ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746933

BLIND MODULO ANALOG-TO-DIGITAL CONVERSION OF VECTOR PROCESSES

Amir Weiss, Everest Huang†, Or Ordentlich◇ and Gregory W. Wornell**

*Dept. of EECS
Massachusetts Institute of Technology
{amirwei,gww}@mit.edu

†MIT Lincoln Laboratory
everest@ll.mit.edu

◇School of CSE
Hebrew University of Jerusalem
or.ordentlich@mail.huji.ac.il

ABSTRACT

In a growing number of applications, there is a need to digitize a (possibly high) number of correlated signals whose spectral characteristics are challenging for traditional analog-to-digital converters (ADCs). Examples, among others, include multiple-input multiple-output systems where the ADCs must acquire at once several signals at a very wide but sparsely and dynamically occupied bandwidth supporting diverse services. In such scenarios, the resolution requirements can be prohibitively high. As an alternative, the recently proposed *modulo-ADC* architecture can in principle require dramatically fewer bits in the conversion to obtain the target fidelity, but requires that spatiotemporal information be known and explicitly taken into account by the analog and digital processing in the converter, which is frequently impractical. Building on our recent work, we address this limitation and develop a *blind* version of the architecture that requires no such knowledge in the converter. In particular, it features an automatic modulo-level adjustment and a fully adaptive modulo-decoding mechanism, allowing it to asymptotically match the characteristics of the unknown input signal. Simulation results demonstrate the successful operation of the proposed algorithm.

Index Terms— data conversion, blind signal processing, adaptive filtering, least-mean-squares algorithm.

1. INTRODUCTION

In a host of applications in communication and signal processing there is often a need to digitize highly correlated analog signals, where each signal, which in general may be temporally correlated in itself, is fed into a separate analog-to-digital converter (ADC). One representative example is in the context of a massive multiple-input multiple-output channels [1–3], where the number of antennas can be of the order of tens and even hundreds, whereas the number of users (/independent sources) it serves is moderate, making the signals received by the antennas highly correlated.

In such scenarios, when the signals are (naively) acquired by standard ADCs, there is a high degree of redundancy in the data. Naturally, this redundancy can be exploited in many ways, depending on the processing phase and the desired objective [4–8]. Particularly attractive is to utilize this redundancy as early as the acquisition phase, namely in the analog to digital conversion. By doing so, in principle, the signals could be digitized at the same fidelity using fewer bits, thus reducing power consumption, which in general grows exponentially with the number of bits [9].

This material is based upon work supported, in part, by the United States Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force. This work was also supported, in part, by ISF under Grant 1791/17, and NSF under Grant CCF-1717610.

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. © 2021 Massachusetts Institute of Technology.

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

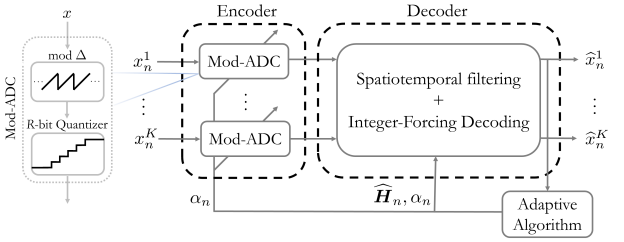

Fig. 1. A schematic illustration of the proposed blind mod-ADCs.

One possible approach to put this notion into practice is to use the recently proposed *modulo ADCs* [10,11]. A modulo ADC first folds each sample of the input process modulo Δ, where Δ is a design parameter, and only then quantizes the result using a traditional uniform quantizer. The modulo operation limits the dynamic range of the signal to be quantized, which in turns results in a quantization error whose magnitude is proportional to Δ, rather than to the dynamic range of the original, unfolded signal. In [10] it is shown that the observed signal can be reliably unfolded, when the second-order statistics (SOSs) of the input signals are known, and Δ is set proportionally to the prediction error standard deviation. More recently, a blind mechanism for a single modulo ADC was proposed [12], which adapts the effective modulo size (analogously to an AGC mechanism in a standard ADC [13]) and learns the required SOSs of the input, while unwrapping the folded signal with the same reliability.

In this work, we extend [12], and develop a *blind* mechanism for *multiple* modulo ADCs working in parallel. Using spatiotemporal correlations of the observed signals, the proposed solution learns the inputs' underlying SOSs, adapts the effective modulo size, and for a given number of bits, considerably decreases the mean square error (MSE) distortion in the reconstruction of the input signals relative to standard ADCs.

2. BRIEF REVIEW ON MODULO ADCS

For a positive number $\Delta \in \mathbb{R}^+$, we define

$$[x] \bmod \Delta \triangleq x - \Delta \cdot \left\lfloor \frac{x}{\Delta} \right\rfloor \in [0, \Delta), \quad \forall x \in \mathbb{R},$$

as the $[\cdot] \bmod \Delta$ operator, where $\lfloor x \rfloor$ is the floor operation, which returns the largest integer smaller than or equal to x. An R-bit modulo ADC with resolution parameter α, termed (R, α) mod-ADC, computes

$$[x]_{R,\alpha} \triangleq [\lfloor \alpha x \rfloor] \bmod 2^R \in \{0, 1, \ldots, 2^R - 1\}, \tag{1}$$

and produces the binary representation of (1) as its output (Fig. 1, left).

Using subtractive dithers [14], an (R, α) mod-ADC can be modeled as a stochastic channel, whose output y for an input x is given by

$$y = [\alpha x + z] \bmod 2^R, \tag{2}$$

where $z \sim \text{Unif}\left((-1, 0]\right)$. Since the modulo operation is a form of *lossy* compression, it is generally impossible to recover the unfolded signal $(\alpha x + z)$ from its folded version y in (2). Nevertheless, under rela-

Authorized licensed use limited to: MIT Libraries. Downloaded on June 22,2022 at 17:23:37 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Oracle Spatiotemporal Modulo Unfolding
$\widehat{\boldsymbol{v}}_{\text{oracle},n} = \texttt{OracleSTModUnfold}\left(\boldsymbol{y}_n, \boldsymbol{v}_{[n]}, \{R_x^{ij}[\ell]\}, \alpha, R\right)$

1 Compute the Linear Minimum MSE (LMMSE) estimate of $\boldsymbol{v}_n$ based on $\boldsymbol{v}_{[n]}$ (denoting $\frac{\mathbf{1}}{\mathbf{2}}$ as the all-$\frac{1}{2}$ vector)

$$\widehat{\boldsymbol{v}}^p_{\text{LMMSE},n} = \boldsymbol{H}^p_{\text{opt}}\left(\boldsymbol{v}_{[n]} + \tfrac{\mathbf{1}}{\mathbf{2}}\right) - \tfrac{\mathbf{1}}{\mathbf{2}}, \tag{3}$$

where $\boldsymbol{H}^p_{\text{opt}} \in \mathbb{R}^{K\times Kp}$ is the matrix filter corresponding the LMMSE predictor, computed based on $\{R_x^{ij}[\ell]\}$ and α;

2 Denote $\boldsymbol{A} = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_K]^{\mathrm{T}}$, and solve [*IF decoding matrix*]

$$\boldsymbol{A} \triangleq \underset{\substack{\bar{\mathbf{A}}\in\mathbb{Z}^{K\times K}\\ \det(\bar{\mathbf{A}})\neq 0}}{\operatorname{argmin}} \max_{k\in\{1,\ldots,K\}} \frac{1}{2}\log_2\left(\bar{\boldsymbol{a}}_k^{\mathrm{T}}\boldsymbol{\Sigma}_p\bar{\boldsymbol{a}}_k\right), \tag{4}$$

where $\boldsymbol{\Sigma}_p$ (6) is computed based on $\{R_x^{ij}[\ell]\}$ and α;

3 Compute $\boldsymbol{w}_{\text{LMMSE},n} = [\boldsymbol{y}_n - \widehat{\boldsymbol{v}}^p_{\text{LMMSE},n}] \bmod 2^R$;

4 Compute $\widetilde{\boldsymbol{g}}_n \in \mathbb{R}^{K\times 1}$ and the estimated prediction error,

$$\begin{cases} g_n^k \triangleq \left[\boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{w}_{\text{LMMSE},n}\right] \bmod 2^R \\ \widetilde{g}_n^k \triangleq \left[g_n^k + \frac{1}{2}2^R\right] \bmod 2^R - \frac{1}{2}2^R \end{cases} \Rightarrow \widehat{\boldsymbol{e}}^p_{\text{LMMSE},n} \triangleq \boldsymbol{A}^{-1}\widetilde{\boldsymbol{g}}_n; \tag{5}$$

5 Return $\widehat{\boldsymbol{v}}_{\text{oracle},n} = \widehat{\boldsymbol{v}}^p_{\text{LMMSE},n} + \widehat{\boldsymbol{e}}^p_{\text{LMMSE},n}$.

tively mild conditions, when the input (possibly high-dimensional) signal is sufficiently temporally- and/or spatially-predictable, e.g., a correlated random vector-process [10] or a deterministic bandlimited signal [15,16], it is in fact possible to *perfectly* recover the unfolded signal[1] from its past samples and its current folded sample via causal processing [17].

More generally, consider K parallel (R,α) mod-ADCs whose input signals $\{x_n^k\}$, collected into a vector $\boldsymbol{x}_n \in \mathbb{R}^{K\times 1}$, are zero-mean jointly stationary processes, with known autocorrelation functions $\{R_x^{ij}[\ell] \triangleq \mathbb{E}\left[x_n^i x_{n-\ell}^j\right] \in \mathbb{R}\}$. The output of the K mod-ADCs is then given by

$$\boldsymbol{y}_n = [\alpha\boldsymbol{x}_n + \boldsymbol{z}_n] \bmod 2^R \in \mathbb{R}^{K\times 1}, \quad \forall n \in \mathbb{N}^+,$$

where $\{z_n^k \sim \text{Unif}((-1,0])\}$, modeling the quantization noises, are independent, identically distributed (i.i.d.) stochastic processes, and the modulo is elementwise. Further, define the unfolded quantized signal,

$$\boldsymbol{v}_n \triangleq \alpha\boldsymbol{x}_n + \boldsymbol{z}_n \in \mathbb{R}^{K\times 1}, \quad \forall n \in \mathbb{N}^+,$$

and assume that the decoder has access to $\{\boldsymbol{v}_{n-1},\ldots,\boldsymbol{v}_{n-p}\}$. Notice that once $\boldsymbol{v}_n$ is recovered, $\boldsymbol{x}_n$ is readily estimated as $\widehat{\boldsymbol{x}}_n = (\boldsymbol{v}_n + \frac{\mathbf{1}}{\mathbf{2}})/\alpha$. Thus, we focus on recovering $\boldsymbol{v}_n$ based on $\boldsymbol{y}_n$ and $\boldsymbol{v}_{[n]} \triangleq \text{vec}\left(\boldsymbol{V}_n\right) \in \mathbb{R}^{Kp\times 1}$, where $\boldsymbol{V}_n \triangleq [\boldsymbol{v}_{n-1}\cdots\boldsymbol{v}_{n-p}] \in \mathbb{R}^{K\times p}$.

The algorithm proposed in [10] for recovering $\boldsymbol{v}_n$ with high probability (w.h.p.) when $\{\boldsymbol{v}_{n-\ell}\}_{\ell=1}^p$ and $\{R_x^{ij}[\ell]\}$ are *known*, here referred to as *oracle* spatiotemporal modulo unfolding, is given in Algorithm 1. The key idea behind this method is that the modulo operation essentially becomes invertible (w.h.p.) when statistical properties of the original (unfolded) signal and correlated measurements to the modulo-folded signal are available, provided that Δ is proportional to the standard deviation of the innovation process. In Algorithm 1, unfolding of the coordinates of $\boldsymbol{v}_n$ is done by using spatiotemporal (causal) linear filtering, followed by *integer-forcing* (IF) decoding [17–20], which exploits spatial correlations in the prediction error vector. In this regard, solving (4) generally requires a complexity exponential in K. However, the optimal integer matrix $\boldsymbol{A}$ needs to be computed only once for the pair $\{\boldsymbol{\Sigma}_p, \alpha\}$, where

$$\boldsymbol{\Sigma}_p \triangleq \mathbb{E}\left[\boldsymbol{e}^p_{\text{LMMSE},n}\left(\boldsymbol{e}^p_{\text{LMMSE},n}\right)^{\mathrm{T}}\right] \in \mathbb{R}^{K\times K} \tag{6}$$

[1]With high probability for random signals, and to an arbitrary precision for deterministic bandlimited signals.

Algorithm 2: Blind Spatiotemporal Modulo Unfolding
$\widehat{\boldsymbol{v}}_n,\ \widehat{\boldsymbol{e}}^p_n = \texttt{BlindSTModUnfold}\left(y_n, \widehat{\bar{\boldsymbol{v}}}_{[n]}, \widehat{\boldsymbol{H}}^p_n, \widehat{\boldsymbol{\Sigma}}_p, \alpha_n, R\right)$

1 Compute the linear estimate of $\boldsymbol{v}_n$ based on $\bar{\boldsymbol{v}}_{[n]}$

$$\widehat{\boldsymbol{v}}^p_n = \widehat{\boldsymbol{H}}^p_n\widehat{\bar{\boldsymbol{v}}}_{[n]} - \tfrac{\mathbf{1}}{\mathbf{2}}; \tag{7}$$

2 Denote $\widehat{\boldsymbol{A}} = [\widehat{\boldsymbol{a}}_1\cdots\widehat{\boldsymbol{a}}_K]^{\mathrm{T}}$, and solve

$$\widehat{\boldsymbol{A}} \triangleq \underset{\substack{\bar{\mathbf{A}}\in\mathbb{Z}^{K\times K}\\ \det(\bar{\mathbf{A}})\neq 0}}{\operatorname{argmin}} \max_{k\in\{1,\ldots,K\}} \frac{1}{2}\log_2\left(\bar{\boldsymbol{a}}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_p\bar{\boldsymbol{a}}_k\right); \tag{8}$$

3 Compute $\boldsymbol{w}_n = [\boldsymbol{y}_n - \widehat{\boldsymbol{v}}^p_n] \bmod 2^R$;

4 Compute $\widehat{\widetilde{\boldsymbol{g}}}_n \in \mathbb{R}^{K\times 1}$ and the estimated prediction error,

$$\begin{cases} \widehat{g}_n^k \triangleq \left[\widehat{\boldsymbol{a}}_k^{\mathrm{T}}\boldsymbol{w}_n\right] \bmod 2^R \\ \widehat{\widetilde{g}}_n^k \triangleq \left[\widehat{g}_n^k + \frac{1}{2}2^R\right] \bmod 2^R - \frac{1}{2}2^R \end{cases} \Rightarrow \widehat{\boldsymbol{e}}^p_n \triangleq \widehat{\boldsymbol{A}}^{-1}\widehat{\widetilde{\boldsymbol{g}}}_n; \tag{9}$$

5 Return $\widehat{\boldsymbol{v}}_n = \widehat{\boldsymbol{v}}^p_n + \widehat{\boldsymbol{e}}^p_n, \widehat{\boldsymbol{e}}^p_n$.

is the covariance matrix of the error $\boldsymbol{e}^p_{\text{LMMSE},n} \triangleq \boldsymbol{v}_n - \widehat{\boldsymbol{v}}^p_{\text{LMMSE},n}$ in optimal linear prediction, and $\widehat{\boldsymbol{v}}^p_{\text{LMMSE},n}$ is defined in (3). Moreover, the LLL algorithm [21] provides a sub-optimal $\boldsymbol{A}$ with a complexity $\mathcal{O}(\text{poly}(K))$.

Of course, in practice, devices such as ADCs usually operate under dynamic conditions, giving rise to a wide range of possible inputs with unknown characteristics, and must still maintain proper operation. Therefore, a significant step towards implementing mod-ADCs for real-life applications can be made by relaxing the (too restrictive) assumption that $\{R_x^{ij}[\ell]\}$ are all known. Building on our recent work, in which a single blind mod-ADC was developed based on temporal correlations, the goal of this work is to take that important step for the general case of K parallel mod-ADCs, exploiting both temporal and spatial correlations.

3. PROBLEM FORMULATION

Consider K parallel (R,α_n) mod-ADC as described in the previous section, with a fixed modulo range $\Delta = 2^R$, but an *adaptable*, possibly *time-varying* resolution parameter $\alpha_n \in \mathbb{R}^+$. The mod-ADCs are fed with the input signal $\{\boldsymbol{x}_n \triangleq \boldsymbol{x}(nT_s)\}_{n\in\mathbb{N}^+}$, acquired by sampling the analog, continuous-time signal $\boldsymbol{x}(t)$ every $T_s = f_s^{-1}$ seconds. We assume that $\{x_n^k\}$ are zero-mean jointly stationary stochastic processes with *unknown* correlation functions $\{R_x^{ij}[\ell]\}$. The observed, distorted signals at the output of the K mod-ADCs are given (in vector form) by

$$\boldsymbol{y}_n = [\alpha_n\boldsymbol{x}_n + \boldsymbol{z}_n] \bmod 2^R, \ \forall n \in \mathbb{N}^+,$$

where, as before, the quantization noise processes $\{z_n^k \sim \text{Unif}((-1,0])\}$ are i.i.d. Further, we redefine the unfolded quantized signal,

$$\boldsymbol{v}_n \triangleq \alpha_n\boldsymbol{x}_n + \boldsymbol{z}_n, \ \forall n \in \mathbb{N}^+, \tag{10}$$

which, in general, is no longer stationary. Nonetheless, whenever α_n is held fixed, $\boldsymbol{v}_n$ can be regarded as stationary on the respective time period.

As explained above, the goal in this context is to estimate the input $\boldsymbol{x}_n$ as accurately as possible based on the observed sequence $\{\boldsymbol{y}_n\}$ at the output of the mod-ADCs using a causal system. However, since $\boldsymbol{v}_n$ is merely a scaled version of $\boldsymbol{x}_n$ contaminated by white noise (10), the problem essentially boils down to recovering $\boldsymbol{v}_n$, and is stated as follows.

Problem Statement: For a fixed number of bits R, design an adaptive mechanism for estimating $\{\boldsymbol{x}_n\}$ from the mod-ADCs' outputs with the lowest possible MSE distortion, *without prior knowledge on the correlation functions* $\{R_x^{ij}[\ell]\}$. The above is interpreted as designing an update mechanism for causally maximizing the resolution parameter α_n, while still allowing for reliable recovery of $\boldsymbol{v}_n$ from $\{\boldsymbol{y}_k\}_{k\leq n}$, and also design the recovery mechanism.

As explained in Section 2, provided $\boldsymbol{v}_n$ is exactly recovered w.h.p.,

Authorized licensed use limited to: MIT Libraries. Downloaded on June 22,2022 at 17:23:37 UTC from IEEE Xplore. Restrictions apply.

i.e., $\widehat{\boldsymbol{v}}_n \overset{\text{w.h.p.}}{=} \boldsymbol{v}_n$, the input signal is readily estimated as $\widehat{\boldsymbol{x}}_n \triangleq \frac{\widehat{\boldsymbol{v}}_n + \frac{\mathbf{1}}{\mathbf{2}}}{\alpha_n}$, where α_n is a known parameter, and $\frac{\mathbf{1}}{\mathbf{2}}$ is to compensate for the quantization noise (non-zero) expectation $\mathbb{E}[\boldsymbol{z}_n] = -\frac{\mathbf{1}}{\mathbf{2}}$ (also in (3) and (7)).

4. BLIND MODULO ADC OF VECTOR PROCESSES

We now present the blind mod-ADC algorithm, which simultaneously estimates the input $\boldsymbol{x}_n$, while learning the (possibly time-varying) SOSs of the unfolded signal (10), necessary for estimation of $\boldsymbol{x}_n$. The structure of the proposed architecture is depicted in Fig. 1. In contrast to the oracle mod-ADC (*cf.* Fig. 6 in [10]), here the encoder and decoder are adaptive, and vary with time according to the statistical properties of the input.

The underlying concept of our approach is the following. For a fixed resolution parameter α_n, given that at any time instance n the unfolded signal $\boldsymbol{v}_n$ can be *exactly* recovered, we may estimate the optimal matrix filter $\boldsymbol{H}^p_{\text{opt}}$, corresponding to the LMMSE predictor of $\boldsymbol{v}_n$ (3) based on the previous samples $\boldsymbol{V}_n$. This can be done using the least mean squares (LMS) algorithm [22], which converges[2] to $\boldsymbol{H}^p_{\text{opt}}$. Upon convergence, α_n can be slightly increased, and as long as the spatiotemporal prediction is sufficiently accurate to allow successful IF decoding as in (5), $\boldsymbol{v}_n$ could still be recovered using the same technique as in Algorithm 1. Fixing α_n again to its new value, the new optimal filter is learned using LMS. The process is repeated until the resolution parameter α_n is sufficiently large.

Conceptually, only appropriate initial conditions and sufficiently smooth transitions from one resolution level to another are required for convergence of this process. Once these are fulfilled, we attain successful steady state operation of the K blind mod-ADCs. We now formalize this concept, and develop the desired adaptive mechanism.

4.1. Phase 1: Initialization and Learning the Optimal Decoding

We begin with a "small" initial value for the resolution parameter, α_0, that ensures a degenerated modulo operation, i.e., $\widetilde{\boldsymbol{y}}_n = \boldsymbol{v}_n$, where

$$\widetilde{\boldsymbol{y}}_n \triangleq \left(\left[\boldsymbol{y}_n + \tfrac{1}{2}2^R\right] \bmod 2^R\right) - \tfrac{1}{2}2^R,$$

such that $\widetilde{y}_n$ is the "modulo-shifted" version of y_n. For purposes that will become clear in the sequel, we further define for convenience

$$\overline{\boldsymbol{v}}_n \triangleq \left(\boldsymbol{v}_n + \tfrac{\mathbf{1}}{\mathbf{2}}\right)/\alpha_n = \boldsymbol{x}_n + \left(\boldsymbol{z}_n + \tfrac{\mathbf{1}}{\mathbf{2}}\right)/\alpha_n, \tag{11}$$

the "α_n-standardized" version of $\boldsymbol{v}_n$. Note that $\mathbb{E}\left[\overline{\boldsymbol{v}}_n\right] = \mathbf{0}$, and the covariance of $\overline{\boldsymbol{v}}_n$ is dominated by the covariance of $\boldsymbol{x}_n$ when α_n is large.

Assuming $\widetilde{y}_n = v_n$ as long as $\alpha_n = \alpha_0$ is fixed, the optimal filter for prediction of $\boldsymbol{v}_n$ (10) based on $\overline{\boldsymbol{v}}_{[n]} \triangleq \text{vec}\left([\overline{\boldsymbol{v}}_{n-1} \cdots \overline{\boldsymbol{v}}_{n-p}]\right) \in \mathbb{R}^{Kp\times 1}$, can be estimated with the LMS algorithm [22], via the updates

$$\widehat{\boldsymbol{H}}^p_n = \widehat{\boldsymbol{H}}^p_{n-1} + \mu \cdot \boldsymbol{e}^p_n \overline{\boldsymbol{v}}^{\text{T}}_{[n]} \in \mathbb{R}^{K\times Kp}.$$

Here, $\widehat{\boldsymbol{H}}^p_n$ is the filter used in Algorithm 2 for the prediction of $\boldsymbol{v}_n$, μ is the learning rate (or step size), and

$$\boldsymbol{e}^p_n \triangleq \boldsymbol{v}_n - \widehat{\boldsymbol{v}}^p_n \in \mathbb{R}^{K\times 1} \tag{12}$$

is the prediction error of the linear predictor $\widehat{\boldsymbol{v}}^p_n$ as in (7). In addition, rather than using $\{\boldsymbol{v}_n\}$ (10), we use the standardized process $\{\overline{\boldsymbol{v}}_n\}$ as the observations in (7), since as the adaptive process evolves and α_n increases, the SOSs of $\overline{\boldsymbol{v}}_n$ become less affected by α_n, alleviating the LMS. Details regarding the convergence of the LMS algorithm, as well as the proper setting of the step size μ which guarantees this convergence, are omitted due to space limitation (see [12] for the same guiding principles). From now on, assume that μ is chosen so as to ensure that [23],

$$\alpha_n = \alpha_0 : \lim_{n\to\infty} \mathbb{E}\left[\widehat{\boldsymbol{H}}^p_n\right] = \boldsymbol{H}^p_{\text{opt}}. \tag{13}$$

After enough iterations, since we assume that α_0 is sufficiently small to ensure that $\widehat{\boldsymbol{v}}_n = \boldsymbol{v}_n$ for every n during initialization, which gives us access to $\widehat{\boldsymbol{e}}^p_n = \boldsymbol{e}^p_n$ (12), the filter $\widehat{\boldsymbol{H}}^p_n$ will approximately converge to an unbiased estimate of $\boldsymbol{H}^p_{\text{opt}}$, as in (13). Further, during the initialization we can compute the empirical covariance $\widehat{\boldsymbol{\Sigma}}_p \triangleq \frac{1}{n}\sum_{\ell=1}^n \widehat{\boldsymbol{e}}^p_\ell \left(\widehat{\boldsymbol{e}}^p_\ell\right)^{\text{T}}$, and use it for IF decoding as in (8). Assuming μ is sufficiently small, the MSE of $\widehat{\boldsymbol{v}}^p_n$ will approximately converge to the MSE of $\widehat{\boldsymbol{v}}^p_{\text{LMMSE},n}$ (with α_n replacing α, according to the definition (10)),

$$\exists N_0 : \forall n > N_0 : \mathbb{E}\left[\boldsymbol{e}^p_n \left(\boldsymbol{e}^p_n\right)^{\text{T}}\right] \underset{\mu\ll 1}{\approx} \boldsymbol{\Sigma}_p. \tag{14}$$

Once (14) is accurate enough and $\widehat{\boldsymbol{\Sigma}}_p \approx \boldsymbol{\Sigma}_p$, by [10, Prop. 5], we have

$$\begin{aligned}\Pr(\mathcal{E}_{\text{OL}_n}) &\triangleq \Pr\left(\widehat{\boldsymbol{v}}_n \neq \boldsymbol{v}_n\right) \approx \Pr\left(\widehat{\boldsymbol{v}}_{\text{oracle},n} \neq \boldsymbol{v}_n\right)\Big|_{\boldsymbol{A}=\widehat{\boldsymbol{A}}} \\ &\leq 2K\exp\left\{-\frac{3}{2}2^{2(R-R_{\text{IF}}(\mathbf{A}))}\right\}\Bigg|_{\boldsymbol{A}=\widehat{\boldsymbol{A}}},\end{aligned} \tag{15}$$

where $\mathcal{E}_{\text{OL}_n}$ is referred to as the overload event,

$$R_{\text{IF}}(\boldsymbol{A}) \triangleq \max_{k\in\{1,\ldots,K\}} \frac{1}{2}\log_2\left(\overline{\boldsymbol{a}}^{\text{T}}_k \boldsymbol{\Sigma}_p \overline{\boldsymbol{a}}_k\right) \triangleq \frac{1}{2}\log_2\left(\sigma^2_{\max}\right),$$

and $\boldsymbol{v}_n$ is recovered w.h.p. The approximation in (15) is mostly due to estimation errors; a sharper analysis is omitted due to lack of space. Having learned the optimal decoding, we can increase the resolution parameter α_n, and use the quantizers' output bits to a finer description of inputs.

4.2. Phase 2: Updating the Resolution Parameter

As shown in [10, Prop. 5], the no overload event $\mathcal{E}^*_{\overline{\text{OL}}_n} \triangleq \{\widehat{\boldsymbol{v}}^p_{\text{oracle},n} = \boldsymbol{v}_n\}$ of the oracle mod-ADCs is equivalent to $\{\max_k |g^k_n| < \frac{\Delta}{2}\}$, where $\boldsymbol{g}^k_n \triangleq \boldsymbol{A}\boldsymbol{e}^p_n$. Therefore, in order to increase α_n, we must somehow detect that the filter $\boldsymbol{H}^p_n$ already approximates the optimal one well enough, such that the *transformed prediction errors* $\{g^k_n\}$ are sufficiently small with respect to the dynamic range Δ. In this case, a small change in the resolution would not affect our ability to recover $\boldsymbol{v}_n$ w.h.p. Due to space limitation, we refer the reader to [12], Subsection IV-B, for a detailed justification for a single mod-ADC, based on similar principles.

Since the resolution update rule is a user-controlled system parameter, it can be easily set to be small enough, so as to maintain sufficiently small transformed prediction errors, and safely continue recovering $\boldsymbol{v}_n$ w.h.p. after the update. It remains only to make sure that before the resolution parameter is increased, we arrive at an intermediate steady state, wherein $\mathcal{E}_{\overline{\text{OL}}_n}$ occurs w.h.p. A plausible way of achieving this, is to update α_n only when the maximum variance of the transformed prediction errors is much smaller than Δ, i.e., when $\kappa \cdot \sigma_{\max} < \frac{\Delta}{2}$ for some large enough non-negative constant $\kappa \in \mathbb{R}^+$. If the prediction errors were Gaussian, this would provide a direct proxy for the overload probability, and increasing κ would exponentially decrease the overload probability. This provides the conditions to re-learn the optimal filter corresponding to the LMMSE predictor of $\boldsymbol{v}_n$ with the updated resolution α_n.

Recall that in practice, though, since $\{R^{ij}_x[\ell]\}$ are unknown, $\sigma_{\max}$ is clearly not known as well. Nevertheless, since $\widehat{\boldsymbol{v}}_n \overset{\text{w.h.p.}}{=} \boldsymbol{v}_n \Leftrightarrow \widehat{\boldsymbol{g}}_n \overset{\text{w.h.p.}}{=} \boldsymbol{g}_n$ throughout the adaptive process, we can estimate $\sigma_{\max}$ online by

$$\widehat{\sigma}^2_{k,n} \triangleq \frac{1}{L_s}\sum_{\ell=0}^{L_s-1}\left(\widehat{g}^k_{n-\ell}\right)^2, \ \widehat{\sigma}_{\max,n} \triangleq \max_{k\in\{1,\ldots,K\}}\sqrt{\widehat{\sigma}^2_{k,n}}.$$

where $L_s \in \mathbb{N}^+$ is a moving-average window length, and is also set to be the minimal time stabilization interval wherein α_n must be kept fixed after its last update. Thus, we increase α_n only when $\mathbb{1}_{\uparrow\alpha,n} = 1$, where

$$\mathbb{1}_{\uparrow\alpha,n} \triangleq \begin{cases} 1, & \kappa\cdot\widehat{\sigma}_{\max,n} < \frac{\Delta}{2} \\ 0, & \kappa\cdot\widehat{\sigma}_{\max,n} > \frac{\Delta}{2} \end{cases}. \tag{16}$$

Whenever $\mathbb{1}_{\uparrow\alpha,n} = 0$, we infer that the prediction errors are not satis-

[2] In the mean sense, under mild conditions; see Subsection 4.1.

Authorized licensed use limited to: MIT Libraries. Downloaded on June 22,2022 at 17:23:37 UTC from IEEE Xplore. Restrictions apply.

factorily small. In these cases, we decrease the resolution so as to resort to a state where $\boldsymbol{v}_n$ is again recovered w.h.p, which allows the LMS to converge to the desired filter, and then safely increase the resolution again.

Upon updating α_n, we also appropriately update the filter $\widehat{\boldsymbol{H}}_n^p$, since the input signal $\bar{\boldsymbol{v}}_n$ is scaled with α_n as well (11). Thus, we update

$$\alpha_{n+1} = \left[\mathbb{1}_{\uparrow\alpha,n} \cdot \delta_\alpha^{-1} + (\mathbb{1}_{\uparrow\alpha,n} - 1) \cdot \delta_\alpha\right] \alpha_n,$$
$$\widehat{\boldsymbol{H}}_{n+1}^p = \left[\mathbb{1}_{\uparrow\alpha,n} \cdot \delta_\alpha^{-1} + (\mathbb{1}_{\uparrow\alpha,n} - 1) \cdot \delta_\alpha\right] \widehat{\boldsymbol{H}}_n^p,$$

where $\delta_\alpha \in (0,1)$ is the update resolution parameter.

It is now straightforward to generalize this adaptive process, since, conceptually, we now only need to repeatedly execute the properly interlaced Phase 1 and Phase 2. In the repeated Phase 1, the "initial" values for the filter and resolution parameter would be the corresponding values of the previous time step. Additionally, $\bar{\boldsymbol{v}}_{[n]}$ will be replaced by $\widehat{\bar{\boldsymbol{v}}}_{[n]} \triangleq \mathrm{vec}\left(\left[\widehat{\bar{\boldsymbol{v}}}_{n-1} \cdots \widehat{\bar{\boldsymbol{v}}}_{n-p}\right]\right) \in \mathbb{R}^{Kp\times 1}$, whose vector-elements[3]

$$\widehat{\bar{\boldsymbol{v}}}_n \triangleq \widehat{\boldsymbol{x}}_n = \left(\widehat{\boldsymbol{v}}_n + \tfrac{\mathbf{1}}{\mathbf{2}}\right)/\alpha_n \tag{17}$$

are $\{\widehat{\bar{\boldsymbol{v}}}_i \overset{\text{w.h.p.}}{=} \bar{\boldsymbol{v}}_i\}_{i=n-1}^{n-p}$. The repeated Phase 2 would then be executed after (at least) L_s time steps with the updated resolution.

4.3. Detecting an Overload, and Preventing Error Propagation

Given there is no overload, which holds w.h.p. when κ is large enough, alternating between these two phases leads to convergence near the limiting fixed point $\kappa\cdot\sigma_{\max,n} = \frac{\Delta}{2}$, as in (16), up to small fluctuations due to the limited-resolution adaptation step δ_α and estimation errors in $\widehat{\sigma}_{\max,n}$. However, the overload event still has non-zero probability, and must be taken into account. Indeed, whenever $\mathcal{E}_{\mathrm{OL}_n} = \{\widehat{\boldsymbol{v}}_n \neq \boldsymbol{v}_n\}$ occurs, an extremely fast error propagation process begins. To detect such errors and prevent the consequent error propagation, we propose a generalized version of the detector developed in [12] (Subsection IV-C therein),

$$\widehat{\mathbb{1}}_{\mathcal{E}_{\mathrm{OL}_n}} \triangleq \begin{cases} 1, & \bigcup_{k=1}^{K}\left\{|\widehat{\bar{v}}_n^k| > \sqrt{2\widehat{\sigma}^2_{\bar{v}^k,n}\log(n)}\right\} \\ 0, & \text{otherwise} \end{cases}, \ \forall n \geq N_s, \tag{18}$$

such that (18) detects an outlier with a large magnitude w.h.p.[4], and

$$\widehat{\sigma}^2_{\bar{v}^k,n} \triangleq \frac{1}{n-1}\sum_{\ell=1}^{n-1}\left(\widehat{\bar{v}}_\ell^k\right)^2, \ \ \forall n \geq 2.$$

Here, $N_s \in \mathbb{N}^+$ is the length of a fixed stabilization time-interval, wherein (18) is still not sufficiently accurate, and we enforce a simple, more conservative condition for the transition phase $n \leq N_s$ (e.g., similarly to Eq. (49) in [12]). The justification for (18) is omitted due to space limitation, though a similar derivation appears in [12], Subsection IV-C. Thus, the error propagation prevention mechanism is implemented by

$$\widehat{\mathbb{1}}_{\mathcal{E}_{\mathrm{OL}_n}} = 1 \Rightarrow \text{Reset: } \alpha_{n+1} = \alpha_0, \ \widehat{\boldsymbol{H}}_{n+1}^p = (\alpha_0/\alpha_n)\cdot\widehat{\boldsymbol{H}}_n^p. \tag{19}$$

To Summarize, our proposed algorithm for the implementation of the parallel mod-ADCs is executed by iteratively repeating Phases 1 and 2, where after each call to `BlindSTModUnfold` (Algorithm 2) in the repeated Phase 1, we check (18) and reset the resolution if necessary (19).

5. SIMULATION RESULTS

We consider $K = 10$ parallel mod-ADCs, operating blindly according to the algorithm described above, whose inputs are the noisy mixtures,

$$\boldsymbol{x}_n = \boldsymbol{\Gamma}\boldsymbol{s}_n + \boldsymbol{\xi}_n \in \mathbb{R}^{K\times 1},$$

[3] We use in (17) the notation $\widehat{\bar{\boldsymbol{v}}}_n$ rather than $\widehat{\boldsymbol{x}}_n$, since in this context we trying to perfectly recover $\bar{\boldsymbol{v}}_n$, rather than to estimate x_n.

[4] Holds when $\{\bar{v}_n^k\}$ are sub-Gaussian random variables.

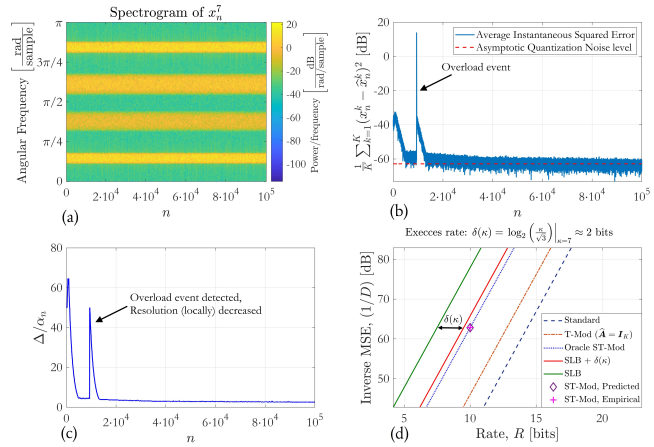

Fig. 2. Input spectrogram, simulation results, and rate-distortion curves.

where $\boldsymbol{s}_n \in \mathbb{R}^{K_s\times 1}$ is a collection of $K_s = 4$ zero-mean, unit-variance spectrally-flat uncorrelated Gaussian bandlimited sources, $\boldsymbol{\Gamma} \in \mathbb{R}^{K\times K_s}$ represents the (spatial) channel response, and $\boldsymbol{\xi}_n$ is an additive white Gaussian noise, with $\mathbb{E}\left[\boldsymbol{\xi}_n\boldsymbol{\xi}_{n-\ell}^{\mathrm{T}}\right] = \mathbb{1}_{\ell=0}\cdot\sigma^2\boldsymbol{I}_K$. We set $R = 10$, $\alpha_0 = \frac{\Delta}{5K}$, $\kappa = 7$, $p = 30$, $L_s = 2.5p$, and the entries of $\boldsymbol{\Gamma}$ were drawn independently from the standard Gaussian distribution. Other system parameters, such as μ and N_s, were set according to the same guiding principles presented in [12], Subsection IV-D. We define SNR $\triangleq 1/\sigma^2$.

For an input of length $N = 10^5$, Fig. 2(a) presents the spectrogram of x_n^7 (as a representative example), a noisy linear mixture of the sources $\boldsymbol{s}_n$, in which the bands of the sources are clearly seen. Fig. 2(b) presents the average instantaneous squared error in dB of $\widehat{\boldsymbol{x}}_n$, attained by our blind mod-ADCs algorithm for SNR $= 30$dB. The empirical error probability in perfectly recovering $\boldsymbol{v}_n$ is $\widehat{\Pr}(\boldsymbol{v}_n \neq \widehat{\boldsymbol{v}}_n) \triangleq \frac{1}{N}\sum_{\ell=1}^{N}\mathbb{1}_{\boldsymbol{v}_n\neq\widehat{\boldsymbol{v}}_n} = 0.001\%$, which allows for accurate estimation of $\boldsymbol{x}_n$, asymptotically matching the limiting quantization noise level (governed by the resolution parameter α_n). Fig. 2(c) presents the effective modulo size Δ/α_n, which is inversely proportional to the resolution parameter. It is seen that due to successful learning of the optimal filter $\widehat{\boldsymbol{H}}_n$ and the decoding matrix $\widehat{\boldsymbol{A}}$, the resolution parameter gradually increases, and converges to an asymptotic value near the theoretical limit for which $\kappa\cdot\sigma_{\max} = \frac{\Delta}{2}$ (see (16)), where $\sigma_{\max}$ is a function of the resolution parameter. Lastly, Fig. 2(d) presents the rate-distortion curves of standard ADCs, individual blind mod-ADCs, exploiting only temporal correlations (T-Mod), which corresponds to the special case $\widehat{\boldsymbol{A}} = \boldsymbol{I}_K$ (i.e., without IF decoding), and the predicted, as well as the empirical, rate-distortion operating point of the proposed spatiotemporal mod-ADCs with $R = 10$ at the asymptotic value of α_n (ST-Mod). We also present the curve of the oracle decoder (Algorithm 1) and Shannon's Lower Bound (SLB, [10]) as theoretical benchmarks. Our blind method achieves the oracle performance, and the gap from SLB is governed mainly by κ, controlling the overload probability as implied by (16) (see also [12], Subsection IV-E), with a small gap due to IF sub-optimality. The substantial gain relative to the competing methods is evident.

6. CONCLUSION

In the context of analog-to-digital conversion, we presented an adaptive algorithm, allowing for a stable and reliable blind operation of K parallel mod-ADCs, i.e., without access to prior knowledge of the input signals. Specifically, we extended our recent work, in which only temporal correlations were exploited, to the more general case where spatiotemporal correlations are utilized for accurate estimation and subsequent decoding. We demonstrated in simulation the successful operation of the proposed solution, corroborating our derivation and the underlying theory.

Authorized licensed use limited to: MIT Libraries. Downloaded on June 22,2022 at 17:23:37 UTC from IEEE Xplore. Restrictions apply.

7. REFERENCES

[1] Erik G. Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[2] Lu Lu, Geoffrey Ye Li, A. Lee Swindlehurst, Alexei Ashikhmin, and Rui Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[3] Thomas L. Marzetta, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[4] David L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] Zhilin Zhang and Bhaskar D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[6] Ali Ahmed and Justin Romberg, "Compressive sampling of ensembles of correlated signals," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1078–1098, 2019.

[7] Ali Ahmed and Fahad Shamshad, "Sub-Nyquist sampling of sparse and correlated signals in array processing," *arXiv preprint arXiv:1910.08792*, 2019.

[8] Nir Shlezinger, Salman Salamatian, Yonina C. Eldar, and Muriel Médard, "Joint sampling and recovery of correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 385–389.

[9] Robert H Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, 1999.

[10] Or Ordentlich, Gizem Tabak, Pavan Kumar Hanumolu, Andrew C. Singer, and Gregory W. Wornell, "A modulo-based architecture for analog-to-digital conversion," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 825–840, 2018.

[11] Luis G. Ordoñez, Paul Ferrand, Melissa Duarte, Maxime Guillaud, and Ganghua Yang, "On full-duplex radios with modulo-ADCs," *IEEE Open J. Commun. Soc.*, 2021.

[12] Amir Weiss, Everest Huang, Or Ordentlich, and Gregory W. Wornell, "Blind modulo analog-to-digital conversion," *arXiv preprint arXiv:2108.08937*, 2021.

[13] Feifei Sun, Jaspreet Singh, and Upamanyu Madhow, "Automatic gain control for ADC-limited communication," in *2010 IEEE Global Telecommunications Conference*, 2010, pp. 1–5.

[14] Stanley P. Lipshitz, Robert A. Wannamaker, and John Vanderkooy, "Quantization and dither: A theoretical survey," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 355–375, 1992.

[15] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar, "On unlimited sampling," in *Proc. Int. Conf. Sampling Theory Appl.*, 2017, pp. 31–35.

[16] Elad Romanov and Or Ordentlich, "Above the Nyquist rate, modulo folding does not hurt," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1167–1171, 2019.

[17] Or Ordentlich and Uri Erez, "Integer-forcing source coding," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1253–1269, 2016.

[18] Amin Sakzad, J. Harshan, and Emanuele Viterbo, "Integer-forcing MIMO linear receivers based on lattice reduction," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4905–4915, 2013.

[19] Liqin Ding, Kimmo Kansanen, Yang Wang, and Jiliang Zhang, "Exact SMP algorithms for integer-forcing linear MIMO receivers," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6955–6966, 2015.

[20] Nuo Huang, Xiaodong Wang, and Ming Chen, "Transceiver design for MIMO VLC systems with integer-forcing receivers," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 1, pp. 66–77, 2017.

[21] Arjen K. Lenstra, Hendrik Willem Lenstra, and László Lovász, "Factoring polynomials with rational coefficients," *Mathematische annalen*, vol. 261, pp. 515–534, 1982.

[22] Simon S. Haykin and Bernard Widrow, *Least-mean-square adaptive filters*, vol. 31, Hoboken, NJ: Wiley, 2003.

[23] Arie Feuer and Ehud Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust, Speech, Signal Process.*, vol. 33, no. 1, pp. 222–230, 1985.

Authorized licensed use limited to: MIT Libraries. Downloaded on June 22,2022 at 17:23:37 UTC from IEEE Xplore. Restrictions apply.