Asymptotic Task-Based Quantization With Application to Massive MIMO

Nir Shlezinger, *Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, and Miguel R. D. Rodrigues, *Senior Member, IEEE*

***Abstract*—Quantizers take part in nearly every digital signal processing system that operates on physical signals. They are commonly designed to accurately represent the underlying signal, regardless of the specific task to be performed on the quantized data. In systems working with high-dimensional signals, such as massive multiple-input multiple-output (MIMO) systems, it is beneficial to utilize low-resolution quantizers, due to cost, power, and memory constraints. In this paper, we study quantization of high-dimensional inputs, aiming at improving performance under resolution constraints by accounting for the system task in the quantizers design. We focus on the task of recovering a desired signal statistically related to the high-dimensional input, and analyze two quantization approaches. We, first, consider vector quantization, which is typically computationally infeasible, and characterize the optimal performance achievable with this approach. Next, we focus on practical systems that utilize hardware-limited scalar uniform analog-to-digital converters (ADCs), and design a task-based quantizer under this model. The resulting system accounts for the task by linearly combining the observed signal into a lower dimension prior to quantization. We then apply our proposed technique to channel estimation in massive MIMO networks. Our results demonstrate that a system utilizing low-resolution scalar ADCs can approach the optimal channel estimation performance by properly accounting for the task in the system design.**

***Index Terms*—Massive MIMO, quantization, hybrid receivers.**

I. Introduction

DIGITAL signal processing and communications systems use quantized representations of continuous-amplitude physical quantities [1]. These digital representations are typically designed to accurately match the original analog signal, by minimizing some distortion measure between the analog signal and the digital representation [2], regardless of the task of the system. Nonetheless, in many cases, the system task is not to recover the analog signal, but to extract some other information from its quantized representation [3]. It is therefore possible that in such systems – which we refer to as *task-based quantizers* – one can obtain further performance improvements in terms of the quantization rate necessary to achieve a certain performance.

Practical quantizers typically utilize scalar uniform analog-to-digital convertors (ADCs) [1]. Recent years have witnessed a growing interest in systems operating with quantized large-scale vectors obtained using low-resolution scalar ADCs. One of the main applications considered is massive multiple-input multiple-output (MIMO) communications [4]–[18], which is a key technology for the realization of next generation wireless networks [19]. In such systems, a wireless base station (BS) is equipped with a large number of antennas [20]–[22]. The BS first quantizes the received signal using a set of ADCs, commonly implementing scalar uniform quantization. Then, the quantized representation is used to estimate the underlying channel [4]–[10] and/or recover the transmitted messages [5]–[17]. For large-scale inputs, i.e., large number of BS antennas, accurate quantizers become costly in terms of power and memory usage, particularly when utilizing a large bandwidth, making low-resolution quantization essential for realizing massive MIMO systems [19]. As the task in massive MIMO is not to recover the input signal, but to estimate the channel or decode the transmitted message, reasonable performance with low-resolution scalar quantizers has been observed [4]–[18]. However, most prior works assume that the quantizers are fixed, commonly assuming one-bit sign quantization [5], [6], [9], [16]. Thus, they do not characterize the achievable performance when the quantizers are designed to account for the system task.

In the presence of multivariate inputs, joint (vector) quantization is known to outperform scalar quantization [23, Ch. 10]. Task-based vector quantization can be considered as an indirect lossy-source coding setup [2]. In such scenarios, one wishes to recover a desired source based on a discrete representation of its noisy version, in the sense of minimizing a given distortion measure [24]. For the mean-squared error (MSE) distortion, it was shown in [25] that the optimal system which achieves the rate-distortion curve, namely, uses the minimal number of bits per input sample required to achieve a fixed distortion, applies vector quantization to the minimum MSE (MMSE) estimate of the desired source. This observation was used in [26], [27] to study sampling and vector quantization of continuous-time signals. Nonetheless, in the presence of high-dimensional inputs, vector quantization becomes infeasible, so that practical task-based quantization approaches are required.

Task-based quantization with scalar uniform ADCs, referred to as *hardware-limited task-based quantization*, can be realized by allowing analog linear processing prior to quantization [28]. MIMO communications systems utilizing both analog and

Manuscript received November 25, 2018; revised March 7, 2019 and May 19, 2019; accepted June 5, 2019. Date of publication June 14, 2019; date of current version July 3, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marios Kountouris. This work was supported in part by the European Unions Horizon 2020 Research and Innovation Program under Grant 646804-ERC-COG-BNYQ, in part by the Israel Science Foundation under Grant 0100101, and in part by the Royal Society International Exchange Scheme IE 160348. This paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, U.K., May 2019. *(Corresponding author: Nir Shlezinger.)*

N. Shlezinger and Y. C. Eldar are with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: nirshlezinger1@gmail.com; yonina@weizmann.ac.il).

M. R. D. Rodrigues is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: m.rodrigues@ucl.ac.uk).

Digital Object Identifier 10.1109/TSP.2019.2923149

digital processing are known as *hybrid architectures* [12], [29], [30], and are the focus of a large amount of recent works. In particular, [31] compared the achievable-rate versus power efficiency tradeoff for various analog combining systems, [12] and [32] designed hybrid architectures aimed at maximizing the achievable rate and signal recovery MSE, respectively, with full channel state information (CSI), while [13] studied bit allocation for minimizing the quantization error when the analog combining is set to the largest channel eigenmodes, using high rate quantization analysis. Additionally, [14] studied the achievable rate with imperfect CSI when distinct sets of inputs are each combined in the analog domain to maximize the receive power, while [33] characterized bounds on the capacity of MIMO communications with analog combining and one-bit quantizers. Most previous works which designed hybrid MIMO receivers, e.g., [12], [13], [32], considered finite-size inputs and required CSI in their design, and thus cannot be utilized for massive MIMO channel estimation. Specifically, the joint design of analog combining, quantization rule, and digital processing, to optimize the accuracy of massive MIMO channel estimation with scalar ADCs has not yet been studied, to the best of our knowledge.

In this work we study task-based quantization for channel estimation in massive MIMO systems operating with scalar ADCs. Our analysis is based on an extension of the hardware-limited task-based quantization framework proposed in our previous work [28], which studied parameter estimation from a finite-sized quantized observed signal. The work [28] proposed to jointly optimize the analog combining, quantization rule, and digital processing, to minimize the MSE in recovering the desired finite-sized vector. Here, we extend the study of [28] to account for asymptotically large data, developing a framework for task-based quantization with high-dimensional inputs, and then apply the resulting analysis to massive MIMO systems, which are commonly studied in the asymptotic number of antennas regime [20], [21]. In particular, we focus on massive MIMO channel estimation, carried out in a time-division duplex (TDD) manner [20]–[22]. Unlike previous works on hybrid architectures optimization with low-resolution quantization, e.g., [12], [13], [32], our work does not require knowledge of the channel. In fact, in the presence of adjustable analog combining hardware, such as dynamic metasurface antennas [34], our analysis can be combined with previously proposed hybrid systems by reconfiguring the analog combining hardware once the channel is accurately estimated. We also note that our analysis can be applied to different tasks, such as signal recovery and noise mitigation.

We begin by studying task-based vector quantization using indirect lossy source coding theory. We characterize the minimal achievable average MSE for any quantization system operating with a fixed quantization rate, namely, a fixed number of bits per input sample. Then, we study the performance when vector quantization is carried out independently from the task, referred to as *task-ignorant vector quantization*. Since the input dimensionality here is asymptotically large, we are able to explicitly obtain the achievable performance, unlike [28], using indirect rate-distortion theory. Studying vector quantizers allows us to quantify the performance bounds of task-based quantization with large-scale inputs, and in particular, understand the fundamental limits of massive MIMO channel estimation.

Next, we study task-based quantization with scalar uniform ADCs, allowing analog combining prior to quantization. While analog combining can contribute in aspects other than improving the performance with finite-resolution quantizers, e.g., reducing the number of costly RF chains in massive MIMO systems [32], we focus here on the achievable performance for a given quantization rate. For this setup we propose a task-based quantization system with linear analog and digital processing which minimizes the average MSE under such hardware-limited structure constraints. We show that, unlike in the fixed size regime studied in [28], for large-scale inputs an important parameter which greatly affects the system performance is the *analog combining ratio*, which determines how the number of scalar quantizers grows as the input size tends to infinity.

Then, we focus on massive MIMO systems, and show how the proposed task-based quantization system can be applied to channel estimation from quantized measurements. We note that in this scenario the inputs are gathered over different antennas as well as over different time instances. Since in some cases, it may be desirable to combine only samples received at the same time instance, to avoid introducing delays in the analog domain, we also derive the system which minimizes the average MSE subject to the constraint that only inputs taken at the same time instance can be combined. This constraint reduces the complexity of the resulting system at the cost of degraded MSE performance. In our numerical study, we illustrate the fundamental performance limits of massive MIMO channel estimation achievable using vector quantizers, and compare these limits to our proposed task-based quantization systems with scalar ADCs, and to massive MIMO channel estimators which operate only in the digital domain. Our results demonstrate that the proposed quantizers, which utilize practical low-resolution scalar ADCs, are capable of approaching the optimal performance, achievable using vector quantizers, and outperform previously proposed estimators.

The rest of this paper is organized as follows: Section II reviews some basics in quantization theory. Section III extends the results of [28] to large-scale data, and Section IV applies them to massive MIMO channel estimation. Section V provides simulation examples. Finally, Section VI concludes the paper.

Throughout the paper, we use boldface lower-case letters for vectors, e.g., $\boldsymbol{x}$; the ith element of $\boldsymbol{x}$ is written as $(\boldsymbol{x})_i$. Matrices are denoted with boldface upper-case letters, e.g., $\boldsymbol{M}$, and we use $(\boldsymbol{M})_{i,j}$ to denote its (i,j)th element. We use $\boldsymbol{I}_n$ to denote the $n \times n$ identity matrix. Sets are expressed with calligraphic letters, e.g., $\mathcal{X}$, and $\mathcal{X}^n$ is the nth order Cartesian power of $\mathcal{X}$. Hermitian transpose, transpose, complex conjugate, stochastic expectation, and mutual information are written as $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, $\mathbb{E}\{\cdot\}$, and $I(\cdot;\cdot)$, respectively. For a real number a, we use $a^+ \triangleq \max(a,0)$; $\langle\cdot\rangle$ denotes the integer divisor (plus one) of the value in the brackets (minus one), namely, $\langle n \rangle_m \triangleq \lfloor \frac{n-1}{m} \rfloor + 1$. We use $\mathrm{Tr}\,(\cdot)$ to denote the trace operator, $\delta_{(\cdot)}$ is the indicator function, $\otimes$ is the Kronecker product, $\mathcal{R}$ and $\mathcal{C}$ are the sets of real and complex numbers, respectively. All logarithms are taken to base-2. Finally, for an $n \times n$ matrix $\boldsymbol{X}$, $\boldsymbol{x} = \mathrm{vec}\,(\boldsymbol{X})$ is the $n^2 \times 1$ vector obtained by stacking the columns of $\boldsymbol{X}$.

II. Preliminaries in Quantization Theory

To formulate the task-based quantization setup. we first briefly review standard quantization notions. While parts of this review also appear in our previous work [28], it is included for completeness. We begin with the definition of a quantizer:

Definition 1 (Quantizer): A quantizer $Q_M^{N,K}(\cdot)$ with $\log M$ bits, input size N, input alphabet $\mathcal{X}$, output size K, and output alphabet $\hat{\mathcal{X}}$, consists of: *1)* An encoding function $g_N^{\mathrm{e}} : \mathcal{X}^N \mapsto \{1,2,\ldots,M\} \triangleq \mathcal{M}$ which maps the input from $\mathcal{X}^N$ into a

discrete index $i \in \mathcal{M}$. *2)* A decoding function $g_K^{\rm d}: \mathcal{M} \mapsto \hat{\mathcal{X}}^K$ which maps each index $i \in \mathcal{M}$ into a codeword $\boldsymbol{q}_i \in \hat{\mathcal{X}}^K$.

The quantizer output for input $\boldsymbol{x}^N = \{\boldsymbol{x}_i\}_{i=1}^N \in \mathcal{X}^N$ is $\hat{\boldsymbol{x}}^K = g_K^{\rm d}\left(g_N^{\rm e}\left(\boldsymbol{x}^N\right)\right) \triangleq Q_M^{N,K}\left(\boldsymbol{x}\right)$. *Scalar quantizers* operate on a scalar input, i.e., $N=1$ and $\mathcal{X}$ is a scalar space, while *vector quantizers* have a multivariate input. Note that when $\mathcal{X}$ is a vector space, then each $\boldsymbol{x}_i$ is a random vector. When the input size and output size are equal, namely, $N=K$, we write $Q_M^N\left(\cdot\right) \triangleq Q_M^{N,N}\left(\cdot\right)$.

In the standard quantization problem, a $Q_M^N\left(\cdot\right)$ quantizer is designed to minimize some distortion measure $d_N: \mathcal{X}^N \times \hat{\mathcal{X}}^N \mapsto \mathcal{R}^+$ between its input and its output. The performance of a quantizer is therefore characterized using two measures: The quantization rate, defined as $R \triangleq \frac{1}{N}\log M$, and the expected distortion $\mathbb{E}\{d_N(\boldsymbol{x}^N, \hat{\boldsymbol{x}}^N)\}$. For a fixed input size N and codebook size M, the optimal quantizer is given by

$$Q_M^{N,\rm opt}\left(\cdot\right) = \min_{Q_M^N\left(\cdot\right)} \mathbb{E}\left\{d_N\left(\boldsymbol{x}^N, Q_M^N\left(\boldsymbol{x}^N\right)\right)\right\}. \tag{1}$$

Characterizing the optimal quantizer via (1) and the optimal tradeoff between distortion and quantization rate is in general a very difficult task. Consequently, optimal quantizers are typically studied assuming either high quantization rate, i.e., $R \to \infty$, see, e.g., [35], or asymptotically large input size, namely, $N \to \infty$, typically with stationary inputs, via rate-distortion theory [23, Ch. 10]. For example, when the quantizer input represents a stationary source, and the distortion measure is subadditive, i.e., for any $N_1, N_2, \boldsymbol{x}^{N_1} \in \mathcal{X}^{N_1}, \hat{\boldsymbol{x}}^{N_1} \in \mathcal{X}^{N_1}, \boldsymbol{x}^{N_2} \in \mathcal{X}^{N_2}, \hat{\boldsymbol{x}}^{N_2} \in \mathcal{X}^{N_2}$, it holds that $d_{N_1+N_2}(\{\boldsymbol{x}^{N_1}, \boldsymbol{x}^{N_2}\}, \{\hat{\boldsymbol{x}}^{N_1}, \hat{\boldsymbol{x}}^{N_2}\}) \leq d_{N_1}(\boldsymbol{x}^{N_1}, \hat{\boldsymbol{x}}^{N_1}) + d_{N_2}(\boldsymbol{x}^{N_2}, \hat{\boldsymbol{x}}^{N_2})$. Then, by [36, Thm. 5.9.1] the optimal distortion in the limit $N \to \infty$ for a fixed rate R is given by the distortion-rate function:

Definition 2 (Distortion-rate function): The distortion-rate function for a stationary source $\{\boldsymbol{x}_i\}_{i=1}^\infty$ with respect to the subadditive distortion measure d_N is defined as

$$D_{\boldsymbol{x}}\left(R\right) = \lim_{N\to\infty} \min_{f_{\hat{\boldsymbol{x}}^N|\boldsymbol{x}^N}: \frac{1}{N} I\left(\hat{\boldsymbol{x}}^N; \boldsymbol{x}^N\right) \leq R} \frac{1}{N}\mathbb{E}\left\{d_N\left(\hat{\boldsymbol{x}}^N, \boldsymbol{x}^N\right)\right\}. \tag{2}$$

The minimization in (2) is carried out over all conditional distributions $f_{\hat{\boldsymbol{x}}^N|\boldsymbol{x}^N}$ which satisfy the given constraint on the resulting mutual information. The marginal output distribution of $\{\hat{\boldsymbol{x}}_i\}$ which obtains the minima in (2) is referred to henceforth as the *optimal marginal distortion-rate distribution*. One scenario where $D_{\boldsymbol{x}}\left(R\right)$ is given in closed-form is when each $\boldsymbol{x}_i$ is a zero-mean $L \times 1$ proper-complex Gaussian random variable (RV) [37, Def. 1], i.e., $\mathcal{X} = \mathcal{C}^L$, such that for each $l \in \{1, 2, \ldots, L\}$, the source $\{(\boldsymbol{x}_i)_l\}_{i=1}^\infty$ is stationary[1] with scalar power spectral density (PSD) $s_{\boldsymbol{x}}: [0, 2\pi) \mapsto \mathcal{R}^+$, thus its multivariate PSD is $\boldsymbol{S}_{\boldsymbol{x}}(\cdot) = \mathbb{E}\{\boldsymbol{x}_i \boldsymbol{x}_i^H\} s_{\boldsymbol{x}}(\cdot)$. The distortion-rate function for this scenario is given in the following example:

Example 1: Let $\{\boldsymbol{x}_i\}_{i=1}^\infty$ be zero-mean proper-complex $L \times 1$ Gaussian source with multivariate PSD $\boldsymbol{S}_{\boldsymbol{x}}(\omega) = \boldsymbol{\Sigma}_{\boldsymbol{x}} s_{\boldsymbol{x}}(\cdot)$, and let the eigenvalue decomposition of $\boldsymbol{\Sigma}_{\boldsymbol{x}} \in \mathcal{C}^{L\times L}$ be given by $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{\Lambda}_{\boldsymbol{x}} \boldsymbol{U}_{\boldsymbol{x}}^H$. The distortion-rate function for $\boldsymbol{x}$ with the MSE distortion is [38, Cor. 1]

$$D_{\rm G}(R, \boldsymbol{\Sigma}_{\boldsymbol{x}}, s_{\boldsymbol{x}}) = \frac{1}{2\pi}\int_0^{2\pi} \sum_{i=1}^L \min\left(\zeta, (\boldsymbol{\Lambda}_{\boldsymbol{x}})_{i,i}\, s_{\boldsymbol{x}}(\omega)\right) d\omega, \tag{3a}$$

where $\zeta > 0$ is the solution to

$$R = \frac{1}{2\pi}\int_0^{2\pi} \sum_{i=1}^L \left(\log \frac{(\boldsymbol{\Lambda}_{\boldsymbol{x}})_{i,i}\, s_{\boldsymbol{x}}(\omega)}{\zeta}\right)^+ d\omega. \tag{3b}$$

The optimal marginal distribution for this setup is a zero-mean proper-complex multivariate Gaussian distribution with PSD $\boldsymbol{S}_{\hat{\boldsymbol{x}}}(\omega) = \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{\Lambda}_{\hat{\boldsymbol{x}}}(\omega) \boldsymbol{U}_{\boldsymbol{x}}^H$, where $\boldsymbol{\Lambda}_{\hat{\boldsymbol{x}}}(\omega)$ is a diagonal matrix with diagonal entries $(\boldsymbol{\Lambda}_{\hat{\boldsymbol{x}}}(\omega))_{i,i} = \left((\boldsymbol{\Lambda}_{\boldsymbol{x}})_{i,i}\, s_{\boldsymbol{x}}(\omega) - \zeta\right)^+$.

Comparing high rate analysis for scalar quantizers and rate-distortion theory for vector quantizers demonstrates the sub-optimality of serial scalar quantization. For example, for quantizing a large-scale real-valued Gaussian random vector with i.i.d. entries and sufficiently large quantization rate R, where one would imagine there is little benefit in quantizing the entries jointly over quantizing each entry independently, vector quantization notably outperforms serial scalar quantization [39, Ch. 23.2].

Finally, we introduce the notion of *dithered quantization*, which will be frequently used in our analysis of hardware-limited task-based quantization systems:

Definition 3 (Dithered quantizer): A scalar quantizer Q_M^1 implements serial non-subtractive uniform dithered quantization [40], referred to henceforth as *dithered quantization*, with support γ and quantization spacing $\Delta = \frac{2\gamma}{M}$, if its output for an input sequence $y_1, y_2, \ldots, y_P$ can be written as $Q_M^1\left(y_i\right) = q\left(\mathrm{Re}\left\{y_i + z_i\right\}\right) + j \cdot q\left(\mathrm{Im}\left\{y_i + z_i\right\}\right)$. Here, $z_1, \ldots, z_P$ are i.i.d. RVs with i.i.d. real and imaginary parts uniformly distributed over $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$, mutually independent of the input, and $q(\cdot)$ implements uniform quantization defined as

$$q(\alpha) = \begin{cases} -\gamma + \Delta\left(l + \frac{1}{2}\right) & \alpha - l\cdot\Delta + \gamma \in [0, \Delta] \\ & l \in \{0, 1, \ldots, M-1\} \\ \mathrm{sign}\left(\alpha\right)\left(\gamma - \frac{\Delta}{2}\right) & |\alpha| > \gamma. \end{cases}$$

Note that when $M=2$, the uniform quantizer $q(y)$ is a standard one-bit sign quantizer of the form $q(\alpha) = c \cdot \mathrm{sign}(\alpha)$, where the $c > 0$ is determined by the support γ.

In the following we study hardware-limited systems assuming dithered quantizers. Our motivation for using dithered quantizers stems from the fact that conventional analysis of uniform quantizers, e.g., [41], does not lead to a tractable model for the quantizer output, nor does it extend to the task-based setup. However, when using dithered quantizers, the digital representation of an input which is in the support of the quantizer can be written as the sum of the quantizer input and an additive uncorrelated white noise signal [40]. This significantly facilitates our analysis and allows to characterize the system which minimizes the MSE. Nonetheless, it is emphasized that this property of dithered quantizers is also *approximately* satisfied in uniform quantization *without dithering* for various input distributions, including Gaussian inputs[2] [42]. Therefore, the rigorous analysis which follows from considering dithered quantization, also holds

[1] Following [36], we use the term *stationary source* for stationary and ergodic signals with time index $i = \{1, 2, \ldots\}$.

[2] For a Gaussian input with magnitude smaller than γ with sufficiently high probability, if the quantization spacing is in the order of the input standard deviation (or smaller), then the output can be modeled as the input corrupted by additive uncorrelated white noise, even without dithering [42, Sec. VII].

TABLE I
MAIN MODEL NOTATIONS

Notation	Type	General Setup (Section III)	Massive MIMO Setup (Section IV)
N	Large integer	Number of observations	Number of antennas
K	Integer	Size of desired signal samples	Number of users in cell
L	Integer	Size of observation samples	Number of pilot symbols
$\underline{\boldsymbol{g}}$	$NK \times 1$ complex vector	Desired signal	Channel coefficients in vector form
$\underline{\boldsymbol{y}}$	$NL \times 1$ complex vector	Observations	Channel outputs in vector form
$\boldsymbol{\Gamma}$	$K \times L$ complex matrix	Linear MMSE matrix	Linear MMSE matrix
$\boldsymbol{\Sigma}_{\boldsymbol{y}}$	$L \times L$ complex matrix	Covariance of each observed sample	Temporal covariance of channel outputs
$\{\phi_i\}$	K real numbers	Singular values of $\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{1/2}$	Singular values of $\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{1/2}$
$c[\cdot]$	Mapping $\mathcal{Z} \mapsto \mathcal{R}$	Entry-wise correlation	Spatial correlation between antennas
$\boldsymbol{C}$	$N \times N$ complex matrix	Toeplitz matrix constructed from $c[\cdot]$	Spatial correlation matrix
$s(\cdot)$	Mapping $[0, 2\pi] \mapsto \mathcal{R}^+$	DTFT of $c[\cdot]$	Spatial PSD of each channel output
R	Real number	Quantization rate	Quantization rate
P	Integer	Number of scalar quantizers	Number of scalar quantizers
r	Real number	Analog combining ratio	Analog combining ratio
$\tilde{M}$	Integer	Number of scalar quantization regions	Number of scalar quantization regions
γ	Real number	Scalar quantizer support	Scalar quantizer support

approximately when using standard uniform quantizers without dithering, as demonstrated in [28].

III. Task-Based Quantization of Large-Scale Data

We now extend the analysis of task-based quantization carried out in our previous work [28], which considered fixed-size signals, to asymptotically large input signals. The motivation of this extension stems from the need to properly design and characterize quantizers for massive MIMO systems, which is our main target application discussed in Section IV. To that aim, we first present the problem formulation in Subsection III-A, and derive the achievable MSE without quantization constraints in Subsection III-B. Then, we study task-based quantization with vector quantizers in Subsection III-C and with hardware-limited quantizers in Subsection III-D. Focusing on the asymptotic regime allows us to rigorously characterize the achievable performance of vector quantizers, for which we were only able to obtain bounds in the finite horizon case studied in [28]. For the hardware-limited case, we formulate the dependency of task-based quantization systems on how the system parameters grow proportionally with the size of the input signal, i.e., the quantization rate and the analog combining ratio.

A. Problem Formulation

We study task-based quantization with asymptotically large observations and a proportionally large desired signal. The design objective of the quantizer is to quantize the observations such that the desired signal can be accurately recovered from the quantized observations in the sense of minimizing the MSE. The desired signal consists of N zero-mean $K \times 1$ random vectors $\{\boldsymbol{g}_i\}_{i=1}^N$, sampled from a stationary source with multivariate autocorrelation function $\mathbb{E}\{\boldsymbol{g}_{i+l}\boldsymbol{g}_i^H\} = \boldsymbol{\Sigma}_{\boldsymbol{g}} c[l]$, where $\boldsymbol{\Sigma}_{\boldsymbol{g}} \in \mathcal{C}^{K\times K}$ is Hermitian and positive semi-definite, while $c[\cdot]$ is an absolutely summable scalar autocorrelation function satisfying $c[0] = 1$. By letting $s(\cdot)$ be the discrete-time Fourier transform (DTFT) of $c[\cdot]$, the corresponding multivariate PSD is given by $\boldsymbol{\Sigma}_{\boldsymbol{g}} s(\cdot)$. The observations are a set of $L \times 1$ random vectors $\{\boldsymbol{y}_i\}_{i=1}^N$ with multivariate PSD $\boldsymbol{\Sigma}_{\boldsymbol{y}} s(\cdot)$, where $L \geq K$, and each vector $\boldsymbol{y}_i$ is related to its corresponding $\boldsymbol{g}_i$ via the same conditional probability measure, denoted $f_{\boldsymbol{y}|\boldsymbol{g}}$. The model assumption that the size of the desired signal is not larger than that of the observed signal allows us to clearly demonstrate the benefits of task-based quantization as noted in [28], and faithfully represent our main target application of channel estimation in massive MIMO systems discussed in Section IV.

We assume that the MMSE estimator which stems from $f_{\boldsymbol{y}|\boldsymbol{g}}$ is linear, i.e., there exists $\boldsymbol{\Gamma} \in \mathcal{C}^{K\times L}$ such that the MMSE estimate of $\boldsymbol{g}_i$ from $\{\boldsymbol{y}_{i'}\}$ can be written as $\tilde{\boldsymbol{g}}_i = \boldsymbol{\Gamma}\boldsymbol{y}_i$, for each $i \in \{1, \ldots, N\}$. Since we focus on large-scale data, N is arbitrarily large. Clearly, this setup specializes to the case in which the desired signal and the observed signal consist of i.i.d. elements. Such scenarios arise, for example, in signal recovery over memoryless channels, where $\boldsymbol{g}_i$ is the channel input at time index i and $\boldsymbol{y}_i$ is the corresponding channel output, or alternatively, in the estimation of fast fading memoryless channels, in which $\boldsymbol{g}_i$ is the unknown channel at time index i and $\boldsymbol{g}_i$ is the channel output. Furthermore, in Section IV we show that this model can also represent channel estimation in massive MIMO systems with correlated antennas.

We write the desired vector and the observed vector as $\underline{\boldsymbol{g}} = \mathrm{vec}([\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N]^T)$ and $\underline{\boldsymbol{y}} = \mathrm{vec}\big([\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]^T\big)$, respectively. By letting $\boldsymbol{C}$ be a $N \times N$ Toeplitz matrix whose entries are given by $(\boldsymbol{C})_{i_1,i_2} = c[i_1 - i_2]$ for each $i_1, i_2 \in \{1, \ldots, N\}$, it holds that the covariance matrices of $\underline{\boldsymbol{g}}$ and $\underline{\boldsymbol{y}}$ are equal to $\boldsymbol{\Sigma}_{\boldsymbol{g}} \otimes \boldsymbol{C}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}} \otimes \boldsymbol{C}$, respectively. The main model notations along with their meaning in the massive MIMO setup considered in Section IV are summarized in Table I. The proposed system forms a quantized representation of $\underline{\boldsymbol{g}}$ based on the observed $\underline{\boldsymbol{y}}$, using up to $\log M$ bits, where the quantization rate $R \triangleq \frac{1}{NK} \log M$ is fixed. An illustration of such a system is depicted in Fig. 1.

The distortion measure for a quantized representation $\hat{\underline{\boldsymbol{g}}}$ is the average MSE, defined as

$$\mu \triangleq \lim_{N\to\infty} \frac{1}{NK} \mathbb{E}\{\|\underline{\boldsymbol{g}} - \hat{\underline{\boldsymbol{g}}}\|^2\}. \tag{4}$$

We consider vector quantizers as well as hardware-limited quantizers. In the following we elaborate on these systems:

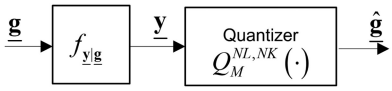

Fig. 1. Task-based quantization system.

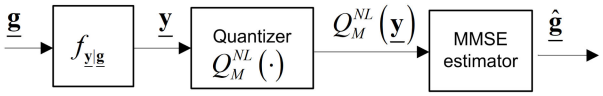

Fig. 2. Task-ignorant quantizer.

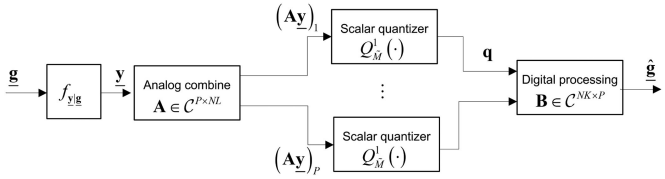

Fig. 3. Hardware-limited task-based quantization system.

Vector Quantizers: Joint (vector) quantization is known to be superior to separate (scalar) quantization [39, Ch. 23]. Thus, analyzing systems utilizing vector quantizers provides the fundamental limits of task-based quantization with large-scale inputs. We consider two different vector quantization systems:

1) **Task-based optimal vector quantization** - in the optimal quantization system, the quantizer $Q_M^{NL,NK}(\cdot)$ in Fig. 1 is the vector quantizer which minimizes the distortion between the quantized representation $\hat{\underline{\boldsymbol{g}}}$ and $\underline{\boldsymbol{g}}$. The performance of this system represents the optimal distortion achievable with any quantization system operating at rate R.
2) **Task-ignorant vector quantization** - here, the quantizer is designed to recover the observed $\underline{\boldsymbol{y}}$ separately from the task, using the optimal vector quantizer for representing $\underline{\boldsymbol{y}}$, namely, the quantizer here is ignorant of the task and is designed to accurately represent the observations. The desired vector $\underline{\boldsymbol{g}}$ is estimated from the quantized representation using the MMSE estimator, as illustrated in Fig. 2. This is a plausible system when the quantizer is ignorant of the task.

Hardware-Limited Quantizers: Vector quantization may be difficult to implement, especially for large input sizes. Consequently, systems utilizing vector quantizers may not be feasible in practice. As discussed in the introduction, practical systems typically implement quantization using scalar ADCs. In such systems, each continuous-amplitude element is converted into a discrete representation using a single quantization rule, which commonly corresponds to uniform quantization. This operation can be modeled using identical scalar uniform quantizers. In particular, we consider the system depicted in Fig. 3. The observed vector $\underline{\boldsymbol{y}}$, is projected into $\mathcal{C}^P$, where $P \leq NL$, using some pre-quantization processing carried out in the analog domain. As arbitrary processing may be difficult to implement in analog, we henceforth restrict our attention to linear pre-quantization processing only. This *analog combining* is modeled via the matrix $\boldsymbol{A} \in \mathcal{C}^{P \times NL}$. We write the number of scalar quantizers P in terms of its integer quotient and remainder with respect to N, denoted P_q and P_r, respectively, i.e.,

$$P = P_q \cdot N + P_r, \quad 0 < P_r < N. \tag{5}$$

The motivation for expressing P using N in (5) stems from the fact that for large-scale inputs, N tends to infinity, and thus P_q and P_r represent how P scales accordingly. These scaling parameters play an important role when analyzing the performance of hardware-limited task-based quantizers, as shown in Subsection III-D.

The real and imaginary parts of each entry of $\boldsymbol{A}\underline{\boldsymbol{y}}$ are quantized using the same scalar quantizer with resolution $\tilde{M} \triangleq \lfloor M^{1/2P} \rfloor$, denoted $Q_{\tilde{M}}^1(\cdot)$. Define the *analog combining ratio*

$$r \triangleq \frac{P}{NL} = \frac{P_q}{L} + \frac{P_r}{NL}. \tag{6}$$

Note that $\tilde{M} = \lfloor 2^{\frac{R}{2 \cdot r}} \rfloor$. The overall quantization rate is $\frac{2 \cdot P}{NL} \log(\tilde{M}) \leq \frac{1}{NL} \log M = R$. The identical scalar quantizers $Q_{\tilde{M}}^1$ implement dithered quantization, as defined in Def. 3. The quantizer is designed to operate within the support γ, namely, the amplitude of the input is not larger than γ with sufficiently large probability. To guarantee this, we fix γ to be some multiple η of the maximal standard deviation of the input. For example, for proper-complex Gaussian inputs, when $\eta \geq \sqrt{2}$ the amplitude of both the real and imaginary parts of the input are smaller than the support with probability over 94%. We assume that $\eta < \sqrt{3/2}\tilde{M}$, such that the variable $\kappa \triangleq \eta^2(1 - \frac{2\eta^2}{3\tilde{M}^2})^{-1}$ is strictly positive. Note that $\eta = 2$ satisfies this requirement for any $\tilde{M} \geq 2$, i.e., the ADC is implemented using scalar quantizers with at least one bit.

Finally, in the digital domain, the system approximates the linear MMSE estimate based on the output of the ADC, denoted $\boldsymbol{q} \in \mathcal{C}^P$, where $(\boldsymbol{q})_i = Q_{\tilde{M}}^1\big((\boldsymbol{A}\underline{\boldsymbol{y}})_i\big)$. Consequently, the estimate can be written as $\hat{\underline{\boldsymbol{g}}} = \boldsymbol{B}\boldsymbol{q}$ for some $\boldsymbol{B} \in \mathcal{C}^{NK \times P}$. We focus on linear digital processing to keep the analysis tractable, and since linear estimators are commonly used in our main application, massive MIMO channel estimation with quantized outputs [5], [7]. This restriction is not expected to notably affect the overall performance, especially when the error due to quantization is small, as the MMSE estimator in the considered setup is linear.

B. No Quantization Constraints

As a preliminary step, we note that the MMSE estimate of $\underline{\boldsymbol{g}}$ from $\underline{\boldsymbol{y}}$, denoted $\tilde{\underline{\boldsymbol{g}}}$ consists of the $K \times 1$ random vectors $\{\tilde{\boldsymbol{g}}_i\}_{i=1}^N$, sampled from a stationary source with multivariate PSD $\boldsymbol{S}_{\tilde{\boldsymbol{g}}}(\cdot) = \boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\Gamma}^H s(\cdot)$. Since $\frac{1}{2\pi}\int_0^{2\pi} s(\omega)d\omega = c[0] = 1$, the average MMSE can be written as

$$\mu^{\text{MMSE}} = \frac{1}{K}\text{Tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{g}} - \boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\Gamma}^H\right). \tag{7}$$

The MMSE in (7) is achievable without quantization, and thus serves as a lower bound on the achievable distortion of the quantization systems discussed in the following subsections.

C. Vector Quantization

We now study the average MSE achievable of the vector quantization systems detailed in Subsection III-A. We note that for fixed size inputs, the achievable performance of vector quantizers can only be obtained in terms of upper and lower bounds, see [28, Prop. 1]. However, as we show next, for large-scale data, we explicitly characterize the minimal achievable average MSE for each system using indirect rate-distortion theory analysis, which considers asymptotically large inputs.

1) Optimal Vector Quantizer: The optimal vector quantizer minimizes the MSE between the unknown desired vector and the system output. Recovering the desired signal $\underline{\boldsymbol{g}}$ from quantized observations is a special case of indirect lossy source coding

[24]. For the MSE distortion, it follows from [25] that the optimal vector quantizer first recovers the MMSE estimate $\tilde{\underline{\boldsymbol{g}}}$, and then uses a vector quantizer to represent $\tilde{\underline{\boldsymbol{g}}}$. The resulting MSE is given in the following theorem:

Theorem 1: The MSE of the optimal vector quantizer is

$$\mu^{\mathrm{Opt}} = \mu^{\mathrm{MMSE}} + \frac{1}{K} D_{\tilde{\underline{\boldsymbol{g}}}}\left(\frac{L}{K}\cdot R\right), \tag{8}$$

where $D_{\tilde{\boldsymbol{g}}}(\cdot)$ is the distortion-rate function, given in Def. 2, of the random vector $\tilde{\boldsymbol{g}}$ with the MSE distortion.

Proof: See Appendix A.

Theorem 1 holds since the MMSE estimate $\tilde{\underline{\boldsymbol{g}}}$ represents a stationary source, thus, in the limit $N \to \infty$, the minimal MSE for a fixed quantization rate is given by the distortion-rate function. The achievable average MSE in (8) constitutes the minimal achievable distortion of any system which recovers $\underline{\boldsymbol{g}}$ from $\underline{\boldsymbol{y}}$ using up to R bits per input sample.

2) Task-Ignorant Vector Quantizer: In task-ignorant quantization, the desired signal is estimated from the quantized observations, which are in turn designed to yield an accurate representation of the input signal. The resulting quantization system, depicted in Fig. 2, first quantizes $\underline{\boldsymbol{y}}$ via a quantizer $Q_M^{NL}(\cdot)$, which minimizes the MSE between its output and $\underline{\boldsymbol{y}}$. Then, $\underline{\boldsymbol{g}}$ is estimated from the output of the quantizer using the MMSE estimator. Characterizing the average MSE of such systems is in general a challenging task, due to difficulty in formulating the conditional distribution of the desired signal given the output of the quantizer $Q_M^{NL}(\cdot)$. However, in the special case where the signals are i.i.d., and thus $s(\omega) = 1$, the resulting average MSE is given in the following theorem:

Theorem 2: When $\{\boldsymbol{y}_i\}$ are i.i.d. the average MSE of the task-ignorant vector quantizer is given by

$$\mu^{\mathrm{Ign}} = \mu^{\mathrm{MMSE}} + \frac{1}{K}\mathrm{Tr}\left(\left(\boldsymbol{\Gamma}\right)^H \boldsymbol{\Gamma}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}} - \boldsymbol{\Sigma}_{\boldsymbol{y},D}(R)\right)\right). \tag{9}$$

Here, $\boldsymbol{\Sigma}_{\boldsymbol{y},D}(R)$ is the covariance matrix of the optimal marginal distribution which achieves the distortion-rate function $D_{\boldsymbol{y}}(R)$ with the MSE distortion, given in Def. 2.

Proof: See Appendix B.

Theorem 2 exploits the fact that when $\underline{\boldsymbol{y}}$ consists of N i.i.d. $L \times 1$ vectors, then, as N grows arbitrarily, the output of the optimal quantizer for representing $\underline{\boldsymbol{y}}$ converges to a set of N i.i.d. vectors, each distributed via the optimal marginal distribution which achieves $D_{\boldsymbol{y}}(R)$. In our numerical study in Section V it is illustrated that for relatively small quantization rates, there is a notable gap between the performance of task-ignorant quantization and the optimal average MSE in (8).

D. Hardware-Limited Quantization

We now characterize the optimal hardware-limited task-based quantization system, using the setup depicted in Fig. 3. We derive the analog combining matrix and digital processing matrix which minimize the average MSE, denoted $\boldsymbol{A}^{\mathrm{o}}$ and $\boldsymbol{B}^{\mathrm{o}}$, respectively, and the corresponding support γ.

To formulate the proposed system, define the $K \times L$ matrix $\tilde{\boldsymbol{\Gamma}} \triangleq \boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{1/2}$, and let $\{\phi_i\}$ be its singular values arranged in descending order. Note that for $i > \mathrm{rank}\big(\tilde{\boldsymbol{\Gamma}}\big)$, $\phi_i = 0$. Let $\{\lambda_i\}$ be the singular values of $\tilde{\boldsymbol{\Gamma}} \otimes \boldsymbol{C}$ arranged in descending order, and define the function $\varphi(\alpha) \triangleq \big(\alpha - 1\big)^+$, $\alpha \in \mathcal{R}^+$. Recall that κ is defined as $\kappa = \eta^2\big(1 - \frac{2\eta^2}{3\tilde{M}^2}\big)^{-1}$, where η is the ratio of the quantizer support to the maximal input standard deviation. The hardware-limited quantization system which minimizes the average MSE is stated in the following theorem:

Theorem 3: In the hardware-limited quantization system which minimizes the average MSE, the analog combining matrix $\boldsymbol{A}^{\mathrm{o}}$ is given by $\boldsymbol{A}^{\mathrm{o}} = \boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}(\boldsymbol{V}_{\boldsymbol{A}}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1/2} \otimes \boldsymbol{C}^{-1/2})$, where

- $\boldsymbol{V}_{\boldsymbol{A}} \in \mathcal{C}^{L\times L}$ is the right singular vectors matrix of $\tilde{\boldsymbol{\Gamma}}$.
- $\boldsymbol{\Lambda}_{\boldsymbol{A}} \in \mathcal{C}^{P\times NL}$ is a diagonal matrix with diagonal entries

$$(\boldsymbol{\Lambda}_{\boldsymbol{A}})_{l,l}^2 = \frac{4\kappa}{3\tilde{M}^2\cdot r}\varphi(\zeta\cdot\lambda_l), \tag{10a}$$

  where ζ is set such that $\frac{4\kappa}{3\tilde{M}^2\cdot P}\sum_{l=1}^{P}\varphi(\zeta\cdot\lambda_l) = 1$, r is defined in (6), and $\tilde{M} = \lfloor 2^{\frac{R}{2r}}\rfloor$.
- $\boldsymbol{U}_{\boldsymbol{A}} \in \mathcal{C}^{P\times P}$ is a unitary matrix which guarantees that $\boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}^H\boldsymbol{U}_{\boldsymbol{A}}^H$ has identical diagonal entries, which can be obtained via [54, Alg. 2.2].

The support of the ADC is given by $\gamma^2 = \frac{\kappa}{r}$, and the digital processing matrix is

$$\boldsymbol{B}^{\mathrm{o}} = \left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}} \otimes \boldsymbol{C}\right)\left(\boldsymbol{A}^{\mathrm{o}}\right)^H \times\left(\boldsymbol{A}^{\mathrm{o}}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}} \otimes \boldsymbol{C}\right)\left(\boldsymbol{A}^{\mathrm{o}}\right)^H + \frac{4\gamma^2}{3\tilde{M}^2}\boldsymbol{I}_P\right)^{-1}. \tag{10b}$$

The corresponding achievable average MSE at the limit $N \to \infty$ when $P_q \geq \mathrm{rank}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\Gamma}^H)$ is given by

$$\mu^{\mathrm{HL}} = \mu^{\mathrm{MMSE}} + \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{K}\sum_{i=1}^{K}\frac{\phi_i^2 s(\omega)}{\varphi(\zeta\cdot\phi_i\sqrt{s(\omega)})+1}d\omega. \tag{10c}$$

Furthermore, when the signals consists of uncorrelated vectors, i.e., $c[\tau] = \delta_\tau$, the asymptotic average MSE for any $P_q \geq 0$ reduces to

$$\mu^{\mathrm{HL}} = \mu^{\mathrm{MMSE}} + \frac{1}{K}\sum_{i=1}^{\min(K,P_q)}\frac{\phi_i^2}{\varphi(\zeta\cdot\phi_i)+1} + \delta_{(P_q<K)} \times\left(\frac{1}{K}\sum_{i=P_q+1}^{K}\phi_i^2 - (r\cdot L - P_q)\frac{\phi_{P_q+1}^2\varphi(\zeta\cdot\phi_{P_q+1})}{\varphi(\zeta\cdot\phi_{P_q+1})+1}\right). \tag{10d}$$

Proof: See Appendix C.

Theorem 3 extends [28, Thm. 1] to asymptotically large complex-valued inputs. A notable difference between Theorem 3 and [28, Thm. 1] is in the performance expression in (10c)–(10d): While [28, Thm. 1] studied the MSE with finite-size inputs, here we consider the asymptotic average MSE. Thus (10c)–(10d) depend on how the number of scalar quantizers grow with the input size, and not on the exact number of inputs and scalar quantizers.

Note that when P_r in (5) does not grow proportionally with N, i.e., $\lim_{N\to\infty}\frac{P_r}{N} = 0$, then by (5), $r\cdot L = P_q$, and the last summand in (10d) vanishes. When P_r equals zero, i.e., P is an integer multiple of N, and $c[\tau] = \delta_\tau$, the optimal system processes $\boldsymbol{y}_i$ using the same transformation for each $i \in \{1,\ldots,N\}$ separately, as stated in the following corollary:

Corollary 1: When $P_r = 0$ and $c[\tau] = \delta_\tau$, the hardware-limited system which minimizes the MSE applies the same mapping to each $\boldsymbol{y}_i$ separately. This mapping includes analog combining via the matrix $\boldsymbol{A}^{\mathrm{o}}$, scalar quantizers with support $\gamma^2 = \frac{\kappa}{r}$, and digital processing with matrix $\boldsymbol{B}^{\mathrm{o}}$. In particular, $\boldsymbol{A}^{\mathrm{o}} = \boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{V}_{\boldsymbol{A}}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1/2}$, where

- $\boldsymbol{V}_{\boldsymbol{A}} \in \mathcal{C}^{L\times L}$ is the right singular vectors matrix of $\tilde{\boldsymbol{\Gamma}}$.

- $\boldsymbol{\Lambda_A} \in \mathcal{C}^{P_q \times L}$ is diagonal with entries $(\boldsymbol{\Lambda_A})_{i,i}^2 = \frac{4\kappa \cdot \varphi(\zeta \cdot \phi_i)}{3\tilde{M}^2 \cdot P_q}$, where ζ is set such that $\frac{4\kappa}{3\tilde{M}^2 \cdot P_q} \sum_{i=1}^{P_q} \varphi(\zeta \cdot \phi_i) = 1$.
- $\boldsymbol{U_A} \in \mathcal{C}^{P_q \times P_q}$ is a unitary matrix for which $\boldsymbol{U_A \Lambda_A \Lambda_A^H U_A^H}$ has identical diagonal entries.

The matrix $\boldsymbol{B}^{\rm o} = \tilde{\boldsymbol{\Gamma}} \boldsymbol{V_A} \boldsymbol{\Lambda_A^H} (\boldsymbol{\Lambda_A \Lambda_A^H} + \frac{4\gamma^2}{3\tilde{M}^2} \boldsymbol{I}_{P_q})^{-1} \boldsymbol{U_A^H}$ represents the digital processing. The achievable average MSE is given by:

$$\mu^{\rm HL} = \mu^{\rm MMSE} + \frac{1}{K} \sum_{i=1}^{\min(K,P_q)} \frac{\phi_i^2}{\varphi(\zeta \cdot \phi_i) + 1} + \frac{\delta_{(P_q<K)}}{K} \sum_{i=P_q+1}^{K} \phi_i^2. \quad (11)$$

Proof: The corollary follows directly from Theorem 3. In particular, (11) and the requirement on ζ are obtained from Theorem 3 since $r \cdot L = P_q$ when $P = P_q \cdot N$. The resulting $\boldsymbol{A}^{\rm o}$ is a special case of $\boldsymbol{A}^{\rm o}$ in Theorem 3 for $P = P_q \cdot N$, and $\boldsymbol{B}^{\rm o}$ is obtained by plugging $\boldsymbol{A}^{\rm o} \otimes \boldsymbol{I}_N$ into (10b). ■

Corollary 1 is quite surprising in light of known results in vector quantization. It is well-known that with unrestricted vector quantizers, jointly processing a set of RVs is beneficial even if they are i.i.d. [39, Ch. 23]. However, Corollary 1 indicates that in the presence of scalar ADCs, if it is possible to process i.i.d. RVs using the same mapping separately, i.e., when $P_r = 0$ and the same number of scalar quantizers can be assigned to each $\boldsymbol{y}_i$, then this strategy minimizes the MSE.

Theorem 3 and Corollary 1 indicate that the analog combining ratio r, and particularly the value of P_q, play an important part in the performance of hardware-limited systems. Guidelines for setting these values are stated in the following corollary:

Corollary 2: In order to minimize the average MSE, P_q must not be larger than the rank of $\tilde{\boldsymbol{\Gamma}} \boldsymbol{\Sigma_y} \tilde{\boldsymbol{\Gamma}}^H$.

Proof: The proof is obtained by repeating the arguments in [28, Appendix D], and is thus omitted for brevity. ■

In order to compare the achievable average MSE in Theorem 3 to the fundamental limit in Theorem 1, one must specify the distribution of the observations, as we do in the following example:

Example 2: Consider the case where the MMSE estimate $\tilde{\boldsymbol{g}}$ has i.i.d. proper-complex Gaussian entries with variance $\sigma_{\tilde{g}}^2$. Here, the excess average MSE of the optimal vector quantizer of Theorem 1 is

$$\mu^{\rm Opt} - \mu^{\rm MMSE} = \frac{1}{K} D_G\left(\frac{L}{K} R, \sigma_{\tilde{g}}^2 \boldsymbol{I}_K, 1\right) \overset{(a)}{=} \sigma_{\tilde{g}}^2 2^{-\frac{L}{K}R}, \quad (12a)$$

where $D_G(\cdot)$ is defined in (3), and (a) follows from the distortion-rate function of Gaussian RVs [39, Ch. 23]. Next, we compute the excess average MSE of a hardware-limited quantizer with analog combining ratio $r = \frac{L}{K}$, namely, $P_r = 0$ and $P_q = K$. By noting that $\phi_i^2 = \sigma_{\tilde{g}}^2$ for each i, it follows from Corollary 1 that

$$\mu^{\rm HL} - \mu^{\rm MMSE} = \frac{\sigma_{\tilde{g}}^2}{\frac{3}{4\kappa}\tilde{M}^2 + 1} \overset{(a)}{=} \frac{\sigma_{\tilde{g}}^2}{\frac{3}{4\kappa}\lfloor 2^{-\frac{L}{2K} \cdot R} \rfloor^2 + 1}, \quad (12b)$$

where (a) holds as $r = \frac{L}{K}$. Note that (12a)–(12b) imply that as R increases, the ratio of the excess average MSEs satisfies

$$\frac{\mu^{\rm HL} - \mu^{\rm MMSE}}{\mu^{\rm Opt} - \mu^{\rm MMSE}} \approxeq \frac{4\kappa}{3} = \frac{4\eta^2}{3 - \frac{2\eta^2}{\tilde{M}^2}}. \quad (12c)$$

As we assume that the quantized input is within the support and each scalar quantizer uses at least one bit, i.e., $\eta \geq 2$ and $\tilde{M} \geq 2$, (12c) is strictly larger than one, as expected.

Example 2 shows that, when the MMSE estimate has i.i.d. entries, the excess average MSE of hardware-limited quantization with large-scale inputs scales with respect to the quantization rate R proportionally to the optimal vector quantizer. This indicates that the proposed hardware-limited quantization system can approach the optimal performance with an average MSE gap that becomes negligible as $\mu^{\rm Opt}$ approaches the average MMSE $\mu^{\rm MMSE}$. A similar relation to (12c) can be obtained for any distribution using the upper bound on the distortion-rate function in [47, Eq. (6)].

Although Example 2 focuses on the case where the MMSE estimate has i.i.d. entries, in the simulations study in Section V we demonstrate that the hardware-limited system of Theorem 3 can also approach the optimal MSE of Theorem 1 in massive MIMO channel estimation with quantized measurements, where the entries of the MMSE estimate are correlated. The application of our results to such setups is described in the following section.

IV. Application: Massive MIMO Channel Estimation

An important application of our study on task-based quantization with large-scale inputs is channel estimation in massive MIMO communications networks. Specifically, in massive MIMO systems, there is a strong need to operate with simple low-resolution quantizers, as increasing quantization rate results in a sharp increase in power consumption and memory usage. The problem of channel estimation from quantized measurements has received considerable attention, most notably in massive MIMO systems with large-scale inputs [4]–[7], but also for finite-scale inputs [44]–[46]. As discussed in the introduction, previous works on massive MIMO channel estimation focus only on the digital processing, while hybrid architectures utilizing analog combiners were designed assuming CSI [12], [13], [32]. By applying the analysis of Section III, we are able to jointly optimize both the analog and the digital processing to improve the channel estimation performance under a given quantization rate constraint.

In the following we first present the massive MIMO system model in Subsection IV-A. Then, we discuss the fundamental limits of massive MIMO channel estimation without quantization in Subsection IV-B. Finally, in Subsection IV-C we show how the results of Section III can be applied to characterize the achievable performance and design the corresponding massive MIMO channel estimators.

A. Massive MIMO System Model

We consider pilot-aided channel estimation in a multi-cell multi-user MIMO system with n_c cells. In each cell, a BS equipped with an array of equally-spaced N antennas serves K single-antenna user terminals (UTs). The antennas are not necessarily half-wavelength spaced, hence, the channel outputs can be spatially correlated. We focus on the *massive MIMO regime*, namely, the number of antennas N is sufficiently large to carry out large-scale (asymptotic) analysis.

The massive MIMO channel follows a block-fading model [20]. To formulate the model, let $\boldsymbol{D}_{l,m}$ be a $K \times K$ diagonal matrix with positive diagonal entries $\{d_{l,m,u}\}_{u=1}^K$ representing the attenuation between the uth UT of the mth cell and the lth BS, $l, m \in \{1, \ldots, n_c\} \triangleq \mathcal{N}_c$. Without loss of generality, we

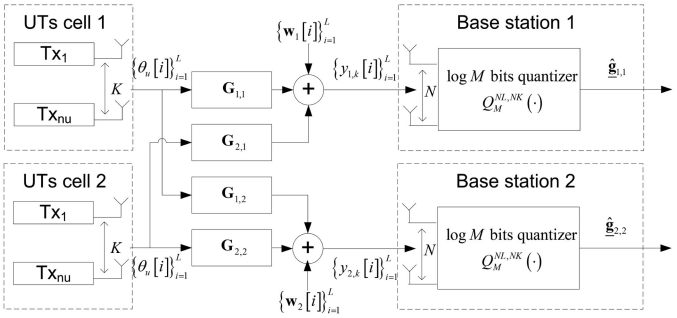

Fig. 4. Massive MIMO channel estimation with $n_c = 2$ cells.

assume that for each $l \in \mathcal{N}_c$, the coefficients $\{d_{l,l,u}\}_{u=1}^{K}$ are arranged in descending order. Furthermore, let $\boldsymbol{H}_{l,m} \in \mathcal{C}^{N\times K}$ be a random proper-complex zero-mean Gaussian matrix with i.i.d. entries of unit variance, representing the instantaneous channel response between the UTs of the mth cell and the lth BS, $l, m \in \mathcal{N}_c$. For each $(l_1, m_1) \neq (l_2, m_2)$, $\boldsymbol{H}_{l_1,m_1}$ and $\boldsymbol{H}_{l_2,m_2}$ are mutually independent, and we assume a block-fading model for $\{\boldsymbol{H}_{l,m}\}_{l,m\in\mathcal{N}_c}$. To account for coupling induced by antenna spacing, we use $\boldsymbol{C}_l \in \mathcal{C}^{N\times N}$ to model the receive side correlation, i.e., $\left(\boldsymbol{C}_l\right)_{k_1,k_2}$ represents the correlation between the antennas of indexes k_1 and k_2. Following conventional models for antenna coupling, e.g., Jakes model [48], the fact that the antennas are equally-spaced implies that $\boldsymbol{C}_l$ is a Toeplitz matrix with unit diagonal entries, and we write $c_l[k_1 - k_2] = \left(\boldsymbol{C}_l\right)_{k_1,k_2}$, and set $s_l(\cdot)$ to be the DTFT of $c_l[\tau]$. The overall random channel matrix from the UTs in the mth cell to the lth BS is given by $\boldsymbol{G}_{l,m} = \boldsymbol{C}_l^{1/2}\boldsymbol{H}_{l,m}\boldsymbol{D}_{l,m}$. Let $\boldsymbol{w}_l[i] \in \mathcal{C}^N$, $l \in \mathcal{N}_c$, be an i.i.d. zero-mean proper-complex Gaussian signal representing the additive channel noise at the lth BS. Due to the antenna coupling at the BS, the noise is also spatially correlated, and its covariance matrix is $\sigma_W^2\boldsymbol{C}_l$, with $\sigma_W^2 > 0$.

Channel estimation is carried out in a TDD fashion. Each UT sends a deterministic orthogonal pilot sequence (PS) consisting of L symbols, where the PSs are the same in all cells and known to the BSs. The BSs use the knowledge of the PSs to estimate the channel. Let $\theta_u[i]$ be the ith pilot symbol of the uth user in each cell, $u \in \{1,\ldots,K\} \triangleq \mathcal{K}$, $i \in \{1,\ldots,L\} \triangleq \mathcal{L}$. The channel output at the kth antenna of the lth BS at time instance $i \in \mathcal{L}$ is

$$y_{l,k}[i] = \sum_{m=1}^{n_c}\sum_{u=1}^{K}\left(\boldsymbol{G}_{l,m}\right)_{k,u}\theta_u[i] + \left(\boldsymbol{w}_l[i]\right)_k. \tag{13}$$

The orthogonality of the PSs implies that for all $l, m \in \mathcal{K}$, $\sum_{i=1}^{L}\theta_l[i]\theta_m^*[i] = L\cdot\delta_{m,k}$. Furthermore, the PS length, L, must not be smaller than the number of UTs, K [20, Sec. III-A]. Each BS uses up to $\log M$ bits to represent the received signal $\{y_{l,k}[i]\}$, from which an estimate of the corresponding channel in vector $\underline{\boldsymbol{g}}_{l,l} \triangleq \mathrm{vec}\left(\boldsymbol{G}_{l,l}\right)$, denoted $\hat{\underline{\boldsymbol{g}}}_{l,l}$, is produced. An illustration of the considered setup with $n_c = 2$ cells is depicted in Fig. 4.

Our goal is to derive the achievable average MSE in estimating the channel matrix at a given cell with index $l \in \mathcal{N}_c$, and to characterize the corresponding quantization scheme. As common in the massive MIMO literature, see, e.g., [20]–[22], we assume that the BS knows: 1) the pilot symbols; 2) the channel input-output relationship, i.e., that the channel output are obtained from the PS via (13); and 3) the statistical model of the channel and the noise. This knowledge is utilized in the design of the quantization system to facilitate the estimation of each realization of the channel. In our analysis, we fix the quantization rate, defined here as $R \triangleq \frac{1}{N\cdot L}\log M$, and derive the achievable MSE in the large number of antennas limit, $\mu_l \triangleq \lim_{N\to\infty}\frac{1}{N\cdot K}\mathbb{E}\{\|\underline{\boldsymbol{g}}_{l,l} - \hat{\underline{\boldsymbol{g}}}_{l,l}\|^2\}$.

B. Achievable MSE without Quantization Constraints

As a preliminary step, we characterize the average MSE without quantization, namely, the average MMSE. As stated in the previous subsection, the BSs use the orthogonal PSs to produce the MMSE estimate of their corresponding channel responses. Define the $N \times L$ random matrices $\boldsymbol{Y}_l$ and $\boldsymbol{W}_l$, such that $(\boldsymbol{Y}_l)_{k,i} = y_{l,k}[i]$ and $(\boldsymbol{W}_l)_{k,i} = (\boldsymbol{w}_l[i])_k$, as well as the $K \times L$ deterministic matrix $\boldsymbol{\Theta}$ with entries $(\boldsymbol{\Theta})_{u,i} = \theta_u[i]$. From (13) we have that for all $l \in \mathcal{N}_c$:

$$\boldsymbol{Y}_l = \sum_{m=1}^{n_c}\boldsymbol{G}_{l,m}\boldsymbol{\Theta} + \boldsymbol{W}_l, \tag{14}$$

or, alternatively, by writing $\underline{\boldsymbol{y}}_l \triangleq \mathrm{vec}(\boldsymbol{Y}_l)$, $\underline{\boldsymbol{g}}_{l,m} \triangleq \mathrm{vec}(\boldsymbol{G}_{l,m})$, and $\underline{\boldsymbol{w}}_l \triangleq \mathrm{vec}(\boldsymbol{W}_l)$, (14) can be written as

$$\underline{\boldsymbol{y}}_l = \sum_{m=1}^{n_c}\left(\boldsymbol{\Theta}^T \otimes \boldsymbol{I}_N\right)\underline{\boldsymbol{g}}_{l,m} + \underline{\boldsymbol{w}}_l. \tag{15}$$

Since the PSs are orthogonal it holds that $\boldsymbol{\Theta}\boldsymbol{\Theta}^H = L\cdot\boldsymbol{I}_K$. The covariance matrix of $\underline{\boldsymbol{y}}_l$ is given by $\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}_l} = \boldsymbol{\Sigma}_{\boldsymbol{y}_l}\otimes\boldsymbol{C}_l$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \triangleq \sum_{m=1}^{n_c}\boldsymbol{\Theta}^T\boldsymbol{D}_{l,m}^2\boldsymbol{\Theta}^* + \sigma_W^2\boldsymbol{I}_L. \tag{16}$$

Next, define the coefficients $\phi_{l,u} \triangleq \sqrt{f_{l,u}}d_{l,l,u}$ where

$$f_{l,u} \triangleq \frac{Ld_{l,l,u}^2}{\sigma_W^2 + L\sum_{m=1}^{n_c}d_{l,m,u}^2}, \quad l \in \mathcal{N}_c, u \in \mathcal{K}, \tag{17}$$

as well as the $K \times K$ diagonal matrices $\{\boldsymbol{\Phi}_l\}_{l\in\mathcal{N}_c}$ and $\{\boldsymbol{F}_l\}_{l\in\mathcal{N}_c}$ with diagonal entries $\{\phi_{l,u}\}_{u=1}^{K}$ and $\{f_{l,u}\}_{u=1}^{K}$, respectively. The MMSE channel estimate and its statistical characterization are stated in the following lemma:

Lemma 1: The MMSE estimate of $\underline{\tilde{\boldsymbol{g}}}_{l,l} \triangleq \mathrm{vec}\left(\tilde{\boldsymbol{G}}_{l,l}\right)$ from $\underline{\boldsymbol{y}}_l$ is given by

$$\underline{\tilde{\boldsymbol{g}}}_{l,l} = L^{-1}\left(\boldsymbol{F}_l\boldsymbol{\Theta}^*\otimes\boldsymbol{I}_N\right)\underline{\boldsymbol{y}}_l. \tag{18}$$

Furthermore, the vector form of the MMSE estimate $\underline{\tilde{\boldsymbol{g}}}_{l,l} \triangleq \mathrm{vec}\left(\tilde{\boldsymbol{G}}_{l,l}\right)$ is a zero-mean $N\cdot K \times 1$ Gaussian random vector with covariance matrix $\mathbb{E}\{\underline{\tilde{\boldsymbol{g}}}_{l,l}\underline{\tilde{\boldsymbol{g}}}_{l,l}^H\} = \left(\boldsymbol{\Phi}_l^2\otimes\boldsymbol{I}_N\right)$.

Proof: The lemma follows from [22, Lem. 1], thus its proof is omitted for brevity. ■

Lemma 1 can be used to obtain the average MMSE in the limit $N \to \infty$, as stated in the following corollary:

Corollary 3: The average MMSE in estimating $\underline{\boldsymbol{g}}_{l,l}$ is

$$\mu_l^{\mathrm{MMSE}} = \frac{1}{K}\sum_{u=1}^{K}\left(d_{l,l,u}^2 - \phi_{l,u}^2\right). \tag{19}$$

Proof: The corollary follows since the covariance matrix of $\underline{\boldsymbol{g}}_{l,l}$ is $\boldsymbol{D}_{l,l}^2\otimes\boldsymbol{C}_l$. Thus, letting $N \to \infty$, it follows from Szego's theorem [50] combined with Lemma 1 and the fact that $c_l[0] = 1$ that the asymptotic average MMSE is given by (19). ■

Having characterized the MMSE channel estimate for the massive MIMO setup without quantization, we are now ready to introduce quantization, and apply the results of Section III.

C. Achievable MSE With Quantized Channel Outputs

We now show how Theorems 1–3 can be used to characterize the achievable average MSE for massive MIMO channel estimation with quantization constraints.

To see that the massive MIMO system model detailed in Subsection IV-A is a special case of the general model described in Subsection III-A, we note that by writing $\boldsymbol{y}_i = [y_{l,i}[1], \ldots, y_{l,i}[L]]^T$, it holds that the set $\{\boldsymbol{y}_i\}_{i=1}^N$ consists of $L \times 1$ zero-mean Gaussian random vectors with autocorrelation $\mathbb{E}\{\boldsymbol{y}_{i_1}\boldsymbol{y}_{i_2}^H\} = \boldsymbol{\Sigma}_{\boldsymbol{y}_l} c_l[i_1 - i_2]$. Similarly, by letting $\boldsymbol{g}_i$ be the ith row of $\boldsymbol{G}_{l,l}$, it holds that $\{\boldsymbol{g}_i\}_{i=1}^N$ are $K \times 1$ zero-mean Gaussian random vectors with autocorrelation $\mathbb{E}\{\boldsymbol{g}_{i_1}\boldsymbol{g}_{i_2}^H\} = \boldsymbol{D}_{l,l}^2 c_l[i_1 - i_2]$. Finally, by Lemma 1 it holds that the MMSE estimate of $\boldsymbol{G}_{l,l}$ from the channel output $\underline{\boldsymbol{y}}_l$ is given by the set of MMSE estimates of $\boldsymbol{g}_i$ from $\boldsymbol{y}_i$ for each $i \in \{1, \ldots, N\}$, which can be written as $\tilde{\boldsymbol{g}}_i = \boldsymbol{\Gamma}\boldsymbol{y}_i$ with $\boldsymbol{\Gamma} = L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*$. We thus conclude that the massive MIMO channel estimation setup is a special case of the general problem formulation stated in Subsection III-A.

In the following, we first show how Theorems 1–2 characterize the achievable average MSE when the BS uses vector quantizers. Then, we apply Theorem 3 to obtain the minimal achievable average MSE when the BS uses hardware-limited quantizers. Finally, we note that in massive MIMO systems, the BS may be able to linearly combine only channel outputs received at the same time instance. By incorporating this constraint into the structure hardware-limited systems, we derive the minimal achievable average MSE and the resulting quantization system for this form of restricted hardware-limited quantization.

1) Vector Quantization: In Subsection III-A we discussed two vector quantization systems: the optimal vector quantizer, which is designed to recover the unknown channel $\underline{\boldsymbol{g}}_{l,l}$, and the task-ignorant vector quantizer, which represents the observed signal $\underline{\boldsymbol{y}}_l$ separately from the task of estimating the channel.

Applying Theorem 1, we obtain the minimal achievable average MSE of any quantization system operating with quantization rate R, as stated in the following proposition:

Proposition 1: The average MSE of the optimal vector quantizer for massive MIMO channel estimation is given by

$$\mu_l^{\mathrm{Opt}} = \mu_l^{\mathrm{MMSE}} + \frac{1}{K} D_{\mathrm{G}}\left(\frac{L}{K} \cdot R, \boldsymbol{\Phi}_l^2, 1\right), \tag{20}$$

where $D_{\mathrm{G}}(\cdot)$ is defined in (3a).

Proof: The proposition follows directly from Theorem 1 by noting that in the limit $N \to \infty$, the MMSE estimate $\tilde{\underline{\boldsymbol{g}}}_{l,l}$ can be represented as an $L \times 1$ Gaussian source with multivariate PSD $\boldsymbol{S}_{\underline{\tilde{\boldsymbol{g}}}}(\omega) = \boldsymbol{\Phi}_l^2$ for each $\omega \in [0, 2\pi]$ by Lemma 1. ■

Using Theorem 2, we characterize the achievable average MSE with vector quantization carried out separately from the task for the case when $\{\boldsymbol{y}_i\}$ are i.i.d., namely, $c_l[\tau] = \delta_\tau$. This is stated in the following proposition:

Proposition 2: When $c_l[\tau] = \delta_\tau$, the average MSE of the task ignorant vector quantizer for massive MIMO channel estimation is given by

$$\mu_l^{\mathrm{Ign}} = \mu_l^{\mathrm{MMSE}} + \frac{1}{K \cdot L^2} \mathrm{Tr}\left(\boldsymbol{\Theta}^T \boldsymbol{F}_l^2 \boldsymbol{\Theta}^* \left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} - \boldsymbol{\Sigma}_{\boldsymbol{y}_l, G}(R)\right)\right), \tag{21}$$

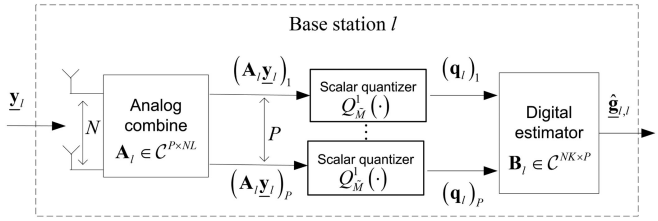

Fig. 5. Massive MIMO channel estimation with scalar ADCs.

where $\boldsymbol{\Sigma}_{\boldsymbol{y}_l}$ is defined in (16), and $\boldsymbol{\Sigma}_{\boldsymbol{y}_l, G}(R)$ is the covariance matrix of the optimal marginal distribution which achieves the distortion-rate function $D_{\mathrm{G}}\big(R, \boldsymbol{\Sigma}_{\boldsymbol{y}_l}, 1\big)$, defined in (3a).

Proof: The proposition is a result of Theorem 2, obtained by substituting $\boldsymbol{\Gamma} = L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*$ in (9), as $\{\boldsymbol{y}_i\}$ are i.i.d. Gaussian with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}_l}$. Therefore, (9) becomes

$$\begin{aligned}\mu_l^{\mathrm{Ign}} &= \mu_l^{\mathrm{MMSE}} \\ &\quad + \frac{1}{K}\mathrm{Tr}\left(\left(L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*\right)^H \left(L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*\right)\left(\boldsymbol{\Sigma}_{\boldsymbol{y}} - \boldsymbol{\Sigma}_{\boldsymbol{y},D}(R)\right)\right) \\ &\overset{(a)}{=} \mu_l^{\mathrm{MMSE}} + \frac{1}{K \cdot L^2}\mathrm{Tr}\left(\boldsymbol{\Theta}^T\boldsymbol{F}_l^2\boldsymbol{\Theta}^*\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} - \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}_l', G}(R)\right)\right),\end{aligned} \tag{22}$$

where (a) holds since $\boldsymbol{F}_l$ is diagonal with non-negative diagonal entries. ■

Note that since $\underline{\boldsymbol{y}}_l$ is Gaussian, $\boldsymbol{\Sigma}_{\boldsymbol{y}_l, G}(R)$ can be obtained using the inverse waterfilling algorithm [23, Ch. 10.3].

2) Hardware-Limited Quantization: Utilizing vector quantization in massive MIMO systems is likely to be infeasible due to its extremely high complexity for large-scale inputs. It is thus desirable to utilize serial scalar uniform ADCs, corresponding to the hardware-limited quantization setup described in Subsection III-A. Here, the linear mapping carried out in the analog domain can be implemented using a fully connected network with complex gains, as considered in [51]–[53]. In some cases, networks with controllable gains may be complex to implement, and more restricted linear structures are desirable. Constrained analog combiners can represent common practical architectures such as phase shifter networks [29], antenna selection structures [11], discrete cosine beamforming [30], and Lorentzian constrained phase combiners, which are encountered when using metasurface antennas [34]. For such scenarios, our analysis constitutes a lower bound on the achievable MSE, and can be used to facilitate the design of restricted analog combiners by approximating the resulting complex gain combiner matrix using a feasible structure, see, e.g., [31], [32], [34]. An illustration of a receiver, representing the lth BS in a massive MIMO network, applying channel estimation with hardware-limited quantization is depicted in Fig. 5.

We note that by setting the analog combining matrix $\boldsymbol{A}_l$ to be the identity matrix, the resulting system specializes the standard model for MIMO channel estimation with quantized measurements, as in, e.g., [5]–[7]. Consequently, the ability to jointly optimize the analog combining, which represents the linear processing of $\underline{\boldsymbol{y}}_l$ carried out in analog, along with the setting of the support and the digital processing, is the main difference between task-based quantization and previously proposed quantizers. In Section V we numerically illustrate that jointly designing the quantization system components significantly improves the estimation accuracy over previously proposed schemes, and that the resulting hardware-limited system can approach the optimal performance achievable with vector quantizers.

Using Theorem 3, we next characterize the minimal achievable average MSE in estimating massive MIMO channels using hardware-limited quantizers. To that aim, let $\{\lambda_{l,u}\}$ be the singular values of $L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*\boldsymbol{\Sigma}_{\boldsymbol{y}_l}^{1/2}\otimes\boldsymbol{C}_l$ arranged in descending order. The resulting optimal hardware-limited quantization system for a fixed quantization rate R and analog combining ratio r, is stated in the following proposition:

Proposition 3: In the hardware-limited quantization system which minimizes the average MSE, the analog combining matrix $\boldsymbol{A}_l^{\mathrm{o}}$ is given by $\boldsymbol{A}_l^{\mathrm{o}}=\boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}(\boldsymbol{V}_{\boldsymbol{A}}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_l}^{-1/2}\otimes\boldsymbol{C}_l^{-1/2})$, where

- $\boldsymbol{V}_{\boldsymbol{A}}\in\mathcal{C}^{L\times L}$ is the right singular vectors matrix of $L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*\boldsymbol{\Sigma}_{\boldsymbol{y}_l}^{1/2}$.
- $\boldsymbol{\Lambda}_{\boldsymbol{A}}\in\mathcal{C}^{P\times L\cdot N}$ is a diagonal matrix with diagonal entries

$$(\boldsymbol{\Lambda}_{\boldsymbol{A}})_{u,u}^2=\frac{4\kappa}{3\tilde{M}^2\cdot r}\varphi(\zeta\cdot\lambda_{l,u}),\tag{23a}$$

  where ζ is set such that $\frac{4\kappa}{3\tilde{M}^2\cdot P}\sum_{u=1}^{P}\varphi(\zeta\cdot\lambda_{l,u})=1$.
- $\boldsymbol{U}_{\boldsymbol{A}}\in\mathcal{C}^{P\times P}$ is a unitary matrix which guarantees that $\boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}^H\boldsymbol{U}_{\boldsymbol{A}}^H$ has identical diagonal entries.

The support of the ADC is given by $\gamma^2=\frac{\kappa}{r}$, and the digital processing matrix is

$$\boldsymbol{B}_l^{\mathrm{o}}=\left(\boldsymbol{D}_{l,l}^2\boldsymbol{\Theta}^*\otimes\boldsymbol{C}_l\right)\left(\boldsymbol{A}_l^{\mathrm{o}}\right)^H\times\left(\boldsymbol{A}_l^{\mathrm{o}}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l}\otimes\boldsymbol{C}_l\right)\left(\boldsymbol{A}_l^{\mathrm{o}}\right)^H+\frac{4\gamma^2}{3\tilde{M}^2}\boldsymbol{I}_P\right)^{-1}.\tag{23b}$$

The corresponding achievable average MSE in the limit $N\to\infty$ when $P_q\geq\mathrm{rank}\left(\boldsymbol{\Phi}_l\right)$ is given by

$$\mu_l^{\mathrm{HL}}=\mu^{\mathrm{MMSE}}+\frac{1}{2\pi K}\int_0^{2\pi}\sum_{u=1}^{K}\frac{\phi_{l,u}^2 s_l(\omega)}{\varphi(\zeta\phi_{l,u}\sqrt{s_l(\omega)})+1}d\omega.\tag{23c}$$

Furthermore, when $c_l[\tau]=\delta_\tau$, the asymptotic average MSE for each $P_q\geq 0$ is given by

$$\mu_l^{\mathrm{HL}}=\mu^{\mathrm{MMSE}}+\frac{1}{K}\sum_{u=1}^{\min(K,P_q)}\frac{\phi_{l,u}^2}{\varphi(\zeta\cdot\phi_{l,u})+1}+\frac{\delta_{(P_q<K)}}{K}\times\left(\sum_{u=P_q+1}^{K}\phi_{l,u}^2-(rL-P_q)\frac{\varphi(\zeta\phi_{l,P_q+1})\phi_{l,P_q+1}^2}{\varphi(\zeta\phi_{l,P_q+1})+1}\right).\tag{23d}$$

Proof: The proposition is a result of Theorem 3. In particular, here $\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\Gamma}^H=\boldsymbol{\Phi}_l^2$. Setting this in Theorem 3 proves (23a), (23c), and (23d). Finally, (23b) is obtained from (10b) by noting that for the massive MIMO setup,

$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{y}}=L^{-1}\boldsymbol{F}_l\boldsymbol{\Theta}^*\left(\sum_{m=1}^{n_c}\boldsymbol{\Theta}^T\boldsymbol{D}_{l,m}^2\boldsymbol{\Theta}^*+\sigma_W^2\boldsymbol{I}_L\right)\overset{(a)}{=}L^{-1}\boldsymbol{F}_l\left(L\sum_{m=1}^{n_c}\boldsymbol{D}_{l,m}^2+\sigma_W^2\boldsymbol{I}_K\right)\boldsymbol{\Theta}^*\overset{(b)}{=}\boldsymbol{D}_{l,l}^2\boldsymbol{\Theta}^*,\tag{24}$$

where (a) follows since $\boldsymbol{\Theta}\boldsymbol{\Theta}^H=L\cdot\boldsymbol{I}_K$, and (b) follows from the definition of $\boldsymbol{F}_l$ in (17). ■

We note that the matrix $\boldsymbol{A}_l^{\mathrm{o}}$ in Proposition 3 linearly combines the vector $\underline{\boldsymbol{y}}_l$, which represents the channel outputs received over the entire channel estimation period. Thus, $\boldsymbol{A}_l^{\mathrm{o}}$ can linearly combine samples taken from different antennas, i.e., spatial combining, and at different time instances, i.e., temporal combining. While spatial combining can be implemented using simple hardware, see, e.g., [32], temporal combining requires storing samples for different durations in analog, which may be difficult when the number of training symbols L is large. Consequently, we next characterize the optimal system when $\boldsymbol{A}_l$ is restricted to implement only spatial combining.

3) Spatial Analog Combining: In Proposition 3 we characterized the achievable average MSE when the input to the scalar ADCs can be written as any linear transformation of all the channel outputs, $\underline{\boldsymbol{y}}_l$. Consequently, we allowed samples from different time instances and different receive antennas to be jointly combined. In fact, it follows from Corollary 1 that if P is an integer multiple of N and the channel outputs are spatially uncorrelated, i.e., $P=P_q\cdot N$ and $c_l[\tau]=\delta_\tau$, then the optimal analog combining matrix is $\boldsymbol{A}_l^{\mathrm{o}}=\boldsymbol{A}_l'\otimes\boldsymbol{I}_N$, for some $\boldsymbol{A}_l'\in\mathcal{C}^{P_q\times L}$. Namely, the optimal matrix $\boldsymbol{A}_l^{\mathrm{o}}$ implements *only temporal combining*, and does not utilize spatial combining. Since in some cases it may be preferable not to combine samples received at different time instances in the analog domain to avoid the need to store data in analog, in the following we restrict the analog combining matrix to operate only on samples received at the same time instance. It should be noted that this is the model used in previous works on analog combining design for MIMO systems [12], [13], [32], which assumed full CSI and fixed quantizers.

To formulate the resulting setup, we use $\tilde{P}$ to denote the number of samples quantized at each time instance, i.e., the number of RF chains, and let $\tilde{\boldsymbol{A}}_l\in\mathcal{C}^{N\times\tilde{P}}$ represent the analog combining, applied to each received channel output. Here, at each time index $i\in\mathcal{L}$, the vector $\tilde{\boldsymbol{A}}_l\boldsymbol{y}_l[i]$ is quantized using $\tilde{P}$ identical scalar quantizers. As the overall number of quantization levels is fixed to M, each scalar quantizer has resolution $\tilde{M}=\lfloor M^{1/(2L\cdot\tilde{P})}\rfloor$.

The considered setup is a special case of the model illustrated in Fig. 5, with analog combining matrix $\boldsymbol{A}_l=\boldsymbol{I}_L\otimes\tilde{\boldsymbol{A}}_l$ and $P=\tilde{P}\cdot L$. The analog combining ratio is thus $r=\frac{P}{L\cdot N}=\frac{\tilde{P}}{N}$. Since r is fixed and positive, letting N grow arbitrarily large implies that $\tilde{P}$ grows proportionally. Let σ_l^2 be the maximal diagonal entry of $\boldsymbol{\Sigma}_{\boldsymbol{y}_l}$, namely, $\sigma_l^2\triangleq\max_{i=1,\ldots,L}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l}\right)_{i,i}$. Under this setting, the optimal system and the corresponding average MSE are stated in the following proposition:

Proposition 4: In the hardware-limited quantization system with spatial analog combining which minimizes the average MSE, the analog combining matrix $\tilde{\boldsymbol{A}}_l$ is given by $\tilde{\boldsymbol{A}}_l=\boldsymbol{U}_{\tilde{\boldsymbol{A}}}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}\boldsymbol{V}_{\tilde{\boldsymbol{A}}}^H\boldsymbol{C}_l^{-1/2}$, where $\boldsymbol{U}_{\tilde{\boldsymbol{A}}}$ guarantees that $\boldsymbol{U}_{\tilde{\boldsymbol{A}}}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}^H\boldsymbol{U}_{\tilde{\boldsymbol{A}}}^H$ has identical diagonal entires [54, Alg. 2.2]; $\boldsymbol{V}_{\tilde{\boldsymbol{A}}}^H$ is the eigenmatrix of $\boldsymbol{C}_l$; and $\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}$ is diagonal with diagonal entries $\{\bar{a}_i\}$, which are the solution to the convex optimization problem:

$$\{\bar{a}_i\}_{i=1}^{\tilde{P}}=\underset{\{a_i\}_{i=1}^{\tilde{P}}}{\arg\max}\sum_{i=1}^{\tilde{P}}\sum_{u=1}^{K}\frac{L\cdot\phi_{l,u}^4\cdot a_i^2\cdot\lambda_{\boldsymbol{C}_l,i}}{L\cdot\phi_{l,u}^2\cdot a_i^2+f_{l,u}^2}\quad\text{subject to}\ \frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2\cdot\tilde{P}}\sum_{i=1}^{\tilde{P}}a_i^2=,.\tag{25a}$$

where $\lambda_{\boldsymbol{C}_l,i}$ is the i-th largest eigenvalue of $\boldsymbol{C}_l$. The support of the ADC is $\gamma^2=\frac{3\tilde{M}^2}{4}$, and the digital processing matrix is

$$\tilde{\boldsymbol{B}}_l^{\mathrm{o}}=\left(\boldsymbol{D}_{l,l}^2\boldsymbol{\Theta}^*\otimes\boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right)\left(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l}\otimes\tilde{\boldsymbol{A}}_l\boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right)+\frac{4\gamma^2}{3\tilde{M}^2}\boldsymbol{I}_{L\tilde{P}}\right)^{-1}.\tag{25b}$$

The corresponding achievable average MSE in the limit $N \to \infty$ is given by

$$\mu_l^{\mathrm{sHL}} = \mu_l^{\mathrm{MMSE}} + \frac{1}{K}\sum_{u=1}^{K}\phi_{l,u}^2 - \frac{r}{K}\sum_{u=1}^{K}\lim_{\tilde{P}\to\infty}\frac{1}{\tilde{P}}\sum_{i=1}^{\tilde{P}}\frac{L\cdot\phi_{l,u}^4\cdot\bar{a}_i^2\cdot\lambda_{\boldsymbol{C}_l,i}}{L\cdot\phi_{l,u}^2\cdot\bar{a}_i^2+f_{l,u}^2}. \quad (25c)$$

Proof: See Appendix D.

The asymptotic average MSE in (25c) can be numerically evaluated by considering a large fixed value of N, for which the set $\{\bar{a}_i\}_{i=1}^{\tilde{P}}$ can be computed by solving the concave optimization problem in (25a). When the BS antennas are not coupled, i.e., $c_l[\tau]=\delta_\tau$, (25c) can be obtained in closed-form, as stated in the following corollary:

Corollary 4: When $c_l[\tau]=\delta_\tau$, the asymptotic achievable average MSE using spatial analog combining is given by

$$\mu_l^{\mathrm{sHL}} = \mu_l^{\mathrm{MMSE}} + \frac{1}{K}\sum_{u=1}^{K}\left(\phi_{l,u}^2 - \frac{r\cdot\phi_{l,u}^4}{\phi_{l,u}^2+\frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2\cdot L}\cdot f_{l,u}^2}\right). \quad (26)$$

Proof: For $c_l[\tau]=\delta_\tau$ it holds that $\lambda_{l,i}=1$ for each i. Thus, as the mapping $\xi(x)\triangleq\sum_{u=1}^{K}\frac{L\cdot\phi_{l,u}^4\cdot x}{L\cdot\phi_{l,u}^2\cdot x+f_{l,u}^2}$ is concave [55, 3.2.1], we have

$$\frac{1}{\tilde{P}}\sum_{i=1}^{\tilde{P}}\xi(a_i) \le \xi\left(\frac{1}{\tilde{P}}\sum_{i=1}^{\tilde{P}}a_i\right) = \xi\left(\frac{3\tilde{M}_{\tilde{P}\cdot L}^2}{4\kappa_{\tilde{P}\cdot L}\cdot\sigma_l^2}\right), \quad (27)$$

so that setting $a_i=\frac{3\tilde{M}^2}{4\kappa\cdot\sigma_l^2}$ maximizes (25a). Substituting into Proposition 4 proves the corollary. ■

The channel output model in (13) implies that, when $c_l[\tau]=\delta_\tau$, the channel outputs received at different antennas for each time instance $i\in\mathcal{L}$, $\{y_{l,k}[i]\}_{k=1}^N$, are i.i.d. Therefore, intuitively, combining $\{y_{l,k}[i]\}_{k=1}^N$ into a smaller set may result in an inaccurate estimation. This is also demonstrated in the numerical study in Subsection V-A, where it is shown that when the antennas are uncorrelated, the proposed quantizer performs better with increased analog combining ratio r (unlike the hardware-limited quantizer with general analog combining, which, as noted in Corollary 2, performs best when $r\le\frac{K}{L}$). Furthermore, it follows from the proof of Corollary 4 that for uncorrelated antennas, the optimal analog spatial combining matrix $\tilde{\boldsymbol{A}}_l$ multiplies each input by a constant, whose purpose is to guarantee that the quantized entries are within the support of the uniform scalar quantizers. This combining is different from conventional hybrid beamforming, which is typically designed assuming full CSI to better capture the energy of the transmitted signal [12], [32]. Consequently, when the channel outputs are not spatially correlated and the quantization system cannot combine samples received at different time instances in the analog domain, most of the performance gain is a result of the processing in the digital domain. This insight is in agreement with a similar conclusion in [31], which considered only spatial analog combining.

Finally, we note that even though the quantizer of Corollary 4 may not reduce the dimensionality of the quantized signal, it does not operate only in digital, as it sets the support based on the statistics of the input. Unlike previous channel estimators for massive MIMO with quantized channel outputs, e.g., [4], [5], [7], which operated only in the digital domain, the proposed quantizer reduces the quantization error by properly setting the support and scaling the channel output.

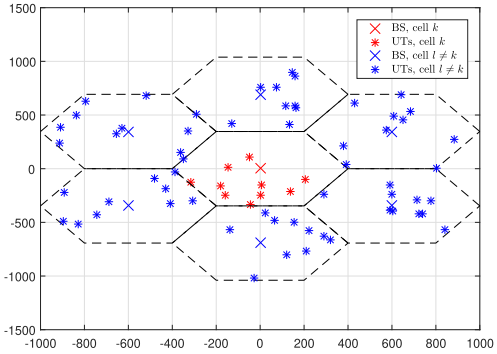

Fig. 6. Massive MIMO network illustration.

V. Numerical Results and Discussion

In this section we numerically evaluate the performance of the quantization systems discussed in Section IV for massive MIMO channel estimation. First, in Subsection V-A, we focus on hardware-limited systems, and demonstrate how to set the number of scalar quantizers, dictated by the ratio r, by numerically computing the value which minimizes the average MSE. Then, in Subsection V-B, we compare the performance of the hardware-limited quantizers to that achievable using vector quantizers, illustrating their ability to approach optimality.

We consider a massive MIMO network consisting of $n_c=7$ hexagonal cells of radius 400 m, with $K=10$ UTs in each cell. As in [20], the UTs are uniformly distributed in the cell, with the exception of a circle with radius 20 m around the BS. The attenuation coefficients $\{d_{l,m,u}\}_{u\in\mathcal{K}}$ are generated as $\left\{\frac{z_{l,m,u}}{\rho_{l,m,u}^2}\right\}_{m\in\mathcal{K}}$, where $\{z_{l,m,u}\}$ are the shadow fading coefficients, independently randomized from a log-normal distribution with standard deviation of 8 dB, and $\{\rho_{l,m,u}\}$ represent the range between the uth UT of the mth cell and the lth BS, $l,m\in\mathcal{N}_c$, $u\in\mathcal{K}$ [20, Sec. II-C]. An illustration of such a network is given in Fig. 6. We focus on the central cell in Fig. 6, and thus drop the subscript l indicating the cell index.

We use two models for the receive side correlation $c_l[\tau]$: *Uncorrelated antennas*, namely, $c_l[\tau]=\delta_\tau$; and *Correlated antennas*, representing spatial correlation induced by antenna spacing of 0.4 wavelength based on Jakes model $c_l[\tau]=J_0\left(0.8\pi|\tau|\right)$, where $J_0(\cdot)$ is the zero-order Bessel function of the first type [48]. Following [5, Sec. II-A], the pilots matrix $\boldsymbol{\Theta}$ is the first K columns of the $L\times L$ discrete Fourier transform matrix. The noise power is $\sigma_W^2=10^{-3}$, and for the scalar quantizers we fix $\eta=2$. In the following all hardware-limited quantization systems are simulated with dithered quantizers, with the exception of the channel estimator of [7], used for comparison in Subsection V-B, which is evaluated in the sequel with standard non-dithered uniform quantizers as derived in [7]. Our results are averaged over 10^3 Monte-Carlo simulations.

A. Selecting the Analog Combining Ratio r

We first numerically evaluate the number of scalar quantizers, dictated by the analog combining ratio $r=\frac{P}{N\cdot L}$, for which the achievable average MSE of the hardware-limited quantization systems studied in Section IV is minimized. To that aim, we fix $L=40$, and evaluate the achievable average MSE versus

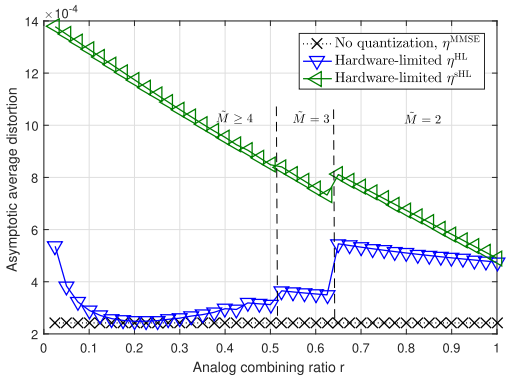

Fig. 7. Asymptotic average MSE vs. r for $R = 2$, uncorrelated antennas.

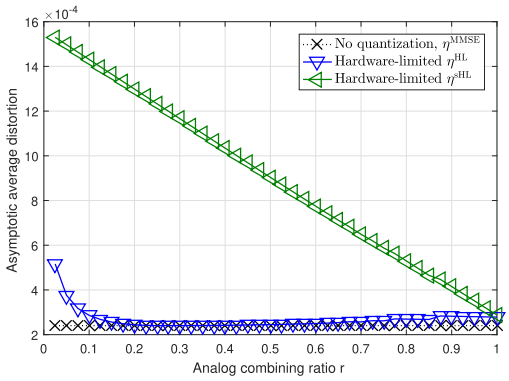

Fig. 8. Average MSE vs. r for $R = 4$, uncorrelated antennas.

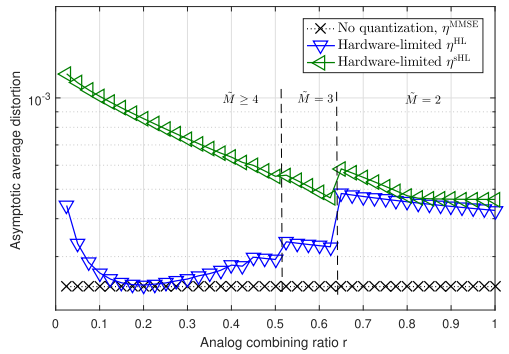

Fig. 9. Average MSE vs. r for $R = 2$, correlated antennas.

$r \in (0, 1]$ for general analog combining via Proposition 3, and for spatial analog combining via Proposition 4. When the asymptotic average MSE is given by a limit expression, e.g., (25c) with correlated antennas, we compute the MSE with $N = 100$ antennas. Note that for $r < \frac{K}{L} = 0.25$, the number of quantized samples is smaller than the number of estimated parameters. The achievable average MSEs for uncorrelated antennas quantization rates $R = 2$ and $R = 4$ are depicted in Figs. 7–8, respectively, and for correlated antennas with quantization rate $R = 2$ in Fig. 9. In Figs. 7–9 we also depict the minimal average MSE achievable without quantization, namely, the average MMSE, computed via Corollary 3.

We first observe in Figs. 7–9 that the analog combining ratio has a notable effect on the average MSE of the considered systems. In particular, for different values of r, the achievable average MSE with quantization rate $R = 2$ and uncorrelated antennas varies from $5.3 \cdot 10^{-4}$ to $2.4 \cdot 10^{-4}$ for general analog combining and from $1.3 \cdot 10^{-3}$ to $4.9 \cdot 10^{-4}$ for spatial analog combining. Furthermore, we note that for hardware-limited quantizers with general analog combining, the analog combining ratio which minimizes the average MSE μ^{HL} is not larger than $\frac{K}{L} = 0.25$, in agreement with Corollary 2. This follows since properly combining correlated samples from different time indexes results in an error which is negligible compared to that induced by the uniform quantizers, hence, hardware-limited quantizers with general analog combining operate best when the analog combining decreases the number of quantized samples to be not larger than the number of channel coefficients, i.e., $r \leq \frac{K}{L}$, allowing the quantization to be carried out with improved resolution.

When the analog combining matrix is restricted to spatial combining, we observe in Figs. 7–8 that for uncorrelated antennas, increasing the combining ratio, namely, increasing the number of scalar quantizers, improves the average MSE μ^{sHL}. This implies that combining only the independent samples received at the same time index induces a more dominant error compared to the quantization error which results from using quantizers with lower resolution. However, when the antennas are correlated, the error induced by combining the correlated samples is less notable compared to the uncorrelated case, and thus setting an analog combining ratio smaller than one can minimize the overall average MSE. In particular, it is noted in Fig. 9 that increasing the analog combining ratio from $r = 0.8$ to $r = 1$, for which the number of bits $\tilde{M} = 2$ does not change, hardly affects the overall performance, even though more samples quantized at the same resolution are processed in the digital domain. Additionally, as expected, for all values of r and for all considered scenarios, the minimal MSE achievable with general analog combining is smaller than the special case where it is restricted to spatial combining.

Finally, recall that the number of quantization levels is $\tilde{M} = \lfloor 2^{\frac{R}{2r}} \rfloor$, thus different values of r may result in the same $\tilde{M}$, most notably when R is small and r is relatively large. Consequently, when increasing r does not reduce $\tilde{M}$, the overall performance is typically improved by increasing r as more samples are processed in digital. However, when increasing r causes the ADC quantization to be less accurate, the average MSE typically increases. For example, in Figs. 7 and 9 we explicitly mark the regions of r for which $\tilde{M} = 2$ and $\tilde{M} = 3$. Observing the average MSEs in these regions, we note that for uncorrelated antennas with a fixed $\tilde{M}$, μ^{sHL} decreases quite sharply as r increases, due to the relationship between μ^{sHL} and r in (26). In both Figs. 7 and 9 we note that μ^{sHL} increases substantially when switching from $\tilde{M} = 3$ to $\tilde{M} = 2$. For general analog combining, increasing r for fixed $\tilde{M}$ has a less notable effect on the average MSE, as in this case (23d) only depends on r through the setting of ζ.

The numerical study in Figs. 7–9 can be used for determining the combining ratio r when using hardware-limited quantizers. In particular, the insights gained in this study are used in the comparison of hardware-limited quantization to task-based vector quantization in the following subsection.

B. Hardware-Limited vs. Vector Quantization

We now compare the average MSE of hardware-limited quantization, which utilizes scalar ADCs, to that achievable using vector quantizers. In particular, we compare the performance

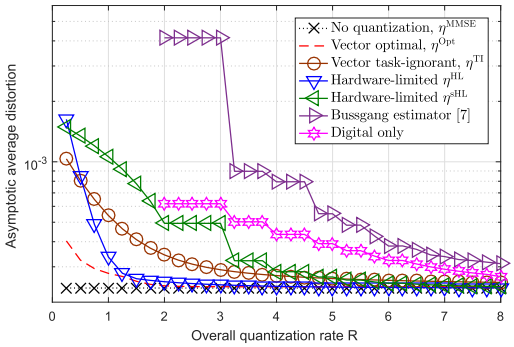

Fig. 10. Average MSE vs. R, uncorrelated antennas.

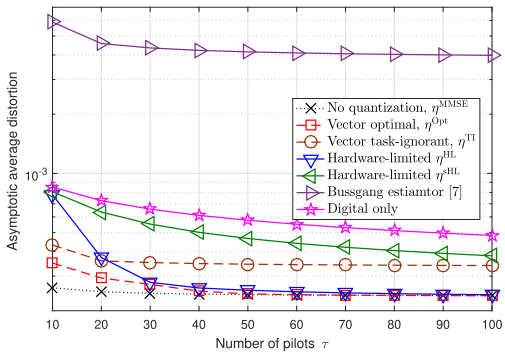

Fig. 11. Average MSE vs. L for $R = 2$, uncorrelated antennas.

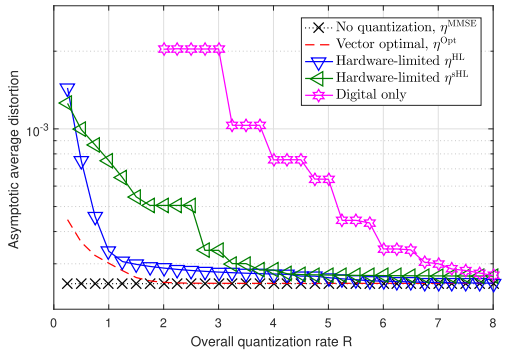

Fig. 12. Average MSE vs. R, correlated antennas.

of the hardware-limited quantizers to the optimal vector quantizer, computed via Proposition 1; to the average MSE achievable using task-ignorant vector quantization, computed via Proposition 2; and to the channel estimator of [7], which extends the 1-bit Bussgang-LMMSE estimator of [5] to multiple bits. The Bussgang estimator of [7] is computed by setting the number of antennas to $N = 100 = 10K$ and the support of the quantizers to $\gamma = 1$. The performance of the estimator of [7] is numerically averaged over 10^3 Monte Carlo simulations in which the estimator processes a uniform non-dithered quantized version of the channel output. Note that [7] considered a single cell thus we expect its channel estimation accuracy to be impaired due to the presence of intercell interference. Finally, we compute the achievable MSE of the linear MMSE digital estimator given in (D.2) with no analog combining and $\gamma = 1$. Comparing this digital only estimator to μ^{sHL} quantifies the gain of properly setting the support and the analog scaling in the spatial-only system of Proposition 4.

Note that the analog combining ratio must satisfy $r \leq \frac{R}{2}$ in order to have $\log \tilde{M} \geq 1$, i.e., to assign at least one bit for each scalar quantizer. Combining this with the numerical study of the values of r in Subsection V-A, we set $r = \min\left(\frac{K}{L}, \frac{R}{2}\right)$ when using the system with general analog combining, and $r = \min\left(1, \frac{R}{2}\right)$ when restricted to spatial analog combining and $c_l[\tau] = \delta_\tau$.

In Fig. 10 we fix the number of pilot symbols to $L = 40$, and evaluate the achievable average MSE versus $R \in [0.5, 8]$ for uncorrelated antennas. Observing Fig. 10, we note that the performance of the hardware-limited quantizer with general analog combining μ^{HL} approaches the optimal performance μ^{Opt}, achievable with vector quantizers, for quantization rates larger than $R = 1.5$. It is emphasized that while μ^{Opt} is smaller than μ^{HL}, both measures are within a gap which is negligible compared to the average MMSE, which constitutes the error floor. The existence of this error floor is an inherent property of task-based quantization problems, in which, unlike standard quantization, the error cannot be made arbitrarily small by increasing the quantization rate, as it cannot be smaller than the average MMSE. Furthermore, the performance of the hardware-limited quantizer with spatial combining μ^{sHL} also approaches μ^{Opt} as R increases, and effectively coincides with the minimal achievable MSE for $R > 5$. The estimator of [7], which operates only in the digital domain and assumes no intercell interference, is outperformed by our proposed systems for all considered quantization rates. The digital only estimator, which is designed for multiple cells yet operates only in the digital domain, is also outperformed by μ^{sHL}, especially at quantization rates $R \in [3, 6]$, where setting the support of the quantizers can notably reduce the quantization error. Furthermore, even for $R = 2$ where one-bit quantizers are used without analog combining, the MSE of the digital only estimator is still larger than μ^{sHL}. This follows since properly setting the support, as done in Proposition 4, is still beneficial here as it controls the energy of the dither signal.

These results indicate that properly designed quantization systems operating with scalar ADCs can approach the optimal performance for channel estimation in massive MIMO systems. Additionally, we note that for nearly all the considered quantization rates, our proposed hardware-limited system with general analog combining outperforms vector quantization carried out separately from the channel estimation task. This demonstrates the clear benefits of taking the task of the system into account when designing quantizers for massive MIMO systems.

Next, we fix $R = 2$. In this case, when no analog combining is applied, each complex sample is represented using two bits, and thus the real and imaginary part are quantized using one-bit sign quantizers. In Fig. 11, we compare the achievable MSEs versus $L \in [10, 100]$ for uncorrelated antennas. From Fig. 11 we note that as L increases, the hardware-limited quantizer with general analog combining approaches the optimal performance for a fixed quantization rate R, as its analog combining ratio $\frac{K}{L}$ decreases. When this happens, uniform quantization can be carried out at more accurately for the same R, reducing the quantization error. Furthermore, the quantizer with spatial analog combining, which, following the results of Subsection V-A, does not decrease its combining ratio as L increases, also demonstrates a steady improvement in the average MSE. This behavior is in agreement with the fact that as $L \to \infty$, μ^{sHL} in (26) approaches μ^{MMSE}.

So far we have considered the case of uncorrelated antennas. In Fig. 12 we compare the achievable average MSEs of the

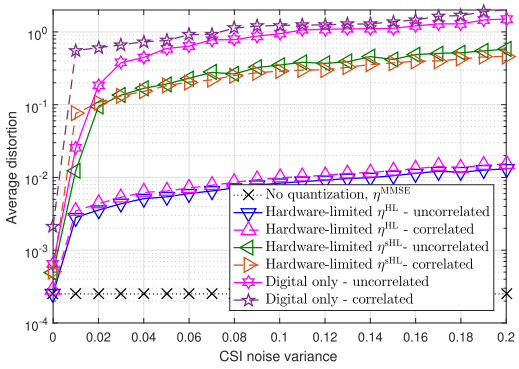

Fig. 13. Average MSE vs. σ_d^2 for $R = 2$.

hardware-limited quantizers to the optimal vector quantizer and to the digital only quantizer for the correlated antennas setup. As in Fig. 10, we compute the average MSE versus $R \in [0.5, 8]$ when the number of pilot symbols is fixed to $L = 40$. Based on the numerical study of the values of r in Subsection V-A, we use here $r = \min\left(\frac{K}{L}, \frac{R}{2}\right)$ when using the system with general analog combining, and $r = \min\left(0.8, \frac{R}{2}\right)$ when restricted to spatial analog combining. Recall that the asymptotic average MSE of the task-ignorant vector quantizer is given in Proposition 2 only for uncorrelated antennas, and is thus not evaluated in this correlated setup. Observing Fig. 12 we note that, similarly to the uncorrelated setup in Fig. 10, μ^{HL} is within a very small gap from optimal performance μ^{Opt} for quantization rates larger than $R = 1.5$. The hardware-limited quantizer with spatial combining, which for the uncorrelated case required the quantization rate to be $R > 5$ to approach μ^{Opt}, is capable of achieving near-optimal performance for $R > 3$ here, due to its ability to exploit the spatial correlation. It is also observed that the average MSE of estimating the channel only in the digital domain is notably higher compared to μ^{sHL}. This indicates that, as noted in [17], spatial correlation in massive MIMO systems with quantized outputs can be exploited by combining the samples received at the same time instance, leading to more accurate recovery.

Finally, we note that our hardware-limited quantizers require accurate knowledge of the channel input-output statistical relationship, from which, e.g., the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ is obtained. In practice, such a-priori knowledge may not be available, and one must utilize noisy estimates of the channel parameters instead of their actual value. In order to evaluate the robustness of the proposed quantization systems to inaccurate knowledge of the underlying channel, we numerically compute the average MSE achieved when using a noisy estimate of the UTs attenuation $\{d_{l,m,u}\}$, given by $d_{l,m,u} + \sigma_d \cdot w_{l,m,u}$, for each $m \in \mathcal{N}_c$ and $u \in \mathcal{K}$, where $\{w_{l,m,u}\}$ are i.i.d. zero mean Gaussian RVs with unit variance. Inaccurate knowledge of $\{d_{l,m,u}\}$ leads to a noisy estimation of the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ and the matrix $\mathbf{\Gamma}$. For each simulated realization of $\{d_{l,m,u}\}$, we evaluate the average MSE over 40 realizations of $\{w_{l,m,u}\}$. We consider both correlated a well as uncorrelated antennas, recalling that the average MMSE in Corollary 3 is identical in both. In Fig. 13 we depict the computed average MSEs of our proposed hardware-limited quantizers with $N = 100$ antennas and fixed quantization rate $R = 2$ compared to the digital only estimator, versus the coefficients noise level $\sigma_d^2 \in [0, 0.2]$. Since the average MSEs here are computed by simulating the proposed quantization systems, and not by computing an analytical expression, we do not simulate vector quantizers, which are very computationally complex to implement at large input sizes. Observing Fig. 13, we note that while the performance of all considered quantizers degrades rapidly as σ_d^2 increases, the relative gain of our proposed quantizers compared to digital only estimation is maintained. This behavior is observed for both uncorrelated as well as correlated antennas. These results indicate that the benefits of the proposed hardware-limited quantizers hold also in the presence of inaccurate CSI.

The simulation results presented in this section demonstrate the fundamental performance limits of channel estimation in massive MIMO systems, and illustrate that properly designed hardware-limited quantization systems are capable of approaching these limits at relatively low quantization rates.

VI. Conclusions

In this work we studied task-based quantization with large-scale inputs. We first derived the average achievable MSE when using vector quantization, and extended our earlier analysis of task-based quantization systems operating with scalar ADCs to large-scale data. Then, we showed how these results can be applied to studying channel estimation in massive MIMO systems with quantized inputs. Our numerical results demonstrate that the minimal achievable average MSE in massive MIMO channel estimation can be approached by properly designed quantization systems utilizing scalar low-resolution ADCs, and that the proposed approach outperforms previous channel estimators operating only in the digital domain.

Appendix

A. Proof of Theorem 1

Recall that the optimal quantizer for finite N quantizes the MMSE estimate [25]. Thus, using the notation $Q_M^{NK}(\cdot) = Q_M^{NK,NK}(\cdot)$, the minimal average MSE is given by

$$\frac{1}{NK} \min_{Q_M^{NL,NK}(\cdot)} \mathbb{E}\left\{\left\|\underline{\boldsymbol{g}} - Q_M^{NL,NK}\left(\underline{\boldsymbol{y}}\right)\right\|^2\right\}$$

$$= \mu^{\mathrm{MMSE}} + \frac{1}{NK} \min_{Q_M^{NK}(\cdot)} \mathbb{E}\left\{\left\|\underline{\tilde{\boldsymbol{g}}} - Q_M^{NK}\left(\underline{\tilde{\boldsymbol{g}}}\right)\right\|^2\right\}. \quad \text{(A.1)}$$

The second summand in (A.1) is the minimal average distortion in quantizing the MMSE estimate $\underline{\tilde{\boldsymbol{g}}}$ at rate $\frac{1}{NK}\log M = \frac{L}{K}\frac{1}{NL}\log M = \frac{L}{K} \cdot R$. Since $\underline{\tilde{\boldsymbol{g}}}$ consists of N zero-mean random vectors sampled from a stationary distribution, it follows from [36, Ch. 5.9] that for $N \to \infty$, the minimal achievable distortion coincides with the distortion-rate function for $\tilde{\boldsymbol{g}}$, namely,

$$\lim_{N\to\infty} \frac{1}{N} \min_{Q_M^{NK}(\cdot)} \mathbb{E}\left\{\left\|\underline{\tilde{\boldsymbol{g}}} - Q_M^{NK}\left(\underline{\tilde{\boldsymbol{g}}}\right)\right\|^2\right\} = D_{\tilde{\boldsymbol{g}}}\left(\frac{L}{K} \cdot R\right).$$

Substituting this in (A.1) proves the theorem. ■

B. Proof of Theorem 2

To prove the theorem, we first express the excess distortion due to quantization. Then, we let $N \to \infty$, and show that the excess distortion coincides with the second summand in (9).

From the orthogonality principle, the resulting distortion in estimating $\tilde{\underline{\boldsymbol{g}}}$ from the quantized $\underline{\boldsymbol{y}}$ is given by

$$\begin{aligned}&\frac{1}{NK}\mathbb{E}\left\{\left\|\underline{\boldsymbol{g}}-\mathbb{E}\left\{\underline{\boldsymbol{g}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right\|^2\right\}\\&=\frac{1}{NK}\mathbb{E}\left\{\left\|\underline{\boldsymbol{g}}-\tilde{\underline{\boldsymbol{g}}}\right\|^2\right\}+\frac{1}{NK}\mathbb{E}\left\{\left\|\tilde{\underline{\boldsymbol{g}}}-\mathbb{E}\left\{\underline{\boldsymbol{g}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right\|^2\right\}\\&\overset{(a)}{=}\mu^{\mathrm{MMSE}}+\frac{1}{NK}\mathbb{E}\left\{\left\|\tilde{\underline{\boldsymbol{g}}}-\mathbb{E}\left\{\tilde{\underline{\boldsymbol{g}}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right\|^2\right\},\end{aligned}\tag{B.1}$$

where (a) follows since $\underline{\boldsymbol{g}}\mapsto\underline{\boldsymbol{y}}\mapsto Q_M^{NL}(\underline{\boldsymbol{y}})$ form a Markov chain, thus, by [43, Prop. 4], $\mathbb{E}\{\underline{\boldsymbol{g}}|Q_M^{NL}(\underline{\boldsymbol{y}})\}=\mathbb{E}\{\tilde{\underline{\boldsymbol{g}}}|Q_M^{NL}(\underline{\boldsymbol{y}})\}$.

Next, we note that $\tilde{\underline{\boldsymbol{g}}}=(\boldsymbol{\Gamma}\otimes\boldsymbol{I}_N)\,\underline{\boldsymbol{y}}$, it thus follows that

$$\begin{aligned}&\mathbb{E}\left\{\left\|\tilde{\underline{\boldsymbol{g}}}-\mathbb{E}\left\{\tilde{\underline{\boldsymbol{g}}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right\|^2\right\}\\&\quad=\mathbb{E}\left\{\left\|(\boldsymbol{\Gamma}\otimes\boldsymbol{I}_N)\left(\underline{\boldsymbol{y}}-\mathbb{E}\left\{\underline{\boldsymbol{y}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right)\right\|^2\right\}\\&\quad\overset{(a)}{=}\mathrm{Tr}\left(\left(\boldsymbol{\Gamma}^H\boldsymbol{\Gamma}\otimes\boldsymbol{I}_N\right)\left(\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}-\boldsymbol{\Sigma}_{Q_M^{NL}(\underline{\boldsymbol{y}})}\right)\right),\end{aligned}\tag{B.2}$$

where (a) holds as the optimal quantizer output is uncorrelated with the quantization error [2, Sec. III]. Since $\underline{\boldsymbol{y}}$ consists here of N i.i.d. $L\times 1$ random vectors distributed as $\boldsymbol{y}$, it follows from [39, Ch. 23.2] that in the limit $N\to\infty$, the output of the optimal quantizer consists of N i.i.d. $L\times 1$ random vectors whose distribution is the marginal distortion-rate distribution which achieves $D_{\boldsymbol{y}}(R)$, i.e., $\boldsymbol{\Sigma}_{Q_M^{NL}(\underline{\boldsymbol{y}})}=\boldsymbol{\Sigma}_{\boldsymbol{y},D}(R)\otimes\boldsymbol{I}_N$. Plugging this into (B.2) and letting $N\to\infty$ yields

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{K\cdot N}\mathbb{E}\left\{\left\|\tilde{\underline{\boldsymbol{g}}}-\mathbb{E}\left\{\tilde{\underline{\boldsymbol{g}}}|Q_M^{NL}\left(\underline{\boldsymbol{y}}\right)\right\}\right\|^2\right\}\\&\quad=\frac{1}{K}\mathrm{Tr}\left(\boldsymbol{\Gamma}^H\boldsymbol{\Gamma}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}}-\boldsymbol{\Sigma}_{\boldsymbol{y},D}(R)\right)\right).\end{aligned}\tag{B.3}$$

Combining (B.3) and (B.1) proves the theorem. ■

C. Proof of Theorem 3

For a finite N, the optimal system and the resulting MSE for the considered setup can be obtained from [28]. Consequently, in the following we formulate the results of [28] (adapted to complex-valued signals), and then let N grow to infinity, obtaining Theorem 3. In particular, under the model detailed in Subsection III-A, the optimal digital processing in (10b) is obtained from [28, Lem. 1]. The analog combining of [28, Thm. 1] is given by $\boldsymbol{A}^{\mathrm{o}}=\boldsymbol{U}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}(\boldsymbol{V}_{\boldsymbol{A}}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1/2}\otimes\boldsymbol{C}^{-1/2})$, where $(\boldsymbol{\Lambda}_{\boldsymbol{A}})_{l,l}^2=\frac{4\kappa}{3\tilde{M}^2\cdot r}\varphi(\zeta\cdot\lambda_l)$. The waterfilling parameter $\zeta>0$ is set such that $\frac{4\kappa}{3\tilde{M}^2\cdot r}\sum_{l=1}^{P}\varphi(\zeta\cdot\lambda_l)=NL$, which can be written as $\frac{4\kappa}{3\tilde{M}^2\cdot P}\sum_{l=1}^{P}\varphi(\zeta\cdot\lambda_l)=1$. The support is set to satisfy

$$\gamma^2=\kappa\max_{l=1,\ldots,P}\mathbb{E}\left\{\left|\left(\boldsymbol{A}^{\mathrm{o}}\underline{\boldsymbol{y}}\right)_l\right|^2\right\},\tag{C.1}$$

and is thus given by $\gamma^2=\frac{\kappa}{P}\mathrm{Tr}(\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}^H)=\frac{\kappa}{r}$.

The resulting optimal average excess MSE compared to the MMSE in [28, Thm. 1] under this setting can be written as

$$\mathrm{MSE}_N\left(\boldsymbol{A}^{\mathrm{o}}\right)=\frac{1}{NK}\sum_{l=1}^{NK}\lambda_l^2-\frac{1}{NK}\sum_{l=1}^{\min(NK,P)}\frac{\varphi(\zeta\cdot\lambda_l)\cdot\lambda_l^2}{\varphi(\zeta\cdot\lambda_l)+1}.\tag{C.2}$$

When both sums in (C.2) have the same number of summands, i.e., $P_q\geq\mathrm{rank}(\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}_{\boldsymbol{y}}\tilde{\boldsymbol{\Gamma}}^H)$, (C.2) yields

$$\mathrm{MSE}_N\left(\boldsymbol{A}^{\mathrm{o}}\right)=\frac{1}{NK}\sum_{l=1}^{NK}\frac{\lambda_l^2}{\varphi(\zeta\cdot\lambda_l)+1}.\tag{C.3}$$

By letting $\lambda_{\boldsymbol{C},k}$ be the k-th largest eigenvalue of $\boldsymbol{C}$, it follows that each singular value λ_l can be written as $\lambda_l=\phi_i\sqrt{\lambda_{\boldsymbol{C},k}}$ for some pair of indexes $i\in\{1,\ldots,K\}$ and $k\in\{1,\ldots,N\}$, where each l corresponds to a different (i,k) pair. The average MSE in (C.3) can thus be written as

$$\mathrm{MSE}_N\left(\boldsymbol{A}^{\mathrm{o}}\right)=\frac{1}{K}\sum_{i=1}^{K}\frac{1}{N}\sum_{k=1}^{N}\frac{\phi_i^2\lambda_{\boldsymbol{C},k}}{\varphi(\zeta\cdot\phi_i\sqrt{\lambda_{\boldsymbol{C},k}})+1}.\tag{C.4}$$

Since the mapping $f(x)\triangleq\frac{x}{\varphi(\zeta\cdot\sqrt{x})+1}$ is continuous over $\mathcal{R}^+$ and since the rows of $\boldsymbol{C}$ are absolutely summable, it follows from Szego's theorem [50, Eq. (1.6)] that in the limit $N\to\infty$, (C.4) becomes

$$\mathrm{MSE}\left(\boldsymbol{A}^{\mathrm{o}}\right)=\frac{1}{K}\sum_{i=1}^{K}\frac{1}{2\pi}\int_0^{2\pi}\frac{\phi_i^2 s(\omega)}{\varphi(\zeta\cdot\phi_i\sqrt{s(\omega)})+1}d\omega,\tag{C.5}$$

thus proving (10c).

Now, when $c[l]=\delta_l$, then $\lambda_l=\phi_{\langle l\rangle_N}$ and $s(\omega)\equiv 1$. In this case, we can write (C.2) for any setting of P as

$$\mathrm{MSE}_N(\boldsymbol{A}^{\mathrm{o}})=\frac{1}{NK}\sum_{l=1}^{P}\frac{\phi_{\langle l\rangle_N}^2}{\varphi(\zeta\cdot\phi_{\langle l\rangle_N})+1}+\frac{1}{NK}\sum_{l=P+1}^{NK}\phi_{\langle l\rangle_N}^2.\tag{C.6}$$

In order to express (C.6) in the limit $N\to\infty$, we recall that by (5), $P<NK$ implies that $P_q<K$, thus, (C.6) becomes

$$\begin{aligned}\mathrm{MSE}_N(\boldsymbol{A}^{\mathrm{o}})&=\frac{1}{NK}\sum_{l=1}^{P_q\cdot N}\frac{\phi_{\langle l\rangle_N}^2}{\varphi(\zeta\cdot\phi_{\langle l\rangle_N})+1}\\&\quad+\frac{1}{NK}\sum_{l=(P_q+1)\cdot N+1}^{K\cdot N}\phi_{\langle l\rangle_N}^2+\frac{1}{NK}\sum_{l=P_q\cdot N+1}^{P_q\cdot N+P_r}\frac{\phi_{\langle l\rangle_N}^2}{\varphi(\zeta\cdot\phi_{\langle l\rangle_N})+1}\\&\quad+\frac{1}{NK}\sum_{l=P_q\cdot N+P_r+1}^{(P_q+1)\cdot N}\phi_{\langle l\rangle_N}^2\\&=\frac{1}{K}\sum_{i=1}^{P_q}\frac{\phi_i^2}{\varphi(\zeta\cdot\phi_i)+1}+\frac{1}{K}\sum_{i=P_q+1}^{K}\phi_i^2\\&\quad-\frac{P_r}{NK}\frac{\phi_{(P_q+1)}^2\varphi(\zeta\cdot\phi_{(P_q+1)})}{\varphi(\zeta\cdot\phi_{(P_q+1)})+1}.\end{aligned}$$

Writing $\frac{P_r}{NK}=r\cdot L-P_q$ yields an expression which does not depend on N, and thus holds for $N\to\infty$. Combining this with (C.5) while setting $s(\omega)\equiv 1$ proves (10d). ■

D. Proof of Proposition 4

To prove the proposition, we first characterize the achievable average MSE for a fixed $\tilde{\boldsymbol{A}}_l$ using [28, Lem. 1]. Then, as in [28, Appendix C], we derive the optimal unitary rotation for a given $\tilde{\boldsymbol{A}}_l$, and obtain the analog combining matrix as well as the resulting average MSE. We characterize the average excess

MSE compared to the average MMSE, from which the overall average MSE can be obtained by adding μ_l^{MMSE}.

Note that spatial analog combining can be written as a special case of the hardware-limited setup by fixing $\boldsymbol{A} = \boldsymbol{I}_L \otimes \tilde{\boldsymbol{A}}_l$ and $P = \tilde{P} \cdot L$. Under this setting, it can be shown that for a given $\tilde{\boldsymbol{A}}_l$, the achievable average MSE for fixed N when setting the digital processing $\tilde{\boldsymbol{B}}$ to the linear MMSE estimator is given by

$$\begin{aligned}&\mathrm{MSE}_N\left(\tilde{\boldsymbol{A}}_l\right)\\&= \frac{1}{K}\mathrm{Tr}\left(\boldsymbol{\Phi}_l^2\right) - \frac{1}{NK}\mathrm{Tr}\Bigg(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \tilde{\boldsymbol{A}}_l\boldsymbol{C}_l^2\tilde{\boldsymbol{A}}_l^H\right)\\&\times\left(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \tilde{\boldsymbol{A}}_l\boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right) + \frac{4\gamma^2}{3\tilde{M}^2}\boldsymbol{I}_{\tilde{P}\cdot L}\right)^{-1}\Bigg). \qquad \text{(D.1)}\end{aligned}$$

Similarly, the optimal digital processing matrix is given by

$$\begin{aligned}\boldsymbol{B}_l^{\mathrm{o}}\left(\tilde{\boldsymbol{A}}_l\right) &= \left(\boldsymbol{D}_{l,l}^2\boldsymbol{\Theta}^* \otimes \boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right)\\&\times\left(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \tilde{\boldsymbol{A}}_l\boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right) + \frac{4\gamma^2}{3\tilde{M}^2}\boldsymbol{I}_{\tilde{P}\cdot L}\right)^{-1}. \qquad \text{(D.2)}\end{aligned}$$

Next, recall that γ is set to η times the maximal standard deviation of the quantizer input. Thus, by (C.1),

$$\begin{aligned}\gamma^2 &= \kappa \max_{i=1,\ldots,\tilde{P}\cdot L}\mathbb{E}\left\{\left|\left(\left(\boldsymbol{I}_L \otimes \tilde{\boldsymbol{A}}_l\right)\underline{\boldsymbol{y}}_l\right)_l\right|^2\right\}\\&\overset{(a)}{=} \kappa\cdot\sigma_l^2\cdot\max_{i=1,\ldots,\tilde{P}}\left(\tilde{\boldsymbol{A}}_l\boldsymbol{C}_l\tilde{\boldsymbol{A}}_l^H\right)_{i,i}^2, \qquad \text{(D.3)}\end{aligned}$$

where (a) holds by writing the covariance of $\underline{\boldsymbol{y}}_l$ and as the maximal diagonal entry of a Kronecker product of positive semi-definite matrices is the product of the maximal diagonal entries [49, Ch. 7.8]. Defining $\bar{\boldsymbol{A}} \triangleq \tilde{\boldsymbol{A}}_l\boldsymbol{C}_l^{1/2}$ and substituting (D.3) in (D.1) results in

$$\begin{aligned}&\mathrm{MSE}_N\left(\bar{\boldsymbol{A}}\right) = \frac{1}{K}\mathrm{Tr}\left(\boldsymbol{\Phi}_l^2\right)\\&-\frac{1}{K\cdot N}\mathrm{Tr}\Bigg(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \bar{\boldsymbol{A}}\boldsymbol{C}_l\bar{\boldsymbol{A}}^H\right)\Bigg(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\\&+\frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2}\max_{i=1,\ldots,\tilde{P}}\left(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)_{i,i}^2\boldsymbol{I}_{\tilde{P}\cdot L}\Bigg)^{-1}\Bigg). \qquad \text{(D.4)}\end{aligned}$$

Using (D.4), we can now characterize the optimal unitary rotation for any given $\bar{\boldsymbol{A}}$, as stated in the following lemma:

Lemma D.1: For every matrix $\bar{\boldsymbol{A}} \in \mathcal{C}^{\tilde{P}\times N}$ there exists a unitary matrix $\boldsymbol{U}_{\bar{\boldsymbol{A}}} \in \mathcal{C}^{\tilde{P}\times\tilde{P}}$ such that

$$\begin{aligned}&\mathrm{MSE}\left(\tilde{\boldsymbol{A}}_l\right) \geq \mathrm{MSE}\left(\boldsymbol{U}_{\bar{\boldsymbol{A}}}\tilde{\boldsymbol{A}}_l\right) = \frac{1}{K}\mathrm{Tr}\left(\boldsymbol{\Phi}_l^2\right)\\&\frac{1}{K\cdot N}\mathrm{Tr}\Bigg(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \bar{\boldsymbol{A}}\boldsymbol{C}_l\bar{\boldsymbol{A}}^H\right)\\&\times\left(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right) + \frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2\cdot\tilde{P}}\mathrm{Tr}\left(\bar{\boldsymbol{A}}^H\right)\boldsymbol{I}_{\tilde{P}\cdot L}\right)^{-1}\Bigg). \qquad \text{(D.5)}\end{aligned}$$

The unitary matrix $\boldsymbol{U}_{\bar{\boldsymbol{A}}}$ is a set such that $\boldsymbol{U}_{\bar{\boldsymbol{A}}}\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\boldsymbol{U}_{\bar{\boldsymbol{A}}}^H$ is weakly majorized by all possible rotations of $\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H$.

Proof: The lemma is obtained by repeating the arguments in [28, Lem. C.1], thus its proof is omitted for brevity. ■

We can now characterize the optimal $\bar{\boldsymbol{A}}$ as the matrix which minimizes (D.5). Note that the right hand side of (D.5) is invariant to replacing $\bar{\boldsymbol{A}}$ with $\alpha\cdot\boldsymbol{U}\bar{\boldsymbol{A}}$ for any $\alpha > 0$ and for any unitary $\boldsymbol{U}$. Consequently, we can fix $\frac{4\kappa_{\tilde{P}\cdot L}\cdot\sigma_l^2}{3\tilde{M}^2_{\tilde{P}\cdot L}\cdot\tilde{P}}\mathrm{Tr}(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H) = 1$. and thus, minimizing (D.5) reduces to solving

$$\begin{aligned}&\arg\max_{\bar{\boldsymbol{A}}}\mathrm{Tr}\Bigg(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \bar{\boldsymbol{A}}\boldsymbol{C}_l\bar{\boldsymbol{A}}^H\right)\\&\left(\left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right) + \boldsymbol{I}_{\tilde{P}L}\right)^{-1}\Bigg),\\&\text{subject to } \frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2\cdot\tilde{P}}\mathrm{Tr}\left(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right) = 1. \qquad \text{(D.6)}\end{aligned}$$

By (D.3), the support is now $\gamma^2 = \frac{\kappa\cdot\sigma_l^2}{\tilde{P}}\mathrm{Tr}(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H) = \frac{3\tilde{M}^2_{\tilde{P}\cdot L}}{4}$. Plugging the resulting γ into (D.2) proves (25b).

In order to solve (D.6), we define the matrix

$$\begin{aligned}\boldsymbol{M} &\triangleq \left(\boldsymbol{\Sigma}_{\boldsymbol{y}_l} \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right) + \boldsymbol{I}_{\tilde{P}\cdot L} = \left(\boldsymbol{I}_L \otimes \left(\boldsymbol{I}_{\tilde{P}} + \sigma_W^2\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\right)\\&+\left(\boldsymbol{\Theta}^T \otimes \boldsymbol{I}_{\tilde{P}}\right)\left(\sum_{m=1}^{n_c}\boldsymbol{D}_{l,m}^2 \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\left(\boldsymbol{\Theta}^* \otimes \boldsymbol{I}_{\tilde{P}}\right). \qquad \text{(D.7)}\end{aligned}$$

Applying the matrix inversion lemma to (D.7), recalling that $\boldsymbol{\Theta}\boldsymbol{\Theta}^H = L\cdot\boldsymbol{I}_K$ results in

$$\begin{aligned}&\mathrm{Tr}\left(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\boldsymbol{M}^{-1}\right)\\&= \mathrm{Tr}\Bigg(\left(L\boldsymbol{D}_{l,l}^4 \otimes \left(\left(\boldsymbol{I}_{\tilde{P}} + \sigma_W^2\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)^{-1}\bar{\boldsymbol{A}}\boldsymbol{C}_l\bar{\boldsymbol{A}}^H\right)\right)\\&\times\left(\left(L\sum_{m=1}^{n_c}\boldsymbol{D}_{l,m}^2 \otimes \left(\boldsymbol{I}_{\tilde{P}} + \sigma_W^2\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)^{-1}\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right) + \boldsymbol{I}_{K\tilde{P}}\right)^{-1}\Bigg). \qquad \text{(D.8)}\end{aligned}$$

We note that (D.8) is invariant to replacing $\bar{\boldsymbol{A}}$ with $\alpha\cdot\boldsymbol{U}\bar{\boldsymbol{A}}$, we henceforth set $\bar{\boldsymbol{A}} = \boldsymbol{\Lambda}\boldsymbol{V}^H$, where $\boldsymbol{\Lambda} \in \mathcal{C}^{\tilde{P}\times N}$ is diagonal with diagonal entries arranged in descending magnitude order, and $\boldsymbol{V} \in \mathcal{C}^{N\times N}$ is unitary. Substituting this in (D.8) and using the invariance of the trace operator to cyclic permutations results in

$$\begin{aligned}&\mathrm{Tr}\left(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^* \otimes \bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\boldsymbol{M}^{-1}\right)\\&= \mathrm{Tr}\Bigg(\left(L\boldsymbol{D}_{l,l}^4 \otimes \left(\boldsymbol{V}^H\boldsymbol{C}_l\boldsymbol{V}\right)\right)\left(\boldsymbol{I}_K \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H\right)\\&\times\left(\left(L\sum_{m=1}^{n_c}\boldsymbol{D}_{l,m}^2 \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H\right) + \left(\boldsymbol{I}_K \otimes \left(\boldsymbol{I}_{\tilde{P}} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H\right)\right)\right)^{-1}\Bigg). \qquad \text{(D.9)}\end{aligned}$$

Note that the matrix $(\boldsymbol{I}_K \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H)((L\sum_{m=1}^{n_c}\boldsymbol{D}_{l,m}^2 \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H) + (\boldsymbol{I}_K \otimes (\boldsymbol{I}_{\tilde{P}} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H)))^{-1}$ is diagonal with non-negative diagonal entries arranged in descending order. Therefore, it follows from [56, Thm. II.1] that (D.9) is maximized by setting $\boldsymbol{V}$ to be the eigenmatrix of $\boldsymbol{C}_l$. Thus, by letting a_i be the diagonal

entries of $\mathbf{\Lambda}$, the objective (D.9) can be written as

$$\begin{aligned}&\mathrm{Tr}\left(\left(\boldsymbol{\Theta}^T\boldsymbol{D}_{l,l}^4\boldsymbol{\Theta}^*\otimes\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^H\right)\boldsymbol{M}^{-1}\right)\\&\quad=\sum_{u=1}^{K}\sum_{i=1}^{\tilde{P}}\frac{L\cdot d_{l,l,u}^4\cdot a_i^2\cdot\lambda_{l,i}}{1+\left(\sigma_W^2+L\sum_{u=1}^{n_c}d_{l,m,u}^2\right)a_i^2}\\&\quad\overset{(a)}{=}\sum_{u=1}^{K}\sum_{i=1}^{\tilde{P}}\frac{L\cdot\phi_{l,u}^4\cdot a_i^2\cdot\lambda_{l,i}}{L\cdot\phi_{l,u}^2\cdot a_i^2+f_{l,u}^2},\end{aligned}\tag{D.10}$$

where (a) follows from the definition of $f_{l,u}$ in (17), and since $\phi_{l,u}^2=f_{l,u}d_{l,l,u}^2$. By combining (D.10) and (D.6) it holds that the analog combining matrix which minimizes the average MSE is given by $\boldsymbol{U}_{\tilde{\boldsymbol{A}}}\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}\boldsymbol{V}_{\tilde{\boldsymbol{A}}}^H$, where $\boldsymbol{U}_{\tilde{\boldsymbol{A}}}$ is given in Lemma VI-D.1, $\boldsymbol{V}_{\tilde{\boldsymbol{A}}}^H$ is the eigenmatrix of $\boldsymbol{C}_l$, and $\boldsymbol{\Lambda}_{\tilde{\boldsymbol{A}}}$ is diagonal with diagonal entries $\{\bar{a}_i\}$, which are the solution to

$$\begin{aligned}\{\bar{a}_i\}_{i=1}^{\tilde{P}}&=\underset{\{a_i\}_{i=1}^{\tilde{P}}}{\arg\max}\sum_{i=1}^{\tilde{P}}\sum_{u=1}^{K}\frac{L\cdot\phi_{l,u}^4\cdot a_i^2\cdot\lambda_{l,i}}{L\cdot\phi_{l,u}^2\cdot a_i^2+f_{l,u}^2}\\&\text{subject to }\frac{4\kappa\cdot\sigma_l^2}{3\tilde{M}^2\cdot\tilde{P}}\sum_{i=1}^{\tilde{P}}a_i^2=1.\end{aligned}\tag{D.11}$$

The concavity of the objective in (D.11) stems from the concavity of the mapping $x\mapsto\frac{L\cdot\phi_{l,u}^4\cdot\lambda_{l,i}\cdot x}{L\cdot\phi_{l,u}^2\cdot x+f_{l,u}^2}$ over $\mathcal{R}^+$.

Combining (D.4) and (D.11), noting that $N\to\infty$ implies that $\tilde{P}\to\infty$, proves (25c), thus concluding the proof. ■

REFERENCES

[1] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[2] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[3] M. R. D. Rodrigues, N. Deligiannis, L. Lai, and Y. C. Eldar, "Rate-distortion trade-offs in acquisition of signal parameters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 6105–6109.

[4] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[5] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[6] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[7] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[8] H. Pirzadeh and A. L. Swindlehurst, "Spectral efficiency of mixed-ADC massive MIMO," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3599–3613, Jul. 2018.

[9] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[10] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.

[11] J. Choi, J. Sung, B. L. Evans, and A. Gatherer, "Antenna selection for large-scale MIMO systems with low-resolution ADCs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 3594–3598.

[12] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[13] J. Choi, B. L. Evans, and A. Gatherer, "Resolution-adaptive hybrid MIMO architectures for millimeter wave communications," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6201–6216, Dec. 2017.

[14] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 484–498, Jun. 2018.

[15] T. C. Zhang, C. K. Wen, S. Jin, and T. Jiang, "Mixed-ADC massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.

[16] Z. Zhang, X. Cai, C. Li, C. Zhong, and H. Dai, "One-bit quantized massive MIMO detection based on variational approximate message passing," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2358–2373, May 2018.

[17] L. G. Ordonez, I. Estella Aguerri, and M. Guillaud, "Integer forcing analog-to-digital conversion for massive MIMO systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 11–15.

[18] Q. Ding and Y. Jing, "Outage probability analysis and resolution profile design for massive MIMO uplink with mixed-ADC," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6293–6306, Sep. 2018.

[19] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[20] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3950–3600, Nov. 2010.

[21] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[22] N. Shlezinger and Y. C. Eldar, "On the spectral efficiency of noncooperative uplink massive MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1956–1971, Mar. 2019.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.

[24] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 5, pp. 518–521, Sep. 1980.

[25] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.

[26] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "Fundamental distortion limits of analog-to-digital compression," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6013–6033, Sep. 2018.

[27] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Analog-to-digital compression: A new paradigm for converting signals to bits," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 16–39, Mar. 2018.

[28] N. Shlezinger, Y. C. Eldar, and M. R. D. Rodrigues, "Hardware-limited task-based quantization," 2018, arXiv:1807.08305.

[29] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. AlKhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[30] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714–2720, Jun. 2015.

[31] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.

[32] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 16, no. 12, pp. 3243–3257, Jun. 2019.

[33] S. Rini, L. Barlett, E. Erkip, and Y. C. Eldar, "A general framework for MIMO receivers with low-resolution quantization," in *Proc. IEEE Inf. Theory Workshop*, Kaohsiung, Taiwan, Nov. 2017, pp. 599–603.

[34] N. Shlezinger, O. Dicker, Y. C. Eldar, I. Yoo, M. F. Imani, and D. R. Smith, "Dynamic metasurfaces for uplink massive MIMO systems," arXiv:1901.01458.

[35] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1082–1091, May 1999.

[36] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer, 2003.

[37] F. D. Nesser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[38] J. Gutierrez-Gutierrez, M. Zarraga-Rodriguez, P, M. Crespo, and X. Insausti, "Rate distortion function of Gaussian asymptotically WSS vector processes," *Entropy*, vol. 20, no. 9, p. 719, Sep. 2018.
[39] Y. Polyanskiy and Y. Wu, *Lecture Notes on Information Theory*. Cambridge, MA, USA: MIT Press, 2015.
[40] R. M. Gray and T. G. Stockholm, "Dithered quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, Mar. 1993.
[41] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
[42] B. Widrow, I. Kollar, and M. C. Liu, "Statistical theory of quantization," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.
[43] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.
[44] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.
[45] O. Dabeer and U. Madhow, "Channel estimation with low-precision analog-to-digital conversion," *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–6.
[46] M. S. Stein, S. Bar, J. A. Nossek, and J. Tabrikian, "Performance analysis for channel estimation with 1-bit ADC and unknown quantization threshold," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2557–2571, May 2018.
[47] J. D. Gibson, "Rate distortion functions and rate distortion function lower bounds for real-world sources," *Entropy*, vol. 19, Nov. 2017, Art. no. 604.
[48] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ, USA: IEEE Press, 1993.
[49] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: SIAM, 2000.
[50] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Delft, The Netherlands: Now Publishers, 2006.
[51] P. Karamalis, N. Skentos, and A. G. Kanatas, "Adaptive antenna subarray formation for MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 2977–2982, Nov. 2006.
[52] J. Nsenga, A. Bourdoux, W. V. Thillo, V. Ramon, and F. Horlin, "Joint Tx/Rx analog linear transformation for maximizing the capacity at 60 GHz," in *Proc. IEEE Int. Conf. Commun.*, Kyoto, Japan, Jun. 2011, pp. 1–5.
[53] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "RF chain reduction for MIMO systems: A hardware prototype," 2019, arXiv:1905.05315.
[54] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory*. Delft, The Netherlands: Now Publishers, 2007.
[55] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[56] J. B. Lassare, "A trace inequality for matrix product," *IEEE Trans. Automat. Control*, vol. 40, no. 8, pp. 1500–1501, Aug. 1995.

Nir Shlezinger (M'17) received the B.Sc., M.Sc., and Ph.D. degrees in 2011, 2013, and 2017, respectively, from Ben-Gurion University, Beer Sheva, Israel, all in electrical and computer engineering. From 2017 to 2019, he was a Postdoctoral Researcher with the Technion, Israel Institute of Technology, Haifa, Israel. He is currently a Postdoctoral Researcher with the Signal Acquisition Modeling and Processing Laboratory, Weizmann Institute of Science, Rehovot, Israel. From 2009 to 2013, he was an Engineer with Yitran Communications. His research interests include information theory and signal processing for communications.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996 both from Tel Aviv University, Tel Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2002 from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was previously a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University, and was a Visiting Professor with Stanford. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology and optics.

Dr. Eldar is a member of the Israel Academy of Sciences and Humanities (elected in 2017) and a EURASIP Fellow. She has received numerous awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (twice). She received several best paper awards and best demo awards together with her research students and colleagues including the SIAM outstanding Paper Prize and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel.

She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor-in-Chief of *Foundations and Trends in Signal Processing*, a member of the IEEE Sensor Array and Multichannel Technical Committee, and is on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE SIGNAL PROCESSING THEORY AND METHODS and *Bio Imaging Signal Processing* technical committees, and was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was the Co-Chair and Technical Co-Chair of several international conferences and workshops.

Miguel R. D. Rodrigues is currently a Reader in Information Theory and Processing with the Department of Electronic and Electrical Engineering, University College London, London, U.K., and a Faculty Fellow with the Turing Institute, London, U.K. He was previously with the Department of Computer Science, University of Porto, Portugal, rising through the ranks from Assistant Professor to Associate Professor. He also held research positions at Princeton University, Cambridge University, and Duke University.

His research interests, which lie in the general areas of information theory and processing, have led to nearly 200 publications in leading journals and conferences in the field, including the prestigious IEEE Communications and Information Theory Societies Joint Paper Award 2011. He is co-author of a book *Information-Theoretic Methods in Data Science* (Cambridge University Press, to be published).