Asymptotic Task-Based Quantization with Application to Massive MIMO

Nir Shlezinger, Yonina C. Eldar, and Miguel R. D. Rodrigues

Abstract

Quantizers take part in nearly every digital signal processing system which 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 work 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 an underlying vector, 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 which utilize 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 dimensions 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.

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

This project has received funding from the European Unions Horizon 2020 research and innovation program under grant No. 646804-ERC-COG-BNYQ, from the Israel Science Foundation under grant No. 0100101, and from the Royal Society International Exchange scheme IE 160348.

N. Shlezinger and Y. C. Eldar are with the department of EE, Technion, Haifa, Israel (e-mail: nirshlezinge@technion.ac.il; yonina@ee.technion.ac.il).

M. R. D. Rodrigues is with the department of EE, University College, London, UK (e-mail: m.rodrigues@ucl.ac.uk).
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]–[13], which is a key technology for the realization of next generation wireless networks [14]. In such systems, a wireless base station (BS) is equipped with a large number of antennas [15], [16]. 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]–[8] and/or recover the transmitted messages [5]–[13]. 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 [14]. 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]–[13]. However, most prior works assume that the quantizers are fixed, 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 [17, 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 [18]. For the mean-squared error (MSE) distortion, it was shown in [19] 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 [20], [21] 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.
MIMO communications systems utilizing both analog and digital processing are known as hybrid architectures [10], and are the focus of a large amount of recent works. In particular, [23] compared the achievable-rate versus power efficiency tradeoff for various analog combining systems, [10] and [25] designed hybrid architectures aimed at maximizing the achievable rate and signal recovery MSE, respectively, with full channel state information (CSI), while [11] 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, [12] studied the achievable rate with imperfect CSI when distinct sets of inputs are each combined in analog to maximize the receive power, while [24] characterized bounds on the capacity of MIMO communications with analog combining and one-bit quantizers. We note that most previous works which designed hybrid MIMO receivers, e.g., [10], [11], [25], 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 the 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 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 [22], which studied parameter estimation from a finite-sized quantized observed signal. The work [22] 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 [22] 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. In particular, we focus on massive MIMO channel estimation, as such wireless channels are block-fading and are commonly estimated periodically in a time-division duplex (TDD) manner [15], [16]. Consequently, unlike previous works on hybrid architectures optimization with low-resolution quantization, e.g., [10], [11], [25], our work does not require knowledge of the channel. In fact, in the presence of adjustable analog combining hardware, such as dynamic metasurface antennas [26], 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 also be applied to different tasks, such as signal recovery and noise mitigation.

We first study 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 [22], 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 [25], 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 [22], for large-scale inputs an important parameter which greatly affects the performance of the system 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. 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 [22] 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., $x$; the $i$th element of $x$ is written as $(x)_i$. Matrices are denoted with boldface upper-case letters, e.g., $M$, and we use $(M)_{i,j}$ to denote its $(i,j)$th element. We use $I_n$ to denote the $n \times n$ identity matrix. Sets are denoted with calligraphic letters, e.g., $X$, and $X^n$ is the $n$th order Cartesian power of $X$. Hermitian transpose, transpose, complex conjugate, stochastic expectation, and mutual information are written as $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, $E\{\cdot\}$, 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 $\text{Tr} (\cdot)$ 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 $X$, $x = \text{vec}(X)$ is the $n^2 \times 1$ vector obtained by stacking the columns of $X$.

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 [22], 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 $g_{n}^{e} : \mathcal{X}^{n} \mapsto \{1,2,\ldots,M\} \triangleq \mathcal{M}$ which maps the input from $\mathcal{X}^{n}$ into a discrete index $i \in \mathcal{M}$. 2) A decoding function $g_{k}^{d} : \mathcal{M} \mapsto \hat{\mathcal{X}}^{k}$ which maps each index $i \in \mathcal{M}$ into a codeword $q_{i} \in \hat{\mathcal{X}}^{k}$.

The quantizer output for input $x \in \mathcal{X}^{n}$ is $\hat{x} = g_{k}^{d}(g_{n}^{e}(x)) \triangleq Q_{M}^{n,k}(x)$. 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. When the input size and output size are equal, namely, $n = k$, we write $Q_{M}^{n}(\cdot) \triangleq Q_{M}^{n,n}(\cdot)$.

In the standard quantization problem, a $Q_{M}^{n}(\cdot)$ 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 \( \mathbb{E}\{d_n (x, \hat{x})\} \). For a fixed input size \( n \) and codebook size \( M \), the optimal quantizer is given by

\[
Q_{M, \text{opt}}^n (\cdot) = \arg \min_{Q_M^n (\cdot)} \mathbb{E}\{d_n (x, Q_M^n (x))\}.
\]  

(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., [28], or asymptotically large input size, namely, \( n \to \infty \), typically with i.i.d. inputs, via rate-distortion theory [17, Ch. 10]. For example, when the quantizer input consists of i.i.d. random variables (RVs) with probability measure \( f_x \), and the distortion measure can be written as \( d_n (x, \hat{x}) = \frac{1}{n} \sum_{i=1}^{n} d((x)_i, (\hat{x})_i) \) for some \( d : \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathbb{R}^+ \), then the optimal distortion in the limit \( n \to \infty \) for a given rate \( R \) is given by the distortion-rate function:

**Definition 2** (Distortion-rate function). The distortion-rate function for input \( x \in \mathcal{X} \) with respect to the distortion measure \( d : \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathbb{R}^+ \) is defined as

\[
D_x (R) = \min_{f_{\hat{x}|x} : I(\hat{x}; x) \leq R} \mathbb{E}\{d(\hat{x}, x)\}.
\]  

(2)

Note that the minimization in (2) is carried out over all conditional distributions \( f_{\hat{x}|x} \) which satisfy the given constraint on the resulting mutual information. The marginal output distribution which obtains the minima in (2) is referred to henceforth as the **optimal marginal distortion-rate distribution**. One scenario where \( D_x (R) \) is given in closed-form is when \( x \) is proper-complex Gaussian [29, Def. 1]. The distortion-rate function for this scenario is given in the following example:

**Example 1.** Let \( x \in \mathbb{C}^n \) be zero-mean proper-complex Gaussian with covariance matrix \( \Sigma_x \), and let \( \Sigma_x = U_x \Lambda_x U_x^H \) be the eigenvalue decomposition of \( \Sigma_x \). The distortion-rate function for \( x \) with the MSE distortion is [17, Ch. 10.3]

\[
D_G (R, \Sigma_x) \triangleq D_x (R) = \sum_{i=1}^{n} \min \left( \Delta, (\Lambda_x)_{i,i} \right),
\]  

(3a)

where \( \Delta > 0 \) is the solution to

\[
R = \sum_{i=1}^{n} \left( \log \frac{(\Lambda_x)_{i,i}}{\Delta} \right)^+.
\]  

(3b)
The optimal marginal distribution for this setup is a zero-mean proper-complex multivariate Gaussian distribution with covariance matrix $\Sigma_{x} = U_{x} \Lambda_{x} U_{x}^{H}$, where $\Lambda_{x}$ is a diagonal matrix with diagonal entries $(\Lambda_{x})_{i,i} = (\Lambda_{x})_{i,i}^{+} - \Delta$.

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 intuitively there is little benefit in quantizing the entries jointly over quantizing each entry independently, vector quantization notably outperforms serial scalar quantization [27, 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 [32], referred to henceforth as dithered quantization, with dynamic range $\gamma$ 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 $Q_{M}^{1}(y_i) = q(\text{Re} \{y_i + z_i\}) + j \cdot q(\text{Im} \{y_i + z_i\})$. Here, $z_1, \ldots, z_p$ are i.i.d. RVs with i.i.d. real and imaginary parts uniformly distributed over $[-\Delta, \Delta]$, mutually independent of the input, and $q(\cdot)$ implements uniform quantization defined as

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

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

In the following we study hardware-limited systems assuming dithered quantizers since, when the input signal is in the dynamic range of the quantizer, dithered quantization results in a digital representation which can be written as the sum of the quantizer input and an additive uncorrelated white noise signal [32]. This significantly facilitates our analysis and allows us to characterize the system which minimizes the MSE. Nonetheless, it is emphasized that this property is also approximately satisfied in uniform quantization without dithering for various input distributions,
including Gaussian inputs\cite{31}. Therefore, the rigorous analysis which follows from considering
dithered quantization, also holds approximately when using standard uniform quantizers without
dithering, as demonstrated in\cite{22}.

III. TASK-BASED QUANTIZATION OF LARGE-SCALE DATA

We now extend the analysis of task-based quantization carried out in our previous work\cite{22},
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 present 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\cite{22}. Furthermore, 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 recover the desired signal from
the quantized observations in the sense of minimizing the MSE. The desired signal consists of
$n_a$ i.i.d. zero-mean $k' \times 1$ random vectors with covariance matrix $\Sigma_{s'}$, denoted $s'_1, \ldots, s'_{n_a}$, where
$n' \geq k'$. The observations are given by a set of $n_a$ i.i.d. $n' \times 1$ random vectors with covariance
matrix $\Sigma_{x'}$, denoted $x'_1, \ldots, x'_{n_a}$, where each vector $x'_i$ is related to its corresponding $s'_i$ via
the same conditional probability measure, denoted $f_{x'|s'}$. We assume that the MMSE estimator
which stems from $f_{x'|s'}$ is linear, i.e., $\exists \Gamma' \in \mathbb{C}^{k' \times n'}$ such that the MMSE estimate of $s'_i$ from
$x'_i$ can be written as $\hat{s}'_i = \Gamma'x'_i$, for each $i \in \{1, \ldots, n_a\}$. Since we focus on large-scale data,
n_a is arbitrarily large. Such scenarios arise, for example, in signal recovery over memoryless
channels, where $s'_i$ is the channel input at time index $i$ and $x'_i$ is the corresponding channel

\footnote{For a Gaussian input with magnitude smaller than $\gamma$ 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\cite{31} Sec. VII}.
output, or alternatively, in the estimation of fast fading memoryless channels, in which $s'_i$ is the unknown channel at time index $i$. Furthermore, in Section [IV] we show that this model can also represent channel estimation in massive MIMO systems.

We write the desired vector and the observed vector as $s = \text{vec}([s'_1, \ldots, s'_{n_a}]^T)$ and $x = \text{vec}([x'_1, \ldots, x'_{n_a}]^T)$, respectively, where the sizes of these stacked vectors are denoted by $k = n_a \cdot k'$ and $n = n_a \cdot n'$, respectively. The conditional distribution of $x$ given $s$, denoted $f_{x|s}$, is given by $\prod_{i=1}^{n_a} f_{x'_i|s'_i}(x'_i|s'_i)$. The considered system forms a quantized representation of $s$ based on the observed $x$, using up to $\log M$ bits, where the quantization rate $R \triangleq \frac{1}{k} \log M$ is fixed. An illustration of such a system is depicted in Fig. 1.

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

$$\mu \triangleq \lim_{n_a \to \infty} \frac{1}{k} \mathbb{E}\{\|s - \hat{s}\|^2\}. \quad (4)$$

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

**Vector Quantizers:** Joint (vector) quantization is known to be superior to separate (scalar) quantization [27, 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 $Q_{n,k,M}^n(\cdot)$ in Fig. 1 is the vector quantizer which minimizes the distortion between the quantized representation $\hat{s}$ and $s$. 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 $x$ separately from the task, using the optimal vector quantizer for representing $x$. The desired vector $s$ is estimated from the quantized representation using the MMSE estimator, as illustrated in Fig. 2. This is intuitively the best system one can construct 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 sample 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 \( \mathbf{x} \), is projected into \( C^p \), where \( p \leq n \), 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 \( \mathbf{A} \in C^{p \times n} \). We write the number of scalar quantizers \( p \) in terms of its integer quotient and remainder with respect to \( n_a \), denoted \( m_p \) and \( m_q \), respectively, i.e.,

\[
p = m_p \cdot n_a + m_q, \quad 0 < m_q < n_a.
\] (5)

Note that for large-scale inputs, \( n_a \) tends to infinity, and thus \( m_p \) and \( m_q \) represent how \( p \) scales accordingly.

The real and imaginary parts of each entry of \( \mathbf{A} \mathbf{x} \) are quantized using the same scalar quantizer with resolution \( \tilde{M}_p = \lceil M^{1/2p} \rceil \), denoted \( Q_{\tilde{M}_p}^L (\cdot) \). Define the *analog combining ratio*

\[
r \triangleq \frac{p}{n} = \frac{m_p}{n} + \frac{m_q}{n}.
\] (6)

Note that \( \tilde{M}_p = \lfloor 2^{R_r} \rfloor \). The overall quantization rate is \( \frac{2p}{n} \log (\tilde{M}_p) \leq \frac{1}{n} \log M = R \). The identical scalar quantizers \( Q_{\tilde{M}_p}^L \) implement dithered quantization, as defined in Def. 3. The quantizer is designed to operate within the dynamic range \( \gamma \), namely, the amplitude of the
input is not larger than \( \gamma \) with sufficiently large probability. To guarantee this, we fix \( \gamma \) to be some multiple \( \eta \) 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 dynamic range with probability over 94\%. We assume that \( \eta < \sqrt{3/2} \tilde{M}_p \), such that the variable \( \kappa_p \triangleq \eta^2 (1 - 2\eta^2/3\tilde{M}_p^2)^{-1} \) is strictly positive. Note that \( \eta = 2 \) satisfies this requirement for any \( \tilde{M}_p \geq 2 \), i.e., the ADC is implemented using scalar quantizers with at least one bit.

Finally, in the digital domain, the system computes the linear MMSE estimate based on the output of the ADC, denoted \( q \in \mathcal{C}^p \), where \( (q)_i = Q_{\tilde{M}_p}^1 ((Ax)_i) \). Consequently, the estimate can be written as \( \hat{s} = Bq \) for some \( B \in \mathcal{C}^{k \times p} \). We focus on linear digital processing to keep the analysis tractable, and as 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 effect the overall performance, especially when the error due to quantization is small, as the MMSE estimator here is linear.

\section*{B. No Quantization Constraints}

As a preliminary step, we note that the MMSE estimate of \( s \) from \( x \), denoted \( \tilde{s} \) consists of \( n_a \) i.i.d. \( k' \times 1 \) random vectors, distributed as the random vector \( \tilde{s}' \). By letting \( \Sigma_{\tilde{s}'} \) denote the covariance matrix of \( \tilde{s}' \), the average MMSE can be written as

\[
\mu_{\text{MMSE}} = \frac{1}{k'} \text{Tr} \left( \Sigma_{\tilde{s}'} - \Sigma_{\tilde{s}'} \right) .
\]

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.

\section*{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 \[22\] 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 $s$ from quantized observations is as a special case of the indirect lossy source coding problem [18]. For the MSE distortion, it follows from [19] that the optimal vector quantizer first recovers the MMSE estimate $\tilde{s}$, and then uses a vector quantizer to represent $\tilde{s}$. The resulting MSE is given in the following theorem:

**Theorem 1.** The MSE of the optimal vector quantizer is

$$\mu^\text{Opt} = \mu^\text{MMSE} + \frac{1}{k'} D_{\tilde{s}'} \left( \frac{n'}{k'} \cdot R \right),$$

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

**Proof:** See Appendix A.

Theorem 1 holds since the MMSE estimate $\tilde{s}$ consists of $n_a$ i.i.d. $k' \times 1$ vectors, thus, in the limit $n_a \to \infty$, 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 $s$ from $x$ 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 $x$ via a quantizer $Q_{nM}(\cdot)$, which minimizes the MSE between its output and $x$. Then, $s$ is estimated from the output of the quantizer using the MMSE estimator. Since the MMSE estimate is linear, the resulting average MSE is given in the following theorem:

**Theorem 2.** The average MSE of the task-ignorant vector quantizer is given by

$$\mu^\text{Ign} = \mu^\text{MMSE} + \frac{1}{k'} \text{Tr} \left( \left( \Gamma' \right)^H \Gamma' \left( \Sigma_{x'} - \Sigma_{x',D}(R) \right) \right),$$

where $\Sigma_{x',D}(R)$ is the covariance matrix of the optimal marginal distribution which achieves the distortion-rate function $D_{x'}(R)$ with the MSE distortion, given in Def. 2.

**Proof:** See Appendix B.

Theorem 2 exploits the fact that $x$ consists of $n_a$ i.i.d. $n' \times 1$ vectors, thus, when $n_a$ grows
arbitrarily, the output of the optimal quantizer for representing $x$ converges to a set of $n_a$ i.i.d. vectors, each distributed via the optimal marginal distribution which achieves $D_{x'}(R)$. 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 $A^o$ and $B^o$, respectively, and the corresponding dynamic range $\gamma$. To formulate the proposed system, define the $k' \times n'$ matrix $\tilde{\Gamma} \triangleq \Gamma' \Sigma_{x'}^{1/2}$, and let $\{\lambda_{F,i}\}$ be its singular values arranged in descending order. Note that for $i > \text{rank}(\tilde{\Gamma})$, $\lambda_{F,i} = 0$. Finally, define the function $\gamma_i(\alpha) \triangleq (\alpha \cdot \lambda_{F,i} - 1)^+ \alpha \in \mathbb{R^+}$. The hardware-limited quantization system which minimizes the average MSE is stated in the following theorem:

**Theorem 3.** The minimal achievable average MSE for hardware-limited quantization is

$$
\mu^{\text{HL}} = \mu^{\text{MMSE}} + \frac{1}{k'} \sum_{i=1}^{\min(k',m_p)} \frac{\lambda_{F,i}^2}{\gamma_i(\zeta) + 1} \gamma_i(\zeta) \left( \frac{r \cdot n' - m_p}{\gamma_{m_p+1}(\zeta) + 1} \right),
$$

(10a)

where $\zeta$ is set such that

$$
\frac{4\kappa_p}{3M_p^2 \cdot r} \left( \frac{1}{n'} \sum_{i=1}^{m_p} \gamma_i(\zeta) + (r \cdot n' - m_p) \gamma_{m_p+1}(\zeta) \right) = 1,
$$

(10b)

and $r$ is defined in (6). The analog combining matrix $A^o$ is given by $A^o = U_A A_A \left( V_A^H \Sigma_{x'}^{-1/2} \otimes I_{n_a} \right)$, where

- $V_A \in \mathbb{C}^{n' \times n'}$ is the right singular vectors matrix of $\tilde{\Gamma}$.
- $A_A \in \mathbb{C}^{p \times n}$ is a diagonal matrix with diagonal entries

$$
(A_A)^2_{l,i} = \frac{4\kappa_p}{3M_p^2 \cdot r} \gamma_{i \cdot n_a}(\zeta).
$$

(10c)

- $U_A \in \mathbb{C}^{p \times p}$ is a unitary matrix which guarantees that $U_A A_A A_A^H U^H_A$ has identical diagonal entries, which can be obtained via [38, Alg. 2.2].

13
The dynamic range of the ADC is given by \( \gamma^2 = \kappa p \), and the digital processing matrix is given by
\[
B^\circ = (\Gamma' \Sigma_x^r \otimes I_{n_a}) (A^\circ)^H \left( A^\circ (\Sigma_x^r \otimes I_{n_a}) (A^\circ)^H + \frac{4\gamma^2}{3M_p^2} I_p \right)^{-1}.
\] (10d)

Proof: See Appendix C.

Theorem 3 extends [22, Thm. 1] to asymptotically large complex-valued inputs. A notable difference between Theorem 3 and [22, Thm. 1] is in the performance expression in (10a): While [22, Thm. 1] studied the MSE with finite-size inputs, here we study the asymptotic average MSE, thus (10a) depends 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 \( m_q \) in (5) does not grow proportionally with \( n_a \), i.e., \( \lim_{n_a \to \infty} \frac{m_q}{n_a} = 0 \), then by (5), \( r \cdot n' = m_p \), and the last summands in (10a) and in (10b) both vanish. In the case where \( m_q \) equals zero, i.e., \( p \) is an integer multiple of \( n_a \), the optimal system processes \( s'_i \) using the same transformation for each \( i \in \{1, \ldots, n_a\} \) separately, as stated in the following corollary:

**Corollary 1.** When \( m_q = 0 \), the hardware-limited system which minimizes the MSE applies the same mapping to each \( s'_i \) separately. This mapping includes analog combining via the matrix \( A' \), scalar quantizing with dynamic range \( \gamma^2 = \kappa p \), and digital processing with matrix \( B' \). In particular, \( A' = U_A' \Lambda_A' V_A' \Sigma_{x'}^{-1/2} \), where

- \( V_A \in \mathbb{C}^{n' \times n'} \) is the right singular vectors matrix of \( \tilde{\Gamma} \).
- \( \Lambda_A' \in \mathbb{C}^{m_p \times n'} \) is diagonal with entries \( (\Lambda_A')_{i,i} = \frac{4\kappa p \gamma_i(\zeta)}{3M_p^2 m_p} \), where \( \zeta \) is set such that \( \sum_{i=1}^{m_p} \gamma_i(\zeta) = 1 \).
- \( U_A' \in \mathbb{C}^{m_p \times m_p} \) is a unitary matrix for which \( U_A' \Lambda_A' V_A' \Sigma_{x'}^{-1/2} A' \) has identical diagonal entries.

The dynamic range of the ADC is \( \gamma^2 = \kappa p \), and the digital processing matrix is
\[
B' = \tilde{\Gamma} V_A \Lambda_A' \left( A' \Lambda_A' + \frac{4\gamma^2}{3M_p^2} I_{m_p} \right)^{-1} U_A'.
\]

The minimal achievable average MSE here is given by:
\[
\mu_{HL} = \mu_{MMSE} + \frac{1}{k'} \sum_{i=1}^{\min(k', m_p)} \frac{\chi_{F,i}^2}{\gamma_i(\zeta) + 1} + \sum_{i=m_p+1}^{k'} \frac{\delta(m_p < k')}{k'} \sum_{i=m_p+1}^{k'} \chi_{F,i}^2.
\] (11)

Proof: The corollary follows directly from Theorem 3. In particular, (11) and the requirement...
on $\zeta$ are obtained from (10a)-(10b) since $r \cdot n' = m_p$ when $p = m_p \cdot n_a$. The resulting $A'$ is a special case of $A^o$ in Theorem 3 for $p = m_p \cdot n_a$, and $B'$ is obtained by plugging $A' \otimes I_{n_a}$ into (10d).

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. [27, 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 $m_q = 0$ and the same number of scalar quantizers can be assigned to each $x_i'$, then this strategy minimizes the MSE.

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

**Corollary 2.** In order to minimize the average MSE, $m_p$ must not be larger than the rank of the covariance of $\tilde{s}'$.

**Proof:** The proof is obtained by repeating the arguments in [22, 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 $\tilde{s}'$, as we do in the following example:

**Example 2.** Consider the case where $\tilde{s}'$ has i.i.d. proper-complex Gaussian entries with variance $\sigma_s^2$. Here, the excess average MSE of the optimal vector quantizer of Theorem 7 is

$$\mu_{\text{Opt}} - \mu_{\text{MMSE}} = \frac{1}{k'} D_G \left( \frac{n'}{k'} \cdot R, \sigma_s^2 I_{k'} \right) \overset{(a)}{=} \frac{\sigma_s^2}{k' - 1},$$

(12a)

where $D_G(\cdot)$ is defined in (3), and (a) follows from the distortion-rate function of Gaussian RVs [27, Ch. 23]. Next, we compute the excess average MSE of a hardware-limited quantizer with analog combining ratio $r = \frac{n'}{k'}$, namely, $m_q = 0$ and $m_p = k'$. By noting that here $\lambda_{R,i}^2 = \sigma_s^2$ for each $i$, it follows from Corollary 1 that

$$\mu_{\text{HL}} - \mu_{\text{MMSE}} = \frac{3}{4\kappa_p} \frac{\sigma_s^2}{M_p^2 + 1} \overset{(a)}{=} \frac{3}{4\kappa_p} \frac{\sigma_s^2}{2 - \frac{n'}{2k'R}} + 1,$$

(12b)

where (a) holds as $r = \frac{n'}{k'}$. Note that (12a)-(12b) imply that as $R$ increases, the ratio of the excess average MSEs satisfies
\[
\frac{\mu^{\text{HL}} - \mu^{\text{MMSE}}}{\mu^{\text{Opt}} - \mu^{\text{MMSE}}} \approx \frac{4\kappa_p}{3} = \frac{4\eta^2}{3 - 2\eta^2}.
\]  

(12c)

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

Example 2 shows that, when \(\tilde{s}'\) 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 \(\mu^{\text{Opt}}\) approaches the average MMSE \(\mu^{\text{MMSE}}\). While (12c) holds for Gaussian \(\tilde{s}'\), a similar relation is obtained for any distribution of the entries of \(\tilde{s}'\) using the upper bound on the distortion-rate function in [36, Eq. (6)].

While Example 2 focuses on the case where \(\tilde{s}'\) 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 \(\tilde{s}'\) 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 [33]–[35]. 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 [10], [11], [25]. 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 \( n_t \) antennas serves \( n_u \) single-antenna user terminals (UTs).

We focus on the massive MIMO regime, namely, the number of antennas \( n_t \) is sufficiently large to carry out large-scale (asymptotic) analysis.

The massive MIMO channel follows a block-fading model \cite{15}. To formulate the model, let \( D_{l,m} \) be an \( n_u \times n_u \) diagonal matrix with positive diagonal entries \( \{ d_{l,m,u} \}_{u=1}^{n_u} \) representing the attenuation between the \( u \)th UT of the \( m \)th cell and the \( l \)th BS, \( l,m \in \{ 1, \ldots, n_c \} = \mathcal{N}_c \).

Without loss of generality, we assume that for each \( l \in \mathcal{N}_c \), the coefficients \( \{ d_{l,l,u} \}_{u=1}^{n_u} \) are arranged in descending order. Furthermore, let \( H_{l,m} \in \mathcal{C}^{n_t \times n_u} \) 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 \( m \)th cell and the \( l \)th BS, \( l,m \in \mathcal{N}_c \).

Channel estimation is carried out in a TDD fashion. Each UT sends a deterministic orthogonal pilot sequence (PS) consisting of \( \tau \) symbols, where the PSs are the same in all cells. The BSs use the knowledge of the PSs to estimate the channel. Let \( \theta_{u}[i] \) be the \( i \)th pilot symbol of the \( u \)th user in each cell, \( u \in \{ 1, \ldots, n_u \} = \mathcal{N}_u \), \( i \in \{ 1, \ldots, \tau \} = \mathcal{T} \). By defining \( \theta[i] = [\theta_1[i], \ldots, \theta_{n_u}[i]]^T \), the channel output at the \( l \)th BS is

\[
y_l[i] = \sum_{m=1}^{n_c} G_{l,m} \theta[i] + w_l[i], \quad l \in \mathcal{N}_c, \ i \in \mathcal{T}. \tag{13}
\]

The orthogonality of the PSs implies that for all \( l, m \in \mathcal{N}_u \), \( \sum_{i=1}^{\tau} \theta_l[i] \theta_{m}^*[i] = \tau \cdot \delta_{m,k} \). Furthermore, the PS length, \( \tau \), must not be smaller than the number of UTs, \( n_u \) \cite[Sec. III-A]{15}. Each BS uses up to \( \log M \) bits to represent the received signal \( \{ y_l[i] \}_{i=1}^{\tau} \), from which an estimate of the corresponding channel \( G_{l,l} \), denoted \( \hat{G}_{l,l} \), 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. In our analysis, we fix the quantization rate, defined here as \( R \triangleq \frac{1}{n_t \cdot \tau} \log M \), and derive the achievable MSE in the large number of antennas limit, \( \mu_l \triangleq \lim_{n_t \to \infty} \frac{1}{n_t \cdot n_u} \mathbb{E}\{\|G_{l,l} - \hat{G}_{l,l}\|^2_F\} \).

**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 \( n_t \times \tau \) random matrices \( Y_l \triangleq [y_l[1], \ldots, y_l[\tau]] \), \( W_l \triangleq [w_l[1], \ldots, w_l[\tau]] \), and the \( n_u \times \tau \) deterministic matrix \( \Theta \triangleq [\theta[1], \ldots, \theta[\tau]] \). From (13) we have that for all \( l \in \mathcal{N}_c \):

\[
Y_l = \sum_{m=1}^{n_c} G_{l,m} \Theta + W_l,
\]

or, alternatively, by writing \( y_l \triangleq \text{vec}(Y_l) \), \( g_{l,m} \triangleq \text{vec}(G_{l,m}) \), and \( w_l \triangleq \text{vec}(W_l) \), (14) can be written as

\[
y_l = \sum_{m=1}^{n_c} (\Theta^T \otimes I_{n_t}) g_{l,m} + w_l.
\]

Since the PSs are orthogonal it holds that \( \Theta \Theta^H = \tau \cdot I_{n_u} \). The covariance matrix of \( y_l \) is given by \( \Sigma_{y_l} = \Sigma_{y_l} \otimes I_{n_t} \), where

\[
\Sigma_{y_l} \triangleq \sum_{m=1}^{n_c} \Theta^T D_{l,m} \Theta^* + \sigma^2_W I_{\tau}.
\]

Next, define the coefficients \( \phi_{l,u} \triangleq \sqrt{\tilde{f}_{l,u} d_{l,u}} \), where

![Fig. 4. Massive MIMO channel estimation with \( n_c = 2 \) cells.](image-url)
\[ f_{l,u} \triangleq \frac{\tau d_{l,u}^2}{\sigma_W^2 + \tau \sum_{m=1}^{n_c} d_{l,m,u}^2}, \quad l \in \mathcal{N}_c, u \in \mathcal{N}_u, \]  

(17)

as well as the \( n_u \times n_u \) diagonal matrices \( \{ \Phi_l \}_{l \in \mathcal{N}_c} \) and \( \{ F_l \}_{l \in \mathcal{N}_c} \) with diagonal entries \( \{ \phi_{l,u} \}_{u=1}^{n_u} \) and \( \{ f_{l,u} \}_{u=1}^{n_u} \), respectively. The MMSE channel estimate and its statistical characterization are stated in the following lemma:

**Lemma 1.** The MMSE estimate of \( G_{l,l} \) from \( Y_l \) is given by

\[ \tilde{G}_{l,l} = \tau^{-1} Y_l \Theta^H F_l. \]  

(18)

Furthermore, the vector form of the MMSE estimate \( \tilde{g}_{l,l} \triangleq \text{vec}(\tilde{G}_{l,l}) \) is a zero-mean \( n_t \cdot n_u \times 1 \) Gaussian random vector with covariance matrix \( \mathbb{E}\{ \tilde{g}_{l,l} \tilde{g}_{l,l}^H \} = (\Phi_l^2 \otimes I_{n_t}) \).

**Proof:** The lemma follows from [16, Lem. 1], thus its proof is omitted for brevity.

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

**Corollary 3.** The average MMSE in estimating \( G_{l,l} \) is

\[ \mu_{\text{MMSE}}^l = \frac{1}{n_u} \sum_{u=1}^{n_u} (d_{l,l,u}^2 - \phi_{l,u}^2). \]  

(19)

**Proof:** The corollary follows since the covariance matrix of \( g_{l,l} \triangleq \text{vec}(G_{l,l}) \) is \( D_{l,l}^2 \otimes I_{n_t} \), thus, the MMSE estimate error covariance matrix is \( (I_{n_u} - F_l) D_{l,l}^2 \otimes I_{n_t} \). As \( \text{Tr} \left( A_1 \otimes A_2 \right) = \text{Tr} (A_1) \cdot \text{Tr} (A_2) \) [37, Ch. 7.8], the average MMSE is given by (19) for any value of \( n_t \).

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 2-3 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, while \( y_i[i] \) is not i.i.d. over time, the channel outputs observed in each antenna are i.i.d., i.e., by writing \( x_i' = [(y_i[1]), \ldots, [(y_i[\tau])]_i^T \), it holds that the set \( \{ x_i' \}_{i=1}^{n_t} \) consists of i.i.d. \( \tau \times 1 \) zero-mean Gaussian random vectors with covariance \( \Sigma_{y_i'} \). Similarly, by letting \( s_i' \) be the \( i \)th row of \( G_{l,l} \), it holds that \( \{ s_i' \}_{i=1}^{n_t} \) are i.i.d. \( n_u \times 1 \) zero-mean Gaussian
random vectors with covariance $D^2_{l,l}$. Lastly, by Lemma 1 it holds that the MMSE estimate of $G_{l,l}$ from the channel output $y_l$ is given by the set of MMSE estimates of $s'_i$ from $x'_i$ for each $i \in \{1, \ldots, n_t\}$, which can be written as $s'_i = \Gamma' x'_i$ with $\Gamma' = \tau^{-1} F_l \Theta^*$. 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 use 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 to 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 $g_{l,l}$, and the task-ignorant vector quantizer, which represents the observed signal $y_l$ separately from the task of estimating the channel.

Using 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

$$\mu_{l, \text{Opt}} = \mu_{l, \text{MMSE}} + \frac{1}{n_u} D_G\left(\frac{\tau}{n_u} R, \Phi^2_{l}\right),$$

where $D_G(\cdot)$ is defined in (3a).

**Proof:** The proposition follows directly from Theorem 1 by noting that here, $k' = n_u$, $n' = \tau$, and $s'$ is a Gaussian random vector with covariance matrix $\Phi^2_{l}$ by Lemma 1.

Using Theorem 2 we characterize the achievable average MSE with vector quantization carried out separately from the task, as stated in the following proposition:

**Proposition 2.** The average MSE of the task ignorant vector quantizer for massive MIMO channel estimation is given by

$$\mu_{l, \text{Ign}} = \mu_{l, \text{MMSE}} + \frac{1}{n_u} \tau^2 \text{Tr}\left(\Theta^T F_l^2 \Theta^* \left(\Sigma_{y'_l} - \Sigma_{y'_l,G}(R)\right)\right),$$

(21)
where $\Sigma_{y_l}$ is defined in (16), and $\Sigma_{y_l,G}(R)$ is the covariance matrix of the optimal marginal distribution which achieves the distortion-rate function $D_G(R, \Sigma_{y_l})$, defined in (3a).

Proof: The proposition is a result of Theorem 2 obtained by substituting $k' = n_u$ and $\Gamma' = \tau^{-1}F_l\Theta^*$ in (9), as $x'$ is Gaussian with covariance matrix $\Sigma_{y_l}'$. Therefore, (9) becomes

$$\mu_{l}^{\text{Ign}} = \mu_{l}^{\text{MMSE}} + \frac{1}{n_u} \text{Tr} \left( (\tau^{-1}F_l\Theta^*)^H (\tau^{-1}F_l\Theta^*) (\Sigma_{x'} - \Sigma_{x',D}(R)) \right)$$

$$= \mu_{l}^{\text{MMSE}} + \frac{1}{n_u} \tau^2 \text{Tr} \left( \Theta^T F_l^H F_l \Theta^* (\Sigma_{y_l}' - \Sigma_{y_l,G}(R)) \right). \quad (22)$$

where (a) holds since $F_l$ is diagonal with non-negative diagonal entries. This proves the proposition.

Note that since $y_l$ is Gaussian, $\Sigma_{y_l,G}(R)$ can be obtained using the inverse waterfilling algorithm [17, 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. An illustration of a receiver, representing the $l$th 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 $A_l$ 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 $y_l$ carried out in analog, along with the setting of the dynamic range and the digital processing, is the main difference between task-based quantization and previously 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, define the mapping $\beta_{l,u}(\alpha) \triangleq (\alpha \cdot \phi_{l,u} - 1)^+$. 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.** The minimal achievable average MSE of the hardware-limited task-based quantization system is given by

$$
\mu_{HL} = \mu_{MMSE} + \frac{1}{n_u} \sum_{u=1}^{\min(n_u,m_p)} \frac{\phi_{l,u}^2}{\beta_{l,u}(\zeta) + 1} + \frac{\delta_{(m_p<n_u)}}{n_u} \left( \sum_{u=m_p+1}^{n_u} \phi_{l,u}^2 - (r \cdot \tau - m_p) \frac{\beta_{l,m_p+1}(\zeta) \phi_{l,m_p+1}^2}{\beta_{l,m_p+1}(\zeta) + 1} \right),
$$

(23a)

where $\zeta$ is set such that

$$
\frac{4\kappa_p}{3M_p^2 \cdot r} \left( \frac{1}{r} \sum_{u=1}^{m_p} \beta_{l,u}(\zeta) + (r \cdot \tau - m_p) \beta_{l,m_p+1}(\zeta) \right) = 1.
$$

(23b)

In the system which achieves (23a), the analog combining matrix $A_l^o$ is given by

$$
A_l^o = U_A \Lambda_A \left( V_A^H \Sigma_y^{-1/2} \otimes I_{n_t} \right),
$$

where

- $V_A \in C^{\tau \times \tau}$ is the right singular vectors matrix of $\tau^{-1} F_i \Theta^* \Sigma_y^{1/2}$.
- $\Lambda_A \in C^{p \times n_t}$ is a diagonal matrix with diagonal entries

  $$(\Lambda_A)_{l,l}^2 = \frac{4\kappa_p}{3M_p^2 \cdot r} \beta_{l,(l)}(\zeta).$$

(23c)

- $U_A \in C^{p \times p}$ is a unitary matrix which guarantees that $U_A \Lambda_A \Lambda_A^H U_A^H$ has identical diagonal entries.

The dynamic range of the ADC is given by $\gamma^2 = \kappa_p$, and the digital processing matrix is given by

$$
B_l^o = \left( D_{l,l}^2 \Theta^* \otimes I_{n_t} \right) (A_l^o)^H \left( A_l^o \left( \Sigma_y^o \otimes I_{n_t} \right) (A_l^o)^H + \frac{4\gamma^2}{3M_p^2 I_p} \right)^{-1}.
$$

(23d)
Proof: The proposition is a result of Theorem 3. In particular, the quantities \( \{ \lambda_{\Gamma, m}^2 \} \) in Theorem 3 are the eigenvalues of the covariance matrix of \( \tilde{s}' \), which equals \( \Phi_l^2 \) here. Thus, \( \lambda_{\tilde{\Gamma}, u} = \phi_{l,u} \) for all \( u \in N_u \). Setting this and \( k' = n_u \) in (10a)-(10c) proves (23a)-(23c). Finally, (23d) is obtained from (10d) by noting that for the massive MIMO setup,

\[
\Gamma' \Sigma_{x'} = \tau^{-1} F_l \Theta^* \left( \sum_{m=1}^{n_c} \Theta^T D_{l,m}^2 \Theta^* + \sigma_w^2 I_c \right)
\]

\[
= (a) \tau^{-1} F_l \left( \tau \sum_{m=1}^{n_c} D_{l,m}^2 + \sigma_w^2 I_{n_u} \right) \Theta^* \quad (b) D_{l,l}^2 \Theta^*,
\]

where (a) follows since \( \Theta \Theta^H = \tau I_{n_u} \), and (b) follows from the definition of \( F_l \) in (17).
each scalar quantizer has resolution $\tilde{M}_{\tau\tilde{p}}$, where $\tilde{M}_{\tau\tilde{p}} = \lceil M^{1/(2\tau\tilde{p})} \rceil$.

We note that the considered setup is a special case of the model illustrated in Fig. 5, with analog combining matrix $A_l = I_\tau \otimes \tilde{A}_l$ and $p = \tilde{p} \cdot \tau$. The analog combining ratio is thus $r = \frac{p}{\tau \cdot n_t} = \frac{\tilde{p}}{n_t}$. Since $r$ is fixed and positive, letting $n_t$ grow arbitrarily large implies that $\tilde{p}$ grows proportionally. Let $\sigma^2_l$ be the maximal diagonal entry of $\Sigma_{y_l}^\prime$, namely, $\sigma^2_l \triangleq \max_{l=1,\ldots,\tau} \left( \Sigma_{y_l}^\prime \right)_{l,l}$. Under this setting, the optimal system and the corresponding average MSE are stated in the following proposition:

**Proposition 4.** The minimal achievable average MSE when only spatial analog combining is carried out is given by

$$
\mu^{sHL}_l = \mu^{MMSE}_l + \frac{1}{n_t} \sum_{u=1}^{n_u} \phi^2_{l,u} - \frac{r}{n_t} \sum_{u=1}^{n_u} \phi^2_{l,u} \frac{\phi^4_{l,u}}{4\kappa_{\tilde{p},\tau}^2 \sigma^2_l} \cdot \tilde{f}_{l,u}^2.
$$

(25a)

In the system which achieves (25a), the analog combining matrix $\tilde{A}_l$ is diagonal with identical diagonal entries $(\tilde{A}_l)_{i,i} = \frac{3\tilde{M}_{\tau\tilde{p}}^2}{4\kappa_{\tilde{p},\tau} \sigma^2_l}$. The dynamic range of the ADC is $\gamma^2 = \frac{3\tilde{M}_{\tau\tilde{p}}^2}{4\kappa_{\tilde{p},\tau}}$, and the digital processing matrix is given by

$$
\tilde{B}_i^o = \left( D_{l,l}^2 \Theta^* \otimes \tilde{A}_l^H \right) \left( \left( \Sigma_{y_l}^\prime \otimes \tilde{A}_l \tilde{A}_l^H \right) + I_{\tau\tilde{p}} \right)^{-1}.
$$

(25b)

**Proof:** See Appendix D

Note that the channel output model in (13) and the fact that $\Sigma_{y_l} = \Sigma_{y_l}^\prime \otimes I_{n_t}$ imply that for each time instance $i \in \mathcal{T}$, the entries of $y_l[i]$ are i.i.d.. Therefore, intuitively, combining the entries of $y_l[i]$ into a lower dimension may result in an inaccurate estimation. This is also demonstrated in the numerical study in Subsection V-A, where it is noted that 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 $r \leq \frac{n_u}{\tau}$). Furthermore, it follows from Proposition 4 that the optimal analog spatial combining matrix $\tilde{A}_l$ merely multiplies each input by a constant, whose purpose is to guarantee that the quantized entries are within the dynamic range of the uniform scalar quantizers. Consequently, when 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 noted in [23], which considered only spatial analog combining.
Finally, we note that even though the quantizer of Proposition 4 may not reduce the dimensionality of the quantized signal, it does not operate only in digital, as it sets the dynamic range 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 dynamic range and scaling the channel output.

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 $n_c = 7$ hexagonal cells of radius 400 m, with $n_u = 10$ UTs in each cell. As in [15], the UTs are uniformly distributed in the cell, with the exception of a circle with radius 20 m around the BS. The attenuation coefficients $\{d_{l,m,u}\}_{u\in N_u}$ are generated as $\{z_{l,m,u} \rho_{l,m,u}^2\}_{m\in N_c}$, 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 $u$th UT of the $m$th cell and the $l$th BS, $l, m \in N_c$, $u \in N_u$ [15 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. Following [5 Sec. II-A], the pilots matrix $\Theta$ is the first $n_u$ columns of the $\tau \times \tau$ discrete Fourier transform.
matrix. The noise power is $\sigma^2_w = 10^{-3}$, and for the scalar quantizers we fix $\eta = 2$. 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 $r = \frac{p}{n_u \cdot \tau}$, for which the achievable average MSE of the hardware-limited quantization systems studied in Section IV is minimized. To that aim, we fix $\tau = 40$, and evaluate the achievable average MSE versus $r \in (0, 1]$ for general analog combining via Proposition 3, and for spatial analog combining via Proposition 4. Note that for $r < \frac{n_u}{\tau} = 0.25$, the number of quantized samples is smaller than the number of estimated parameters. The achievable average MSEs for quantization ratios $R = 2$ and $R = 4$ are depicted in Figs. 7-8, respectively. In both figures we also depict the minimal average MSE achievable without quantization, namely, the average MMSE, computed via Corollary 3.

Observing Figs. 7-8 we first note 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 $R = 2$ 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 $\mu_{HL}$ is not larger than $\frac{n_u}{\tau} = 0.25$, 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., $r \leq \frac{n_u}{\tau}$, allowing the quantization to be carried out with improved resolution.

When the analog combining matrix is restricted to spatial combining, we note in Figs. 7-8 that increasing the combining ratio, namely, increasing the number of scalar quantizers, improves the average MSE $\mu^{\text{SHL}}$. 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. Additionally, as expected, for all values of $r$, 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 $\tilde{M}_p = \left\lfloor \frac{2^R}{r} \right\rfloor$, thus different values of $r$ may result in the same $\tilde{M}_p$, most notably when $R$ is small and $r$ is relatively large. For example, in Fig. 7 we explicitly mark the regions of $r$ for which $\tilde{M}_p = 2$ and $\tilde{M}_p = 3$. Observing the average MSEs in these regions, we note that for a fixed $\tilde{M}_p$, $\mu^{\text{SHL}}$ decreases quite sharply as $r$ increases, due to the relationship between $\mu^{\text{SHL}}$ and $r$ in (25a). For general analog combining, increasing $r$ for fixed $\tilde{M}_p$ has a less notable effect on the average MSE, as in this case (23a) only depends on $r$ through the setting of $\zeta$ in (23b).

The numerical study in Figs. 7-8 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 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 Busgang-LMMSE estimator of [5] to multiple bits. The Busgang estimator of [7] is computed by setting the number of antennas to $n_t = 10n_u$ and the dynamic range 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 $\mu_{sHL}$ quantifies the gain of properly setting the dynamic range 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 $\log \tilde{M}_p \geq 1$, 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 $r = \min \left( \frac{n_u}{\tau}, \frac{R}{2} \right)$ when using the system with general analog combining, and $r = \min \left( 1, \frac{R}{2} \right)$ when restricted to spatial analog combining.

In Fig. 9 we fix the number of pilot symbols to $\tau = 40$, and evaluate the achievable average MSE versus $R \in [0.5, 8]$. Observing Fig. 9 we note that the performance of the hardware-limited quantizer with general analog combining $\mu^{HL}$ approaches the optimal performance $\mu^{Opt}$, achievable with vector quantizers, for quantization rates larger than $R = 1.5$. Furthermore, the performance of the hardware-limited quantizer with spatial combining $\mu^{sHL}$ also approaches $\mu^{Opt}$ 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 $\mu^{sHL}$, especially at quantization rates $R \in [3, 6]$, where setting the dynamic range 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 $\mu^{sHL}$. This follows since properly setting the dynamic range, 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.

Finally, 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. 10, we compare the achievable MSEs versus $\tau \in [10, 100]$. From Fig. 10 we note that as $\tau$ increases, the hardware-limited quantizer with general analog combining approaches the optimal performance for a fixed quantization rate $R$, as its analog combining ratio $\frac{n_u}{\tau}$ 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 $\tau$ increases, also demonstrates a steady improvement in the average MSE. This behavior is in agreement with the fact that as $\tau \to \infty$, $\mu^{sHL}$ in (25a) approaches $\mu^{MMSE}$.

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.
APPENDIX

A. Proof of Theorem 1

Recall that the optimal quantizer for finite $n$ quantizes the MMSE estimate $[19]$, thus, the minimal average MSE is

$$
\frac{1}{k} \min_{Q_{M}^{n,k} (\cdot)} \mathbb{E} \left\{ \| s - Q_{M}^{n,k} (x) \|^{2} \right\} = \frac{1}{k} \mathbb{E} \left\{ \| s - \bar{s} \|^{2} \right\} + \frac{1}{k} \min_{Q_{M}^{n,k} (\cdot)} \mathbb{E} \left\{ \| \bar{s} - Q_{M}^{k} (\bar{s}) \|^{2} \right\}
$$

$$
= \mu_{\text{MMSE}} + \frac{1}{k} \min_{Q_{M}^{n,k} (\cdot)} \mathbb{E} \left\{ \| \bar{s} - Q_{M}^{k} (\bar{s}) \|^{2} \right\}.
$$

The second summand in (A.1) is the minimal average distortion in quantizing the MMSE estimate $\bar{s}$ at rate $1/k \log M = \frac{1}{n} \frac{n'}{k'} \log M = \frac{n'}{k'} \cdot R$. Since $\bar{s}$ consists of $n_a$ i.i.d. $k' \times 1$ zero-mean random vectors distributed as $\bar{s}'$, it follows from [27, Ch. 23.2] that for $n_a \to \infty$, the minimal achievable distortion coincides with the distortion-rate function for $\bar{s}'$, namely,

$$
\lim_{n_a \to \infty} \frac{1}{n_a} \min_{Q_{M}^{n,k} (\cdot)} \mathbb{E} \left\{ \| \bar{s} - Q_{M}^{k} (\bar{s}) \|^{2} \right\} = D_{\bar{s}'} \left( \frac{n'}{k'} \cdot R \right).
$$

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 $n_a \to \infty$, and show that the excess distortion coincides with the second summand in (9).

From the orthogonality principle, the resulting distortion in estimating $\bar{s}$ is given by

$$
\frac{1}{k} \mathbb{E} \left\{ \| s - \mathbb{E} \{ s | Q_{M}^{n,k} (x) \} \|^{2} \right\} = \frac{1}{k} \mathbb{E} \left\{ \| s - \bar{s} \|^{2} \right\} + \frac{1}{k} \mathbb{E} \left\{ \| \bar{s} - \mathbb{E} \{ s | Q_{M}^{n,k} (x) \} \|^{2} \right\}
$$

$$
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad (a)
$$

$$
= \mu_{\text{MMSE}} + \frac{1}{k} \mathbb{E} \left\{ \| \bar{s} - \mathbb{E} \{ s | Q_{M}^{n,k} (x) \} \|^{2} \right\},
$$

where $(a)$ follows since $s \mapsto x \mapsto Q_{M}^{n,k} (x)$ form a Markov chain, thus, by [30, Prop. 4], $
\mathbb{E} \{ s | Q_{M}^{n,k} (x) \} = \mathbb{E} \{ \bar{s} | Q_{M}^{n,k} (x) \}$.

Next, we note that $\bar{s} = (\Gamma' \otimes I_{n_a}) x$, it thus follows that

$$
\mathbb{E} \left\{ \| \bar{s} - \mathbb{E} \{ s | Q_{M}^{n,k} (x) \} \|^{2} \right\} = \mathbb{E} \left\{ \| (\Gamma' \otimes I_{n_a}) (x - \mathbb{E} \{ x | Q_{M}^{n,k} (x) \}) \|^{2} \right\}
$$

$$
= \text{Tr} \left( (\Gamma')^{H} (\Gamma' \otimes I_{n_a}) \left( x - \mathbb{E} \{ x | Q_{M}^{n,k} (x) \} \right) (x - \mathbb{E} \{ x | Q_{M}^{n,k} (x) \})^{H} \right)
$$

$$
= \text{Tr} \left( (\Gamma')^{H} (\Gamma' \otimes I_{n_a}) \left( \Sigma_{x} - \Sigma_{Q_{M}^{n,k} (x)} \right) \right),
$$

(B.2)

30
where (a) holds as the optimal quantizer output is uncorrelated with the quantization error \[2\] Sec. III]. Since \(x\) consists of \(n_a\) i.i.d. \(n' \times 1\) random vectors distributed as \(x'\), it follows from \[27\] Ch. 23.2 that in the limit \(n_a \to \infty\), the output of the optimal quantizer consists of \(n_a\) i.i.d. \(n' \times 1\) random vectors whose distribution is the marginal distortion-rate distribution which achieves \(D_x'(R)\), i.e., \(\Sigma_{x,M}^n(x) = \Sigma_{x',D}(R) \otimes I_{n_a}\). Plugging this into (B.2) and letting \(n_a \to \infty\) yields

\[
\lim_{n_a \to \infty} \frac{1}{k' \cdot n_a} E\left\{ \| \tilde{s} - E\{ \tilde{s} | Q_{n,n}^M(x) \} \|^2 \right\} = \frac{1}{k'} \text{Tr} \left( (\Gamma')^H \Gamma' (\Sigma_{x'} - \Sigma_{x',D}(R)) \right). \tag{B.3}
\]

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

\[\Box\]

\[\text{C. Proof of Theorem 3}\]

For a finite \(n_a\), the optimal system and the resulting MSE for the considered setup can be obtained from [22]. Consequently, in the following we formulate the results of [22] (adapted to complex-valued signals), and then modify the expression to be independent of \(n_a\), which grows to infinity, obtaining Theorem 3. In particular, under the model detailed in Subsection III-A, the optimal digital processing in [10b] is obtained from [22, Lem. 1]. The analog combining of [22, Thm. 1] is given by

\[
A^\circ = U_A \Sigma_A V_A^H \Sigma_x^{-1/2},
\]

where (a) follows since by (5) it holds that \(m_q = r \cdot n' - m_p\). Note that (C.1) coincides with (10b). The dynamic range is set to satisfy

\[
\gamma^2 = \kappa_p \max_{l=1,\ldots,p} E \left\{ |(Ax)_l|^2 \right\}, \tag{C.2}
\]

and is thus given by

\[
\gamma^2 = \frac{\kappa_p}{p} \text{Tr} \left( A^H A \right) = \frac{\kappa_p}{p}.
\]

The resulting optimal average excess MSE compared to the MMSE in [22, Thm. 1] under this setting can be written as
MSE \( (\mathbf{A}^o) = \frac{1}{k} \sum_{i=1}^{k'} \lambda_{\Gamma,i}^2 - \frac{1}{k} \sum_{l=1}^{\min(k,p)} \frac{\gamma(l)_{n_a}(\zeta) \cdot \lambda_{\Gamma,l}^2}{\gamma(l)_{n_a}(\zeta)+1} \)

\[ = \frac{1}{k'} \sum_{i=1}^{k'} \lambda_{\Gamma,i}^2 - \frac{1}{k} \sum_{l=1}^{\min(k,p)} \frac{\gamma(l)_{n_a}(\zeta) \cdot \lambda_{\Gamma,l}^2}{\gamma(l)_{n_a}(\zeta) + 1}. \quad (C.3) \]

Now, if \( p \geq k \), then, as \( k = n_a \cdot k' \), \((C.3)\) can be written as

\[ \text{MSE} \left( \mathbf{A}^o \right) = \frac{1}{k} \sum_{l=1}^{k} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2} + \frac{1}{k} \sum_{l=k+1}^{k} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2}. \quad (C.4) \]

When \( p < k \), we can write \((C.3)\) as

\[ \text{MSE} \left( \mathbf{A}^o \right) = \frac{1}{k} \sum_{l=1}^{p} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2} + \frac{1}{k} \sum_{l=p+1}^{k} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2}. \quad (C.5) \]

In order to express \((C.5)\) in the limit \( n_a \to \infty \), we recall that by \((5)\), \( p < k \) implies that \( m_p < k' \), thus, \((C.5)\) becomes

\[ \text{MSE} \left( \mathbf{A}^o \right) = \frac{1}{k} \sum_{l=1}^{m_p \cdot n_a} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2} + \frac{1}{k} \sum_{l=m_p \cdot n_a+1}^{m_q \cdot n_a+q} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2} + \frac{1}{k} \sum_{l=(m_p+1) \cdot n_a+1}^{k' \cdot n_a} \frac{\gamma(l)_{n_a}(\zeta)+1}{\lambda_{\Gamma,l}^2} \]

\[ = \frac{1}{k'} \sum_{i=1}^{m_p} \frac{\lambda_{\Gamma,i}^2}{\gamma_i(\zeta)+1} + \frac{1}{k'} \sum_{i=m_p+1}^{k' \cdot m_p+1} \frac{\lambda_{\Gamma,i}^2}{\gamma_{m_p+1}(\zeta)+1} + \frac{1}{k'} \sum_{i=m_p+1}^{k' \cdot m_p+2} \frac{\lambda_{\Gamma,i}^2}{\gamma_{m_p+1}(\zeta)+1}. \quad (C.6) \]

Writing \( \frac{m_q}{k} = r \cdot n' - m_p \) yields an expression which does not depend on \( n_a \), and thus holds for \( n_a \to \infty \). Combining this with \((C.4)\) proves \((10a)\), concluding the proof of the theorem. \( \square \)

D. Proof of Proposition [4]

To prove the proposition, we first characterize the achievable average MSE for fixed \( \tilde{\mathbf{A}}_l \) using [22, Lem. 1]. Then, as in [22, Appendix C], we derive the optimal unitary rotation for a given \( \tilde{\mathbf{A}}_l \), and characterize 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 \( \mu_i^{\text{MMSE}} \).
Note that spatial analog combining can be written as a special case of the hardware-limited setup by fixing $A = I_r \otimes \tilde{A}_l$ and $p = \tilde{p} \cdot \tau$. Under this setting, it can be shown that for a given $\tilde{A}_l$, the achievable average MSE when setting the digital processing $\tilde{B}$ to the linear MMSE estimator is given by

$$\text{MSE} \left( \tilde{A}_l \right) = \frac{1}{n_u} \text{Tr} \left( \Phi_l^2 \right) - \frac{1}{n_u \cdot n_t} \text{Tr} \left( \Theta^T D_{l,l}^4 \Theta^* \otimes \tilde{A}_l \tilde{A}_l^H \right) \times \left( \Sigma_{y'} \otimes \tilde{A}_l \tilde{A}_l^H + \frac{4\gamma^2}{3M_{\tilde{p} \cdot \tau}^2} I_{\tilde{p} \cdot \tau} \right)^{-1}. \quad \text{(D.1)}$$

Similarly, the optimal digital processing matrix is given by

$$B_{l}^o \left( \tilde{A}_l \right) = \left( \Theta^T D_{l,l}^4 \Theta^* \otimes \tilde{A}_l \tilde{A}_l^H \right) \left( \Sigma_{y'} \otimes \tilde{A}_l \tilde{A}_l^H + \frac{4\gamma^2}{3M_{\tilde{p} \cdot \tau}^2} I_{\tilde{p} \cdot \tau} \right)^{-1}. \quad \text{(D.2)}$$

Next, recall that $\gamma$ is set to $\eta$ times the maximal standard deviation of the quantizer input. Thus, by (C.2),

$$\gamma^2 = \kappa_{\tilde{p} \cdot \tau} \max_{l=1,\ldots,\tilde{p} \cdot \tau} \mathbb{E} \left\{ \left( \left( I_{\tau} \otimes \tilde{A}_l \right) y_l \right)^2 \right\} \overset{(a)}{=} \kappa_{\tilde{p} \cdot \tau} \cdot \sigma^2 \max_{l=1,\ldots,\tilde{p} \cdot \tau} \left( \tilde{A}_l \tilde{A}_l^H \right)_{l,l}^2, \quad \text{(D.3)}$$

where $(a)$ holds by writing the covariance of $y_l$ and as the maximal diagonal entry of a Kronecker product of positive semi-definite matrices is the product of the maximal diagonal entries [37, Ch. 7.8]. Substituting (D.3) in (D.1) results in

$$\text{MSE} \left( \tilde{A}_l \right) = \frac{1}{n_u} \text{Tr} \left( \Phi_l^2 \right) - \frac{1}{n_u \cdot n_t} \text{Tr} \left( \Theta^T D_{l,l}^4 \Theta^* \otimes \tilde{A}_l \tilde{A}_l^H \right) \left( \Sigma_{y'} \otimes \tilde{A}_l \tilde{A}_l^H \right) + \frac{4\kappa_{\tilde{p} \cdot \tau} \cdot \sigma^2}{3M_{\tilde{p} \cdot \tau}^2} \max_{l=1,\ldots,\tilde{p} \cdot \tau} \left( \tilde{A}_l \tilde{A}_l^H \right)_{l,l} I_{\tilde{p} \cdot \tau} \right)^{-1}. \quad \text{(D.4)}$$

Using (D.4), we can now characterize the optimal unitary rotation for any given $\tilde{A}_l$, as stated in the following lemma:
Lemma D.1. For every matrix $\tilde{A}_l \in \mathbb{C}^{\tilde{p} \times n_t}$ there exists a unitary matrix $U_{\tilde{A}} \in \mathbb{C}^{\tilde{p} \times \tilde{p}}$ such that

$$\text{MSE} \left( \tilde{A}_l \right) \geq \text{MSE} \left( U_{\tilde{A}} \tilde{A}_l \right) \geq \frac{1}{n_u} \text{Tr} \left( \Phi_l^2 \right) - \frac{1}{n_u \cdot n_l} \text{Tr} \left( \left( \Theta^T D_{l,\theta}^4 \Theta^* \otimes \tilde{A}_l \tilde{A}_l^H \right) \right) \times \left( \left( \Sigma_{y_l} \otimes \tilde{A}_l \tilde{A}_l^H \right) + \frac{2\kappa_{\bar{p} \tau} \cdot \sigma_{\tau}^2}{3M_{\bar{p} \tau}^2 \cdot \bar{p}} \text{Tr} \left( \tilde{A}_l \tilde{A}_l^H \right) I_{\bar{p} \tau} \right)^{-1} \right).$$

(D.5)

The unitary matrix $U_{\tilde{A}}$ is a set such that $U_{\tilde{A}} \tilde{A}_l \tilde{A}_l^H U_{\tilde{A}}^H$ is weakly majorized by all possible rotations of $\tilde{A}_l \tilde{A}_l^H$.

**Proof:** The lemma is obtained by repeating the arguments in [22, Lem. C.1], thus its proof is omitted for brevity.

We can now characterize the optimal $\tilde{A}_l$ as the matrix which minimizes (D.5). Note that the right hand side of (D.5) is invariant to replacing $\tilde{A}_l$ with $\alpha \cdot U \tilde{A}_l V$ for any $\alpha > 0$ and for any unitary $U, V$. Consequently, we can fix $\frac{4\kappa_{\bar{p} \tau} \cdot \sigma_{\tau}^2}{3M_{\bar{p} \tau}^2 \cdot \bar{p}} \text{Tr} \left( \tilde{A}_l \tilde{A}_l^H \right) = 1$, and write $\tilde{A}_l \tilde{A}_l^H = \Lambda$, where $\Lambda \in \mathbb{C}^{\bar{p} \times \bar{p}}$ is diagonal with non-negative diagonal entries arranged in descending order. Consequently, minimizing (D.5) reduces to solving

$$\arg \max_\Lambda \text{Tr} \left( \left( \Theta^T D_{l,\theta}^4 \Theta^* \otimes \Lambda \right) \left( \Sigma_{y_l} \otimes \Lambda \right) + I_{\bar{p} \tau} \right)^{-1} \right),$$

subject to $\frac{4\kappa_{\bar{p} \tau} \cdot \sigma_{\tau}^2}{3M_{\bar{p} \tau}^2 \cdot \bar{p}} \text{Tr} \left( \Lambda \right) = 1, \ (\Lambda)_{i,i} \geq 0, \ \forall i \in \{1, \ldots, \bar{p}\}$.

By (D.5), the dynamic range is now $\gamma^2 = \frac{\kappa_{\bar{p} \tau} \cdot \sigma_{\tau}^2}{\bar{p}} \text{Tr} \left( \Lambda \right) = \frac{3M_{\bar{p} \tau}^2}{4}$. Plugging the resulting $\gamma$ into (D.2) proves (255b).

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

$$M \triangleq \left( \Sigma_{y_l} \otimes \Lambda \right) + I_{\bar{p} \tau} = \left( \left( \Theta^T \sum_{m=1}^{n_c} D_{l,m}^2 \Theta^* + \sigma_W^2 I_{\tau} \right) \otimes \Lambda \right) + I_{\bar{p} \tau}$$

$$= \left( I_{\tau} \otimes \left( I_{\bar{p}} + \sigma_W^2 \Lambda \right) \right) + \left( \Theta^T \otimes I_{\bar{p}} \right) \left( \sum_{m=1}^{n_c} D_{l,m}^2 \otimes \Lambda \right) \left( \Theta^* \otimes I_{\bar{p}} \right).$$

(D.7)

Applying the matrix inversion lemma to (D.7), recalling that $\Theta^T \Theta = \tau \cdot I_{n_u}$ results in

$$M^{-1} = \left( I_{\tau} \otimes \left( I_{\bar{p}} + \sigma_W^2 \Lambda \right)^{-1} \right) - \left( \Theta^T \otimes \left( I_{\bar{p}} + \sigma_W^2 \Lambda \right)^{-1} \right) \times \left( \left( \sum_{m=1}^{n_c} D_{l,m}^2 \otimes \Lambda \right)^{-1} + \left( \tau \cdot I_{n_u} \otimes \left( I_{\bar{p}} + \sigma_W^2 \Lambda \right)^{-1} \right) \right)^{-1} \left( \Theta^* \otimes \left( I_{\bar{p}} + \sigma_W^2 \Lambda \right)^{-1} \right).$$

(D.8)
Since $I_p + \sigma^2_W \Lambda$ is diagonal, it follows from (D.8) that

$$\text{Tr} \left( (\Theta^T D^4_{l,\ell} \Theta^* \otimes \Lambda) M^{-1} \right) = \tau \cdot \text{Tr} \left( D^4_{l,\ell} \right) \text{Tr} \left( (I_p + \sigma^2_W \Lambda)^{-1} \Lambda \right)$$

$$- \text{Tr} \left( \left( \tau \cdot D^4_{l,\ell} \otimes (I_p + \sigma^2_W \Lambda)^{-1} \Lambda \right) \left( \left( \tau \cdot \sum_{m=1}^{n_c} D^2_{l,m} \right) \otimes (I_p + \sigma^2_W \Lambda)^{-1} \Lambda \right)^{-1} + I_{n_u,p} \right)^{-1} \right). \quad (D.9)$$

Next, we let $a_i$ be diagonal entries of $\Lambda$ and $t_i = \frac{a_i}{1 + \sigma^2_W a_i}$ be the diagonal entries of $(I_p + \sigma^2_W \Lambda)^{-1} \Lambda$.

Using these notations, (D.9) can be written as

$$\text{Tr} \left( (\Theta^T D^4_{l,\ell} \Theta^* \otimes \Lambda) M^{-1} \right) = \sum_{u=1}^{n_u} \sum_{i=1}^{\tilde{p}} \frac{\tau \cdot \phi^4_{l,u} \cdot a_i}{1 + \tau \sum_{u=1}^{n_u} d^2_{l,m,u} \cdot t_i}$$

$$\overset{(a)}{=} \sum_{u=1}^{n_u} \sum_{i=1}^{\tilde{p}} \frac{\tau \cdot \phi^4_{l,u} \cdot a_i}{\tau \cdot \phi^2_{l,u} \cdot a_i + f^2_{l,u} \cdot t_i}, \quad (D.10)$$

where $(a)$ follows from the definition of $f_{l,u}$ in (17), and since $\phi^2_{l,u} = f_{l,u} d^2_{l,u}$. By combining (D.10) and (D.6) it holds that the analog combining matrix which minimizes the average MSE is given by $U_A \Lambda_A$, where $U_A$ is given in Lemma D.1, and $\Lambda_A$ is diagonal such that the diagonal entries of $\Lambda_A \Lambda^H_A$, denoted $\{a^o_i\}$, are the solution to

$$\{a^o_i\}_{i=1}^{\tilde{p}} = \arg \max_{\{a_i\}_{i=1}^{\tilde{p}}} \sum_{i=1}^{\tilde{p}} \sum_{u=1}^{n_u} \frac{\tau \cdot \phi^4_{l,u} \cdot a_i}{\tau \cdot \phi^2_{l,u} \cdot a_i + f^2_{l,u}}$$

subject to

$$\frac{4 \kappa_{\tilde{p},\tau} \cdot \sigma^2_t}{3 M^2_{\tilde{p},\tau} \cdot \tilde{p}} \sum_{i=1}^{\tilde{p}} a_i = 1, \quad a_i \geq 0, \quad \forall i \in \{1, \ldots, \tilde{p}\}. \quad (D.11)$$

The solution of (D.11) is stated in the following lemma:

**Lemma D.2.** The entries $\{a^o_i\}$ which solve (D.11) are given by $a^o_i = \frac{3 M^2_{\tilde{p},\tau}}{4 \kappa_{\tilde{p},\tau} \cdot \sigma^2_t}$, $\forall i \in \{1, \ldots, \tilde{p}\}$.

**Proof:** Note that $\forall m \in N_u$, the mapping $x \mapsto \frac{\tau \cdot \phi^4_{l,u} \cdot x}{\tau \cdot \phi^2_{l,u} \cdot x + f^2_{l,u}}$ is concave over $\mathcal{R}^+$. Consequently, $\xi(x) \triangleq \sum_{u=1}^{n_u} \frac{\tau \cdot \phi^4_{l,u} \cdot x}{\tau \cdot \phi^2_{l,u} \cdot x + f^2_{l,u}}$ is concave [39, 3.2.1], and thus

$$\frac{1}{\tilde{p}} \sum_{i=1}^{\tilde{p}} \xi(a_i) \leq \xi \left( \frac{1}{\tilde{p}} \sum_{i=1}^{\tilde{p}} a_i \right) = \xi \left( \frac{3 M^2_{\tilde{p},\tau}}{4 \kappa_{\tilde{p},\tau} \cdot \sigma^2_t} \right). \quad (D.12)$$

It follows from (D.12) that setting $a_i = \frac{3 M^2_{\tilde{p},\tau}}{4 \kappa_{\tilde{p},\tau} \cdot \sigma^2_t}$ maximizes (D.11), thus proving the lemma. □
By substituting (D.10) in (D.5), we have that the resulting average MSE is given by

\[ \text{MSE} \left( \tilde{A}_l \right) = \frac{1}{n_u} \sum_{u=1}^{n_u} \phi_{l,u}^2 - \frac{1}{n_u} \sum_{u=1}^{n_u} \frac{\tau \cdot \phi_{l,u}^4 \cdot \phi_{l,u}^4}{\phi_{l,u}^2 + 4 \kappa \cdot \sigma_i^2 \cdot f_{l,u}^2} \]

where (a) follows from Lemma D.2 and since \( \tilde{p}_n = r \), thus proving (25a). As the resulting \( \tilde{A}_l \) is a diagonal matrix with identical diagonal matrix, \( U \tilde{A} \) in Lemma D.1 can be set to the identity matrix, concluding the proof of the proposition.

REFERENCES

[1] Y. C. Eldar. *Sampling Theory: Beyond Bandlimited Systems*. Cambridge Press, 2015.
[2] R. M. Gray and D. L. Neuhoff. “Quantization”. *IEEE Trans. Inform. Theory*, vol. 44, no. 6, Oct. 1998, pp. 2325-2383.
[3] M. R. D. Rodrigues, N. Deligiannis, L. Lai, and Y. C. Eldar. “Rate-distortion trade-offs in acquisition of signal parameters”. *Proc. IEEE ICASSP*, New-Orleans, LA, 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, Mar. 2018, pp. 1141-1154.
[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, Aug. 2017, pp. 4075-4089.
[6] J. Choi, J. Mo, and R. W. Heath. “Peak maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs”. *IEEE Trans. Commun.*, vol. 64, no. 5, May 2016, pp. 2005-2018.
[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, Jun. 2017, pp. 4038–4051.
[8] H. Pirzadeh and A. L. Swindlehurst. “Spectral efficiency of mixed-ADC massive MIMO”. *IEEE Trans. Signal Process.*, vol. 66, no. 13, Jul. 2018, pp. 3599-3613.
[9] J. Choi, J. Sung, B. L. Evans, and A. Gatherer. “Antenna selection for large-scale MIMO systems with low-resolution ADCs”. *Proc. IEEE ICASSP*, Calgary, Canada, Apr. 2018.
[10] 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, Apr. 2017, pp. 2274-2287.
[11] 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, Dec. 2017, pp. 6201-6216.
[12] 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. Top. Signal Process.*, vol. 12, no. 3, Jun. 2018, pp. 484-498.
[13] 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, Nov. 2016, pp. 7738-7752.
[14] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. “What will 5G be?”. *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, Jun. 2014, pp. 1065–1082.
[15] T. L. Marzetta. “Noncooperative cellular wireless with unlimited numbers of base station antenna”. *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, Nov. 2010, pp. 3950–3600.
[16] N. Shlezinger and Y. C. Eldar. “On the spectral efficiency of noncooperative uplink massive MIMO systems”. IEEE Trans. Commun., early access, 2018.
[17] T. M. Cover and J. A. Thomas. Elements of Information Theory. Wiley, 2006.
[18] H. Witsenhausen. “Indirect rate distortion problems”. IEEE Trans. Inform. Theory, vol. 26, no. 5, Sep. 1980, pp. 518-521.
[19] J. K. Wolf and J. Ziv. “Transmission of noisy information to a noisy receiver with minimum distortion”. IEEE Trans. Inform. Theory, vol. 16, no. 4, Jul. 1970, pp. 406-411.
[20] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar. “Fundamental distortion limits of analog-to-digital compression”. Accepted to the IEEE Trans. Inform. Theory, 2018.
[21] 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, Mar. 2018, pp. 16-39.
[22] N. Shlezinger, Y. C. Eldar, and M. R. D. Rodrigues. “Hardware-limited task-based quantization”. arXiv preprint, arXiv:1807.08305, 2018.
[23] 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, Dec. 2017, pp. 8131-8146.
[24] S. Rini, L. Barlett, E. Erkip, and Y. C. Eldar. “A general framework for MIMO receivers with low-resolution quantization”. Proc. IEEE ITW, Kaohsiung, Taiwan, Nov. 2017.
[25] S. Stein and Y. C. Eldar. “Hybrid analog-digital beamforming for massive MIMO systems”. arXiv preprint, arXiv:1712.03485, 2017.
[26] N. Shlezinger, O. Dicker, Y. C. Eldar, M. F. Imani, and D. R. Smith. “Dynamic metasurfaces for massive MIMO networks”. Submitted to IEEE ICASSP, 2019.
[27] Y. Polyanskiy and Y. Wu. Lecture Notes on Information Theory. 2015.
[28] J. Li, N. Chaddha, and R. M. Gray. “Asymptotic performance of vector quantizers with a perceptual distortion measure”. IEEE Trans. Inform. Theory, vol. 45, no. 4, May 1999, pp. 1082-1091.
[29] F. D. Nesser and J. L. Massey. “Proper complex random processes with applications to information theory,” IEEE Trans. Inform. Theory, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
[30] O. Rioul. “Information theoretic proofs of entropy power inequalities”. IEEE Trans. Inform. Theory, vol. 57, no. 1, Jan. 2011, pp. 33-55.
[31] B. Widrow, I. Kollar, and M. C. Liu. “Statistical theory of quantization”. IEEE Trans. Instrumentation and Measurement, vol. 45, no. 2, Apr. 1996, pp. 353-361.
[32] R. M. Gray and T. G. Stockholm. “Dithered quantization”. IEEE Trans. Inform. Theory, vol. 39, no. 3, Mar. 1993, pp. 805-812.
[33] G. Zeitler, G. Kramer, and A. C. Singer. “Bayesian parameter estimation using single-bit dithered quantization”. IEEE Trans. Signal Process., vol. 60, no. 6, Jun. 2012, pp. 2713-2726.
[34] O. Dabeer and U. Madhow. “Channel estimation with low-precision analog-to-digital conversion”. Proc. IEEE ICC, May 2010.
[35] 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, May 2018, pp. 2557-2571.
[36] J. D. Gibson. “Rate distortion functions and rate distortion function lower bounds for real-world sources”. Entropy, vol. 19, no. 604, Nov. 2017.
[37] C. D. Meyer. Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics, 2000.
[38] D. P. Palomar and Y. Jiang. MIMO Transceiver Design via Majorization Theory. Now Publishers, 2007.
[39] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, 2004.