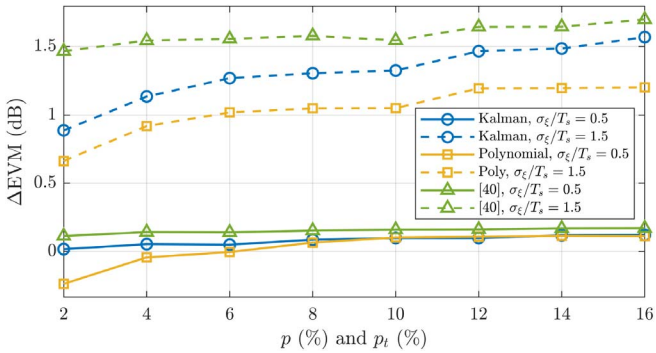


Fig. 10. EVM performances. On the x axis: pilot-symbol density for Algorithms 1 and 2, and pilot-tone fractional power for [40]. $\varphi = 0.999$.

V. CONCLUSION AND FUTURE RESEARCH

In this paper, we propose two algorithms to address stochastic distortions in baseband signals caused by local-oscillator timing jitter in analog-to-digital converters. We modeled the jitter as a first-order autoregressive process. The Kalman smoother approach shows robust and nearly optimal performance across a wide range of scenarios (cf. Fig. 4), and both methods display significant improvements in SINADR even with low pilot-sample densities (Fig. 5). We also assess their impact when pushing the linear approximation (5) to its limits, and we find that for very large jitter the polynomial compensation algorithm could be more suitable, as it is less sensitive to outliers and model mismatches (Fig. 6). Future research will focus on incorporating additional distortions (e.g. phase noise when the signal is upconverted to radio frequency, strong nearby blockers already contaminated with Digital-to-Analog Converter (DAC) jitter, power-amplifier nonlinearities, harmonic distortions, non-zero ISI) in the compensation routines.

APPENDIX

Proof of Proposition II.1: This follows from basic properties of random variables. Indeed if $\text{var}(x'_n) = \sigma^2_{x'}$, it is well-known [27] that $\text{var}(x''_n) = 12\pi^2\sigma^2_{x'}W^2/5$, under the assumption that $\sigma^2_x = 1$ and $\mathcal{S}_x(f) = 1/(2W)\chi_{\{|f|\le W\}}(f)$.

The two processes ξ^2_n and x''_n originate from unrelated sources and can thus be considered independent. Therefore we

have

$$\begin{aligned}\operatorname{var}(\xi_n^2 x_n'') &= \operatorname{var}(\xi_n^2)\operatorname{var}(x_n'') + \operatorname{var}(\xi_n^2)\underbrace{\mathbb{E}[x_n'']^2}_{=0} + \operatorname{var}(x_n'')\mathbb{E}[\xi_n^2]^2 \\ &= 36\pi^2\sigma_\xi^4\sigma_{x'}^2 W^2/5;\end{aligned}$$

similarly

$$\begin{aligned}\operatorname{var}(\xi_n x_n') &= \operatorname{var}(\xi_n)\operatorname{var}(x_n') + \operatorname{var}(\xi_n)\underbrace{\mathbb{E}[x_n']^2}_{=0} + \operatorname{var}(x_n')\underbrace{\mathbb{E}[\xi_n]^2}_{=0} \\ &= \sigma_\xi^2\sigma_{x'}^2.\end{aligned}$$

The conclusion follows if $36\pi^2\sigma_\xi^2 W^2/5 \ll 1$, which is true under the "small jitter" hypothesis (SJH) and the fact that $WT_s < 1/2$, as a consequence of the Shannon-Nyquist sampling theorem. □

Proof of Proposition II.2: We define $\mathbf{u} := \boldsymbol{\xi} \odot \mathbf{x}'$ and we assume without loss of generality that $\sigma_x^2 = 1$. Moreover, to highlight the dependence on the factor WT_s, we use here normalized frequencies $\omega \in [-\pi, \pi]$. The spectral density of the product of two WSS discrete-time processes can be written as convolution [25]

$$\mathcal{S}_u(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{S}_\xi(e^{i\nu})\mathcal{S}_{x'}(e^{i(\omega-\nu)})\,d\nu; \tag{31}$$

moreover a discrete autoregressive process of order 1 has power spectral density [19]

$$\mathcal{S}_\xi(e^{i\nu}) = \frac{\sigma_\epsilon^2}{1 - 2\varphi\cos(\nu) + \varphi^2},\ \nu \in [-\pi, \pi].$$

The power spectral density of our continuous-time signal model is $\mathcal{S}_x(f) = 1/(2W)\chi_{\{|f|\le W\}}$ which after sampling becomes $\mathcal{S}_x(e^{i\omega}) = 1/(2WT_s)\chi_{\{|(\omega)_{2\pi}|\le 2\pi WT_s\}}$, where the operator $(\cdot)_{2\pi}$ denotes wrapping into $[-\pi, \pi]$. In addition, the continuous-time frequency response of an ideal differentiator is $H_D(\Omega) = i\Omega$ (cf. example 4.5 in [25]), which becomes $H_D(e^{i\omega}) = i\omega/T_s$ in our setting. Thus

$$\mathcal{S}_{x'}(e^{i\omega}) = |H_D(e^{i\omega})|^2\mathcal{S}_x(e^{i\omega}) = \frac{((\omega)_{2\pi})^2}{2WT_s^3}\chi_{\{|(\omega)_{2\pi}|\le 2\pi WT_s\}}$$

and by applying once more the Wiener-Khinchin theorem we can conclude that

$$\begin{aligned}\mathcal{R}_{D\mathbf{u}}[0] &= \operatorname{var}((D\mathbf{u})_n) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H_D(e^{i\omega})|^2\mathcal{S}_u(e^{i\omega})\,d\omega \\ &= B\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\omega^2((\omega-\nu)_{2\pi})^2}{1 - 2\varphi\cos(\nu) + \varphi^2} \\ &\qquad \chi_{\{|(\omega-\nu)_{2\pi}|\le 2\pi WT_s\}}\,d\omega\,d\nu\end{aligned} \tag{32}$$

with $B = \sigma_\epsilon^2/(8\pi^2 WT_s^5)$ and (12) follows by recalling that $\sigma_{x'}^2 = 4\pi^2 W^2/3$ and $\sigma_\xi^2 = \sigma_\epsilon^2/(1-\varphi^2)$. □

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the Associate Editor for the many insightful and helpful comments and suggestions.

We acknowledge the use of ChatGPT (OpenAI) for grammar and clarity suggestions and as a sounding board to improve exposition. The authors are solely responsible for all technical content and conclusions.

REFERENCES

[1] A. Aitken, "On least squares and linear combinations of observations," in *Proc. Roy. Soc. Edinburgh*, 1936, vol. 55, pp. 42–48.
[2] H. Araghi, M. A. Akhaee, and A. Amini, "Joint compensation of jitter noise and time-shift errors in multichannel sampling system," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3932–3941, Oct. 2019.
[3] Y. K. Belyaev, "Local properties of the sample functions of stationary Gaussian processes," *Theory Probability Appl.*, vol. 5, no. 1, pp. 117–123, 1960.
[4] T. Blu and M. Unser, "Optimal interpolation of fractional Brownian motion given its noisy samples," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2006, pp. III–III.
[5] V. I. Bogachev, *Measure Theory*, Berlin, Germany: Springer-Verlag, 2006.
[6] Y.-S. Chang, C.-L. Lin, W.-S. Wang, C.-C. Lee, and C.-Y. Shih, "An analytical approach for quantifying clock jitter effects in continuous-time sigma–delta modulators," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 9, pp. 1861–1868, Sep. 2006.
[7] F. Colone, D. W. O'Hagan, P. Lombardo, and C. J. Baker, "A multistage processing algorithm for disturbance removal and target detection in passive bistatic radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 2, pp. 698–722, Apr. 2009.
[8] N. Da Dalt, M. Harteneck, C. Sandner, and A. Wiesbauer, "Numerical modeling of PLL jitter and the impact of its non-white spectrum on the SNR of sampled signals," in *Proc. Southwest Symp. Mixed-Signal Design (Cat. No.01EX475)*, 2001, pp. 38–44.
[9] V. Divi and G. W. Wornell, "Blind calibration of timing skew in time-interleaved analog-to-digital converters," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 3, pp. 509–522, Jun. 2009.
[10] J. Elbornsson, "Analysis, estimation and compensation of mismatch effects in A/D converters," Ph.D. thesis, Dept. of Elect. Eng., Linköping Univ., 2003.
[11] F. Eng, "Non-uniform sampling in statistical signal processing," Ph.D. thesis, Dept. of Elect. Eng., Linköping Univ., 2007.
[12] B. Gävert, M. Coldrey, and T. Eriksson, "Phase noise estimation in OFDM systems," *IEEE Trans. Commun.*, vol. 73, no. 6, pp. 4349–4362, Jun. 2025.
[13] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta*, vol. 21, no. 9, pp. 709–720, 1974.
[14] D. Gerosa, L. Anttila, and T. Eriksson, "Compensation of correlated autoregressive clock jitter in arrays of analog-to-digital converters," in *Proc. 59th Asilomar Conf. Signals, Syst., Comput.*, to be published.
[15] J. A. Grubner, *Probability and Random Processes for Electrical and Computer Engineers.* Cambridge, U.K.: Cambridge Univ. Press, 2006.
[16] L. Hernandez, A. Wiesbauer, S. Paton, and A. Di Giandomencio, "Modelling and optimization of low pass continuous-time sigma delta modulators for clock jitter noise reduction," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2004, pp. I–1072.
[17] O. G. Ibarra-Manzano, J. A. Andrade-Lucio, M. A. Vazquez Olguin, and Y. S. Shmaliy, "Kalman filter for discrete processes with timing jitter," *IEEE Trans. Signal Process.*, vol. 73, pp. 219–229, 2025.
[18] T. Kailath, A. H. Sayed, and N. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ, USA: Prentice Hall, 2000.
[19] S. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory.* Upper Saddle River, NJ, USA: Pearson, 1993.
[20] M. R. Khanzadi, R. Krishnan, D. Kuylenstierna, and T. Eriksson, "Oscillator phase noise and small-scale channel fading in higher frequency bands," in *Proc. IEEE Globecom Workshops*, 2014, pp. 410–415.
[21] B. P. Lathi, *Modern Digital and Analog Communication Systems*, 3rd. ed. London, U.K.: Oxford Univ. Press, 1998.
[22] J. Ma, R. Tao, Y. Li, and X. Kang, "Fractional spectrum analysis for nonuniform sampling in the presence of clock jitter and timing offset," *IEEE Trans. Signal Process.*, vol. 68, pp. 4148–4162, 2020.
[23] F. Marvasti, *Nonuniform Sampling: Theory and Practice.* New York, NY, USA: Plenum, 2001.
[24] M. H. Moghaddam, S. R. Aghdam, and T. Eriksson, "An additive noise modeling technique for accurate statistical study of residual RF

hardware impairments," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2019, Shanghai, China, pp. 1–5.
[25] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 1999.
[26] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.*, vol. 22, no. 9, pp. 735–742, Sep. 1975.
[27] A. Papoulis, *Probability, Random Variables and Stochastic Processes.* New York, NY, USA: McGraw-Hill, 1984.
[28] B. Petrovic, W. Rave, and G. Fettweis, "Phase noise suppression in OFDM Using a Kalman filter," Legacy Vodafone chair publication. [Online]. Available: https://www.vodafone-chair.org/pbls/legacy/d-petrovic/Phase_Noise_Suppression_in_OFDM_using_a_Kalman_Filter.pdf
[29] J. G. Proakis, *Digital Communications.* New York, NY, USA: McGraw-Hill, 2001.
[30] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics.* New York, NY, USA: Springer-Verlag, 2007.
[31] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Signal reconstruction errors in jittered sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4711–4718, Dec. 2009.
[32] C. R. Rao and H. Toutenberg, *Linear Models: Least Squares and Alternatives.* New York, NY, USA: Springer-Verlag, 1999.
[33] R. Rutten, L. J. Breems, and R. H. M. van Veldhoven, "Digital jitter-cancellation for narrowband signals," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2008, pp. 1444–1447.
[34] A. Salib, M. F. Flanagan, and B. Cardiff, "A high-precision time skew estimation and correction technique for time-interleaved ADCs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 10, pp. 3747–3760, Oct. 2019.
[35] M. Shinagawa, Y. Akazawa, and T. Wakimoto, "Jitter analysis of high-speed sampling systems," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 220–224, Feb. 1990.
[36] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Hoboken, NJ, USA: Wiley, 2006.
[37] V. Syrjälä and M. Valkama, "Sampling jitter estimation and mitigation in direct RF sub-sampling receiver architecture," in *Proc. 6th Int. Symp. Wireless Commun. Syst.*, 2006, pp. 323–327.
[38] J. Sung and J. Choi, "A new interpretation of the time-interleaved ADC mismatch problem: A tracking-based hybrid calibration approach," *IEEE Signal Process. Lett.*, vol. 32, pp. 3710–3714, 2025.
[39] N. Testoni, N. Speciale, A. Ridolfi, and C. Pouzat, "Adaptive Wavelet-based signal dejittering," in *Proc. Ph.D. Res. Microelectron. Electron. Conf.*, 2007, pp. 257–260.
[40] Z. J. Towfic, S.-K. Ting, and A. H. Sayed, "Clock jitter compensation in high-rate ADC circuits," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5738–5753, Nov. 2012.
[41] M. Valkama, A. Springer, and G. Hueber, "Digital signal processing for reducing the effects of RF imperfections in radio devices - An overview," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2010, pp. 813–816.
[42] D. S. Weller and V. K. Goyal, "Bayesian post-processing methods for jitter mitigation in sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2112–2123, May 2011.
[43] Y. Yang, X. Zhou, and M. Wang, "Taylor's Theorem and mean value theorem for random functions and random variables," 2025. [Online]. Available: https://arxiv.org/abs/2102.10429v2
[44] A. Zanchi and C. Samori, "Analysis and characterization of the effects of clock jitter in A/D converters for subsampling," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 2, pp. 522–534, Mar. 2008.

Daniele Gerosa was born in Sondrio, Italy, in 1992. He received the B.Sc. and M.Sc. degrees in mathematics from the University of Padova, Italy, in 2015 and 2017 respectively, and the Ph.D. degree in mathematics from Lund University, Sweden, in 2022. From 2022 to 2024, he was with Ericsson AB, Göteborg, Sweden, where he worked on algorithms for the mitigation of hardware impairments in analog-to-digital converters and power amplifiers. Since 2024, he has been a Postdoctoral Fellow with the Department of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden. His research interests include compressive sensing, sampling theory and statistical signal processing for telecommunication and radar applications.

Rui Hou received the M.S. degree in microelectronics and the Ph.D. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2008 and 2017, respectively. From 2008 to 2011, he was with the Space Research Institute of the Netherlands, Utrecht, The Netherlands. Since 2015, he has been with Ericsson AB, Stockholm, Sweden, where he is currently the Expert in RF Technology.

Vimar Björk received the M.Sc. degree in electrical engineering from Chalmers University, Göteborg, in 1991. He joined Ericsson, Mölndal, in 1991. Since 2006, he has been an Expert in radio architecture with Ericsson AB, Göteborg.

Ulf Gustavsson received the M.Sc. degree in electrical engineering from Örebro University, Örebro, Sweden, in 2006, the Ph.D. degree from Chalmers University of Technology, Gothenburg, Sweden, in 2011, and a Docent in communication systems with Chalmers University of Technology in 2023. His background ranges from power amplifier design to radio signal processing. He is currently a Senior Specialist with Ericsson Research, where his research interests include novel radio signal processing techniques for hardware impairment mitigation and behavioral modeling of radio hardware for future advanced antenna systems and communication engineering. He is currently the Lead Scientist from Ericsson Research with Marie Skodowska-Curie European Industrial Doctorate Innovative Training Networks, SILIKA, My-Wave, and ANTERRA.

Thomas Eriksson (Member, IEEE) received the Ph.D. degree in information theory from Chalmers University of Technology, Gothenburg, Sweden, in 1996. From 1990 to 1996, he was with Chalmers. In 1997 and 1998, he was with AT&T Labs - Research, Murray Hill, NJ, USA. In 1998 and 1999, he was with Ericsson Radio Systems AB, Kista, Sweden. Since 1999, he has been with Chalmers University, where he is currently a Professor of communication systems. Further, he was a Guest Professor with Yonsei University, South Korea, from 2003 to 2004. He has authored or coauthored more than 300 journal and conference papers, and holds 16 patents. He is leading the research on hardware-constrained communications and radar processing with Chalmers University of Technology. His research interests include communication, radar signal processing, and modeling and compensation of non-ideal hardware components (e.g. amplifiers, oscillators, ADC/DAC, and modulators).