Asymptotic SEP Analysis and Optimization of Linear-Quantized Precoding in Massive MIMO Systems

Zheyu Wu, *Graduate Student Member, IEEE*, Junjie Ma, Ya-Feng Liu, *Senior Member, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*— A promising approach to deal with the high hardware cost and energy consumption of massive MIMO transmitters is to use low-resolution digital-to-analog converters (DACs) at each antenna element. This leads to a transmission scheme where the transmitted signals are restricted to a finite set of voltage levels. This paper is concerned with the analysis and optimization of a low-cost quantized precoding strategy, referred to as linear-quantized precoding, for a downlink massive MIMO system under Rayleigh fading. In linear-quantized precoding, the signals are first processed by a linear precoding matrix and subsequently quantized component-wise by the DAC. In this paper, we analyze both the signal-to-interference-plus-noise ratio (SINR) and the symbol error probability (SEP) performances of such linear-quantized precoding schemes in an asymptotic framework where the number of transmit antennas and the number of users grow large with a fixed ratio. Our results provide a rigorous justification for the heuristic arguments based on the Bussgang decomposition that are commonly used in prior works. Based on the asymptotic analysis, we further derive the optimal precoder within a class of linear-quantized precoders that includes several popular precoders as special cases. Our numerical results demonstrate the excellent accuracy of the asymptotic analysis for finite systems and the optimality of the derived precoder.**

Manuscript received 4 February 2023; revised 13 July 2023; accepted 18 September 2023. Date of publication 2 October 2023; date of current version 19 March 2024. The work of Zheyu Wu and Ya-Feng Liu was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 12371314, Grant 12022116, and Grant 12288201. The work of Junjie Ma was supported in part by the National Natural Science Foundation of China under Grant 12101592 and in part by the Key Research Program of the Chinese Academy of Sciences under Grant XDA27010102. The work of A. Lee Swindlehurst was supported in part by the U.S. National Science Foundation under Grant CCF-2008724 and Grant CCF-2225575. An earlier version of this paper was presented in part at the 2023 Signal Processing Advances in Wireless Communications (SPAWC) [1]. *(Corresponding author: Ya-Feng Liu.)*

Zheyu Wu is with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: wuzy@lsec.cc.ac.cn).

Junjie Ma and Ya-Feng Liu are with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: majunjie@lsec.cc.ac.cn; yafliu@lsec.cc.ac.cn).

A. Lee Swindlehurst is with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697 USA (e-mail: swindle@uci.edu).

Communicated by A. Sabharwal, Associate Editor for Communications.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3321340.

Digital Object Identifier 10.1109/TIT.2023.3321340

***Index Terms*— Massive MIMO, quantized precoding, linear precoding, asymptotic analysis, Haar random matrix, random matrix theory.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) is a key technology for 5G wireless communication systems. By equipping the base station (BS) with many antennas, massive MIMO can significantly improve the channel capacity, energy efficiency, and spectral efficiency of wireless communication systems [2], [3], [4]. Despite the great potential of massive MIMO systems, high power consumption and hardware cost are serious practical challenges for their commercial deployment.

One of the main power-hungry components in a massive MIMO system is the digital-to-analog-converter (DAC) [5], the number of which scales linearly with the number of antennas at the BS. To reduce circuit complexity and power consumption, low-resolution DACs have been considered for massive MIMO systems [6], [7], [8], [9], [10], [11], [12]. Unlike conventional precoding schemes where at the symbol sampling points the transmitted signals can be freely chosen from a continuous set, only a small finite set of signals can be transmitted to convey information when low-resolution DACs are employed. The analysis and design of quantized precoding with the use of low-resolution DACs has become an active research topics in recent years [6], [7], [8], [9], [10], [11], [12].

Power ampifiers (PAs) are another main source of power consumption in massive MIMO systems. To achieve the highest power efficiency, the PAs need to operate close to saturation, but for continuous-valued signals this incurs nonlinear distortion and causes difficulties for signals with high peak-to-average power ratio (PAPR) [13]. A popular way to handle such difficulty is to restrict the transmitted signal from each antenna to have the same amplitude [14], [15], [16], which minimizes the PAPR and enables the employment of the most efficient and cheapest PAs. Combining such a constant envelope (CE) constraint with the use of low-resolution DACs motivates a new quantized precoding scheme, quantized constant envelope (QCE) precoding, where at the symbol sampling points the transmitted signals are restricted to have a fixed amplitude and their phases are limited to finite values. The QCE precoding scheme has attracted significant research interest [17], [18], [19], [20], [21], [22], [23] as it combines the advantage of using low-resolution DACs and energy-

efficient PAs. In particular, as an extreme case of both QCE and traditional quantized precoding, one-bit precoding (where one-bit DACs are employed) has been widely and extensively studied [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. The power efficiency gain can be several dB, and in many cases is sufficient to overcome the loss in fidelity due to the coarse quantization.

We note here that traditional quantized precoding (without the CE constraint) and QCE precoding are both realized by the use of low-resolution DACs and have a common feature that the transmitted signals are only allowed to be selected from a finite set. In the following discussion, we will refer to them (and possibly other precoding schemes with the finite transmission set feature) collectively as quantized precoding.

Existing quantized precoding schemes can be broadly categorized into two classes: *linear-quantized* precoding and *nonlinear* precoding. A linear-quantized precoding scheme[1] simply quantizes the output of a linear precoder [6], [7], [8], [9], [24], [25], [26], [27], [28], [29]. In contrast, nonlinear precoders do not have this simple structure and are typically obtained by solving appropriate optimization problems [10], [11], [12], [18], [19], [20], [21], [22], [23], [30], [31], [32], [33], [34], [35].

In this paper, we focus on the analysis and optimization of linear-quantized precoding schemes, which are arguably more practical than the computationally expensive nonlinear schemes. Unlike existing works that focus on either traditional quantized precoding or QCE precoding, this paper deals with these two types of quantized precoding in a unified framework. In what follows, we first give a brief review of related works and then present the main contributions of this paper.

A. Related Work

1) Linear-Quantized Precoding: A direct approach to obtain linear-quantized precoders is to quantize the output of classical linear precoders such as the matched filter (MF) and zero-forcing (ZF) precoders [6]. However, this approach does not take into account the effect of quantization and thus yields precoders that, although simple, are suboptimal in the context of quantized precoding. Noting this, the authors in [7] characterized the mean square error (MSE) between the desired symbol and the received signal with the presence of low-resolution DACs and proposed the quantized transmit Wiener filter (TxWFQ) precoder that minimizes the MSE. Under the same setup as [7], the authors of [8] proposed a gradient-based approach to maximize the weighted sum rate of the system. For the one-bit case, the authors in [24] proposed a minimum mean square error (MMSE) based precoder. Later, a higher-rank linear precoder was designed in [25] for a downlink one-bit massive MIMO system, showing superior performance to traditional linear-quantized precoders of channel rank. We remark here that all existing works on the design of linear-quantized precoding focus on traditional quantized precoding or the special one-bit case. To the best of our knowledge, no existing work considers the design of linear-quantized precoding in the QCE context.

There are also some works focusing on the performance analysis of linear-quantized schemes, e.g., [6] and [9] for traditional quantized precoding, [26], [27], [28], [29] for one-bit precoding, and [17] and [36] for QCE precoding. Specifically, the authors in [6] and [9] derived lower-bounds on the downlink achievable rates of linear-quantized precoding for a flat-fading and a frequency-selective channel, respectively. For a one-bit massive MIMO system, [26] derived a lower bound on the achievable rate for MF precoding with estimated channel state information (CSI). The performance of the one-bit ZF precoder was investigated in [27], in which a closed-form expression of the symbol error probability (SEP) was derived in the asymptotic setting where the numbers of transmit antennas and users both tend to infinity with a fixed ratio. The same problem was considered in [28] and [29], where the input-output correlation relationship was expanded up to third-order instead of first-order as in [27], and the derived SEP expression shows better accuracy than that in [27] when the number of users is small relative to the number of transmit antennas.

A widely used technique for analyzing the performances of linear-quantized precoding schemes is the *Bussgang decomposition* [37], which decomposes a non-linear function of a Gaussian signal as the sum of a linear signal term and an uncorrelated distortion term. We remark that although the Bussgang decomposition is per se rigorous, it is often used in conjunction with various heuristics to analyze the performance of linear-quantized precoding. For instance, the distortion term is often treated as a random variable that is independent of all other random variables in the system. Although there is strong numerical evidence that the heuristic treatments can yield accurate predictions (e.g., SEP performance) for large systems [27], [28], [29], a rigorous analysis of such heuristics in the context of linear-quantized precoding is still lacking. Please refer to Section II-C for a detailed discussion of the Bussgang decomposition technique.

Beyond the analyses of traditional quantized precoding and one-bit precoding, there are also some preliminary attempts to analyze the performance of QCE precoding. Specifically, [36] studied the statistical properties of the CE quantizer (which models the overall operation of low-resolution DACs and the CE constraint and generates signals satisfying the QCE constraint) and derived closed-form expressions of the cross-correlation factors between the input and output signals of the CE quantizer. Very recently, the authors in [17] considered QCE precoding for a multiple-input single-output (MISO) system and derived the diversity order of the MF precoder, which characterized how fast the system SEP tends to zero as the signal-to-noise ratio (SNR) grows [38].

2) Nonlinear Precoding: Besides linear-quantized precoding schemes, various nonlinear precoders based on different criteria have been proposed in recent years, see, e.g., [10], [11], and [12] for traditional quantized precoding and [18], [19], [20], [21], [22], and [23] for QCE precoding. There are also many algorithms designed specifically for one-bit precoding, see, e.g., [30], [31], [32], [33], and [34]. Nonlinear precoding

[1]Note that the overall operation of a linear-quantized scheme is not linear due to the presence of the quantization step, but for convenience we will use this term throughout this paper.

schemes (especially symbol-level nonlinear precoders) usually have better symbol error rate (SER) performance than their linear counterparts but their computational complexity is much higher.

Although there have been substantial progress in the design of nonlinear precoding schemes, the performance analysis of nonlinear precoding remains an open problem. This is because nonlinear precoders are typically solutions to complicated optimization problems without closed-form expressions. In addition, the discrete nature of the transmitted signals in quantized precoding leads to discrete constraints in the corresponding optimization problem, which further complicates the analysis. Analyzing and designing nonlinear precoding schemes are beyond the scope of this paper and can be considered as a future work.

B. Our Contributions

In this paper, we analyze the performance of a broad class of linear-quantized precoding schemes for a downlink massive MIMO system. Our results rigorously justify and substantially generalize existing results for MF and ZF based schemes derived using heuristic Bussgang decomposition arguments. The main contributions are summarized as follows.

1) *Statistically equivalent model*: By exploiting a recursive characterization of the Haar random matrix [39], we derive a model that is statistically equivalent to the original system model. The statistically equivalent model is close to a "signal plus independent Gaussian noise" form and is more amenable to analysis. This step is non-asymptotic and the technique we use may be applicable to other problems as well.
2) *Asymptotic analysis*: We further consider the large system limit as in [27] and show that in the asymptotic regime, the statistically equivalent model is exactly in a "signal plus independent Gaussian noise" form. This provides a rigorous justification for the heuristic analyzes based on the Bussgang decomposition. We also prove that the signal-to-interference-plus-noise ratio (SINR) and SEP of the original model converge to those of the asymptotic model. Simulations show that the asymptotic results are accurate for realistic systems with finite dimensions.
3) *Optimal linear-quantized precoder*: Based on the asymptotic analysis, we derive the optimal linear-quantized precoder that optimizes both the asymptotic SINR and the asymptotic SEP performance. We show that the optimal linear-quantized precoder is a regularized ZF (RZF) precoder, whose regularization parameter is determined by the quantization type/level as well as the system parameters. To the best of our knowledge, the optimal RZF precoder derived in this paper is the first linear-quantized precoder applicable to general forms of quantization.

C. Organization and Notations

The remaining parts of the paper are organized as follows. Section II describes the system model and the problem formulation. Some preliminaries for analysis are introduced in Section III. Section IV derives the statistically equivalent model and gives the asymptotic analysis. The optimal linear-quantized precoder is then given in Section V. Simulation results are shown in Section VI and the paper is concluded in Section VII.

Notation: Throughout the paper, we use the typefaces x, $\mathbf{x}$, $\mathbf{X}$, and $\mathcal{X}$ to denote scalar, vector, matrix, and set, respectively. For a vector $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}[i_1 : i_2]$ denotes a sub-vector of $\mathbf{x}$ consisting of its i_1-th to i_2-th elements, where $1 \leq i_1 \leq i_2 \leq n$; in particular, $\mathbf{x}[i]$ is the i-th entry of $\mathbf{x}$, and x_i is also used if it does not cause any ambiguity. For a matrix $\mathbf{X}$, $\mathbf{X}[i_1, i_2]$ is the (i_1, i_2)-th entry of $\mathbf{X}$. The operators $\arg(\cdot)$, $\mathcal{R}(\cdot)$, $\mathcal{I}(\cdot)$, $(\cdot)^\dagger$, $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$, and $(\cdot)^{-1}$ return the angle, the real part, the imaginary part, the conjugate, the transpose, the conjugate transpose, and the inverse of their corresponding arguments, respectively. We use $\|\cdot\|$ to denote the ℓ_2 norm of the corresponding vector or the spectral norm of the corresponding matrix. For $m, n \in \mathbb{N}$, we denote the $m \times m$ identity matrix by $\mathbf{I}_m$ and the $m \times n$ matrix of all zero entries by $\mathbf{0}_{m\times n}$. We use $\mathrm{diag}(x_1, x_2, \ldots, x_n)$ to refer to a diagonal matrix with $\{x_i\}_{i=1}^n$ as its diagonal entries. We use $\mathcal{U}(n)$ to denote the set of $n \times n$ unitary matrices over $\mathbb{C}$. The operators $\mathbb{E}[\cdot]$, $\mathrm{var}(\cdot)$, and $\mathbb{P}(\cdot)$ return the expectation, the variance, and the probability of their corresponding argument, respectively. For two random variables X and Y, $X \overset{d}{=} Y$ means that they have the same distribution. We denote almost sure convergence by $\xrightarrow{a.s.}$. We use $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ to denote the zero-mean circularly symmetric complex Gaussian distribution with covariance matrix $\sigma^2\mathbf{I}$, and $\mathrm{Unif}(\mathcal{S})$ to denote uniform distribution on set $\mathcal{S}$. We reserve the sans serif font (e.g., g) for vectors with i.i.d. standard complex Gaussian random variables. Finally, j is the imaginary unit satisfying $j^2 = -1$.

II. System Model and Problem Formulation

A. Linear-Quantized Precoding

Consider the downlink of a multiuser massive MIMO system in which an N-antenna BS simultaneously serves K single-antenna users, where $K < N$. The received signals at the users can be modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

where $\mathbf{y} \in \mathbb{C}^K$ is the received signal vector of the users; $\mathbf{x} \in \mathbb{C}^N$ is the transmitted signal vector from the BS; $\mathbf{H} \in \mathbb{C}^{K\times N}$ models the channel matrix between the BS and the users, and $\mathbf{n} \in \mathbb{C}^K$ is the additive noise. We assume that the analog-to-digital converters (ADCs) equipped at the user side are ideal and have infinite resolution and that perfect CSI is available at the BS. We model the DAC as a quantization function and ignore various practical effects such as glitches, element mismatch, slewing, thermal noise, clipping, etc [40], [41], [42].

In this paper, we consider the linear-quantized precoding scheme, where the signal vector to be transmitted at the BS has the following form

$$\mathbf{x} = \eta\, q(\mathbf{P}\mathbf{s}). \tag{1}$$

In the above expression, $\mathbf{s} \in \mathbb{C}^K$ is the desired data vector; $\mathbf{P} \in \mathbb{C}^{N\times K}$ is a precoding matrix; $q(\cdot) : \mathbb{C} \to \mathcal{X}_L$ is a

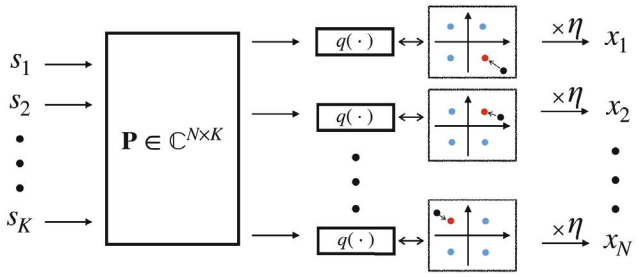


Fig. 1. An illustration of the linear-quantized precoding scheme.

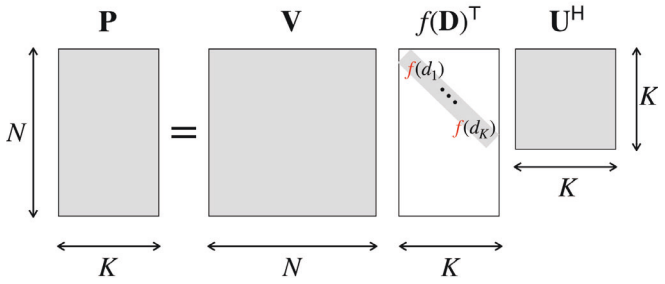


Fig. 2. The structure of the precoding matrix in consideration.

quantization function that acts component-wise on its input vector, where $\mathcal{X}_L$ is a finite set with L elements and L is referred to as the quantization level; η is a scaling factor to ensure that the following average transmit power constraint is satisfied:

$$\frac{1}{N}\mathbb{E}[\|\mathbf{x}\|_2^2] \le P_T, \tag{2}$$

where $P_T > 0$ is the maximum average transmit power. See Fig. 1 for an illustration of the linear-quantized precoding scheme.

We now introduce the quantization function corresponding to traditional quantized precoding and QCE precoding, which are most relevant for applications.

- *Traditional quantized precoding*: In this case, the real and imaginary parts of the input signal are quantized independently with a pair of low-resolution DACs and $\mathcal{X}_L$ can be expressed as

$$\mathcal{X}_L = \left\{ x \mid \mathcal{R}(x), \mathcal{I}(x) \in \left\{ \frac{\Delta}{2}\left(2\ell - 1 - \sqrt{L}\right), \ell = 1, \ldots, \sqrt{L} \right\} \right\}, \tag{3}$$

  where Δ is the quantization interval. The corresponding quantization function maps its input to the nearest point in (3). In the following, we call it independent quantizer since the quantizer acts independently on the real and imaginary parts of its input, and denote it by $q_{\mathrm{I}}(\cdot)$. For an L-level independent quantizer, the resolution of the DACs is $\frac{1}{2}\log_2 L$ bits, where $L \ge 4$ and is a power of 2.
- *QCE precoding*: In this case, the CE constraint is combined with the use of low-resolution DACs and $\mathcal{X}_L$ has the following expression:

$$\mathcal{X}_L = \left\{ e^{j\frac{(2\ell-1)\pi}{L}} \mid \ell = 1, 2, \ldots, L \right\}. \tag{4}$$

  The corresponding quantization function maps its input to the nearest point in (4). In the following, we call it CE quantizer and denote it by $q_{\mathrm{CE}}(\cdot)$. The resolution of the DACs is $\log_2 \frac{L}{2}$ bits for an L-level CE quantizer, where $L \ge 4$ and is a power of 2.

Note that when $L = 4$ and $\Delta = 2$, the independent quantizer and the CE quantizer are the same, both reducing to the one-bit precoding case.

Let $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}$ be the singular value decomposition (SVD) of $\mathbf{H}$, where $\mathbf{U} \in \mathcal{U}(K), \mathbf{V} \in \mathcal{U}(N)$, and $\mathbf{D} = \left(\mathrm{diag}(d_1, d_2, \ldots, d_K) \quad \mathbf{0}_{K\times(N-K)}\right) \in \mathbb{R}^{K\times N}$ with $d_1, d_2, \ldots, d_K$ representing the non-zero singular values[2] of $\mathbf{H}$. In this paper, we focus on precoding matrices with the following structure:

$$\mathbf{P} = \mathbf{V} f(\mathbf{D})^{\mathsf{T}} \mathbf{U}^{\mathsf{H}}, \tag{5}$$

where $f(\cdot)$ acts independently on the nonzero singular values of $\mathbf{H}$, i.e.,

$$f(\mathbf{D}) = \left(\mathrm{diag}(f(d_1), f(d_2), \ldots, f(d_K)) \quad \mathbf{0}_{K\times(N-K)}\right).$$

See Fig. 2 for an illustration of the structure of $\mathbf{P}$. The motivations to consider the special class of the precoding matrix in (5) are twofold. First, as will be shown in Section III, the structure of $\mathbf{P}$ in (5) enables us to apply existing results in random matrix theory for performance analysis. Second, the structure of $\mathbf{P}$ is fairly general and includes the following popular precoders as special cases:

- *MF*: $\mathbf{P} = \mathbf{H}^{\mathsf{H}}$, which corresponds to $f(x) = x$ in (5);
- *ZF*: $\mathbf{P} = \mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{H}^{\mathsf{H}})^{-1}$, which corresponds to $f(x) = x^{-1}$ in (5);
- *RZF*: $\mathbf{P} = \mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \rho\mathbf{I}_K)^{-1}$, which corresponds to $f(x) = \frac{x}{x^2+\rho}$ in (5).

With the above linear-quantized precoding scheme, the received signals at the user side read

$$\mathbf{y} = \eta\mathbf{H}q(\mathbf{P}\mathbf{s}) + \mathbf{n} = \eta\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}q(\mathbf{V}f(\mathbf{D})^{\mathsf{T}}\mathbf{U}^{\mathsf{H}}\mathbf{s}) + \mathbf{n}. \tag{6}$$

As in [19], [20], [21], and [22], we assume that each user is able to rescale the received signal y_k by a factor $\beta_k \in \mathbb{C}$, i.e., $r_k = \beta_k y_k$. (This corresponds to removing the effective channel gain.) After the rescaling step, the users employ symbol-wise nearest-neighbor decoding, i.e., each user k maps r_k to the nearest constellation point.

As will be shown below, the nonlinear function $f(\cdot)$ in $\mathbf{P}$ has a major impact on the performance of the overall system. In this paper, we will first analyze the performance of the linear-quantized scheme in the asymptotic regime where N and K tend to infinity simultaneously, and then optimize $f(\cdot)$ based on the asymptotic analysis.

B. Assumptions

In this subsection, we specify our assumptions on the system model in (6). We first make a few standard assumptions on $\mathbf{H}$, $\mathbf{n}$, and $\mathbf{s}$.

[2]We assume throughout the paper that $\mathbf{H}$ is of full row rank. This holds with probability one if Assumption 1 further ahead is satisfied, i.e., if the entries of $\mathbf{H}$ are i.i.d. following $\mathcal{CN}(0, \frac{1}{N})$.

Assumption 1: The entries of $\mathbf{H}$ and $\mathbf{n}$ are independently drawn from $\mathcal{CN}\left(0, \frac{1}{N}\right)$ and $\mathcal{CN}\left(0, \sigma^2\right)$, respectively. The entries of $\mathbf{s}$ are independently and uniformly drawn from a finite set $\mathcal{S}_M$ with nonzero elements (i.e., $0 \notin \mathcal{S}_M$), and $\mathbb{E}[|s_1|^2] = \sigma_s^2$. Furthermore, $\mathbf{H}, \mathbf{s}$, and $\mathbf{n}$ are mutually independent.

The i.i.d. Gaussian assumption on the channel $\mathbf{H}$ is widely adopted in the massive MIMO literature for ease of analysis, see, e.g., [27], [28], and [29]. This assumption is reasonable in a rich scattering environment where the number of scattered components is large and independent. Such a scenario arises when the antennas are widely spaced or when the physical environment exhibits scattering in all directions [43]. Note that we have assumed $H_{ij} \sim \mathcal{CN}(0, \frac{1}{N})$ instead of $H_{ij} \sim \mathcal{CN}(0, 1)$. This normalization is introduced as in [44], [45], and [46] to ensure that the received power of the users does not grow with N. We would like to emphasize that the i.i.d. Gaussian assumption on the channel $\mathbf{H}$ is not essential to our analysis. Our results can be extended to a broader class of channel models, as discussed in Remark 3 below Theorem 2.

The assumption on $\mathbf{s}$ is quite general and is satisfied by common constellation schemes including phase shift keying (PSK) and quadrature amplitude modulation (QAM).

For technical reasons, we impose the following assumption on the nonlinear function $f(\cdot)$ in $\mathbf{P}$ in (5).

Assumption 2: The function $f(\cdot)$ is positive, continuous almost everywhere (a.e.), and bounded on any compact set of $(0, \infty)$.

Notice that the f functions corresponding to the MF, ZF, and RZF precoders discussed in the previous subsection all satisfy Assumption 2. We emphasize that the positivity assumption on f is not essential and can be relaxed to $\mathbb{P}\left(f(\mathbf{D}) = \mathbf{0}\right) = 0$. Further, as will be shown in Section V, there always exists an optimal precoder satisfying $f > 0$, implying that the positivity assumption does not impose any restriction in terms of the best achievable performance.

Finally, we assume that the quantization function q in (1) satisfies some regularity conditions, as stated in Assumption 3 below.

Assumption 3: The quantization function $q : \mathbb{C} \to \mathcal{X}_L$ is continuous a.e. and bounded.

It is straightforward to verify that the independent quantizer $q_{\text{I}}(\cdot)$ and the CE quantizer $q_{\text{CE}}(\cdot)$ both satisfy Assumption 3 (as they are piecewise constant). We emphasize that some of our results can be simplified for QCE precoding, i.e., when $q(\cdot) = q_{\text{CE}}(\cdot)$. In the following, we will first present our results in the most general form and then discuss the case of QCE precoding separately.

C. Heuristic Analysis Via Bussgang Decomposition

The nonlinear quantization function $q(\cdot)$ causes some difficulties for performance analysis. A popular technique to deal with it is the Busggang decomposition [6], [9], [26], [27], [28], [29], which decomposes a nonlinear function of Gaussian random variables into a linear signal term and an uncorrelated nonlinear distortion term. We now outline a *heuristic* analysis of the problem using the Bussgang decomposition technique.

Consider the nonlinear quantization process $q(\mathbf{Ps})$, where $\mathbf{P} = \mathbf{V} f(\mathbf{D})^{\mathsf{T}} \mathbf{U}^{\mathsf{H}}$. Under Assumption 1, $\mathbf{V}$ and $\mathbf{U}$ are Haar distributed (see Definition 1 and Lemma 1 further ahead) and it can be shown that $\mathbf{Ps}$ is approximately distributed as

$$\mathbf{Ps} \overset{d}{\approx} \mathcal{CN}(\mathbf{0}, \bar{\alpha}^2 \mathbf{I}_N),$$

where

$$\bar{\alpha}^2 = \frac{1}{N} \mathbb{E}[\|\mathbf{Ps}\|^2].$$

Based on the Busggang decomposition technique [37], we can write $q(\mathbf{Ps})$ as

$$q(\mathbf{Ps}) = \overline{C}_1 (\mathbf{Ps}) + \mathbf{q}_\perp,$$

where $\overline{C}_1 = \mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]/\bar{\alpha}$, $\mathsf{Z} \sim \mathcal{CN}(0, 1)$, and $\mathbf{q}_\perp$ is the residual nonlinear distortion which is approximately orthogonal to $\mathbf{Ps}$. Substituting this decomposition into (6) gives

$$\begin{aligned} \mathbf{y} &= \eta\, \mathbf{H} q(\mathbf{Ps}) + \mathbf{n} \\ &= \eta\, \overline{C}_1 \mathbf{HPs} + \eta\, \mathbf{Hq}_\perp + \mathbf{n} \\ &= \eta \frac{\overline{C}_1}{K} \mathrm{tr}(\mathbf{HP}) \mathbf{s} \\ &\quad + \eta\, \overline{C}_1 \left(\mathbf{HP} - \frac{1}{K} \mathrm{tr}(\mathbf{HP}) \mathbf{I} \right) \mathbf{s} + \eta \mathbf{Hq}_\perp + \mathbf{n}. \end{aligned} \tag{7}$$

In the above decomposition, the first term is a signal term and the last three terms are effective noise terms. This demonstrates the advantage of the Bussgang decomposition: it can transform a nonlinear system into a linear one so that the useful signal and the effective noise can be distinguished. However, the problem is that the distribution of $\mathbf{q}_\perp$ and its correlation with $(\mathbf{H}, \mathbf{s})$ are hard to characterize, which makes it still highly non-trivial to analyze the performance (e.g., SEP performance) of the system with (7). Heuristically, one may treat $\mathbf{q}_\perp$ as if it is independent of both $\mathbf{H}$ and $\mathbf{s}$, so that $\mathbf{Hq}_\perp$ can be approximated as independent Gaussian noise. It turns out that this treatment, though heuristic, leads to very accurate predictions [27]. Developing a new analytical framework that can rigorously justify the above heuristics is a main motivation behind this work.

III. Preliminaries

Our analysis is based on Householder dice [39], a technique for recursively generating Haar random matrices. Before presenting our main results, in this section we first give some preliminaries on the Haar random matrix and the Householder dice technique.

We begin with the definition of the Haar measure and the Haar random matrix.

Definition 1 (Haar Measure [47]): The Haar measure on $\mathcal{U}(N)$ is defined as the unique probability measure μ on $\mathcal{U}(N)$ that satisfies the following translation invariant property: for any measurable subset $\mathcal{A} \subset \mathcal{U}(N)$ and any fixed $\mathbf{M} \in \mathcal{U}(N)$,

$$\mu(\mathbf{M}\mathcal{A}) = \mu(\mathcal{A}\mathbf{M}) = \mu(\mathcal{A}),$$

where $\mathbf{M}\mathcal{A}$ denotes the set obtained by taking all the elements of $\mathcal{A}$ and multiplying them by $\mathbf{M}$.

In the following, we denote by Haar(N) the ensemble of random unitary matrices drawn from the Haar measure on $\mathcal{U}(N)$.

Lemma 1 below is a well-known fact in random matrix theory [48] and suggests the crucial role of the Haar random matrix plays in our analysis.

Lemma 1: Let $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}$ be the SVD of $\mathbf{H}$. Under Assumption 1, $\mathbf{U}, \mathbf{D}, \mathbf{V}$ are mutually independent and $\mathbf{U}, \mathbf{V}$ are Haar distributed random matrices.

We now introduce the Householder dice (HD) technique proposed in [39], which deals with an iterative process involving Haar matrices as follows:

$$\mathbf{x}_{t+1} = f_t(\mathbf{Q}\mathbf{x}_t),\ 0 \le t \le T-1, \tag{8}$$

where $\mathbf{Q} \sim$ Haar(N) and $\mathbf{x}_0 \in \mathbb{C}^N$ are independent. HD was originally proposed as an efficient numerical method for simulating iterations like (8) of large dimension. In our paper, we employ it as a tool for performance analysis, which is a novel application of this technique. Specifically, using HD, one can show that the sequence $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ generated by (8) is statistically equivalent to another sequence that is fully determined by the initial vector $\mathbf{x}_0$ and a sequence of independent standard Gaussian random vectors. Compared to the original sequence that exhibits complicated correlation through the Haar matrix $\mathbf{Q}$, the new sequence is more amenable to analysis, particularly in the high-dimensional case, which enables us to study the statistical properties of the original sequence with greater ease.

To get some insight on how HD facilitates analysis, we consider the following simple example that contains only two iterations:

$$\begin{cases} \mathbf{x}_1 = f_0(\mathbf{Q}\mathbf{x}_0), \\ \mathbf{x}_2 = f_1(\mathbf{Q}\mathbf{x}_1), \end{cases} \tag{9}$$

where $\mathbf{Q} \sim$ Haar(N) and $\mathbf{x}_0 \in \mathbb{C}^N$ are independent. Using HD, one can show that $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$ is statistically equivalent to $(\mathbf{x}_0, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2)$ given below:

$$\begin{cases} \tilde{\mathbf{x}}_1 = f_0(a_0^1 \mathrm{g}_1), \\ \tilde{\mathbf{x}}_2 = f_1(a_1^1 \mathrm{g}_1 + a_1^2 \mathrm{g}_2), \end{cases} \tag{10}$$

where $\mathrm{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathrm{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ are independent and both are independent of $\mathbf{x}_0$, and the random variables $\{a_0^1, a_1^1, a_1^2\}$ are defined by

$$a_0^1 = \frac{\|\mathbf{x}_0\|}{\|\mathrm{g}_1\|},$$

$$a_1^1 = \frac{\mathbf{x}_0^{\mathsf{H}}\tilde{\mathbf{x}}_1}{\|\mathrm{g}_1\|} - \frac{\mathrm{g}_1^{\mathsf{H}}\mathrm{g}_2}{\|\mathrm{g}_1\|\|\mathbf{x}_0\|} \frac{\sqrt{\|\tilde{\mathbf{x}}_1\|^2\|\mathbf{x}_0\|^2 - \left|\mathbf{x}_0^{\mathsf{H}}\tilde{\mathbf{x}}_1\right|^2}}{\sqrt{\|\mathrm{g}_1\|^2\|\mathrm{g}_2\|^2 - |\mathrm{g}_1^{\mathsf{H}}\mathrm{g}_2|^2}},$$

$$a_1^2 = \frac{\|\mathrm{g}_1\|\sqrt{\|\tilde{\mathbf{x}}_1\|^2\|\mathbf{x}_0\|^2 - \left|\mathbf{x}_0^{\mathsf{H}}\tilde{\mathbf{x}}_1\right|^2}}{\|\mathbf{x}_0\|\sqrt{\|\mathrm{g}_1\|^2\|\mathrm{g}_2\|^2 - |\mathrm{g}_1^{\mathsf{H}}\mathrm{g}_2|^2}}.$$

The statistical equivalence between (9) and (10) can be proved using a technique similar to that in [39, Section 3.3] and thus we omit the details here. Clearly, $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ are fully specified by the initial vector $\mathbf{x}_0$ and the two Gaussian vectors g_1 and g_2. The scaling factors $\{a_0^1, a_1^1, a_1^2\}$, though correlated with $\{\mathbf{x}_0, \mathrm{g}_1, \mathrm{g}_2\}$ in a complicated way, converge in many cases to deterministic values as the matrix dimension N tends to infinity. For instance, when $f_0(\cdot)$ is separable and satisfies some mild regularity conditions and the entries of $\mathbf{x}_0$ are i.i.d., the convergence of $\{a_0^1, a_1^1, a_1^2\}$ can be easily proved via the law of large numbers.

The above example illustrates the strength of the HD technique: it transforms the original sequence, which is specified by the $N \times N$ Haar matrix $\mathbf{Q}$, into another sequence that is determined by only a few Gaussian vectors (e.g., two Gaussian vectors of dimension N for the above example) with an explicit form. The new sequence is usually more convenient for analysis. The interested reader is referred to [39] for a detailed description of the HD technique. We will provide a comprehensive description of the HD technique for handling our specific problem in Appendix A and Appendix B.

IV. Statistically Equivalent Model and Asymptotic Analysis

In this section, we first use the HD technique to derive a statistically equivalent model for (6), which is close to a "signal plus independent Gaussian noise" form. This step is non-asymptotic and the equivalence holds for any finite dimension when $N, K \ge 3$. The "signal plus Gaussian noise" insight is made precise by further considering the large system limit where N and K tend to infinity at a fixed ratio. We will derive sharp asymptotic expressions for the SINR and SEP performances of the linear-quantized precoding scheme.

A. Statistically Equivalent Model

Recall that our system model is

$$\mathbf{y} = \eta\mathbf{H}q(\mathbf{P}\mathbf{s}) + \mathbf{n} = \eta\mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}q(\mathbf{V}f(\mathbf{D})^{\mathsf{T}}\mathbf{U}^{\mathsf{H}}\mathbf{s}) + \mathbf{n},$$

where $\mathbf{U} \sim$ Haar$(K), \mathbf{V} \sim$ Haar(N), and $\{\mathbf{U}, \mathbf{V}, \mathbf{D}, \mathbf{s}, \mathbf{n}\}$ are mutually independent. The received signal $\mathbf{y}$ can be seen as being obtained by performing the following iterations:

$$\begin{cases} \mathbf{s}_1 = f(\mathbf{D})^{\mathsf{T}}\mathbf{U}^{\mathsf{H}}\mathbf{s}, \\ \mathbf{s}_2 = q(\mathbf{V}\mathbf{s}_1), \\ \mathbf{s}_3 = \mathbf{D}\mathbf{V}^{\mathsf{H}}\mathbf{s}_2, \\ \mathbf{y} = \eta\mathbf{U}\mathbf{s}_3 + \mathbf{n}, \end{cases} \tag{11}$$

The above iterative process has a form similar to (8). At each iteration, it involves one multiplication of a Haar random matrix and a random vector, while the other operations can be modeled as $f_t(\cdot)$ in (8), since $\{\mathbf{D}, \mathbf{n}\}$ are independent of $\{\mathbf{U}, \mathbf{V}, \mathbf{s}\}$. Specifically, $f_0(\mathbf{x}) = f(\mathbf{D})^{\mathsf{T}}\mathbf{x}, f_1(\mathbf{x}) = q(\mathbf{x}), f_2(\mathbf{x}) = \mathbf{D}\mathbf{x}, f_3(\mathbf{x}) = \eta\mathbf{x} + \mathbf{n}$, where $\mathbf{x}$ is a vector of an appropriate dimension. The only minor difference with (8) is that two different Haar matrices and their conjugate transposes are included in the above iterations. However, this difference is not important and the HD technique can still be applied. With the help of the HD technique, we can obtain the following statistically equivalent model, which is more convenient for analysis.

Theorem 1 (Statistically Equivalent Model): When $N \ge 3, K \ge 3$, the distribution of $(\mathbf{y}, \mathbf{s})$ in the original model (6) is the same as that of $(\hat{\mathbf{y}}, \mathbf{s})$ specified by the following model:

$$\hat{\mathbf{y}} = \eta T_{\mathbf{s}}\,\mathbf{s} + \eta T_{\mathrm{g}}\,\mathrm{g}_2 + \mathbf{n}, \tag{12}$$

where

$$T_{\mathbf{s}} = \frac{\mathsf{g}_1^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathbf{s}}_1)\mathsf{z}_2[2:N]\}}{\|\mathsf{g}_1\|\|\mathbf{s}\|} - T_{\mathsf{g}}\frac{(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[1]}{\|\mathbf{s}\|},$$
$$T_{\mathsf{g}} = \frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathbf{s}}_1)\mathsf{z}_2[2:N]\}\|}{\|(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[2:K]\|},$$
$$C_1 = \frac{\mathsf{z}_1^{\mathsf{H}} q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathsf{z}_1\|}\mathsf{z}_1\right)}{\|\hat{\mathbf{s}}_1\|\|\mathsf{z}_1\|},\ C_2 = \frac{\left\|\mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathsf{z}_1\|}\mathsf{z}_1\right)\right\|}{\|\mathsf{z}_2[2:N]\|},$$
$$\hat{\mathbf{s}}_1 = \frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|} f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1. \tag{13}$$

In the above expressions, $\mathsf{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathsf{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathsf{z}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, $\mathsf{z}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ are mutually independent standard Gaussian random vectors, which are further independent of the signal vector $\mathbf{s}$, the singular value matrix $\mathbf{D}$, and the noise vector $\mathbf{n}$; $f(\cdot)$ is a processing function involved in the precoder (5); $\mathbf{R}(\cdot)$ denotes the Householder transform of the input vector and $\mathbf{B}(\cdot)$ represents the submatrix of $\mathbf{R}(\cdot)$ with the first column removed (see (39) and (40) in Appendix A).

Proof: See Appendix B. □

Before proceeding, let us take a look at the statistically equivalent model in (12): the first term is the (scaled) signal vector, the second term is an equivalent noise that captures both the multi-user interference and the distortion caused by quantization, and the last term is the channel noise. Note, however, it is still difficult to exactly analyze the performance of the system (e.g., SEP performance) based on the statistically equivalent model in (12). This is because $T_{\mathbf{s}}$ and T_{g} therein are correlated with $\mathbf{s}$ and g_2 in a complicated way. Fortunately, as $N, K \to \infty$ and $N/K \to \gamma \in (1, \infty)$, both $T_{\mathbf{s}}$ and T_{g} converge to deterministic quantities, enabling us to derive sharp asymptotic formulas for both the SINR and SEP performance.

B. Asymptotic Analysis

In this subsection, we consider the large system limit where both N and K tend to infinity while keeping a finite ratio $\frac{N}{K} \to \gamma \in (1, \infty)$. This is a common assumption in the performance analysis of massive MIMO systems, and such asymptotic analyses can usually provide tight approximations for realistic systems with finite N, K (see, e.g., [44], [45], [46], [49]). In what follows, all vectors and matrices should be understood as sequences of vectors and matrices of growing dimensions. For simplicity, their dependence on N and K is not explicitly shown.

Our main asymptotic result is summarized in the following theorem. Its proof is given in Appendix C.

Theorem 2 (Asymptotic Model): Define the following asymptotic model:

$$\bar{\mathbf{y}} := \eta\overline{T}_{\mathbf{s}}\,\mathbf{s} + \eta\overline{T}_{\mathsf{g}}\,\mathsf{g}_2 + \mathbf{n}, \tag{14}$$

where

$$\overline{T}_{\mathbf{s}} = \overline{C}_1\,\mathbb{E}[d\,f(d)],$$
$$\overline{T}_{\mathsf{g}} = \sqrt{\sigma_s^2\,|\overline{C}_1|^2\mathrm{var}[d\,f(d)] + \overline{C}_2^2},$$
$$\overline{C}_1 = \frac{\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]}{\bar{\alpha}},$$
$$\overline{C}_2 = \sqrt{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] - |\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2},$$
$$\bar{\alpha} = \sqrt{\frac{\sigma_s^2\,\mathbb{E}[f^2(d)]}{\gamma}}, \tag{15}$$

$\mathsf{Z} \sim \mathcal{CN}(0,1)$, $d = \sqrt{\lambda}$, and λ follows the Marchenko-Pastur distribution, whose probability density function is given by

$$p_\lambda(x) = \frac{\sqrt{(x-a)_+(b-x)_+}}{2\pi c x} \tag{16}$$

with $a = (1-\sqrt{c})^2, b = (1+\sqrt{c})^2, c = \frac{1}{\gamma}; (x)_+ = \max\{x, 0\}$. Then under Assumptions 1-3, the following holds as $N, K \to \infty$, and $\frac{N}{K} \to \gamma \in (1, \infty)$:

$$(\hat{y}_k, s_k) \xrightarrow{a.s.} (\bar{y}_k, s_k), \quad \forall\, k \in [K],$$

where $(\hat{y}_k, s_k)$ and $(\bar{y}_k, s_k)$ are given in (12) and (14), respectively.

Several remarks on Theorems 1 and 2 are in order.

Remark 1: It is worth noting that the conditioning technique developed in [50] and [51] may be used to derive the asymptotic model in Theorem 2, though the model analyzed in the current paper is more complicated than that in [51]; compare (6) and [51, Eq. (15)]. Note that both the HD technique and the conditioning method of [50] and [51] heavily rely on the rotational invariance of the Haar random matrix. For our problem, the HD technique is more direct and transparent.

Remark 2 (Connection With Bussgang Decomposition): The asymptotic model (14) gives a precise characterization of the Bussgang decomposition and results in (7):

$$\mathbf{y} = \eta\frac{\overline{C}_1}{K}\mathrm{tr}(\mathbf{HP})\mathbf{s} + \eta\,\overline{C}_1\left(\mathbf{HP} - \frac{1}{K}\mathrm{tr}(\mathbf{HP})\mathbf{I}\right)\mathbf{s} + \eta\mathbf{H}\mathbf{q}_\perp + \mathbf{n}. \tag{17}$$

Loosely speaking, we have the following correspondence between (14) and (17):

$$\eta\overline{T}_{\mathbf{s}}\,\mathbf{s} \ \leftrightarrow\ \eta\frac{\overline{C}_1}{K}\mathrm{tr}(\mathbf{HP})\mathbf{s}$$
$$\eta\overline{T}_{\mathsf{g}}\,\mathsf{g}_2 \ \leftrightarrow \eta\,\overline{C}_1\left(\mathbf{HP} - \frac{1}{K}\mathrm{tr}(\mathbf{HP})\mathbf{I}\right)\mathbf{s} + \eta\mathbf{H}\mathbf{q}_\perp$$

The above correspondence is in a distributional sense, i.e., the corresponding terms have the same (asymptotic) distribution, and is implied by our proof of Theorem 2.

Remark 3 (Assumption on Channel $\mathbf{H}$): Theorems 1 and 2 are stated under the assumption that the elements of $\mathbf{H}$ are i.i.d. Gaussian, but can be extended to the following more general models.

- (Unitarily invariant model) For the unitarily invariant model, the SVD of $\mathbf{H}$ satisfies $\mathbf{U} \sim \mathrm{Haar}(K), \mathbf{V} \sim \mathrm{Haar}(N)$, and $\{\mathbf{U}, \mathbf{D}, \mathbf{V}\}$ are independent [48], and hence Theorem 1 holds. In addition, Theorem 2 holds as long as the empirical spectral distribution (see Definition 3 in Appendix C) of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ further has a continuous limiting distribution with a bounded support, with λ in Theorem 2 following the limiting e.s.d. of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$.
- (Large scale fading) In the case where the users have different large scale fading variances, the channel can be

modeled as $\mathbf{H} = \mathbf{\Sigma}_0\tilde{\mathbf{H}}$, where $\tilde{\mathbf{H}}$ satisfies Assumption 1 and $\mathbf{\Sigma}_0 = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K)$ is a diagonal matrix with σ_k representing the standard deviation of the large scale fading of user k, $k = 1, 2, \ldots, K$. Let $\tilde{\mathbf{H}} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{H}}$ be the SVD of $\tilde{\mathbf{H}}$. Theorems 1 and 2 can be directly generalized to this more general channel model with precoding matrices of the following form:

$$\mathbf{P} = \mathbf{V} f(\mathbf{D})^{\mathsf{T}} \mathbf{U}^{\mathsf{H}} g(\mathbf{\Sigma}_0),$$

where $\mathbf{\Sigma}_0$ and g satisfy some mild regularity conditions. The above class of precoding matrices still includes MF precoding and ZF precoding as special cases.

Theorem 2 shows that in the asymptotic regime, the system model is in a simple "signal plus independent Gaussian noise" form as (14). In the following, we will characterize the individual performance of the K users with the help of Theorem 2. Two commonly used performance measures are the SINR:

$$\text{SINR}_k := \frac{|\rho_k|^2 \, \mathbb{E}[|s_k|^2]}{\mathbb{E}[|y_k|^2] - |\rho_k|^2 \, \mathbb{E}[|s_k|^2]}, \tag{18}$$

where $\rho_k = \mathbb{E}[s_k^\dagger y_k]/\mathbb{E}[|s_k|^2]$, and the SEP:

$$\text{SEP}_k(\beta) := \mathbb{P}\left(\text{dec}(\beta y_k) \neq s_k\right), \tag{19}$$

where $\text{dec}(\cdot)$ is the decision function that maps its argument to the nearest constellation point in $\mathcal{S}_M$.

We note that the SINR and SEP performance of the K users in the asymptotic model (14) are the same and can be characterized by the following asymptotic scalar model:

$$\bar{y} := \eta \overline{T}_{\mathbf{s}} \, s + \eta \overline{T}_{\mathbf{g}} \, \mathsf{g} + n, \tag{20}$$

where $s \sim \text{Unif}(\mathcal{S}_M)$, $\mathsf{g} \sim \mathcal{CN}(0,1)$, $n \sim \mathcal{CN}(0, \sigma^2)$ are independent. Following the definitions in (18) and (19), the SINR and SEP of the above scalar asymptotic model are given by

$$\overline{\text{SINR}} = \frac{\sigma_s^2 \, \eta^2 \, \overline{T}_{\mathbf{s}}^2}{\eta^2 \, \overline{T}_{\mathbf{g}}^2 + \sigma^2} = \frac{\mathbb{E}^2[d\, f(d)]}{\text{var}[d\, f(d)] + \phi(\bar{\alpha}, \eta) \, \frac{\mathbb{E}[f^2(d)]}{\gamma}}, \tag{21}$$

where $\bar{\alpha} = \sqrt{\sigma_s^2 \, \mathbb{E}[f^2(d)]/\gamma}$ and

$$\phi(\bar{\alpha}, \eta) := \frac{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] - |\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2 + \sigma^2/\eta^2}{|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2}, \tag{22}$$

and

$$\overline{\text{SEP}}(\beta) = \mathbb{P}\left(\text{dec}(\beta \bar{y}) \neq s\right).$$

The second step of (21) is obtained based on the definitions of $\overline{T}_{\mathbf{s}}$ and $\overline{T}_{\mathbf{g}}$ in (15).

Theorem 3 below shows that both the SINR and the SEP performance of the original model converge to those of the scalar asymptotic model in (20). Its proof can be found in Appendix D.

Theorem 3: Denote $\widehat{\text{SINR}}_k$ and $\widehat{\text{SEP}}_k(\beta)$ as the SINR and SEP of user k of the model in (12), respectively. Under the asymptotic setting in Theorem 2, the following hold for any $k \in [K]$ and $\beta \in \mathbb{C}$:

(i) $\lim_{N,K\to\infty} \widehat{\text{SINR}}_k = \overline{\text{SINR}}$;
(ii) $\lim_{N,K\to\infty} \widehat{\text{SEP}}_k(\beta) = \overline{\text{SEP}}(\beta)$.

With the above theorem, we can give predictions of the SINR and SEP performance of the original model based on the asymptotic scalar model in (20). Note that $\overline{\text{SEP}}(\beta)$ is defined for a scalar additive white Gaussian noise (AWGN) channel and therefore can be easily computed for specific constellation types. Clearly, a natural scaling factor should be

$$\beta = \frac{\overline{T}_{\mathbf{s}}^\dagger}{\eta |\overline{T}_{\mathbf{s}}|^2}. \tag{23}$$

We denote the corresponding asymptotic SEP as $\overline{\text{SEP}}$. When M-PSK modulation is adopted, $\overline{\text{SEP}}$ can be tightly approximated as [52, Section 4.3-2]:

$$\overline{\text{SEP}} \approx 2Q\left(\sqrt{2}\sin\frac{\pi}{M}\sqrt{\overline{\text{SINR}}}\right), \tag{24}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}} dt$. For M-QAM modulation, the SEP has an explicit expression [52, Section 4.3-3]:

$$\overline{\text{SEP}} = 4\left(1 - \frac{1}{\sqrt{M}}\right) E_0 - 4\left(1 - \frac{1}{\sqrt{M}}\right)^2 E_0^2 \tag{25}$$

where

$$E_0 = Q\left(\sqrt{\frac{3}{M-1}\,\overline{\text{SINR}}}\right). \tag{26}$$

In the rest of this paper, we assume that β in (23) is used unless otherwise stated. It is worth noting that the asymptotic SEP formulas in (24) and (25), though derived under the asymptotic assumption that the system dimension grows to infinity, are also accurate for practical systems with moderate dimensions; see the simulation results in Section VI.

C. Example: QCE Precoding

So far, our results hold for general $q(\cdot)$ satisfying Assumption 3. In this subsection, we specify our results to the case of CE quantizer $q_{\text{CE}}(\cdot)$. We start with a few properties of $q_{\text{CE}}(\cdot)$.

Lemma 2 (Properties of $q_{\text{CE}}(\cdot)$):

(i) $|q_{\text{CE}}(x)| = 1, \ \forall\, x \in \mathbb{C}$;
(ii) $q_{\text{CE}}(\alpha x) = q_{\text{CE}}(x), \ \forall\, x \in \mathbb{C}, \alpha > 0$;
(iii) $\mathbb{E}[\mathsf{Z}^\dagger q_{\text{CE}}(\mathsf{Z})] = \frac{L \sin\frac{\pi}{L}}{2\sqrt{\pi}}$, where $\mathsf{Z} \sim \mathcal{CN}(0,1)$ and L is the number of quantization levels [36].

Under the above properties of $q_{\text{CE}}(\cdot)$, the asymptotic SINR in (21) simplifies to

$$\overline{\text{SINR}}_{\text{QCE}} = \frac{\mathbb{E}^2[d\, f(d)]}{\text{var}[d\, f(d)] + \frac{C_{L,\sigma,\eta}}{\gamma}\,\mathbb{E}[f^2(d)]}, \tag{27}$$

where

$$C_{L,\sigma,\eta} = \frac{1 - \frac{L^2 \sin^2\frac{\pi}{L}}{4\pi} + \sigma^2/\eta^2}{\frac{L^2\sin^2\frac{\pi}{L}}{4\pi}}. \tag{28}$$

The asymptotic SEP is a simple function of the SINR, as discussed above.

By further specifying the nonlinear function $f(\cdot)$ as $f(x) = x$ and $f(x) = x^{-1}$ respectively (see discussions in Section II-A) and using [48, Eq. (2.104)], we can obtain the

following asymptotic SINR formulas for the quantized MF and ZF precoders:

$$\overline{\mathrm{SINR}}_{\mathrm{QCE}}^{\mathrm{MF}} = \frac{\gamma}{C_{L,\sigma,\eta}+1},$$
$$\overline{\mathrm{SINR}}_{\mathrm{QCE}}^{\mathrm{ZF}} = \frac{\gamma-1}{C_{L,\sigma,\eta}},$$

where $C_{L,\sigma,\eta}$ is defined in (28). The above SINR formula for quantized ZF precoding has been obtained for the one-bit case (i.e., $L = 4$) using the Bussgang decomposition technique in [27].

V. Optimal Linear-Quantized Precoding

For the precoding scheme in (5), the function $f(\cdot)$ can be designed to optimize the system SEP performance. In this section, we derive the optimal $f(\cdot)$ based on the asymptotic characterization developed in Theorem 2.

A. Optimal Linear-Quantized Precoding

Our goal is to find the optimal $f(\cdot)$ in terms of the asymptotic SEP performance. First, maximizing $\overline{\mathrm{SINR}}$ is equivalent to minimizing $\overline{\mathrm{SEP}}$, as shown in Appendix E. In what follows, we shall focus on the following SINR maximization problem (see (21)):

$$\begin{aligned}\zeta^* := \sup_{f,\eta>0,\bar{\alpha}>0} \ & \frac{\mathbb{E}^2[d\,f(d)]}{\mathrm{var}[d\,f(d)]+\phi(\bar{\alpha},\eta)\,\frac{\mathbb{E}[f^2(d)]}{\gamma}} \\ \text{s.t.} \ & \eta^2\,\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] \le P_T, \\ & \mathbb{E}[f^2(d)] = \frac{\gamma}{\sigma_s^2}\,\bar{\alpha}^2, \end{aligned} \tag{29}$$

where $\mathsf{Z} \sim \mathcal{CN}(0,1)$, $P_T > 0$ is the maximum average transmit power in (2), d is defined in Theorem 2, and $\phi(\bar{\alpha},\eta)$ is defined in (22). If $\bar{\alpha}$ in the second constraint is eliminated and substituted into the first constraint, then the obtained constraint represents the asymptotic counterpart of the actual average power constraint in (2). Therefore, the function $f(\cdot)$ and the transmit power η are jointly optimized in problem (29).

The SINR maximization problem (29) may seem challenging to solve as it involves optimization over a function $f(\cdot)$. Moreover, the variables $(\bar{\alpha},\eta,f)$ are coupled by the constraints, which further complicates the problem. Fortunately, the optimal solution of (29) has a simple structure, as shown by the following theorem.

Theorem 4: Suppose that the following infimum is attained by some $\bar{\alpha}^* > 0$:

$$\phi^* := \left(1+\frac{\sigma^2}{P_T}\right)\inf_{\bar{\alpha}>0}\frac{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2]}{|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2}-1. \tag{30a}$$

Then, $(\bar{\alpha}^*,\eta^*,f^*)$ is an optimal solution to the problem in (29) where

$$\eta^* := \sqrt{\frac{P_T}{\mathbb{E}[|q(\bar{\alpha}^*\mathsf{Z})|^2]}}, \tag{30b}$$
$$f^*(d) := \pm\frac{d}{\tau^*(d^2+\frac{\phi^*}{\gamma})}, \tag{30c}$$
$$\tau^* := \sqrt{\frac{\sigma_s^2}{\gamma(\bar{\alpha}^*)^2}\,\mathbb{E}\left[\left(\frac{d}{d^2+\frac{\phi^*}{\gamma}}\right)^2\right]}. \tag{30d}$$

Furthermore, the maximum SINR in (29) is equal to

$$\zeta^* = \frac{1}{1-\mathbb{E}\left[\frac{d^2}{d^2+\frac{\phi^*}{\gamma}}\right]}-1. \tag{30e}$$

Proof: See Appendix F. □

Remark 4: Some comments on Theorem 4 are in order.

1) In Theorem 4, we do not impose the positivity constraint on f, and an optimal f satisfying $f > 0$ always exists; see (30c). This supports our claim below Assumption 2 that the positivity assumption on f does not impose any restriction in terms of the best achievable performance.
2) We assume that the infimum in (30a) is attainable. Our numerical results suggest that this holds for commonly used quantization functions $q(\cdot)$. In cases where the infimum in (30a) is not attainable, we may modify the theorem using a simple truncation argument. Specifically, we add an additional constraint $\frac{1}{M} \le \bar{\alpha} \le M$ to (30a) and let $\bar{\alpha}_M^*$ be any optimal solution to the constrained problem. We define ϕ_M^*, η_M^* and f_M^* as in (30) with $\bar{\alpha}^*$ replaced by $\bar{\alpha}_M^*$. Then, it can be shown that the SINR achieved by $(\bar{\alpha}_M^*,\eta_M^*,f_M^*)$ tends to ζ^* as $M \to \infty$.

We now take a closer look at (30a). Loosely speaking, it may be interpreted as finding the optimal input power for the quantization function $q(\cdot)$. More precisely, using the Bausssgang decomposition technique, we can decompose $q(\bar{\alpha}\mathsf{Z})$ as

$$q(\bar{\alpha}\mathsf{Z}) \overset{d}{=} \mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]\,\mathsf{Z}+d,$$

where $\mathsf{Z} \sim \mathcal{CN}(0,1)$ and d models the quantization error which is uncorrelated with Z. Then, problem (30a) can be viewed as maximizing the signal to quantization error plus noise ratio (SQNR):

$$\begin{aligned}\mathrm{SQNR} &:= \frac{|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2\,\mathbb{E}[|\mathsf{Z}|^2]}{\mathbb{E}[|d|^2]} \\ &= \frac{|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2}{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2]-|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2} \\ &= \frac{1}{\frac{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2]}{|\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2}-1}.\end{aligned}$$

For a general $q(\cdot)$, problem (30a) can be solved by a one-dimensional search. For the special case of the CE quantizer, the objective function of (30a) is actually a constant (see Lemma 2) and the problem is trivial; see details in the next subsection.

Remark 5 (Connection With the WFQ Precoder [7]): Theorem 4 shows that for precoding matrices of the form $\mathbf{P} = \mathbf{V}f(\mathbf{D})^{\mathsf{T}}\mathbf{U}^{\mathsf{H}}$, the asymptotically optimal one is precisely the RZF precoder:

$$\mathbf{P}^* = \frac{1}{\tau^*}\mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}+\frac{\phi^*}{\gamma}\,\mathbf{I}_K\right)^{-1}, \tag{31}$$

where τ^* and ϕ^* can be obtained by solving the one-dimensional optimization problem in (30a). Interestingly,

(31) is closely related to the WFQ precoder [7] designed for traditional quantized precoding:

$$\mathbf{P}_{\text{WFQ}} = \frac{1}{g_0}\left(\mathbf{H}^{\mathsf{H}}\mathbf{H} - \rho_q \text{nondiag}(\mathbf{H}^{\mathsf{H}}\mathbf{H}) + \frac{\sigma^2}{\gamma P_T}\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathsf{H}},$$

where g_0 is a scaling factor that ensures a certain power constraint to be satisfied. As $N, K \to \infty$, the diagonal entries of $\mathbf{H}^{\mathsf{H}}\mathbf{H}$ converge to

$$(\mathbf{H}^{\mathsf{H}}\mathbf{H})[i,i] = \sum_{j=1}^{K} |\mathbf{H}[j,i]|^2 \xrightarrow{a.s.} \frac{1}{\gamma}, \quad i = 1, 2, \ldots, N. \tag{32}$$

Therefore, $\mathbf{P}_{\text{WFQ}}$ can be approximated as

$$\begin{aligned}
\mathbf{P}_{\text{WFQ}} &= \frac{1}{g_0}\left(\mathbf{H}^{\mathsf{H}}\mathbf{H} - \rho_q \text{nondiag}(\mathbf{H}^{\mathsf{H}}\mathbf{H}) + \frac{\sigma^2}{\gamma P_T}\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathsf{H}} \\
&= \frac{1}{g_0}\left((1-\rho_q)\,\mathbf{H}^{\mathsf{H}}\mathbf{H} + \rho_q \text{diag}(\mathbf{H}^{\mathsf{H}}\mathbf{H}) + \frac{\sigma^2}{\gamma P_T}\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathsf{H}} \\
&\overset{(a)}{\approx} \frac{1}{g_0}\left((1-\rho_q)\,\mathbf{H}^{\mathsf{H}}\mathbf{H} + \frac{\rho_q + \frac{\sigma^2}{P_T}}{\gamma}\mathbf{I}_N\right)^{-1}\mathbf{H}^{\mathsf{H}} \\
&\overset{(b)}{=} \frac{1}{g_0}\mathbf{H}^{\mathsf{H}}\left((1-\rho_q)\,\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\rho_q + \frac{\sigma^2}{P_T}}{\gamma}\mathbf{I}_K\right)^{-1} \\
&\overset{(c)}{=} \frac{1}{g_0(1-\rho_q)}\mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\rho_q + \frac{\sigma^2}{P_T}}{\gamma(1-\rho_q)}\mathbf{I}_K\right)^{-1},
\end{aligned} \tag{33}$$

where (a) is due to (32), (b) uses the fact that $\left(\mathbf{X}^{\mathsf{H}}\mathbf{X} + \rho\mathbf{I}_N\right)^{-1}\mathbf{X}^{\mathsf{H}} = \mathbf{X}^{\mathsf{H}}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} + \rho\mathbf{I}_K\right)^{-1}$ for any $\mathbf{X} \in \mathbb{C}^{K\times N}$ and $\rho > 0$, and (c) is obtained by extracting the coefficient $1-\rho_q$ from the inverse matrix. Comparing the last line of (33) with (31), we see that both $\mathbf{P}_{\text{WFQ}}$ and $\mathbf{P}^*$ are RZF precoders, and are identical if the regularization parameters and the scaling factors are the same:

$$\phi^* = \frac{\rho_q + \frac{\sigma^2}{P_T}}{1-\rho_q}, \quad \tau^* = g_0(1-\rho_q).$$

Note that $\mathbf{P}_{\text{WFQ}}$ and $\mathbf{P}^*$ are derived based on different criteria and motivations, and not directly comparable. For the WFQ precoder, the independent quantizer $q_{\text{I}}(\cdot)$ is optimized together with the linear precoder (see also [53], [54] for optimization of the independent quantizer $q_{\text{I}}(\cdot)$), and the quantization intervals for $q_{\text{I}}(\cdot)$ have to be chosen carefully to satisfy several conditions (see [7, Eqs. (7)-(9)]). On the other hand, $q(\cdot)$ is a generic fixed function in this work, and the input power of $q(\cdot)$ (dictated by $\bar{\alpha}$) is properly optimized.

B. Example: QCE Precoding

In this subsection, we consider the special case of QCE precoding. As a direct corollary of Theorem 4, we have the following result.

Corollary 1: When $q(\cdot)$ is specified as $q_{\text{CE}}(\cdot)$ in problem (29), the optimal η is $\eta^* = \sqrt{P_T}$, and the optimal f has a closed-form expression:

$$f^*(d) = \frac{d}{d^2 + \frac{C_{L,\sigma,\eta^*}}{\gamma}}, \tag{34}$$

where C_{L,σ,η^*} is given in (28). In this case, the asymptotically optimal precoding matrix is given by

$$\mathbf{P}^*_{\text{QCE}} = \mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{C_{L,\sigma,\eta^*}}{\gamma}\mathbf{I}_K\right)^{-1}, \tag{35}$$

and its asymptotic SINR is

$$\overline{\text{SINR}}^*_{\text{QCE}} = \frac{\sqrt{u^2 + 4C_{L,\sigma,\eta^*}} + u}{2C_{L,\sigma,\eta^*}} - 1,$$

where $u = C_{L,\sigma,\eta^*} + \gamma - 1$.

Proof: When $q(\cdot) = q_{\text{CE}}(\cdot)$, it follows from Lemma 2 and (30b) that $\eta^* = \sqrt{P_T}$ and the objective function of problem (30a) is a constant C_{L,σ,η^*}, which is independent of $\bar{\alpha}$. Hence, the optimal solution set of problem (30a) is $\{\bar{\alpha}^* \mid \bar{\alpha}^* > 0\}$ and τ^* can be any positive number. Without loss of generality, here we set $\tau^* = 1$. This further gives the optimal $f^*(\cdot)$ and the optimal precoding matrix $\mathbf{P}^*_{\text{QCE}}$ in (34) and (35), respectively. Finally, $\overline{\text{SINR}}^*_{\text{QCE}}$ can be obtained by plugging $\phi^* = C_{L,\sigma,\eta^*}$ into (30e) and using [48, Eq. (2.42)]. □

We note that when γ is large, $\sqrt{u^2 + 4C_{L,\sigma,\eta^*}} \approx u$, which implies that

$$\overline{\text{SINR}}^*_{\text{QCE}} \approx \frac{u}{C_{L,\sigma,\eta^*}} - 1 = \frac{\gamma - 1}{C_{L,\sigma,\eta^*}} = \overline{\text{SINR}}^{\text{ZF}}_{\text{QCE}}, \tag{36}$$

i.e., the performance of quantized ZF precoding is nearly optimal when the antenna-user ratio is large.

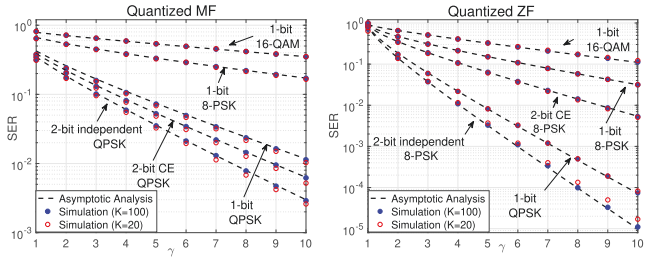
VI. Simulation Results

In this section, we present simulation results to demonstrate the theoretical results obtained in previous sections. We consider both traditional quantized precoding (where the independent quantizer is used) and QCE precoding (where the CE quantizer is used), and assume that the precoding factor (namely, the linear scaling applied at the receiver side before detection) in (23) is used. We also assume that the precoder adopted in this section has the optimal input power for the corresponding quantization function, i.e., $\bar{\alpha} = \bar{\alpha}^*$, where $\bar{\alpha}^*$ is an optimal solution to problem (30a). The transmit power is set as $P_T = 1$. All results are averaged over 10^5 channel realizations.

A. Numerical Validation of SEP Formulas

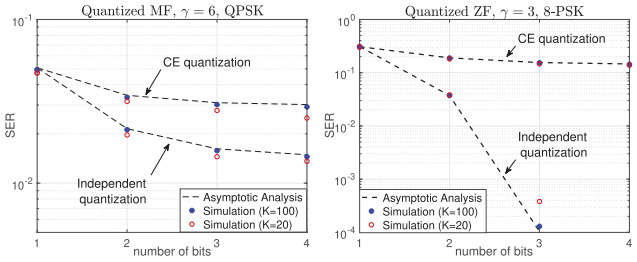
We first verify the accuracy of the SEP formulas given in (24)–(26).

In Fig. 3, we plot the symbol error rate (SER) as a function of the ratio of the number of antennas to users γ for the quantized MF and ZF precoding schemes. Both the cases of a large system with $K = 100$ and a more practical system with $K = 20$ are investigated, where the number of transmit antennas is set as $N = \gamma K$. We consider three types of signal constellations: QPSK, 8-PSK, and 16-QAM, and two types of quantization: CE quantization and independent quantization. The channel noise is set as $\sigma = 0$. As shown in the figure, there is a slight mismatch between simulations and asymptotic predictions when $K = 20$, and the differences become indistinguishable when $K = 100$.



(a) Quantized MF. (b) Quantized ZF.

Fig. 3. SER performance versus γ for quantized MF and ZF precoding with $\sigma = 0$, where PSK and QAM represent the type of constellation and "independent" and "CE" represent the type of quantization. The quantization interval for the independent quantizer in (3) is $\Delta = 2$.



(a) Quantized MF. (b) Quantized ZF.

Fig. 4. SER performance versus DAC resolution for quantized MF with $\gamma = 6$ and QPSK modulation and quantized ZF with $\gamma = 3$ and 8-PSK modulation; $\sigma = 0$.

TABLE I

ϕ^* IN (30A) FOR SCALAR AND CE QUANTIZERS WITH $\sigma = 0$

	1 bit	2 bits	3 bits	4 bits
scalar	0.57	0.14	0.04	0.01
CE	0.57	0.34	0.29	0.28

We can also draw some interesting observations from Fig. 3. First, the logarithm of SEP/SER decreases linearly with γ, i.e., the antenna-user ratio, which reveals the gain of increasing the number of transmit antennas at the BS, and the slope of the decrease is determined by the precoder, the quantizer, and the constellation that are employed. Second, the independent quantizer and the CE quantizer have different behaviors under different precoding schemes. For example, the superiority (in terms of SEP/SER) of the 2-bit independent quantizer over the 2-bit CE quantizer under ZF precoding is much more prominent than that under MF precoding.

In fact, the above observations are clear from our asymptotic analysis. To be specific, applying the approximation $Q(x) \approx \frac{1}{2}e^{-\frac{1}{2}x^2}$ [52] to the SEP predictions for M-PSK and M-QAM constellations in (24) and (25), we get

$$\ln \overline{\text{SEP}} \approx \begin{cases} -2\sin^2 \frac{\pi}{M} \overline{\text{SINR}}, & \text{for PSK;} \\ -\frac{3}{2(M-1)} \overline{\text{SINR}} + \log(2 - \frac{2}{\sqrt{M}}), & \text{for QAM,} \end{cases} \tag{37}$$

i.e., the logarithm of $\overline{\text{SEP}}$ decreases linearly with $\overline{\text{SINR}}$. According to (21), $\overline{\text{SINR}}$ for MF and ZF precoding can be expressed as

$$\overline{\text{SINR}} = \begin{cases} \dfrac{\gamma}{\phi^* + 1}, & \text{for MF;} \\ \dfrac{\gamma - 1}{\phi^*}, & \text{for ZF,} \end{cases} \tag{38}$$

where ϕ^* is given in (30a) and is determined by the quantization type. The values of ϕ^* for scalar and CE quantizers with $\sigma = 0$ are given in Table I. Eqs (37) and (38) demonstrate the linear decrease in the logarithm of the SEP with γ, and quantitatively characterize the effects of precoding, quantization, and constellation on the slope of the decrease. In particular, we can see that ϕ^* has a stronger impact on ZF than on MF, since the slope of decrease is proportional to $\frac{1}{\phi^*}$ and $\frac{1}{\phi^*+1}$ for ZF and MF, respectively, and ϕ^* is on the order of $10^{-2} - 10^{-1}$ as shown in Table I. This explains why the 2-bit independent quantizer has a greater performance gain compared with the 2-bit CE quantizer under ZF precoding than under MF precoding.

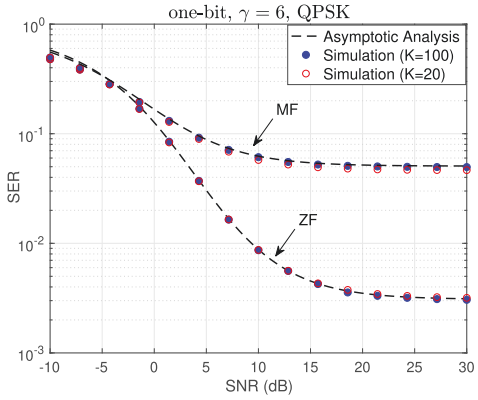


Fig. 5. SER performance versus SNR for one-bit quantized MF and ZF precoding with $\gamma = 6$ and QPSK modulation.

In Fig. 4, we plot the SER performance of quantized MF and ZF for both traditional quantized precoding and QCE precoding as a function of the DAC resolution. For MF, the performance gain of both quantizers is small as the DAC resolution increases, especially when the resolution is larger than 2 bits. The same happens for ZF precoding with CE quantization. However, there is a remarkable gain for ZF precoding with independent quantization as the DAC resolution increases. These observations can also be interpreted by our previous discussions.

Finally, Fig. 5 shows the SER of quantized MF and ZF as a function of the channel SNR (i.e., $1/\sigma^2$) for a one-bit system with $\gamma = 6$ and QPSK modulation. We see that ZF has a noticeable performance gain compared with MF in the high SNR region.

B. Optimality of Quantized RZF Precoding

In this subsection, we present some simulation results to demonstrate the optimality of the quantized RZF precoding matrix given in (31).

In Fig. 6, we consider quantized RZF precoding and depict the SER as a function of the regularization parameter for a one-bit system with $\gamma = 3$, $\sigma = 0$, and QPSK modulation. As shown in the figure, the errors between the asymptotic SER and the actual SER are within 0.002 and 0.005 for $K = 100$ and $K = 20$, respectively, which again validates the accuracy of our analytical results. In addition, it can

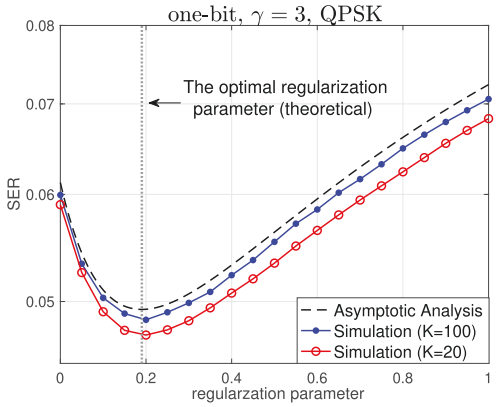
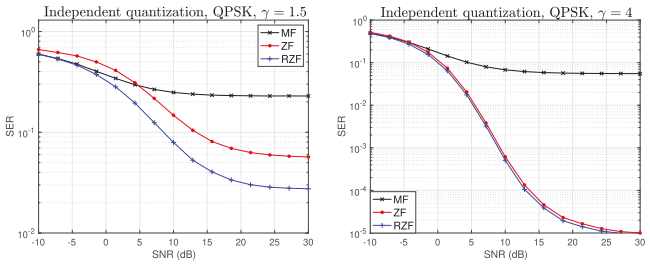
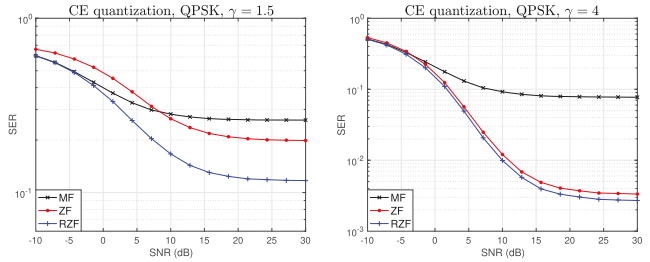


Fig. 6. SER performance versus regularization parameter for a one-bit system with $\gamma = 3$ and QPSK modulation.



(a) $\gamma = 1.5$. (b) $\gamma = 4$.

Fig. 7. Comparison of the SER performance between different linear-quantized precoders with $K = 20$, for 2-bit independent quantization with QPSK.



(a) $\gamma = 1.5$. (b) $\gamma = 4$.

Fig. 8. Comparison of the SER performance between different linear-quantized precoders with $K = 20$, for 2-bit CE quantization with QPSK.

be observed that the simulation curves and the (asymptotic) analysis curve show almost the same trends and all attain their minimum at approximately 0.2, which agrees well with the predicted value of 0.19 for the optimal regularization parameter. This demonstrates the optimality of the RZF precoding matrix $\mathbf{P}^*$ in (31).

In Figs. 7 and 8, we further consider realistic systems with $K = 20$. We plot the SER performance of quantized MF, ZF, and the proposed RZF precoding as a function of the SNR for both independent quantization and CE quantization. We investigate two different cases: $\gamma = 1.5$ and $\gamma = 4$, which corresponds to small and large antenna-user ratio, respectively. Compared with the quantized MF and ZF precoder, the proposed quantized RZF precoder enjoys a substantial performance gain for small γ. When γ is large, the proposed quantized RZF precoder performs similarly to the quantized ZF precoder (which is consistent with our discussion in (36)), and they both yield a much lower SER than the quantized MF precoder.

VII. Conclusion

In this paper, we studied the performance of linear-quantized precoding in massive MIMO downlink systems. Assuming an i.i.d. Gaussian channel matrix, we showed that the linear-quantized precoding scheme is statistically equivalent to a simple scalar model in the asymptotic sense when the number of antennas N and the number of users K tend to infinity with a fixed ratio. We further derived the optimal precoding strategy within a class of linear-quantized precoders, and found that it is precisely the RZF precoder where the optimal regularization parameter depends on the type and level of quantization and various system parameters. For future work, it would be interesting to extend the current analysis to encompass more general scenarios, such as general correlated channels and imperfect CSI. It is also interesting to give performance analysis of the more challenging nonlinear precoding schemes.

Appendix A
Preliminaries of the Householder Dice Technique [39]

In this appendix, we collect some useful results about the HD technique. We begin with the definition of the Householder transform.

Definition 2 (Householder Transform): For a given vector $\mathbf{v} = (v_1, v_2, \ldots, v_N)^{\mathsf{T}} \in \mathbb{C}^N \backslash \{\mathbf{0}\}$, denote

$$p_{\mathbf{v}} = -e^{j \arg(v_s)}, \quad \text{where } s = \min_i \{i \mid v_i \neq 0\}.$$

Let $\mathbf{H}(\mathbf{v})$ be the Householder transform of $\mathbf{v}$, i.e.,

$$\mathbf{H}(\mathbf{v}) = \mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^{\mathsf{H}}}{\|\mathbf{u}\|^2}$$

with $\mathbf{u} = \mathbf{v} - p_{\mathbf{v}}\|\mathbf{v}\|\mathbf{e}_1$, $\mathbf{e}_1 = (1, 0, \ldots, 0)^{\mathsf{T}}$. Further, define

$$\mathbf{R}(\mathbf{v}) = p_{\mathbf{v}}\,\mathbf{H}(\mathbf{v}). \tag{39}$$

With a slight abuse of notation, $\mathbf{R}(\mathbf{v})$ will also be called the Householder transform matrix associated with $\mathbf{v}$. We further define a generalized Householder transform as [39]

$$\mathbf{R}_k(\mathbf{v}) = \begin{pmatrix} \mathbf{I}_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\mathbf{v}[k:N]) \end{pmatrix}.$$

The following lemma collects some useful properties of $\mathbf{R}(\mathbf{v})$ that will be used in the subsequent analysis. The proofs are straightforward and thus omitted.

Lemma 3 (Facts and Properties of $\mathbf{R}(\mathbf{v})$): For a given vector $\mathbf{v} \in \mathbb{C}^N \backslash \{\mathbf{0}\}$, the Householder transform $\mathbf{R}(\mathbf{v})$ defined in (39) satisfies:

(i) $\mathbf{R}(\mathbf{v}) \in \mathcal{U}(N)$;
(ii) $\mathbf{R}(\mathbf{v})^{\mathsf{H}}\mathbf{v} = \|\mathbf{v}\|\mathbf{e}_1$;
(iii) $\mathbf{R}(\mathbf{v})\mathbf{e}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|}$, i.e., $\mathbf{R}(\mathbf{v})$ can be expressed as

$$\mathbf{R}(\mathbf{v}) = \left(\frac{\mathbf{v}}{\|\mathbf{v}\|} \quad \mathbf{B}(\mathbf{v})\right), \tag{40}$$

where $\mathbf{B}(\mathbf{v}) \in \mathbb{C}^{N \times (N-1)}$ is a basis matrix for $\{\mathbf{v}\}^{\perp}$.

The generalized Householder transform $\mathbf{R}_k(\mathbf{v})$ has similar properties as $\mathbf{R}(\mathbf{v})$, except that it leaves the first $k-1$ entries of $\mathbf{v}$ unchanged and applies a Householder transform on $\mathbf{v}[k:N]$.

The following lemma is a recursive characterization of the Haar random matrix introduced in [39, Lemma 1], which serves as the theoretical basis of the HD technique. The only difference is that we consider complex unitary matrices instead of real orthogonal matrices.

Lemma 4: Let $\mathrm{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, $\mathbf{Q}_{N-1} \sim \text{Haar}(N-1)$, and $\mathbf{v} \in \mathbb{C}^N \backslash \{\mathbf{0}\}$, all of which are independent, where $N \geq 2$. Then

$$\mathbf{Q}_N := \mathbf{R}_1(\mathrm{g}) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{N-1} \end{pmatrix} \mathbf{R}_1(\mathbf{v})^{\mathsf{H}} \sim \text{Haar}(N) \tag{41}$$

and

$$\widetilde{\mathbf{Q}}_N := \mathbf{R}_1(\mathbf{v}) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{N-1} \end{pmatrix} \mathbf{R}_1(\mathrm{g})^{\mathsf{H}} \sim \text{Haar}(N). \tag{42}$$

Moreover, $\mathbf{Q}_N$ and $\widetilde{\mathbf{Q}}_N$ are independent of $\mathbf{v}$.

Proof: We first note that the first column of $\mathbf{R}_1(\mathrm{g})$ is $\frac{\mathrm{g}}{\|\mathrm{g}\|}$, which is uniformly distributed on the unit sphere $\mathbb{S}^{N-1} \subseteq \mathbb{C}^N$. Then according to [47, page 21], we have

$$\mathbf{R}_1(\mathrm{g}) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{N-1} \end{pmatrix} \sim \text{Haar}(N).$$

Moreover, since a Haar matrix is both left and right translation invariant, we are free to multiply unitary matrices (either deterministic or independent of the Haar matrix) from left or right, hence (41) is correct. The above discussions imply that the conditional distribution of $\mathbf{Q}_N$ given $\mathbf{v}$ is the same as the distribution of $\mathbf{Q}_N$ (both are Haar distributed), i.e.,

$$\mu_{\mathbf{Q}_N|\mathbf{v}} = \mu_{\mathbf{Q}_N} = \mu, \quad \forall\, \mathbf{v} \in \mathbb{C}^N \backslash \{\mathbf{0}\}, \tag{43}$$

where μ denotes the Haar measure on $\mathcal{U}(N)$. Therefore, $\mathbf{Q}_N$ and $\mathbf{v}$ are independent. Finally, (42) is also true since $\mathbf{Q} \sim \text{Haar}(N)$ implies $\mathbf{Q}^{\mathsf{H}} \sim \text{Haar}(N)$, and the independence between $\widetilde{\mathbf{Q}}_N$ and $\mathbf{v}$ can be justified in a similar way as (43). □

APPENDIX B
PROOF OF THEOREM 1

In this section, we provide the detailed proof of Theorem 1. This section is long and is organized as follows:

- Section B-A contains the main proof of Theorem 1, relying on two auxiliary results: Lemma 5 and Lemma 6;
- Lemma 5 is critical to the proof of Theorem 1. For better understanding, we present some high-level ideas about the proof of Lemma 5 in Section B-B;
- Section B-C contains the complete proof of Lemma 5;
- Section B-D contains the proof of Lemma 6.

A. Proof of Theorem 1

The proof of Theorem 1 contains two major steps. The first step is to apply the HD technique [39] to our model in (11) to obtain a statistically equivalent model that is more amenable to analysis. The result of this step is summarized in the following lemma and the proof is given in Appendix B-C.

Lemma 5: The distribution of $(\mathbf{s}, \mathbf{y})$ given in (11) is the same as that of $(\mathbf{s}, \tilde{\mathbf{y}})$ specified by the following model:

$$\begin{cases} \tilde{\mathbf{s}}_1 = f(\mathbf{D})^{\mathsf{T}} \mathbf{R}_1(\mathrm{g}_1) \mathbf{R}_1(\mathbf{s})^{\mathsf{H}} \mathbf{s}, \\ \tilde{\mathbf{s}}_2 = q(\mathbf{R}_1(\mathrm{z}_1) \mathbf{R}_1(\tilde{\mathbf{s}}_1)^{\mathsf{H}} \tilde{\mathbf{s}}_1), \\ \tilde{\mathbf{s}}_3 = \mathbf{D} \mathbf{R}_1(\tilde{\mathbf{s}}_1) \mathbf{R}_2(\mathrm{z}_2) \mathbf{R}_1(\mathbf{v}_1)^{\mathsf{H}} \mathbf{v}_1, \\ \tilde{\mathbf{y}} = \eta \mathbf{R}_1(\mathbf{s}) \mathbf{R}_2(\mathrm{g}_2) \mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}} \mathbf{v}_2 + \mathbf{n}, \end{cases} \tag{44}$$

where $\mathbf{v}_1 = \mathbf{R}_1(\mathrm{z}_1)^{\mathsf{H}} \tilde{\mathbf{s}}_2$ and $\mathbf{v}_2 = \mathbf{R}_1(\mathrm{g}_1)^{\mathsf{H}} \tilde{\mathbf{s}}_3$; $\mathrm{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathrm{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathrm{z}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, $\mathrm{z}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$; $\{\mathrm{g}_1, \mathrm{g}_2, \mathrm{z}_1, \mathrm{z}_2, \mathbf{s}, \mathbf{n}, \mathbf{D}\}$ are mutually independent.

The second step is to simplify the above statistically equivalent model $(\mathbf{s}, \tilde{\mathbf{y}})$ using basic properties of the Householder transform in Lemma 3 to obtain the explicit model (12) in Theorem 1. This step requires careful calculation and we leave the details to Appendix B-D.

Lemma 6: The distribution of $(\mathbf{s}, \tilde{\mathbf{y}})$ given in (44) is the same as that of $(\mathbf{s}, \bar{\mathbf{y}})$ given in (12) in Theorem 1.

Combining Lemmas 5 and 6, we get the desired result in Theorem 1.

B. Discussions on Lemma 5

Since Lemma 5 is critical to the proof of Theorem 1, we would like to provide some high-level ideas and informal discussions before we present its full proof in Appendix B-C.

Recall that our original model in (11) reads

$$\begin{cases} \mathbf{s}_1 = f(\mathbf{D})^{\mathsf{T}} \mathbf{U}^{\mathsf{H}} \mathbf{s}; \\ \mathbf{s}_2 = q(\mathbf{V} \mathbf{s}_1); \\ \mathbf{s}_3 = \mathbf{D} \mathbf{V}^{\mathsf{H}} \mathbf{s}_2; \\ \mathbf{y} = \mathbf{U} \mathbf{s}_3 + \mathbf{n}. \end{cases}$$

Our goal is to characterize the joint distribution of $(\mathbf{s}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{y})$. Roughly speaking, for $N, K \geq 3$, fixing the distribution of $(\mathbf{s}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{y})$ does not fully fix the randomness of the Haar matrices $\mathbf{U}$ and $\mathbf{V}$, and we still have freedom to generate the remaining randomness in a convenient way. A systematic way of carrying out this process is via the HD technique in [39]. *To be clear, we use* $(\mathbf{s}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{y})$ *to denote the random vectors from the original model, and use* $(\mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3, \tilde{\mathbf{y}})$ *to denote the corresponding vectors generated via the HD method in the following discussion.*

The main idea of HD is to generate the Haar random matrices $\mathbf{U}$ and $\mathbf{V}$ involved in the above iterations in a recursive way by repeatedly applying Lemma 4. More specifically, the HD technique tells that at each iteration we only need to generate a single Gaussian vector to unfold the randomness of $\mathbf{U}$ (or $\mathbf{V}$) in one dimension, and the resulting sequence will only depend on the initial condition $\{\mathbf{s}, \mathbf{D}, \mathbf{n}\}$ and the exposed Gaussian vectors.

Here we take the first iteration of (11) as an example to shed some light on this. To compute $\mathbf{U}^{\mathsf{H}}\mathbf{s}$, we can construct a Haar random matrix $\mathbf{U}^{\mathsf{H}}$ according to Lemma 4 as

$$\widetilde{\mathbf{U}}^{\mathsf{H}} = \mathbf{R}_1(\mathrm{g}_1) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{K-1} \end{pmatrix} \mathbf{R}_1(\mathbf{s})^{\mathsf{H}}, \tag{45}$$

where $\mathsf{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ and $\mathbf{Q}_{K-1} \sim \text{Haar}(K-1)$ are independent of each other and of $\mathbf{s}$, $\mathbf{D}$, and $\mathbf{n}$. Then we have

$$\begin{aligned}\tilde{\mathbf{s}}_1 &:= f(\mathbf{D})^{\mathsf{T}}\widetilde{\mathbf{U}}^{\mathsf{H}}\mathbf{s} \\ &= f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}\end{pmatrix}\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s} \\ &= f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s},\end{aligned}$$

where the last equality holds since $\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s}$ is only nonzero in its first element due to Lemma 3. *An important observation is that* $\tilde{\mathbf{s}}_1$ *depends on the Haar matrix* $\widetilde{\mathbf{U}}$ *only through a Gaussian vector* g_1*, and is invariant to the remaining Haar matrix* $\mathbf{Q}_{K-1}$. Since $\mathbf{Q}_{K-1}$ involved in (45) is Haar distributed and independent of all the other random variables generated up to this point, we can apply the same technique to $\mathbf{Q}_{K-1}$ when another multiplication involving $\mathbf{U}$ is required, and a new Gaussian vector $\mathsf{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ will be exposed (see the expression of $\tilde{\mathbf{y}}$ in (44)). The Haar matrix $\mathbf{V}$ can be constructed similarly when we deal with the second and third iterations, and two Gaussian vectors $\mathsf{z}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathsf{z}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ will be exposed in these two iterations after applying the HD technique (see the expression of $\tilde{\mathbf{s}}_2$ and $\tilde{\mathbf{s}}_3$ in (44)). A detailed derivation of (11) using the HD technique is provided in Appendix B-C.1.

To gain some further insight, we directly give the form of the two Haar matrices constructed using the HD technique without proof (see Appendix B-C.1 for a detailed proof):

$$\widetilde{\mathbf{U}} = \mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}} \quad (46)$$

and

$$\widetilde{\mathbf{V}} = \mathbf{R}_1(\mathsf{z}_1)\mathbf{R}_2(\mathbf{v}_1)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-2}\end{pmatrix}\mathbf{R}_2(\mathsf{z}_2)^{\mathsf{H}}\mathbf{R}_1(\tilde{\mathsf{s}}_1)^{\mathsf{H}}. \quad (47)$$

In the above expressions, $\mathbf{Q}_{K-2}$ and $\mathbf{P}_{N-2}$ are Haar matrices independent of all the other random variables, which are the unexposed random matrices that are absent in the final result; $\mathsf{g}_1, \mathsf{g}_2, \mathsf{z}_1, \mathsf{z}_2$ are the exposed Gaussian vectors; $\tilde{\mathsf{s}}_1, \mathbf{v}_1$, and $\mathbf{v}_2$ are some intermediate random vectors generated due to the recursive nature of the HD technique. Equation (44) can be interpreted as replacing the Haar random matrices $\mathbf{U}$ and $\mathbf{V}$ in the original model (11) by the two unitary matrices $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ given above. Here, we use the notation $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ to emphasize the fact that their distributional properties are yet to be proved. To show $(\mathbf{s}, \mathbf{y}) \overset{d}{=} (\mathbf{s}, \tilde{\mathbf{y}})$, it remains to check that $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ have the desired properties, i.e., $\widetilde{\mathbf{U}} \sim \text{Haar}(K)$, $\widetilde{\mathbf{V}} \sim \text{Haar}(N)$, and $\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \mathbf{D}, \mathbf{s}, \mathbf{n}$ are mutually independent. We relegate the details to Appendix B-C.2.

C. Proof of Lemma 5

In this subsection, we give the complete proof of Lemma 5, which consists of the following two steps:

- We first show that (44) is the sequence generated by applying the HD technique to (11);
- We then prove that (44) is statistically equivalent to (11).

1) Derivation of (44): We first give a detailed derivation of how (44) is obtained via the HD technique. Recall that our original model (6) can be written as the following iterative process (see (11)):

$$\begin{cases}\mathbf{s}_1 = f(\mathbf{D})^{\mathsf{T}}\mathbf{U}^{\mathsf{H}}\mathbf{s}; \\ \mathbf{s}_2 = q(\mathbf{V}\mathbf{s}_1); \\ \mathbf{s}_3 = \mathbf{D}\mathbf{V}^{\mathsf{H}}\mathbf{s}_2; \\ \mathbf{y} = \mathbf{U}\mathbf{s}_3 + \mathbf{n}.\end{cases}$$

Next, we apply the HD technique to deal with the above model.

For the first iteration, we construct $\mathbf{U}^{\mathsf{H}}$ according to Lemma 4 as

$$\left(\mathbf{U}^{(1)}\right)^{\mathsf{H}} = \mathbf{R}_1(\mathsf{g}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}\end{pmatrix}\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}, \quad (48)$$

where $\mathsf{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ and $\mathbf{Q}_{K-1} \sim \text{Haar}(K-1)$ are independent of each other and of $\mathbf{s}$, $\mathbf{D}$, and $\mathbf{n}$. Then we have

$$\begin{aligned}\tilde{\mathbf{s}}_1 &= f(\mathbf{D})^{\mathsf{T}}\left(\mathbf{U}^{(1)}\right)^{\mathsf{H}}\mathbf{s} \\ &= f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}\end{pmatrix}\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s} \\ &= f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s},\end{aligned}$$

where the last equality holds since $\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s}$ is only nonzero in its first element.

For the second iteration, we use a similar technique to construct the Haar matrix $\mathbf{V}$ according to Lemma 4 as

$$\mathbf{V}^{(1)} = \mathbf{R}_1(\mathsf{z}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-1}\end{pmatrix}\mathbf{R}_1(\tilde{\mathsf{s}}_1)^{\mathsf{H}}, \quad (49)$$

where $\mathsf{z}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{P}_{N-1} \sim \text{Haar}(N-1)$ are independent of each other and of all the existing random variables. It follows immediately that

$$\begin{aligned}\mathbf{V}^{(1)}\tilde{\mathbf{s}}_1 &= \mathbf{R}_1(\mathsf{z}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-1}\end{pmatrix}\mathbf{R}_1(\tilde{\mathsf{s}}_1)^{\mathsf{H}}\tilde{\mathsf{s}}_1 \\ &= \mathbf{R}_1(\mathsf{z}_1)\mathbf{R}_1(\tilde{\mathsf{s}}_1)^{\mathsf{H}}\tilde{\mathsf{s}}_1,\end{aligned}$$

and thus

$$\tilde{\mathbf{s}}_2 = q(\mathbf{V}^{(1)}\tilde{\mathbf{s}}_1) = q(\mathbf{R}_1(\mathsf{z}_1)\mathbf{R}_1(\tilde{\mathsf{s}}_1)^{\mathsf{H}}\tilde{\mathsf{s}}_1).$$

For the third iteration, we need to calculate $\mathbf{D}\mathbf{V}^{\mathsf{H}}\tilde{\mathbf{s}}_2$. From (49), we get

$$\left(\mathbf{V}^{(1)}\right)^{\mathsf{H}} = \mathbf{R}_1(\tilde{\mathsf{s}}_1)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-1}^{\mathsf{H}}\end{pmatrix}\mathbf{R}_1(\mathsf{z}_1)^{\mathsf{H}}.$$

Let $\mathbf{v}_1 = \mathbf{R}_1(\mathsf{z}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_2$. Since $\mathbf{P}_{N-1}^{\mathsf{H}}$ is Haar distributed and independent of $\mathbf{v}_1$, we can still apply the above technique to construct $\mathbf{P}_{N-1}^{\mathsf{H}}$ as

$$\mathbf{P}_{N-1}^{\mathsf{H}} = \mathbf{R}_1(\mathsf{z}_2[2:N])\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-2}^{\mathsf{H}}\end{pmatrix}\mathbf{R}_1(\mathbf{v}_1[2:N])^{\mathsf{H}},$$

where $\mathsf{z}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{P}_{N-2} \sim \text{Haar}(N-2)$ are independent of each other and of all the existing random variables. Then we have

$$\left(\mathbf{V}^{(2)}\right)^{\mathsf{H}} = \mathbf{R}_1(\tilde{\mathsf{s}}_1)\mathbf{R}_2(\mathsf{z}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-2}^{\mathsf{H}}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{R}_1(\mathsf{z}_1)^{\mathsf{H}}$$

and

$$\begin{aligned}\tilde{\mathbf{s}}_3 &= \mathbf{D}\left(\mathbf{V}^{(2)}\right)^{\mathsf{H}}\tilde{\mathbf{s}}_2\\ &= \mathbf{D}\mathbf{R}_1(\tilde{\mathbf{s}}_1)\mathbf{R}_2(\mathsf{z}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{N-2}^{\mathsf{H}}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1\\ &= \mathbf{D}\mathbf{R}_1(\tilde{\mathbf{s}}_1)\mathbf{R}_2(\mathsf{z}_2)\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1,\end{aligned}$$

where $\mathbf{P}_{N-2}^{\mathsf{H}}$ disappears in the second equality since $\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1$ is only nonzero in its first two elements.

Finally, we calculate $\tilde{\mathbf{y}} = \mathbf{U}\tilde{\mathbf{s}}_3 + \mathbf{n}$. According to (48),

$$\mathbf{U}^{(1)} = \mathbf{R}_1(\mathbf{s})\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}^{\mathsf{H}}\end{pmatrix}\mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}.$$

Similarly, let $\mathbf{v}_2 = \mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_3$ and construct $\mathbf{Q}_{K-1}^{\mathsf{H}}$ as

$$\mathbf{Q}_{K-1}^{\mathsf{H}} = \mathbf{R}_1(\mathsf{g}_2[2:K])\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{pmatrix}\mathbf{R}_1(\mathbf{v}_2[2:K])^{\mathsf{H}},$$

where $\mathsf{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ and $\mathbf{Q}_{K-2} \sim \text{Haar}(K-2)$ are independent of each other and of all the existing random variables. Then we have

$$\mathbf{U}^{(2)} = \mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}$$

and

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{U}^{(2)}\tilde{\mathbf{s}}_3 + \mathbf{n}\\ &= \mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_3 + \mathbf{n}\\ &= \mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{v}_2 + \mathbf{n}.\end{aligned}$$

This gives the sequence $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3, \tilde{\mathbf{y}})$ in (44), and the two constructed Haar matrices are $\widetilde{\mathbf{U}} = \mathbf{U}^{(2)}$ and $\widetilde{\mathbf{V}} = \mathbf{V}^{(2)}$, which are exactly those given in (46) and (47).

2) Statistical Equivalence of (44) and (11): The proof follows the general principle proposed in [39, Theorem 2]. Here we provide a complete proof to make the paper self-contained.

First, it is easy to check that $(\mathbf{s}, \tilde{\mathbf{y}})$ given in (44) can be obtained by substituting $\widetilde{\mathbf{U}}$ in (46) and $\widetilde{\mathbf{V}}$ in (47) into (11). To show the statistical equivalence, we still need to prove that $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ in (46) and (47) have the following desired properties:

- $\widetilde{\mathbf{U}} \sim \text{Haar}(K), \widetilde{\mathbf{V}} \sim \text{Haar}(N)$; (50a)
- $\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \mathbf{s}, \mathbf{D}, \mathbf{n}$ are mutually independent. (50b)

Next, we prove the above properties for $\widetilde{\mathbf{U}}$, and those for $\widetilde{\mathbf{V}}$ can be proved by similar arguments. We first analyze the inner term $\mathbf{R}_2(\mathsf{g}_2)\left(\begin{smallmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{smallmatrix}\right)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}$ in $\widetilde{\mathbf{U}}$:

$$\begin{aligned}&\mathbf{R}_2(\mathsf{g}_2)\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{pmatrix}\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\\ &= \begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_1(\mathsf{g}_2[2:K])\left(\begin{smallmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-2}\end{smallmatrix}\right)\mathbf{R}_1(\mathbf{v}_2[2:K])^{\mathsf{H}}\end{pmatrix}\\ &:= \begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}\end{pmatrix}.\end{aligned}$$

By the definition of $\mathbf{v}_2$ and from the method of generating $\mathbf{Q}_{K-2}$ and g_2, it is clear that $\mathsf{g}_2, \mathbf{Q}_{K-2}$, and $\mathbf{v}_2$ are mutually independent. It follows immediately from Lemma 4 that $\mathbf{Q}_{K-1} \sim \text{Haar}(K-1)$ and is independent of $\mathbf{v}_2$, and hence independent of all other random variables except $\mathbf{Q}_{K-2}$ and g_2 that construct $\mathbf{Q}_{K-1}$. We then investigate

$$\tilde{\mathbf{U}} = \mathbf{R}_1(\mathbf{s})\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{K-1}\end{pmatrix}\mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}.$$

Again from Lemma 4, we know that $\widetilde{\mathbf{U}} \sim \text{Haar}(K)$ and is independent of all other random variables (except $\mathbf{Q}_{K-2}$, g_2, and g_1), i.e., $\widetilde{\mathbf{U}}$ has the desired properties in (50). This completes our proof of the statistical equivalence between $(\mathbf{s}, \mathbf{y})$ in (11) and $(\mathbf{s}, \tilde{\mathbf{y}})$ in (44).

D. Proof of Lemma 6

In this subsection, we give the proof of Lemma 6, i.e., we derive (12) from (44). For clarity, we copy the expressions of $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3, \tilde{\mathbf{y}}$ in (44) here.

$$\begin{cases}\tilde{\mathbf{s}}_1 = f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s};\\ \tilde{\mathbf{s}}_2 = q(\mathbf{R}_1(\mathsf{z}_1)\mathbf{R}_1(\tilde{\mathbf{s}}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_1);\\ \tilde{\mathbf{s}}_3 = \mathbf{D}\mathbf{R}_1(\tilde{\mathbf{s}}_1)\mathbf{R}_2(\mathsf{z}_2)\mathbf{R}_1(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1;\\ \tilde{\mathbf{y}} = \eta\mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{v}_2 + \mathbf{n},\end{cases}$$

where $\mathbf{v}_1 = \mathbf{R}_1(\mathsf{z}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_2$ and $\mathbf{v}_2 = \mathbf{R}_1(\mathsf{g}_1)^{\mathsf{H}}\tilde{\mathbf{s}}_3$. In the subsequent analysis, we will frequently encounter Householder transform-vector multiplications of the following form:

$$\mathbf{R}_1(\mathsf{g})\mathbf{R}_1(\mathbf{v})^{\mathsf{H}}\mathbf{v},$$

where $\mathsf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{v}$ is a random vector independent of g. According to the properties of the Householder transform, i.e., (ii) and (iii) of Lemma 3, we have

$$\mathbf{R}_1(\mathsf{g})\mathbf{R}_1(\mathbf{v})^{\mathsf{H}}\mathbf{v} = \|\mathbf{v}\|\,\mathbf{R}_1(\mathsf{g})\mathbf{e}_1 = \frac{\|\mathbf{v}\|}{\|\mathsf{g}\|}\,\mathsf{g}. \tag{51}$$

It follows immediately that

$$\tilde{\mathbf{s}}_1 = f(\mathbf{D})^{\mathsf{T}}\mathbf{R}_1(\mathsf{g}_1)\mathbf{R}_1(\mathbf{s})^{\mathsf{H}}\mathbf{s} = \frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|}\,f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1 = \hat{\mathbf{s}}_1, \tag{52}$$

where the last equality is due to the definition of $\hat{\mathbf{s}}_1$ in (13).

Next we begin our derivation of (12). First, we compute $\mathbf{R}_2(\mathsf{g}_2)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{v}_2$ in $\tilde{\mathbf{y}}$ using (51):

$$\begin{aligned}&\mathbf{R}_2(\mathsf{g}_2)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{v}_2\\ &= \begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_1(\mathsf{g}_2[2:K])\mathbf{R}_1(\mathbf{v}_2[2:K])^{\mathsf{H}}\end{pmatrix}\begin{pmatrix}\mathbf{v}_2[1]\\ \mathbf{v}_2[2:K]\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{v}_2[1]\\ \mathbf{R}_1(\mathsf{g}_2[2:K])\mathbf{R}_1(\mathbf{v}_2[2:K])^{\mathsf{H}}\mathbf{v}_2[2:K]\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{v}_2[1]\\ \frac{\|\mathbf{v}_2[2:K]\|}{\|\mathsf{g}_2[2:K]\|}\,\mathsf{g}_2[2:K]\end{pmatrix}.\end{aligned}$$

Combining the above equation with the definition $\mathbf{R}_1(\mathbf{s}) = \begin{pmatrix}\frac{\mathbf{s}}{\|\mathbf{s}\|} & \mathbf{B}(\mathbf{s})\end{pmatrix}$, we can express $\tilde{\mathbf{y}}$ as

$$\begin{aligned}\tilde{\mathbf{y}} &= \eta\mathbf{R}_1(\mathbf{s})\mathbf{R}_2(\mathsf{g}_2)\mathbf{R}_2(\mathbf{v}_2)^{\mathsf{H}}\mathbf{v}_2 + \mathbf{n}\\ &= \eta\begin{pmatrix}\frac{\mathbf{s}}{\|\mathbf{s}\|} & \mathbf{B}(\mathbf{s})\end{pmatrix}\begin{pmatrix}\mathbf{v}_2[1]\\ \frac{\|\mathbf{v}_2[2:K]\|}{\|\mathsf{g}_2[2:K]\|}\,\mathsf{g}_2[2:K]\end{pmatrix} + \mathbf{n}\\ &= \eta\frac{\mathbf{v}_2[1]}{\|\mathbf{s}\|}\mathbf{s} + \eta\frac{\|\mathbf{v}_2[2:K]\|}{\|\mathsf{g}_2[2:K]\|}\mathbf{B}(\mathbf{s})\mathsf{g}_2[2:K] + \mathbf{n}.\end{aligned} \tag{53}$$

By further noting that

$$\mathbf{B}(\mathbf{s})\mathrm{g}_2[2:K] = \mathbf{R}(\mathbf{s})\mathrm{g}_2 - \frac{\mathrm{g}_2[1]}{\|\mathbf{s}\|}\mathbf{s}$$

and substituting it into (53), we get

$$\tilde{\mathbf{y}} = \eta\left(\frac{\mathbf{v}_2[1]}{\|\mathbf{s}\|} - \frac{\mathrm{g}_2[1]}{\|\mathbf{s}\|}\frac{\|\mathbf{v}_2[2:K]\|}{\|\mathrm{g}_2[2:K]\|}\right)\mathbf{s} + \eta\frac{\|\mathbf{v}_2[2:K]\|}{\|\mathrm{g}_2[2:K]\|}\mathbf{R}(\mathbf{s})\mathrm{g}_2 + \mathbf{n}. \quad (54)$$

Note that if $\mathrm{g} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$, then for any unitary matrix $\mathbf{U}$ independent of g, $\mathbf{U}\mathrm{g} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$, i.e., $\mathbf{U}\mathrm{g} \overset{d}{=} \mathrm{g}$, and $\mathbf{U}\mathrm{g}$ is independent of $\mathbf{U}$. Therefore, since g_2 is independent of all other random variables in (54), i.e., $\mathbf{s}, \mathbf{v}_2$, and $\mathbf{n}$, $\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2 \overset{d}{=} \mathrm{g}_2$ and $\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2$ is also independent of $\mathbf{s}$, $\mathbf{v}_2$ and $\mathbf{n}$, and hence we can replace g_2 with $\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2$ in (54), which yields

$$\tilde{\mathbf{y}} \overset{d}{=} \eta\left(\frac{\mathbf{v}_2[1]}{\|\mathbf{s}\|} - \frac{(\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2)[1]}{\|\mathbf{s}\|}\frac{\|\mathbf{v}_2[2:K]\|}{\|(\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2)[2:K]\|}\right)\mathbf{s} + \eta\frac{\|\mathbf{v}_2[2:K]\|}{\|(\mathbf{R}(\mathbf{s})^{-1}\mathrm{g}_2)[2:K]\|}\mathrm{g}_2 + \mathbf{n}. \quad (55)$$

We next analyze

$$\mathbf{v}_2 = \mathbf{R}_1(\mathrm{g}_1)^{\mathsf{H}}\tilde{\mathrm{s}}_3 = \mathbf{R}_1(\mathrm{g}_1)^{\mathsf{H}}\mathbf{D}\mathbf{R}_1(\tilde{\mathrm{s}}_1)\mathbf{R}_2(\mathrm{z}_2)\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1 \quad (56)$$

step by step. First, from the definitions of $\mathbf{v}_1$ and $\tilde{\mathrm{s}}_2$, we have

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{R}_1(\mathrm{z}_1)^{\mathsf{H}}q(\mathbf{R}_1(\mathrm{z}_1)\mathbf{R}_1(\tilde{\mathrm{s}}_1)^{\mathsf{H}}\tilde{\mathrm{s}}_1)\\
&= \mathbf{R}_1(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\tilde{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right) \quad \text{(from (51))}\\
&= \mathbf{R}_1(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right) \quad \text{(from (52))}\\
&= \begin{pmatrix}\frac{\mathrm{z}_1^{\mathsf{H}}}{\|\mathrm{z}_1\|}\\ \mathbf{B}(\mathrm{z}_1)^{\mathsf{H}}\end{pmatrix} q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right) \quad \text{(from (40))}\\
&= \begin{pmatrix}\frac{\mathrm{z}_1^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)}{\|\mathrm{z}_1\|}\\ \mathbf{B}(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)\end{pmatrix}. \quad (57)
\end{aligned}$$

Then we have

$$\begin{aligned}
&\mathbf{R}_2(\mathrm{z}_2)\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1\\
&= \begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_1(\mathrm{z}_2[2:N])\mathbf{R}_1(\mathbf{v}_1[2:N])^{\mathsf{H}}\end{pmatrix}\begin{pmatrix}\mathbf{v}_1[1]\\ \mathbf{v}_1[2:N]\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{v}_1[1]\\ \mathbf{R}_1(\mathrm{z}_2[2:N])\mathbf{R}_1(\mathbf{v}_1[2:N])^{\mathsf{H}}\mathbf{v}_1[2:N]\end{pmatrix}\\
&= \begin{pmatrix}\frac{\mathrm{z}_1^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)}{\|\mathrm{z}_1\|}\\ \frac{\left\|\mathbf{B}(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)\right\|}{\|\mathrm{z}_2[2:N]\|}\mathrm{z}_2[2:N]\end{pmatrix},
\end{aligned}$$

where the last equality comes from (51) and (57). It follows from the above equality, (40), and (52) that

$$\begin{aligned}
&\mathbf{R}_1(\tilde{\mathrm{s}}_1)\mathbf{R}_2(\mathrm{z}_2)\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1\\
&= \mathbf{R}_1(\hat{\mathrm{s}}_1)\mathbf{R}_2(\mathrm{z}_2)\mathbf{R}_2(\mathbf{v}_1)^{\mathsf{H}}\mathbf{v}_1\\
&= \begin{pmatrix}\frac{\hat{\mathbf{s}}_1}{\|\hat{\mathbf{s}}_1\|} & \mathbf{B}(\hat{\mathrm{s}}_1)\end{pmatrix}\begin{pmatrix}\frac{\mathrm{z}_1^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)}{\|\mathrm{z}_1\|}\\ \frac{\left\|\mathbf{B}(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)\right\|}{\|\mathrm{z}_2[2:N]\|}\mathrm{z}_2[2:N]\end{pmatrix}\\
&= \frac{\mathrm{z}_1^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)}{\|\mathrm{z}_1\|\|\hat{\mathbf{s}}_1\|}\hat{\mathbf{s}}_1 + \frac{\left\|\mathbf{B}(\mathrm{z}_1)^{\mathsf{H}}q\left(\frac{\|\hat{\mathbf{s}}_1\|}{\|\mathrm{z}_1\|}\mathrm{z}_1\right)\right\|}{\|\mathrm{z}_2[2:N]\|}\mathbf{B}(\hat{\mathrm{s}}_1)\mathrm{z}_2[2:N]\\
&:= C_1\hat{\mathbf{s}}_1 + C_2\mathbf{B}(\hat{\mathrm{s}}_1)\mathrm{z}_2[2:N].
\end{aligned}$$

Finally, from the above equality, (40), and (56), we have

$$\begin{aligned}
\mathbf{v}_2 &= \begin{pmatrix}\frac{\mathrm{g}_1^{\mathsf{H}}}{\|\mathrm{g}_1\|}\\ \mathbf{B}(\mathrm{g}_1)^{\mathsf{H}}\end{pmatrix}(C_1\mathbf{D}\hat{\mathbf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathrm{s}}_1)\mathrm{z}_2[2:N])\\
&= \begin{pmatrix}\frac{\mathrm{g}_1^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathrm{s}}_1)\mathrm{z}_2[2:N])\}}{\|\mathrm{g}_1\|}\\ \mathbf{B}(\mathrm{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathrm{s}}_1)\mathrm{z}_2[2:N]\}\end{pmatrix}. \quad (58)
\end{aligned}$$

Plugging (58) into (55), we can get the desired model (12).

APPENDIX C
PROOF OF THEOREM 2

In this section, we provide a detailed proof of Theorem 2. In Section C-A, we first give some definitions and preliminary results of the asymptotic analysis. Section C-B gives two useful auxiliary results that are important for the proof, and Section C-C contains the main proof of Theorem 2.

A. Preliminaries

Lemma 7: Let $\{X_N\}$ and $\{Y_N\}$ be sequences of random variables. If

$$X_N \xrightarrow{a.s.} X, \quad Y_N \xrightarrow{a.s.} Y,$$

then

$$X_N + Y_N \xrightarrow{a.s.} X + Y, \quad X_N Y_N \xrightarrow{a.s.} XY,$$

and

$$\frac{X_N}{Y_N} \xrightarrow{a.s.} \frac{X}{Y} \text{ (if } Y_N, Y \neq 0).$$

Lemma 8 (Kolmogorov's Strong Law of Large Numbers [55]): Assume that $X_1, X_2, \ldots$ are independent with means $\mu_1, \mu_2, \ldots$ and variances $\sigma_1^2, \sigma_2^2, \ldots$ such that $\sum_N \frac{\sigma_N^2}{N^2} < \infty$. Then

$$\frac{X_1 + X_2 + \cdots + X_N - (\mu_1 + \mu_2 + \cdots + \mu_N)}{N} \xrightarrow{a.s.} 0.$$

As a corollary, if $X_1, X_2, \ldots$ are i.i.d. with mean μ, then

$$\frac{X_1 + X_2 + \cdots + X_N}{N} - \mu \xrightarrow{a.s.} 0.$$

Lemma 9 [56, Theorem 3]: Let $\{X_N\}$ be a sequence of random variables. If $\{X_N\}$ converges in distribution to X, then

$$\mathbb{E}[g(X_N)] \to \mathbb{E}[g(X)]$$

for all bounded measurable functions g such that $\mathbb{P}\{X \in C(g)\} = 1$, where $C(g) = \{x \mid g \text{ is continuous at } x\}$ denotes the continuity set of g.

Definition 3 (Empirical Spectral Distribution [57]): Consider an $N \times N$ Hermitian matrix $\mathbf{T}_N$. Define its empirical spectral distribution (e.s.d.) $F^{\mathbf{T}_N}$ to be the distribution function of the eigenvalues of $\mathbf{T}_N$, i.e., for $x \in \mathbb{R}$,

$$F^{\mathbf{T}_N}(x) = \frac{1}{N}\sum_{j=1}^{N} 1_{\{\lambda_j \le x\}}(x),$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\mathbf{T}_N$.

Lemma 10 [57, Section 3.2 and Section 7.1]: Consider a matrix $\mathbf{X} \in \mathbb{C}^{K\times N}$ with i.i.d. entries following $\mathcal{CN}\left(0, \frac{1}{N}\right)$. As $K, N \to \infty$ with $\frac{K}{N} \to c \in (0,1)$, the following results hold:

(i) the e.s.d. of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ converges weakly and almost surely to a distribution function F with density given by:

$$p(x) = \frac{1}{2\pi c x}\sqrt{(x-a)_+(b-x)_+},$$

where $a = (1-\sqrt{c})^2, b = (1+\sqrt{c})^2$, and $(x)_+ = \max\{x, 0\}$;

(ii) the largest eigenvalue of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, denoted by $\lambda_{\max}$, satisfies

$$\lambda_{\max} \xrightarrow{a.s.} (1+\sqrt{c})^2.$$

B. Auxiliary Lemmas

This subsection introduces two auxiliary lemmas used in the main proof in Section C-C.

Lemma 11: Define $\widetilde{\mathbf{D}} = \operatorname{diag}(d_1, d_2, \ldots, d_K)$, where $d_1, d_2, \ldots, d_K$ are the nonzero singular values of $\mathbf{H}$ (satisfying Assumption 1). Assume that $\sigma(\cdot)$ is a function that is continuous a.e. and bounded on any compact set of $(0, \infty)$; $\mathsf{g}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\mathsf{g}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, and $\widetilde{\mathbf{D}}$ are mutually independent. Then as $N, K \to \infty$ and $\frac{N}{K} \to \gamma > 1$, it holds that

(i) $\dfrac{\mathsf{g}_1^{\mathsf{H}}\sigma(\widetilde{\mathbf{D}})\mathsf{g}_1}{K} \xrightarrow{a.s.} \mathbb{E}[\sigma(d)]$;

(ii) $\dfrac{\mathsf{g}_1^{\mathsf{H}}\sigma(\widetilde{\mathbf{D}})\mathsf{g}_2}{K} \xrightarrow{a.s.} 0$,

where $d = \sqrt{\lambda}$ and λ follows the Marchenko-Pastur distribution, whose density is given in (16).

Proof: First, from the definition of $\sigma(\cdot)$ and (ii) of Lemma 10, we know that for sufficiently large N and K, there exists a constant $M > 0$ such that

$$\sup_{1\le i\le K} |\sigma(d_i)| \le M$$

with probability one. Then according to the strong law of large numbers in Lemma 8, we have

$$\begin{aligned}&\frac{\mathsf{g}_1^{\mathsf{H}}\sigma(\widetilde{\mathbf{D}})\mathsf{g}_1}{K} - \frac{1}{K}\sum_{i=1}^{K}\sigma(d_i)\\ &= \frac{1}{K}\sum_{i=1}^{K}\sigma(d_i)|\mathsf{g}_1[i]|^2 - \frac{1}{K}\sum_{i=1}^{K}\sigma(d_i) \xrightarrow{a.s.} 0,\end{aligned}$$

and

$$\frac{\mathsf{g}_1^{\mathsf{H}}\sigma(\widetilde{\mathbf{D}})\mathsf{g}_2}{K} = \frac{1}{K}\sum_{i=1}^{K}\sigma(d_i)\mathsf{g}_1[i]^{\dagger}\mathsf{g}_2[i] \xrightarrow{a.s.} 0,$$

which immediately gives (ii) of Lemma 11. Next we continue to prove (i) of Lemma 11. Note that

$$\frac{1}{K}\sum_{i=1}^{K}\sigma(d_i) = \frac{1}{K}\sum_{i=1}^{K}\sigma(\sqrt{\lambda_i}),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_K$ are the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$. Let X_K be a random variable following the e.s.d. of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$, then

$$\frac{1}{K}\sum_{i=1}^{K}\sigma(\sqrt{\lambda_i}) = \mathbb{E}\left[\sigma(\sqrt{X_K})\right].$$

Define

$$g(x) = \begin{cases}\sigma(\sqrt{x}), & \text{if } x \in [0, (1+\sqrt{c})^2 + 1];\\ 0, & \text{otherwise,}\end{cases}$$

where $c = \frac{1}{\gamma}$. Then $g(\cdot)$ is continuous a.e. and bounded. It follows from (ii) of Lemma 10 that with probability one, the largest eigenvalue of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ is bounded by $(1+\sqrt{c})^2 + 1$ for sufficiently large K, which implies that

$$\mathbb{E}\left[\sigma(\sqrt{X_K})\right] \to \mathbb{E}\left[g(X_K)\right].$$

Let λ be a random variable following the Marchenko-Pastur distribution. Then we have $\mathbb{P}\{\lambda \in C(g)\} = 1$ since λ is a continuous random variable and g is continuous a.e., where the definition of $C(g)$ is given in Lemma 9. Therefore, according to (i) of Lemma 10, we can apply Lemma 9 to $\{X_K\}$ and λ to obtain

$$\mathbb{E}[g(X_K)] \to \mathbb{E}[g(\lambda)] = \mathbb{E}\left[\sigma(\sqrt{\lambda})\right].$$

Combining the above discussions, we get the desired result

$$\frac{\mathsf{g}_1^{\mathsf{H}}\sigma(\widetilde{\mathbf{D}})\mathsf{g}_1}{K} \xrightarrow{a.s.} \mathbb{E}\left[\sigma(\sqrt{\lambda})\right] = \mathbb{E}\left[\sigma(d)\right].$$

The proof is completed. □

Lemma 12: Assume that $\mathsf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, $\alpha \xrightarrow{a.s.} \bar{\alpha}$, and $q(\cdot)$ satisfies Assumption 3. Then,

$$\frac{\mathsf{z}^{\mathsf{H}}q(\alpha\mathsf{z})}{N} \xrightarrow{a.s.} \mathbb{E}\left[\mathsf{Z}^{\dagger}q(\bar{\alpha}\mathsf{Z})\right], \quad \text{where } \mathsf{Z} \sim \mathcal{CN}(0,1).$$

Proof: Note that

$$\begin{aligned}&\frac{\mathsf{z}^{\mathsf{H}}q(\alpha\mathsf{z})}{N}\\ &= \frac{1}{N}\sum_{i=1}^{N} z_i^{\dagger}q(\alpha z_i)\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathcal{R}(z_i)\mathcal{R}(q(\alpha z_i)) + \frac{1}{N}\sum_{i=1}^{N}\mathcal{I}(z_i)\mathcal{I}(q(\alpha z_i))\\ &\quad + j\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}(z_i)\mathcal{I}(q(\alpha z_i)) - j\frac{1}{N}\sum_{i=1}^{N}\mathcal{I}(z_i)\mathcal{R}(q(\alpha z_i)).\end{aligned}$$

To show the almost sure convergence, it suffices to show that each of the above four terms converges almost surely to a constant. Next, we will prove that

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}(z_i)\mathcal{R}(q(\alpha z_i)) \xrightarrow{a.s.} \mathbb{E}\left[\mathcal{R}(\mathsf{Z})\mathcal{R}(q(\bar{\alpha}\mathsf{Z}))\right],$$

where $\mathsf{Z}\sim\mathcal{CN}(0,1)$, and the convergence of the other three terms can be proved using similar arguments. Let $g(x)=\mathcal{R}(q(x))$. For any given $\epsilon>0$, we construct the following lower and upper bounds of g:

$$l_\epsilon(x)=\inf_y\left\{g(y)+\frac{1}{\epsilon}|y-x|\right\},$$
$$u_\epsilon(x)=\sup_y\left\{g(y)-\frac{1}{\epsilon}|y-x|\right\}.$$

According to [56, page 15], l_ϵ and u_ϵ satisfy

(i) $l_\epsilon(x)\le g(x)\le u_\epsilon(x), \forall\, x\in\mathbb{C}$.
(ii) l_ϵ and u_ϵ are Lipschitz continuous with Lipschitz constant $\frac{1}{\epsilon}$, i.e., $|l_\epsilon(x_1)-l_\epsilon(x_2)|\le\frac{1}{\epsilon}|x_1-x_2|$ and $|u_\epsilon(x_1)-u_\epsilon(x_2)|\le\frac{1}{\epsilon}|x_1-x_2|$.
(iii) $l_\epsilon(x)$ and $u_\epsilon(x)$ are bounded, i.e., $\exists\, B>0$ not depending on ϵ such that $|l_\epsilon(x)|\le B$ and $|u_\epsilon(x)|\le B$, $\forall\, x\in\mathbb{C}$.
(iv) If x is a continuous point of $g(\cdot)$, then $\lim_{\epsilon\downarrow 0} l_\epsilon(x)=g(x)=\lim_{\epsilon\downarrow 0} u_\epsilon(x)$.

Denote $G(x,\alpha)=\mathcal{R}(x)\mathcal{R}(q(\alpha x))=\mathcal{R}(x)g(\alpha x)$. Based on the above upper and lower bounds of $g(x)$, we can upper and lower bound $G(x,\alpha)$ as

$$G(x,\alpha)\le\mathcal{R}(x)_+u_\epsilon(\alpha x)+\mathcal{R}(x)_-l_\epsilon(\alpha x)\triangleq U_\epsilon(x,\alpha)$$

and

$$G(x,\alpha)\ge\mathcal{R}(x)_+l_\epsilon(\alpha x)+\mathcal{R}(x)_-u_\epsilon(\alpha x)\triangleq L_\epsilon(x,\alpha),$$

where $x_+=\max\{x,0\}$ and $x_-=\min\{x,0\}$. It follows that for any $\epsilon>0$,

$$\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\alpha)\le\frac{1}{N}\sum_{i=1}^{N}G(z_i,\alpha)\le\frac{1}{N}\sum_{i=1}^{N}U_\epsilon(z_i,\alpha). \quad (59)$$

To show that $\frac{1}{N}\sum_{i=1}^{N}G(z_i,\alpha)\xrightarrow{a.s.}\mathbb{E}[G(\mathsf{Z},\bar{\alpha})]$, we will first prove that for a given $\epsilon>0$,

$$\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\alpha)\xrightarrow{a.s.}\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})], \quad (60a)$$
$$\frac{1}{N}\sum_{i=1}^{N}U_\epsilon(z_i,\alpha)\xrightarrow{a.s.}\mathbb{E}[U_\epsilon(\mathsf{Z},\bar{\alpha})], \quad (60b)$$

which, together with (59), implies

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}G(z_i,\alpha)\ge\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]\quad a.s. \quad (61a)$$
$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}G(z_i,\alpha)\le\mathbb{E}[U_\epsilon(\mathsf{Z},\bar{\alpha})]\quad a.s. \quad (61b)$$

Then we will show that

$$\lim_{\epsilon\downarrow 0}\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]=\lim_{\epsilon\downarrow 0}\mathbb{E}[U_\epsilon(\mathsf{Z},\bar{\alpha})]=\mathbb{E}[G(\mathsf{Z},\bar{\alpha})]. \quad (62)$$

Letting $\epsilon\downarrow 0$ in (61) and using (62) give the desired result:

$$\frac{1}{N}\sum_{i=1}^{N}G(z_i,\alpha)\xrightarrow{a.s.}\mathbb{E}[G(\mathsf{Z},\bar{\alpha})].$$

It remains to prove (60) and (62). We will only provide the proof of (60) and (62) for the lower bound L_ϵ since the upper bound can be proved in exactly the same way. First,

$$\begin{aligned}&\left|\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\alpha)-\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]\right|\\ &\le\left|\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\alpha)-\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\bar{\alpha})\right|\\ &\quad+\left|\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\bar{\alpha})-\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]\right|.\end{aligned}$$

It follows immediately from the strong law of large numbers (i.e., Lemma 8) that the second term tends to zero almost surely. For the first term, we have

$$\begin{aligned}&\left|\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\alpha)-\frac{1}{N}\sum_{i=1}^{N}L_\epsilon(z_i,\bar{\alpha})\right|\\ &\le\frac{1}{N}\sum_{i=1}^{N}|L_\epsilon(z_i,\alpha)-L_\epsilon(z_i,\bar{\alpha})|\\ &\le\frac{1}{N}\sum_{i=1}^{N}|\mathcal{R}(z_i)|\,\frac{|z_i(\alpha-\bar{\alpha})|}{\epsilon}\\ &\le\frac{|\alpha-\bar{\alpha}|}{\epsilon}\,\frac{1}{N}\sum_{i=1}^{N}|z_i|^2\xrightarrow{a.s.}0,\end{aligned}$$

where the second inequality holds due to Lipschitz continuity (i.e., property (ii))) of $l_\epsilon(x)$ and $u_\epsilon(x)$, and the almost convergence is due to the assumption $\alpha\xrightarrow{a.s.}\bar{\alpha}$, Lemma 8, and Lemma 7. This proves (60). Since l_ϵ and u_ϵ are bounded (i.e., property (iii))), we have

$$|L_\epsilon(x,\bar{\alpha})|\le B|\mathcal{R}(x)|.$$

According to the dominated convergence theorem [55],

$$\lim_{\epsilon\downarrow 0}\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]=\mathbb{E}\left[\lim_{\epsilon\downarrow 0}L_\epsilon(\mathsf{Z},\bar{\alpha})\right].$$

Property (iv) of l_ϵ implies that $\lim_{\epsilon\downarrow 0}L_\epsilon(x,\bar{\alpha})=G(x,\bar{\alpha})$, if $\bar{\alpha}x$ is a continuous point of $g(x)$, i.e., $\lim_{\epsilon\downarrow 0}L_\epsilon(x,\bar{\alpha})=G(x,\bar{\alpha})$ almost everywhere. Therefore,

$$\lim_{\epsilon\downarrow 0}\mathbb{E}[L_\epsilon(\mathsf{Z},\bar{\alpha})]=\mathbb{E}\left[\lim_{\epsilon\downarrow 0}L_\epsilon(\mathsf{Z},\bar{\alpha})\right]=\mathbb{E}[G(\mathsf{Z},\bar{\alpha})],$$

which completes the proof.

□

C. Main Proof of Theorem 2

To prove Theorem 2, it suffices to prove (see Lemma 7)

$$T_{\mathbf{s}}\xrightarrow{a.s.}\overline{T}_{\mathbf{s}}\quad\text{and}\quad T_{\mathbf{g}}\xrightarrow{a.s.}\overline{T}_{\mathbf{g}},$$

where

$$T_{\mathbf{s}} = \frac{\mathsf{g}_1^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathsf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N]\}}{\|\mathsf{g}_1\|\|\mathbf{s}\|} - T_{\mathsf{g}}\frac{(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[1]}{\|\mathbf{s}\|},$$
$$T_{\mathsf{g}} = \frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathsf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N]\}\|}{\|(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[2:K]\|},$$

and

$$\overline{T}_{\mathbf{s}} = \overline{C}_1\,\mathbb{E}[d\,f(d)],$$
$$\overline{T}_{\mathsf{g}} = \sqrt{\sigma_s^2\,|\overline{C}_1|^2\,\mathrm{var}[d\,f(d)] + \overline{C}_2^2}.$$

In the above equations,

$$C_1 = \frac{\mathsf{z}_1^{\mathsf{H}} q\left(\frac{\|\hat{\mathsf{s}}_1\|}{\|\mathsf{z}_1\|}\mathsf{z}_1\right)}{\|\hat{\mathsf{s}}_1\|\|\mathsf{z}_1\|},\ C_2 = \frac{\left\|\mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} q\left(\frac{\|\hat{\mathsf{s}}_1\|}{\|\mathsf{z}_1\|}\mathsf{z}_1\right)\right\|}{\|\mathsf{z}_2[2:N]\|},$$
$$\overline{C}_1 = \frac{\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]}{\bar{\alpha}},\quad \overline{C}_2 = \sqrt{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] - |\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2}.$$

Further, $\hat{\mathsf{s}}_1 = \frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|} f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1$ and $\bar{\alpha} = \sqrt{\frac{\sigma_s^2\,\mathbb{E}[f^2(d)]}{\gamma}}$.

We start with analyzing the limiting values of C_1 and C_2. From the strong law of large numbers in Lemma 8, we have

$$\frac{\|\mathsf{z}_1\|^2}{N} \xrightarrow{a.s.} 1,\quad \frac{\|\mathsf{g}_1\|^2}{K} \xrightarrow{a.s.} 1,\quad \frac{\|\mathbf{s}\|^2}{K} \xrightarrow{a.s.} \sigma_s^2. \tag{63}$$

Furthermore, applying Lemma 11 in Section C-B with $\sigma(x) = f^2(x)$ yields

$$\frac{\|f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\|^2}{K} = \frac{\mathsf{g}_1^{\mathsf{H}} f(\mathbf{D}) f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{K} \xrightarrow{a.s.} \mathbb{E}[f^2(d)]. \tag{64}$$

It follows immediately from (63), (64), and Lemma 7 that

$$\alpha := \frac{\|\hat{\mathsf{s}}_1\|}{\|\mathsf{z}_1\|} = \frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|}\frac{\|f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\|}{\|\mathsf{z}_1\|} \xrightarrow{a.s.} \sqrt{\frac{\mathbb{E}[f^2(d)]\,\sigma_s^2}{\gamma}} = \bar{\alpha}. \tag{65}$$

Note that C_1 can be expressed as

$$C_1 = \frac{\mathsf{z}_1^{\mathsf{H}} q\left(\frac{\|\hat{\mathsf{s}}_1\|}{\|\mathsf{z}_1\|}\mathsf{z}_1\right)}{\|\hat{\mathsf{s}}_1\|\|\mathsf{z}_1\|} = \frac{\mathsf{z}_1^{\mathsf{H}} q(\alpha\mathsf{z}_1)}{\alpha\|\mathsf{z}_1\|^2}.$$

According to Lemma 12 in Section C-B and noting that $\alpha \xrightarrow{a.s.} \bar{\alpha}$, we have

$$\frac{\mathsf{z}_1^{\mathsf{H}} q(\alpha\mathsf{z}_1)}{N} \xrightarrow{a.s.} \mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})], \tag{66}$$

where $\mathsf{Z} \sim \mathcal{CN}(0,1)$. This, together with (63), (65), and Lemma 7, proves the convergence of C_1:

$$C_1 = \frac{\mathsf{z}_1^{\mathsf{H}} q(\alpha\mathsf{z}_1)}{\alpha\|\mathsf{z}_1\|^2} \xrightarrow{a.s.} \frac{\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]}{\bar{\alpha}} = \overline{C}_1.$$

To show the convergence of $C_2 = \frac{\|\mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} q(\alpha\mathsf{z}_1)\|}{\|\mathsf{z}_2[2:N]\|}$, we first note that

$$\begin{aligned}
&\frac{\left\|\mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} q\left(\alpha\mathsf{z}_1\right)\right\|^2}{N} \\
&= \frac{\left\|\begin{pmatrix} \frac{\mathsf{z}_1^{\mathsf{H}}}{\|\mathsf{z}_1\|} \\ \mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} \end{pmatrix} q\left(\alpha\mathsf{z}_1\right)\right\|^2 - \left|\frac{\mathsf{z}_1^{\mathsf{H}} q(\alpha\mathsf{z}_1)}{\|\mathsf{z}_1\|}\right|^2}{N} \\
&= \frac{\|q(\alpha\mathsf{z}_1)\|^2}{N} - \left|\frac{\mathsf{z}_1^{\mathsf{H}} q(\alpha\mathsf{z}_1)}{N}\right|^2 \frac{N}{\|\mathsf{z}_1\|^2},
\end{aligned} \tag{67}$$

where the second equality holds because $\mathbf{R}(\mathsf{z}_1)^{\mathsf{H}} = \begin{pmatrix} \frac{\mathsf{z}_1^{\mathsf{H}}}{\|\mathsf{z}_1\|} \\ \mathbf{B}(\mathsf{z}_1)^{\mathsf{H}} \end{pmatrix}$ and $\|\cdot\|$ is rotationally invariant. Similar to Lemma 12, we can show that

$$\frac{\|q(\alpha\mathsf{z}_1)\|^2}{N} \xrightarrow{a.s.} \mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2].$$

Combining this with (63), (66), and (67), and noticing that $\frac{\|\mathsf{z}_2[2:N]\|^2}{N} \xrightarrow{a.s.} 1$, we have

$$C_2 \xrightarrow{a.s.} \sqrt{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] - |\mathbb{E}[\mathsf{Z}^\dagger q(\bar{\alpha}\mathsf{Z})]|^2} = \overline{C}_2.$$

Next, we analyze the convergence of $T_{\mathbf{s}}$ and T_{g}. We begin with $T_{\mathbf{s}}$:

$$\begin{aligned}
T_{\mathbf{s}} &= \frac{\mathsf{g}_1^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathsf{s}}_1 + C_2\mathbf{D}\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N]\}}{\|\mathsf{g}_1\|\|\mathbf{s}\|} - T_{\mathsf{g}}\frac{(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[1]}{\|\mathbf{s}\|} \\
&= C_1\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{\|\mathsf{g}_1\|^2} + C_2\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N]}{\|\mathsf{g}_1\|\|\mathbf{s}\|} \\
&\quad - T_{\mathsf{g}}\frac{(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2)[1]}{\|\mathbf{s}\|} \\
&:= T_{11} + T_{12} + T_{13},
\end{aligned}$$

where the second equality uses the definition of $\hat{\mathsf{s}}_1$ in (13). In what follows, we will show

$$T_{11} \xrightarrow{a.s.} \overline{C}_1\,\mathbb{E}\left[d\,f(d)\right], \tag{68a}$$
$$T_{12} \xrightarrow{a.s.} 0, \tag{68b}$$
$$T_{13} \xrightarrow{a.s.} 0. \tag{68c}$$

Substituting $\sigma(x) = xf(x)$ into Lemma 11, we have

$$\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{K} \xrightarrow{a.s.} \mathbb{E}\left[d\,f(d)\right],$$

which, together with $\frac{\|\mathsf{g}_1\|^2}{K} \xrightarrow{a.s.} 1$ and $C_1 \xrightarrow{a.s.} \overline{C}_1$, implies (68a). We next show that T_{12} vanishes asymptotically. Recalling the definition $\mathbf{R}(\hat{\mathsf{s}}_1) = \left(\frac{\hat{\mathsf{s}}_1}{\|\hat{\mathsf{s}}_1\|}\ \ \mathbf{B}(\hat{\mathsf{s}}_1)\right)$, we then have

$$\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N] = \mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2 - \mathsf{z}_2[1]\frac{\hat{\mathsf{s}}_1}{\|\hat{\mathsf{s}}_1\|}. \tag{69}$$

Using (69) and the definition of $\hat{\mathsf{s}}_1$ in (13), we can upper bound T_{12} by

$$\begin{aligned}
|T_{12}| &= C_2\left|\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{B}(\hat{\mathsf{s}}_1)\mathsf{z}_2[2:N]}{\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| \\
&\le C_2\left|\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2}{\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| + C_2\left|\frac{\mathsf{z}_2[1]\,\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\frac{\hat{\mathsf{s}}_1}{\|\hat{\mathsf{s}}_1\|}}{\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| \\
&= C_2\left|\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2}{\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| + C_2\left|\mathsf{z}_2[1]\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{\|f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\|\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| \\
&\le C_2\left|\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2}{\|\mathsf{g}_1\|\|\mathbf{s}\|}\right| + C_2\left|\frac{\mathsf{z}_2[1]}{\|\mathbf{s}\|}\|\mathbf{D}\|\right|.
\end{aligned}$$

Since both $\hat{\mathsf{s}}_1$ and g_1 are independent of z_2, $\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2 \overset{d}{=} \mathsf{z}_2$ and $\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2$ is independent of g_1. It follows immediately from (63) and Lemmas 7 and 11 that

$$\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathbf{R}(\hat{\mathsf{s}}_1)\mathsf{z}_2}{\|\mathsf{g}_1\|\|\mathbf{s}\|} \xrightarrow{a.s.} 0.$$

In addition, since $\frac{\mathsf{z}_2[1]}{\|\mathbf{s}\|} \xrightarrow{a.s.} 0$ and $\|\mathbf{D}\| \xrightarrow{a.s.} 1+\sqrt{1/\gamma}$ (see (ii) of Lemma 10), we have $\left|\frac{\mathsf{z}_2[1]}{\|\mathbf{s}\|}\|\mathbf{D}\|\right| \xrightarrow{a.s.} 0$. Noting that $C_2 \xrightarrow{a.s.} \overline{C}_2$, we further have (68b). Finally, since $\frac{\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2[1]}{\|\mathbf{s}\|} \xrightarrow{a.s.} 0$ and T_{g} converges almost surely to a constant (which will be shown in the sequel), we have (68c). Combining the above discussions, we get

$$T_{\mathbf{s}} = T_{11}+T_{12}+T_{13} \xrightarrow{a.s.} \overline{C}_1\,\mathbb{E}\left[d\,f(d)\right].$$

We next show the almost sure convergence of T_{g} to a constant. Note that $\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2 \overset{d}{=} \mathsf{g}_2$ due to the independence between $\mathbf{s}$ and g_2. Hence the following holds for the denominator of T_{g}:

$$\frac{\left\|\left(\mathbf{R}(\mathbf{s})^{-1}\mathsf{g}_2\right)[2:K]\right\|}{\sqrt{K}} \xrightarrow{a.s.} 1.$$

For the numerator, using (69) and the definition of $\hat{\mathbf{s}}_1$ in (13), we have

$$\begin{aligned}
&\left|\frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathbf{B}(\hat{\mathbf{s}}_1)\mathsf{z}_2[2:N]\}\|}{\sqrt{K}}\right.\\
&\qquad\left.-\frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathbf{R}(\hat{\mathbf{s}}_1)\mathsf{z}_2\}\|}{\sqrt{K}}\right|\\
&\le C_2\,\frac{\|\mathsf{z}_2[1]\,\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\mathbf{D}\frac{\hat{\mathbf{s}}_1}{\|\hat{\mathbf{s}}_1\|}\|}{\sqrt{K}}\\
&= C_2\,\frac{|\mathsf{z}_2[1]|\,\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\|}{\sqrt{K}\,\|f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\|}\\
&\le C_2\,\frac{|\mathsf{z}_2[1]|\|\mathbf{B}(\mathsf{g}_1)\|\|\mathbf{D}\|}{\sqrt{K}} \xrightarrow{a.s.} 0,
\end{aligned}$$

where the almost sure convergence uses the facts that $\|\mathbf{D}\| \xrightarrow{a.s.} 1+\sqrt{1/\gamma}$, $\|\mathbf{B}(\mathsf{g}_1)\| \le 1$, and $C_2 \xrightarrow{a.s.} \overline{C}_2$. Furthermore,

$$\begin{aligned}
&\frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathbf{R}(\hat{\mathbf{s}}_1)\mathsf{z}_2\}\|^2}{K}\\
&\overset{d}{=}\frac{\|\mathbf{B}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathsf{z}_2\}\|^2}{K}\\
&=\frac{\left\|\mathbf{R}(\mathsf{g}_1)^{\mathsf{H}}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathsf{z}_2\}\right\|^2}{K}\\
&\quad-\frac{\left\|\frac{\mathsf{g}_1^{\mathsf{H}}}{\|\mathsf{g}_1\|}\{C_1\mathbf{D}\hat{\mathbf{s}}_1+C_2\mathbf{D}\mathsf{z}_2\}\right\|^2}{K}\\
&=\frac{\|C_1\frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1+C_2\mathbf{D}\mathsf{z}_2\|^2}{K}\\
&\quad-\frac{1}{K}\left|\frac{\mathsf{g}_1^{\mathsf{H}}\{C_1\frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1+C_2\mathbf{D}\mathsf{z}_2\}}{\|\mathsf{g}_1\|}\right|^2\\
&:= T_{21}-T_{22},
\end{aligned} \tag{70}$$

where the first equality holds since z_2 is independent of $\mathsf{g}_1, \mathbf{D}$, and $\hat{\mathbf{s}}_1$; the second equality is due to the definition of $\mathbf{B}(\cdot)$ in (40); the third equality uses the definition of $\hat{\mathbf{s}}_1$ in (13) and the rotational invariance of $\|\cdot\|$. The first term T_{21} in (70) can be expressed as

$$\begin{aligned}
T_{21} =&|C_1|^2\frac{\|\mathbf{s}\|^2}{\|\mathsf{g}_1\|^2}\frac{\mathsf{g}_1^{\mathsf{H}}f(\mathbf{D})\mathbf{D}^{\mathsf{T}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{K}\\
&+C_2^2\frac{\mathsf{z}_2^{\mathsf{H}}\mathbf{D}^{\mathsf{T}}\mathbf{D}\mathsf{z}_2}{K}+2\frac{\|\mathbf{s}\|}{\|\mathsf{g}_1\|}\frac{\mathcal{R}\left(C_1C_2\mathsf{z}_2^{\mathsf{H}}\mathbf{D}^{\mathsf{T}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1\right)}{K}.
\end{aligned}$$

Applying Lemma 11 with $\sigma(x)=x^2f^2(x)$, $\sigma(x)=x^2$, and $\sigma(x)=x^2f(x)$, and using the almost sure convergence of C_1 and C_2, we have

$$T_{21} \xrightarrow{a.s.} \sigma_s^2\,|\overline{C}_1|^2\,\mathbb{E}[d^2f^2(d)]+\overline{C}_2^2\,\mathbb{E}[d^2].$$

Similarly, for the second term T_{22} in (70), we have

$$\begin{aligned}
T_{22} &= \left|C_1\frac{\sqrt{K}\|\mathbf{s}\|}{\|\mathsf{g}_1\|^2}\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}f(\mathbf{D})^{\mathsf{T}}\mathsf{g}_1}{K}+C_2\frac{\sqrt{K}}{\|\mathsf{g}_1\|}\frac{\mathsf{g}_1^{\mathsf{H}}\mathbf{D}\mathsf{z}_2}{K}\right|^2\\
&\xrightarrow{a.s.} \sigma_s^2\,|\overline{C}_1|^2\,\mathbb{E}^2[d\,f(d)].
\end{aligned}$$

Combining the above discussions and noting that $\mathbb{E}[d^2]=1$ [48], we can conclude that

$$T_{\mathsf{g}} \xrightarrow{a.s.} \sqrt{\sigma_s^2\,|\overline{C}_1|^2\,\mathrm{var}[d\,f(d)]+\overline{C}_2^2},$$

which completes our proof.

Appendix D
Proof of Theorem 3

In this section, we give the proof of Theorem 3. We first prove the convergence of $\widehat{\mathrm{SINR}}_k$ and then show the convergence of $\widehat{\mathrm{SEP}}_k(\beta)$.

A. Proof of Convergence of $\widehat{SINR}_k$

Note that $\mathbb{E}[|s_k|^2]=\sigma_s^2$ is a constant (see Assumption 1). If we can show that $\mathbb{E}[|\hat{y}_k|^2] \to \mathbb{E}[|\bar{y}_k|^2]$ and $\mathbb{E}[s_k^{\dagger}\hat{y}_k] \to \mathbb{E}[s_k^{\dagger}\bar{y}_k]$, then according to the definition of the SINR in (18), we have

$$\widehat{\mathrm{SINR}}_k \to \overline{\mathrm{SINR}}.$$

In the following, we will only prove the convergence of $\mathbb{E}[|\hat{y}_k|^2]$, and the convergence of $\mathbb{E}[s_k^{\dagger}\hat{y}_k]$ can be shown similarly.

Firstly, it follows from Theorem 2 that $\hat{y}_k \xrightarrow{a.s.} \bar{y}_k$, and thus

$$|\hat{y}_k|^2 \xrightarrow{a.s.} |\bar{y}_k|^2.$$

To show the convergence of expectation, we only need to prove that $\{|\hat{y}_k|^2\}$ is uniformly integrable [55]. Note that for a sequence of random variables $\{X_N, N\ge 1\}$, the boundedness of $\mathbb{E}[|X_N|^2]$ can imply uniform integrability of $\{X_N\}$ (see [55, Section 9.5] for details about uniform integrability of random variables). Therefore, it suffices to prove that the sequence $\{\mathbb{E}[|y_k|^4\}$ is bounded. According to (12), we have

$$\begin{aligned}
|\hat{y}_k|^4 &= |\eta T_{\mathbf{s}}s_k+\eta T_{\mathsf{g}}\mathsf{g}_2[k]+n_k|^4\\
&\le 27|\eta T_{\mathbf{s}}s_k|^4+27|\eta T_{\mathsf{g}}\mathsf{g}_2[k]|^4+27|n_k|^4\\
&\le \frac{27}{2}(\eta^8|T_{\mathbf{s}}|^8+|s_k|^8+\eta^8|T_{\mathsf{g}}|^8+|\mathsf{g}_2[k]|^8+2|n_k|^4),
\end{aligned} \tag{71}$$

where the first inequality holds since $|a+b+c|^2 \le 3|a|^2 + 3|b|^2 + 3|c|^2$. From Theorem 2 and Lemma 7, we have

$$|T_{\mathsf{s}}|^8 \xrightarrow{a.s.} |\overline{T}_{\mathsf{s}}|^8 \quad \text{and} \quad |T_{\mathsf{g}}|^8 \xrightarrow{a.s.} |\overline{T}_{\mathsf{g}}|^8,$$

where $\overline{T}_{\mathsf{s}}$ and $\overline{T}_{\mathsf{g}}$ are both constants, and thus

$$\mathbb{E}[|T_{\mathsf{s}}|^8] \to |\overline{T}_{\mathsf{s}}|^8 \quad \text{and} \quad \mathbb{E}[|T_{\mathsf{g}}|^8] \to |\overline{T}_{\mathsf{g}}|^8. \tag{72}$$

Taking expectations over both the left-hand and the right-hand sides of (71) and using (72), we can conclude that

$$\sup_k \mathbb{E}[|\hat{y}_k|^4] < +\infty,$$

which completes the proof.

B. Proof of Convergence of $\widehat{SEP}_k(\beta)$

Given a constellation symbol s, we use $\mathcal{D}_s$ to denote its decision region, i.e., the region within which s will be recovered. With this notation, $\widehat{\mathrm{SEP}}_k(\beta)$ can be expressed as

$$\begin{aligned}\widehat{\mathrm{SEP}}_k(\beta) &= 1 - \mathbb{P}\left(\beta\hat{y}_k \in \mathcal{D}_{s_k}\right) \\ &= 1 - \sum_{m=1}^{M} \mathbb{P}\left(s_k = s^{(m)}\right) \mathbb{P}\left(\beta\hat{y}_k \in \mathcal{D}_{s_k} \mid s_k = s^{(m)}\right) \\ &= 1 - \frac{1}{M}\sum_{m=1}^{M} \mathbb{P}\left(\beta\hat{y}_k \in \mathcal{D}_{s_k} \mid s_k = s^{(m)}\right), \end{aligned} \tag{73}$$

where the third equality holds since s_k is uniformly drawn from $\mathcal{S}_M$. Similarly,

$$\overline{\mathrm{SEP}}_k(\beta) = 1 - \frac{1}{M}\sum_{m=1}^{M} \mathbb{P}\left(\beta\bar{y}_k \in \mathcal{D}_{s_k} \mid s_k = s^{(m)}\right).$$

From Theorem 2, we have $(s_k, \hat{y}_k) \xrightarrow{a.s.} (s_k, \bar{y}_k)$. Hence, the joint distribution of $(s_k, \hat{y}_k)$ converges weakly to that of $(s_k, \bar{y}_k)$. If we can further show

$$\mathbb{P}\left(\beta\bar{y}_k \in \delta\mathcal{D}_{s_k} \mid s_k = s^{(m)}\right) = 0, \tag{74}$$

where $\delta\mathcal{D}_{s_k}$ denotes the boundary of $\mathcal{D}_{s_k}$, then according to [58, Lemma 2.2],

$$\mathbb{P}\left(\beta\hat{y}_k \in \mathcal{D}_{s_k} \mid s_k = s^{(m)}\right) \to \mathbb{P}\left(\beta\bar{y}_k \in \mathcal{D}_{s_k} \mid s_k = s^{(m)}\right),$$

and hence $\widehat{\mathrm{SEP}}_k(\beta) \to \overline{\mathrm{SEP}}_k(\beta) = \overline{\mathrm{SEP}}(\beta)$.

We next show (74). When nearest-neighbor decoding is employed as assumed in this paper, the decision region of s can be expressed as[3]

$$\mathcal{D}_s = \left\{ r \mid |r-s|^2 < \left|r - s^{(i)}\right|^2,\ \forall\, s^{(i)} \in \mathcal{S}_M,\ s^{(i)} \neq s \right\},$$

or equivalently,

$$\begin{aligned}\mathcal{D}_s \;=\; \Big\{ r \mid\ & 2\mathcal{R}((s^{(i)}-s)^\dagger r) + |s|^2 - |s^{(i)}|^2 < 0, \\ & \forall\, s^{(i)} \in \mathcal{S}_M,\ s^{(i)} \neq s \Big\}. \end{aligned} \tag{75}$$

[3]A minor problem is that the decision regions given here are all open sets and cannot cover the whole complex space. This is not an essential problem: if r lies on the boundary of two decision regions, we can just randomly choose one of the corresponding constellation point as the recovered symbol.

It follows that the boundary of $\mathcal{D}_s$ can be expressed as

$$\delta\mathcal{D}_s = \left\{ r \mid \max_{i, s^{(i)} \neq s} \left\{ 2\mathcal{R}((s^{(i)}-s)^\dagger r) + |s|^2 - |s^{(i)}|^2 \right\} = 0 \right\}.$$

Recall that $\beta\bar{y}_k = \beta(\eta\overline{T}_{\mathsf{s}}\, s + \eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k)$. Then

$$\begin{aligned}&\mathbb{P}\left(\beta\bar{y}_k \in \delta\mathcal{D}_{s_k} \mid s_k = s^{(m)}\right) \\ &= \mathbb{P}\left(\beta\left(\eta\overline{T}_{\mathsf{s}}\, s^{(m)} + \eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k\right) \in \delta\mathcal{D}_{s^{(m)}}\right) \\ &= \mathbb{P}\left(\max_{i\neq m}\left\{\mathcal{R}\left(a_{m,i}^\dagger(\eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k)\right) + C_{m,i}\right\} = 0\right), \end{aligned}$$

where $a_{m,i} = 2\beta(s^{(i)} - s^{(m)})$ and $C_{m,i} = 2\mathcal{R}(\eta\overline{T}_{\mathsf{s}} a_{m,i}^\dagger s^{(m)}) + |s^{(m)}|^2 - |s^{(i)}|^2$ are both constants. The last probability can further be upper bounded as

$$\begin{aligned}&\mathbb{P}\left(\max_{i\neq m}\left\{\mathcal{R}\left(a_{m,i}^\dagger(\eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k)\right) + C_{m,i}\right\} = 0\right) \\ &\le \sum_{i\neq m} \mathbb{P}\left(\mathcal{R}\left(a_{m,i}^\dagger(\eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k)\right) + C_{m,i} = 0\right). \end{aligned}$$

Since $\eta\overline{T}_{\mathsf{g}}\, \mathsf{g}_2[k] + n_k$ is Gaussian, each term of the right-hand side of the above inequality is 0, which further implies that $\mathbb{P}\left(\beta\bar{y}_k \in \delta D_{s_k} \mid s_k = s^{(m)}\right) = 0$. This completes our proof.

Appendix E
Equivalence of Maximizing the Asymptotic SINR and Minimizing the Asymptotic SEP

For the scalar asymptotic model (20) with precoding factor β in (23), the received signal is

$$\beta\bar{y} = s + \frac{\overline{T}_{\mathsf{s}}^\dagger}{\eta|\overline{T}_{\mathsf{s}}|^2}(\eta\overline{T}_{\mathsf{g}}\, \mathsf{g} + n) \triangleq s + \bar{n}, \tag{76}$$

where $\bar{n} \sim \mathcal{CN}\left(0, \frac{\eta^2\overline{T}_{\mathsf{g}}^2 + \sigma^2}{\eta^2|\overline{T}_{\mathsf{s}}|^2}\right)$ since $\mathsf{g} \sim \mathcal{CN}(0,1)$ and $n \sim \mathcal{CN}(0,\sigma^2)$ are independent. Following (73) in Appendix D, we can express the asymptotic SEP as

$$\overline{\mathrm{SEP}} = 1 - \frac{1}{M}\sum_{m=1}^{M} \mathbb{P}\left(\beta\bar{y} \in D_s \mid s = s^{(m)}\right).$$

Each probability in the above summation can further be expressed as

$$\begin{aligned}\mathbb{P}\left(\beta\bar{y} \in D_s \mid s = s^{(m)}\right) &= \mathbb{P}\left(\bar{n} \in D_{s^{(m)}} - s^{(m)}\right) \\ &= \mathbb{P}\left(\mathsf{Z} \in \frac{\overline{\mathrm{SINR}}}{\sigma_s^2}\left(D_{s^{(m)}} - s^{(m)}\right)\right), \end{aligned}$$

where we have rewritten $\bar{n}$ as $\bar{n} = \frac{\sigma_s^2}{\overline{\mathrm{SINR}}}\mathsf{Z}$ with $\mathsf{Z} \sim \mathcal{CN}(0,1)$. From (75), it is easy to check that $D_{s^{(m)}} - s^{(m)}$ is a polyhedron containing 0, and thus the above probability is increasing in $\overline{\mathrm{SINR}}$, i.e., $\overline{\mathrm{SEP}}$ is a decreasing function of $\overline{\mathrm{SINR}}$.

APPENDIX F
PROOF OF THEOREM 4

In this section, we give the proof of Theorem 4, i.e., we prove that $(\bar{\alpha}^*, \eta^*, f^*)$ given in (30) is the optimal solution to the following problem:

$$\begin{aligned}\zeta^* := \sup_{f,\eta>0,\bar{\alpha}>0} \quad & \frac{\mathbb{E}^2[d\,f(d)]}{\mathrm{var}[d\,f(d)] + \phi(\bar{\alpha},\eta)\,\frac{\mathbb{E}[f^2(d)]}{\gamma}} \\ \text{s.t.} \quad & \eta^2\,\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] \le P_T, \\ & \mathbb{E}[f^2(d)] = \frac{\gamma}{\sigma_s^2}\,\bar{\alpha}^2. \end{aligned} \tag{77}$$

We first note that $\phi^* = \infty$ (see the definition in (30a)) is a pathological case where the SINR in (77) is identically zero[4] and any $(\bar{\alpha}, \eta, f)$ is optimal. This happens, e.g., when $q(\cdot)$ is a constant function. In the rest of the proof, we assume $\phi^* < \infty$.

It is convenient to work with the inverse of the objective function in (77) and consider the following equivalent problem:

$$\begin{aligned}\Phi^* := \inf_{\eta>0,\bar{\alpha}>0,f} \quad & \frac{\mathbb{E}[d^2 f^2(d)] + \frac{\phi(\bar{\alpha},\eta)}{\gamma}\,\mathbb{E}[f^2(d)]}{\mathbb{E}^2[d\,f(d)]} - 1 \\ \text{s.t.} \quad & \eta^2\,\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] \le P_T, \\ & \mathbb{E}[f^2(d)] = \frac{\gamma}{\sigma_s^2}\,\bar{\alpha}^2. \end{aligned} \tag{78}$$

We first ignore the constraints. By the Cauchy-Swarchz inequality, we can upper bound the denominator of the objective function of (78) by

$$\mathbb{E}^2[d\,f(d)] \le \mathbb{E}\left[\left(d^2 + \frac{\phi(\bar{\alpha},\eta)}{\gamma}\right) f^2(d)\right] \mathbb{E}\left[\frac{d^2}{d^2 + \frac{\phi(\bar{\alpha},\eta)}{\gamma}}\right],$$

and thus

$$\begin{aligned}\frac{\mathbb{E}[d^2 f^2(d)] + \frac{\phi(\bar{\alpha},\eta)}{\gamma}\,\mathbb{E}[f^2(d)]}{\mathbb{E}^2[d\,f(d)]} &= \frac{\mathbb{E}\left[\left(d^2 + \frac{\phi(\bar{\alpha},\eta)}{\gamma}\right) f^2(d)\right]}{\mathbb{E}^2[d\,f(d)]} \\ &\ge \frac{1}{\mathbb{E}\left[\frac{d^2}{d^2 + \phi(\bar{\alpha},\eta)/\gamma}\right]},\end{aligned}$$

where the inequality holds with equality when

$$f(x) := \frac{x}{\tau\left(x^2 + \frac{\phi(\bar{\alpha},\eta)}{\gamma}\right)}, \quad \forall\,\tau \ne 0.$$

By the above inequality, the constrained infimum in (78) is further lower bounded by

$$\begin{aligned}\Phi^* \ge \inf_{\eta>0,\bar{\alpha}>0} \quad & \frac{1}{\mathbb{E}\left[\frac{d^2}{d^2 + \phi(\bar{\alpha},\eta)/\gamma}\right]} - 1 \\ \text{s.t.} \quad & \eta^2\,\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] \le P_T, \\ & \mathbb{E}[f^2(d)] = \frac{\gamma}{\sigma_s^2}\,\bar{\alpha}^2. \end{aligned} \tag{79}$$

Note that the objective function of (79) is independent of $f(\cdot)$, and thus the second constraint can be removed. In addition, $\mathbb{E}[d^2/(d^2 + x)]$ is decreasing in $x \in [0, \infty)$ and it is straightforward to check that $\phi(\bar{\alpha}, \eta) \ge 0$ (see (22)). Hence, the optimization problem in the right-hand side of (79) is equivalent to

$$\begin{aligned}\inf_{\eta>0,\bar{\alpha}>0} \quad & \phi(\bar{\alpha},\eta) \\ \text{s.t.} \quad & \eta^2\,\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2] \le P_T. \end{aligned} \tag{80}$$

Since $\phi(\bar{\alpha}, \eta)$ is decreasing in η (see (22)), the power constraint in (80) must be satisfied with equality at the infimum, namely,

$$\eta = \sqrt{\frac{P_T}{\mathbb{E}[|q(\bar{\alpha}\mathsf{Z})|^2]}}. \tag{81}$$

Substituting (81) into (80) yields the optimization problem in (30a), whose optimal value is given by ϕ^*. Hence,

$$\zeta^* = \frac{1}{\Phi^*} \le \frac{1}{1 - \mathbb{E}\left[\frac{d^2}{d^2 + \frac{\phi^*}{\gamma}}\right]} - 1.$$

The proof is complete by further verifying that $(\bar{\alpha}^*, \eta^*, f^*)$ is feasible for (78) and Φ^* is attained at $(\bar{\alpha}^*, \eta^*, f^*)$.

[4]This is not the case for quantization functions used in practice.

ACKNOWLEDGMENT

The authors would like to thank Prof. Yue M. Lu from Harvard University, Dr. Mingjie Shao from Shandong University, and the anonymous reviewers for their helpful comments on the article.

REFERENCES

[1] Z. Wu, J. Ma, Y.-F. Liu, and A. L. Swindlehurst, "Linear one-bit precoding in massive MIMO: Asymptotic SEP analysis and optimization," in *Proc. IEEE 24th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2023, pp. 1–5.

[2] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] B. Razavi, *Principles of Data Conversion System Design*, 1st ed. Hoboken, NJ, USA: Wiley, 1994.

[6] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

[7] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *Proc. 16th IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, Dec. 2009, pp. 683–686.

[8] A. Kakkavas, J. Munir, A. Mezghani, H. Brunner, and J. A. Nossek, "Weighted sum rate maximization for multi-user MISO systems with low resolution digital to analog converters," in *Proc. 20th Int. ITG Workshop Smart Antennas (WSA)*, Mar. 2016, pp. 1–8.

[9] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1595–1609, Mar. 2019.

[10] A. Haqiqatnejad, F. Kayhan, S. ShahbazPanahi, and B. Ottersten, "Finite-alphabet symbol-level multiuser precoding for massive MU-MIMO downlink," *IEEE Trans. Signal Process.*, vol. 69, pp. 5595–5610, 2021.

[11] C. G. Tsinos, A. Kalantari, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding with low resolution DACs for large-scale array MU-MIMO systems," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[12] C.-E. Chen, "A fast quantized precoding algorithm for massive MU-MISO systems with low-resolution DACs and PSK signaling: A safety margin fitting approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11267–11271, Oct. 2022.

[13] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA, USA: Artech House, 2006.

[14] S. K. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, Nov. 2012.

[15] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[16] J. Pan and W.-K. Ma, "Constant envelope precoding for single-user large-scale MISO channels: Efficient precoding and optimal designs," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 982–995, Oct. 2014.

[17] Z. Wu, J. Wu, W.-K. Chen, and Y.-F. Liu, "Diversity order analysis for quantized constant envelope transmission," *IEEE Open J. Signal Process.*, vol. 4, pp. 21–30, 2023.

[18] M. Kazemi, H. Aghaeinia, and T. M. Duman, "Discrete-phase constant envelope precoding for massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2011–2021, May 2017.

[19] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Finite-alphabet precoding for massive MU-MIMO with low-resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4706–4720, Jul. 2018.

[20] J.-C. Chen, "Efficient constant envelope precoding with quantized phases for massive MU-MIMO downlink systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4059–4063, Apr. 2019.

[21] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Quantized constant envelope precoding with PSK and QAM signaling," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8022–8034, Dec. 2018.

[22] M. Shao, Q. Li, W.-K. Ma, and A. M. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, Oct. 2019.

[23] Z. Wu, Y.-F. Liu, B. Jiang, and Y.-H. Dai, "Efficient quantized constant envelope precoding for multiuser downlink massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[24] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 3381–3385.

[25] O. De Candido, H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Reconsidering linear transmit signal processing in 1-bit quantized multi-user MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 254–267, Jan. 2019.

[26] Y. Li, C. Tao, A. Lee Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.

[27] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.

[28] A. K. Saxena et al., "Asymptotic performance of downlink massive MIMO with 1-bit quantized zero-forcing precoding," in *Proc. 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[29] A. K. Saxena, A. Mezghani, and R. W. Heath, "Linear CE and 1-bit quantized precoding with optimized dithering," *IEEE Open J. Signal Process.*, vol. 1, pp. 310–325, 2020.

[30] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit massive MU-MIMO precoding in VLSI," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 508–522, Dec. 2017.

[31] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7559–7575, Nov. 2018.

[32] A. Li, F. Liu, C. Masouros, Y. Li, and B. Vucetic, "Interference exploitation 1-bit massive MIMO precoding: A partial branch-and-bound solution with near-optimal performance," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3474–3489, May 2020.

[33] F. Sohrabi, Y.-F. Liu, and W. Yu, "One-bit precoding and constellation range design for massive MIMO with QAM signaling," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 557–570, Jun. 2018.

[34] Z. Wu, B. Jiang, Y.-F. Liu, M. Shao, and Y.-H. Dai, "Efficient CI-based one-bit precoding for multiuser downlink massive MIMO systems with PSK modulation," *IEEE Trans. Wireless Commun.*, to be published. [Online]. Available: https://arxiv.org/abs/2110.11628

[35] X. Ma, A. Kammoun, A. M. Alrashdi, T. Ballal, M.-S. Alouini, and T. Y. Al-Naffouri, "Asymptotic analysis of RLS-based digital precoder with limited PAPR in massive MIMO," *IEEE Trans. Signal Process.*, vol. 70, pp. 5488–5503, 2022.

[36] H. Jedda and J. A. Nossek, "On the statistical properties of constant envelope quantizers," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 1006–1009, Dec. 2018.

[37] J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," MIT Res. Lab. Electron., Cambridge, MA, USA, Tech. Rep. 216, vol. 216, 1952.

[38] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[39] Y. M. Lu, "Householder dice: A matrix-free algorithm for simulating dynamics on Gaussian and random orthogonal ensembles," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, pp. 8264–8272, Dec. 2021.

[40] P. Hendriks, "Specifying communications DACs," *IEEE Spectr.*, vol. 34, no. 7, pp. 58–69, Jul. 1997.

[41] M. Gustavsson, J. J. Wikner, and N. Tan, *CMOS Data Converters for Communications*. New York, NY, USA: Springer, 2000.

[42] R. J. van de Plassche, *CMOS Integrated Analog-to-Digital and Digitalto-Analog Converters*. Norwell, MA, USA: Kluwer Academic, 2003.

[43] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[44] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.

[45] R. Couillet, J. Hoydis, and M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6392–6425, Oct. 2012.

[46] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[47] E. S. Meckes, *The Random Matrix Theory of the Classical Compact Groups*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[48] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

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

[50] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[51] A. Khalili, E. Erkip, and S. Rangan, "Quantized MIMO: Channel capacity and spectrospatial power distribution," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2303–2308.

[52] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008.

[53] D. Hui and D. L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 957–977, Mar. 2001.

[54] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-6, no. 1, pp. 7–12, Mar. 1960.

[55] K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*. New York, NY, USA: Springer, 2006.

[56] T. Ferguson, *A Course in Large Sample Theory*. Evanston, IL, USA: Routledge, 2017.

[57] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2011.

[58] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

Zheyu Wu (Graduate Student Member, IEEE) received the B.Sc. degree in applied mathematics from Jilin University, Changchun, China, in 2019. She is currently pursuing the Ph.D. degree with the Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. Her current research interests include optimization algorithm and its applications to signal processing and wireless communications.

Junjie Ma received the B.E. degree from Xidian University, Xi'an, China, in 2010, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2015. He was a Post-Doctoral Researcher with the City University of Hong Kong from 2015 to 2016, Columbia University from 2016 to 2019, and Harvard University from 2019 to 2020. He has been an Assistant Professor with the Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, since July 2020. His current research interests include message passing algorithms and their applications for high-dimensional signal processing.

Ya-Feng Liu (Senior Member, IEEE) received the B.Sc. degree in applied mathematics from Xidian University, Xi'an, China, in 2007, and the Ph.D. degree in computational mathematics from the Chinese Academy of Sciences (CAS), Beijing, China, in 2012. During the Ph.D. study, he was supported by the Academy of Mathematics and Systems Science (AMSS), CAS, to visit the Prof. Zhi-Quan (Tom) Luo with the University of Minnesota (Twins Cities) from 2011 to 2012. After his graduation, he joined the Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, CAS, in 2012, where he became an Associate Professor in 2018. His current research interests include nonlinear optimization and its applications to signal processing, wireless communications, and machine learning. He is an elected member of the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC) of the IEEE Signal Processing Society from 2020 to 2022 and from 2023 to 2025. He received the Best Paper Award from the IEEE International Conference on Communications (ICC) in 2011, the Chen Jingrun Star Award from the AMSS in 2018, the Science and Technology Award for Young Scholars from the Operations Research Society of China in 2018, the 15th IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award in 2020, and the Science and Technology Award for Young Scholars from the China Society for Industrial and Applied Mathematics in 2022. Students supervised and co-supervised by him won the Best Student Paper Award from the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), in 2015, and the Best Student Paper Award of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), in 2022. He currently serves as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and *Journal of Global Optimization* and a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2019 to 2022 and an Associate Editor for IEEE SIGNAL PROCESSING LETTERS from 2019 to 2023.

A. Lee Swindlehurst (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Brigham Young University (BYU) in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University in 1991. He was with the Department of Electrical and Computer Engineering, BYU, from 1990 to 2007, where he was the Department Chair from 2003 to 2006. From 1996 to 1997, he was a Visiting Scholar with Uppsala University and the Royal Institute of Technology, Sweden. From 2006 to 2007, he was on leave working as the Vice President of Research for ArrayComm LLC, San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science (EECS) Department, University of California at Irvine, where he worked as the Associate Dean of Research and Graduate Studies with the Samueli School of Engineering from 2013 to 2016. He is currently the Chair of the EECS Department. From 2014 to 2017, he was a Hans Fischer Senior Fellow with the Institute for Advanced Studies, Technical University of Munich. His current research interests include array signal processing for radar, wireless communications, and biomedical applications. He has more than 400 publications in these areas. In 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). He received the 2000 IEEE W. R. G. Baker Prize Paper Award; the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory; the 2006, 2010, and 2021 IEEE Signal Processing Society's best paper awards; the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award; the Best Paper Award at the 2020 IEEE International Conference on Communications; and the 2022 Claude Shannon-Harry Nyquist Technical Achievement Award from the IEEE Signal Processing Society. He was the Inaugural Editor-in-Chief of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.