Quantized MIMO: Channel Capacity and Spectrospatial Power Distribution

Abbas Khalili, Elza Erkip, Sundeep Rangan

Dept. of Electrical and Computer Engineering, NYU Tandon School of Engineering.

Abstract-Millimeter wave systems suffer from high power consumption and are constrained to use low resolution quantizers -digital to analog and analog to digital converters (DACs and ADCs). However, low resolution quantization leads to reduced data rate and increased out-of-band emission noise. In this paper, a multiple-input multiple-output (MIMO) system with linear transceivers using low resolution DACs and ADCs is considered. An information-theoretic analysis of the system to model the effect of quantization on spectrospatial power distribution and capacity of the system is provided. More precisely, it is shown that the impact of quantization can be accurately described via a linear model with additive independent Gaussian noise. This model in turn leads to simple and intuitive expressions for spectrospatial power distribution of the transmitter and a lower bound on the achievable rate of the system. Furthermore, the derived model is validated through simulations and numerical evaluations, where it is shown to accurately predict both spectral and spatial power distributions.

I. INTRODUCTION

Digital to analog and analog to digital converters (DACs and ADCs) are essential components of any digital communication system. While sub-6GHz systems use high resolution DACs and ADCs, millimeter wave (mmWave) rely on communication across wide bandwidths with large numbers of antennas thereby making high resolution DACs and ADCs very costly in terms of the power consumption [1], [2]. Use of low resolution DACs and ADCs (typically 3-4 bits in I and Q) have been suggested as energy-efficient approaches for next-generation mmWave systems [3]–[21]

Unlike its high resolution counterpart, low resolution quantization leads to high quantization noise reducing the achievable rate of the system and adding out-of-band (OOB) emission noise which results in high adjacent carrier leakage ratio (ACLR). Therefore, it is of great importance to model this quantization noise and understand its impacts accurately.

There is a large body of work on low resolution DACs and ADCs [3]–[24]. The most common model used to approximate the effect of quantization is the additive Gaussian noise (AGN) model [25], [26]. This model has been rigorously analyzed in several works in the high rate quantization regime or under dithered quantization [25], [27]–[30]. For low resolution mmWave systems, it is observed that the AGN and other Gaussian noise predictions match simulations and therefore provide an accurate model of quantization; however, the accuracy is not rigorously justified [15]–[21].

In [31], we considered a single-input single-output (SISO) communication system with a linear transceiver and low resolution quantization, and using information theoretic tools, proved the validity of the AGN model. In this paper, we extend the analysis of [31] to multiple-input multiple-output (MIMO) communications. More precisely, we consider a MIMO system using hybrid transceivers with low resolution DACs and ADCs as illustrated in Fig. 1. The transmitter modulates a set of transmit streams through a unitary transform \mathbf{V}^{H} prior to the DACs. We model the spectrum of the transmitted signal through the transform $\mathbf{r} = \mathbf{V}\mathbf{x}$, where \mathbf{x} is the output of the transmitter. Note that if we set the unitary matrix V to an FFT-matrix, we would have a MIMO-OFDM system and can capture filtering and spectral constraints which arise form practical constraint. The MIMO extension of [31] presented here requires new results on empirical convergence and mutual information of two random matrices. In addition, it allows us to study spatial distribution of power as argued below. A summary of our contributions are as follows:

- *Rigorous linear model:* We analyse a certain large random limit of the system where $\mathbf{V} \in \mathbb{C}^{N \times N}$ is selected uniformly from $N \times N$ unitary matrices and provide a rigorous framework for a linear model of quantization. This generalizes our result in [31] which considers only SISO systems (Sec. III).
- *Rate and spectrospatial power distribution:* We use the derived linear model and provide a simple asymptotically exact expression for per sample spectral covariance matrix of the system in Fig. 1. Such covariance matrices can be computed at both transmitter or receiver leading to spectrospatial power distributions and lower bounds on the system's capacity (Sec. IV). Furthermore, through simulations and numerical evaluations we show that the derived linear model accurately predicts ACLR and spatial power distribution (Sec. V).

Notation: x is scalar, x is a column vector whose i^{th} element is x_i , and X is a matrix whose $(i, j)^{\text{th}}$ element is x_{ij} . We denote the column i of X with $\mathbf{x}_{:i}$ and with a slight abuse of notation, we denote the transpose of row j with $\mathbf{x}_{j:}$. Also, X is a random column vector and [N] is the set $\{1, 2, \ldots N\}$.

II. SYSTEM MODEL AND PRELIMINARIES

A. Transceiver with Linear Transform

We consider the general transceiver structure with linear transform modulation and analog precoder and combiner

The work supported in part by NSF grants 1952180, 1925079, 1564142, 1547332, SRC, and the industrial affiliates of NYU WIRELESS.



Fig. 1: System model with linear modulation, demodulation and quantization at both the transmitter and receiver. The linear modulation is modeled as a multiplication by V^{H} prior to quantization at the transmitter, while a spectrum analyzer and receiver employ the inverse transform V.

shown in Fig. 1. In this system, the transmitter generates N_{st} streams of length N denoted by matrix \mathbf{Z} , whose columns $\mathbf{z}_{:i} = [z_{1i}, z_{2i}, \cdots, z_{Ni}]^{\mathsf{T}} \in \mathbb{C}^{N}, i \in [N_{st}]$ are linearly modulated as $\mathbf{u}_{:i} = \mathbf{V}^{\mathsf{H}} \mathbf{z}_{:i}$, where $\mathbf{V} \in \mathbb{C}^{N \times N}$ is a unitary matrix. The modulated signals are converted into analog using a set of DACs $\mathbf{Q}_{tx}(\cdot)$, and then passed through a linear analog precoder determined by the matrix $\mathbf{W}_{ ext{tx}} \in \mathbb{C}^{N_t imes N_{st}}$ and transmitted. If V were an FFT matrix, we could consider the symbols $z_{ni}, i \in [N_{st}], n \in [N]$ as the values of the information carrying signal in frequency domain and u_{ni} , $i \in$ $[N_{st}], n \in [N]$ the digital values in time-domain. The modulation can thus be regarded as a simplified version of the OFDM (where we ignore the cyclic prefix). In addition, if we zero-pad the input frequency-domain symbols $z_{ni}, i \in [N_{st}], n \in [N]$, the transformed vector $\mathbf{u}_{:i} = \mathbf{V}^{\mathsf{H}} \mathbf{z}_{:i}$ can be seen as a linearly up-sampled version of \mathbf{z}_{ii} . The transmitted signal goes through a general MIMO channel of the form

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}, \boldsymbol{\Xi}),\tag{1}$$

where $\mathbf{X} = [\mathbf{x}_{:1}, \mathbf{x}_{:2}, \cdots, \mathbf{x}_{:N_t}], \mathbf{Y} = [\mathbf{y}_{:1}, \mathbf{y}_{:2}, \cdots, \mathbf{y}_{:N_r}], \mathbf{\Xi} \in \mathbb{C}^{N \times N_r}$ with $\xi_{ni} \sim \mathcal{CN}(0, 1)$, and $\mathbf{F}(\cdot)$ is a row-wise function. This channel can also model certain non-linearites in the RF front-end [5]. For the case of MIMO AWGN channel,

$$\mathbf{y}_{n:} = \mathbf{H}\mathbf{x}_{n:} + \boldsymbol{\xi}_{n:}, n \in [N], \tag{2}$$

where $\mathbf{y}_{n:}^{\mathsf{T}}, \mathbf{x}_{n:}^{\mathsf{T}}$, and $\boldsymbol{\xi}_{n:}^{\mathsf{T}}$ are the n^{th} rows of matrices \mathbf{Y}, \mathbf{X} , and $\boldsymbol{\Xi}$, respectively. At the receiver, the channel output is first passed through a linear analog combiner $\mathbf{W}_{\text{rx}} \in \mathbb{C}^{N_{sr} \times N_r}$, and quantized using a set of ADCs $\mathbf{Q}_{\text{rx}}(\cdot)$, and then the receiver performs the inverse transform operation to obtain $\widehat{\mathbf{z}}_{:j} = \mathbf{V}\widehat{\mathbf{u}}_{:j}, j \in [N_{sr}]$. Let $\widehat{\mathbf{Z}} = [\widehat{\mathbf{z}}_{:1}, \widehat{\mathbf{z}}_{:2}, \cdots, \widehat{\mathbf{z}}_{:N_{sr}}]$ using (1), we have

$$\widehat{\mathbf{Z}} = \mathbf{VG}(\mathbf{V}^{\mathsf{H}}\mathbf{Z}, \boldsymbol{\Xi}),$$
 (3a)

$$\mathbf{G}(\mathbf{U}, \mathbf{\Xi}) = \mathbf{Q}_{\mathrm{rx}} \left(\mathbf{F} \left(\mathbf{Q}_{\mathrm{tx}} \left(\mathbf{U} \right) \mathbf{W}_{\mathrm{tx}}^{\mathsf{T}}, \mathbf{\Xi} \right) \right) \mathbf{W}_{\mathrm{rx}}^{\mathsf{T}}.$$
 (3b)

B. Spectral Covariance and Spectrospatial Power

To derive per sample spectral covariance matrix at the transmitter, consider $\mathbf{r}_{:k} = \mathbf{V}\mathbf{x}_{:k}, k \in [N_t]$ which is the transform of the transmitted signal from the k^{th} antenna. The

component $|r_{nk}|^2, n \in [N]$ can be regarded as the energy of the transmit signal from k^{th} antenna at frequency n.

We assume the frequency is divided into M sub-bands. Let $\ell_n \in [M]$ be the variable that indicates which sub-band frequency n of the transmitted signals (rows of matrix **X**) belong to. We call $\ell = (\ell_1, \ldots, \ell_N)$ the *sub-band selection vector* and define,

$$\delta_m(\boldsymbol{\ell}) := \frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\{\ell_n = m\}},\tag{4}$$

which represents the fraction of the frequency components in sub-band m. We will call δ_m the bandwidth fraction for sub-band m. Define $\mathbf{R} = [\mathbf{r}_{:1}, \mathbf{r}_{:2}, \cdots, \mathbf{r}_{:N_t}] = [\mathbf{r}_{1:}, \mathbf{r}_{2:}, \cdots, \mathbf{r}_{N:t}]^{\mathsf{T}}$. The per sample covariance matrix in subband m is

$$\mathbf{S}_{m}(N) := \frac{1}{N} \sum_{n=1}^{N} \mathbf{r}_{n:} \mathbf{r}_{n:}^{\mathsf{H}} \mathbb{1}_{\{l_{n}=m\}},$$
(5)

As a result, for sub-band m, the fraction of spectral power, ν_m , and spatial (beamforming) power towards angle $\psi \in (0, 2\pi]$, BF_m(ψ), respectively, are

$$\nu_m = \frac{\operatorname{Tr} \left(\mathbf{S}_m(N) \right)}{\operatorname{Tr} \left(\mathbf{S}(N) \right)}, \quad \mathbf{S}(N) = \sum_m \mathbf{S}_m(N), \text{ and } (6)$$

$$BF_m(\psi) = \mathbf{e}(\psi)^{\mathsf{H}} \mathbf{S}_m(N) \mathbf{e}(\psi), \qquad (7)$$

where $\mathbf{e}(\psi)$ is the transmit array response. For the case of uniform linear array (ULA) with half wavelength spacing $\mathbf{e}(\psi) = [1, e^{i\pi\cos(\psi)}, \dots, e^{(N_t-1)i\pi\cos(\psi)}]^{\mathsf{T}}$. Note that the analysis can also be applied to two dimensional beamforming.

C. Achievable Rate

An achievable rate for the system can be computed by fixing the distributions of $\mathbf{z}_{:i}, i \in [N_{st}]$ and calculating the mutual information $I(\mathbf{Z}; \widehat{\mathbf{Z}})$ between the transmitted and received frequency-domain matrices $\mathbf{Z}, \widehat{\mathbf{Z}}$. For the input distribution, we will use an independent complex Gaussian in each frequency and stream. Specifically, we will assume the components $\mathbf{z}_{n:}, n \in [N]$ ($\mathbf{z}_{n:}^{\mathsf{T}}$ is the n^{th} row of matrix \mathbf{Z}) are i.i.d.

$$\mathbf{z}_{n:} \sim C\mathcal{N}(\mathbf{0}, \mathbf{P}_m), \text{ when } \ell_n = m,$$
 (8)

where \mathbf{P}_m is the covariance matrix of the components in subband m. As a result, the average per symbol (row) covariance matrix for \mathbf{Z} is

$$\mathbf{P} = \frac{1}{N} \mathbb{E} \left[\mathbf{Z}^{\mathsf{T}} \mathbf{Z}^{*} \right] = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E} \left[\mathbf{z}_{n:} \mathbf{z}_{n:}^{\mathsf{H}} \right] = \sum_{m=1}^{M} \delta_{m} \mathbf{P}_{m}, \quad (9)$$

where δ_m s are the bandwidth fractions (4). We note that using the Gaussian input distribution is not necessarily optimal in systems with quantization. Finding the optimal input distribution is left for future work.

D. Assumptions

For tractability of the analysis, we consider a certain large system limit of random instances of the system indexed by the dimension N with $N \to \infty$. For each N, We consider $\mathbf{V} = \mathbf{V}(N)$ is a random unitary matrix \mathbf{V} that is uniformly distributed on the $N \times N$ unitary matrices (i.e. Haar distributed) instead of considering the deterministic FFT matrix. We assume that the sub-band selection vectors $\boldsymbol{\ell} = \boldsymbol{\ell}(N)$ are fixed sequence satisfying,

$$\lim_{N \to \infty} \frac{1}{N} |\{\ell_n(N) = m\}| = \delta_m.$$
(10)

This condition ensures that a fraction δ_m of the components are in sub-band m.

We consider that the channel $\mathbf{F}(\cdot)$ is Lipschitz continuous and acts as a *row-wise separable* function. More specifically,

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}, \mathbf{\Xi}) \Longleftrightarrow \mathbf{y}_{n:} = F(\mathbf{x}_{n:}, \boldsymbol{\xi}_{n:}), \quad (11)$$

where $F(\cdot)$ is a Lipschitz vector-input, vector-output function. Note that these conditions are satisfied for the MIMO AWGN channel in (2) as it is performs the same operation on each row of the pair $(\mathbf{X}, \boldsymbol{\xi})$ and has the Lipschitz constant equal to the maximum singular value of **H**. Similarly, we assume that the DAC and ADC functions, $\mathbf{Q}_{\mathbf{x}}(\mathbf{u}), \mathbf{x} \in \{\mathrm{tx}, \mathrm{rx}\}$ are Lipschitz continuous and *component-wise separable* for some scalarinput, scalar-output function $Q_{\mathbf{x}}(\cdot), \mathbf{x} \in \{\mathrm{tx}, \mathrm{rx}\}$, respectively. We also assume that these functions are deterministic and do not change over time. We note that typical quantizers are not Lipschitz continuous. However, we assume they can be approximated arbitrarily closely by a Lipschitz function. Through simulations, we will show that the predictions hold true even for standard quantizers.

For our proofs, we use results on empirical convergence. The analysis framework was developed by Bayati and Montanari [32] and also used in the vector approximate massage passing (VAMP) analysis [33]. In next section, we will provide a definition of empirical convergence along with the necessary result for our proofs.

III. EMPIRICAL CONVERGENCE OF RANDOM VECTORS

For a given $p \ge 1$, a mapping $\Phi : \mathbb{C}^{d_i} \to \mathbb{C}^{d_o}$ is called *pseudo-Lipschitz* of order p if

$$\|\Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_2)\| \le C \|\mathbf{x}_1 - \mathbf{x}_2\| \left(1 + \|\mathbf{x}_1\|^{p-1} + \|\mathbf{x}_2\|^{p-1}\right),$$

for some constant C > 0. Note that when p = 1, we obtain the standard definition of Lipschitz continuity. Now, suppose that for each N, $\mathbf{X}(N)$ is a matrix $\mathbf{X}(N) = [\mathbf{x}_{1:}, \dots, \mathbf{x}_{N:}]^{\mathsf{T}}$ with rows $\mathbf{x}_{n:}^{\mathsf{T}} \in \mathbb{C}^{1 \times d_i}$ for some fixed dimension d_i . Let $X \in \mathbb{C}^{d_i}$ be a random vector. We say that the rows of $\mathbf{X}(N)$ converge empirically to X^{T} with p-th order moments if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \Phi(\mathbf{x}_{n:}(N)) = \mathbb{E}\left[\Phi(X)\right], \quad (12)$$

for all pseduo-Lipschitz functions of order p. Loosely speaking, the condition requires that the empirical distribution of the rows of $\mathbf{X}(N)$ converge to that of the random vector X^{T} which is satisfied when $\mathbf{x}_{n:s}$ are i.i.d. with distribution X. We will denote this by

$$\lim_{N \to \infty} \{\mathbf{x}_{n:}\} \stackrel{PL(p)}{=} X.$$
(13)

Next proposition describes the distribution of the matrices under random unitary transform and generalizes [31, Prop. 1] which only considers the vector case. Let us consider a sequence of systems indexed by N, and for each N suppose that $\mathbf{V} \in \mathbb{C}^{N \times N}$ is uniformly distributed on the unitary matrices. Let $\mathbf{X} = \mathbf{X}(N)$ be a sequence of matrices with

$$\lim_{N \to \infty} \{\mathbf{x}_{n:}\} \stackrel{PL(2)}{=} X.$$
(14)

Proposition 1. Define $\mathbf{U} = \mathbf{VX}$. Given (14) and Haar distributed matrix \mathbf{V} , the sequence $\{\mathbf{u}_{n:}\}$ converge empirically to random vector $U \sim C\mathcal{N}(\mathbf{0}, \mathbf{P})$, with $\mathbf{P} = \mathbb{E}[XX^{\mathsf{H}}]$.

This proposition shows that random unitary transformation effectively creates a Gaussian distribution in the sense that for input matrices whose rows converge empirically, the rows of the output matrix converge empirically to a Gaussian distribution. Now, consider a matrix \mathbf{Y} generated by,

$$\mathbf{Y} = \mathbf{V} \boldsymbol{\Phi} (\mathbf{V}^{\mathsf{H}} \mathbf{X}, \boldsymbol{\Xi}), \tag{15}$$

where $\Phi(\cdot)$ is some function that operates row-wise in that

$$\mathbf{Y} = \mathbf{\Phi}(\mathbf{U}, \mathbf{\Xi}) \Longleftrightarrow \mathbf{y}_{n:} = \Phi(\mathbf{u}_{n:}, \boldsymbol{\xi}_{n:}), \quad (16)$$

where $\Phi(\cdot)$ is a pseudo-Lipschitz vector-input, vector-output function. Assume that rows of Ξ also converge empirically in that $\lim_{N\to\infty} \{\xi_n\} \stackrel{PL(2)}{=} \Xi$, for some random vector Ξ . To analyze the statistics on \mathbf{Y} , we define the quantities:

$$\mathbf{A} := \mathbb{E}\left[\frac{\partial \Phi(U, \Xi)}{\partial U}\right],\tag{17a}$$

$$\mathbf{T} := \mathbb{E}\left[(\Phi(U, \Xi) - \mathbf{A}U) (\Phi(U, \Xi) - \mathbf{A}U)^{\mathsf{H}} \right].$$
(17b)

where $U \sim C\mathcal{N}(\mathbf{0}, \mathbf{P})$, $\mathbf{P} := \mathbb{E}[XX^{\mathsf{H}}]$, and $\partial \Phi(U,\Xi)/\partial U$ denotes the Jacobian of $\Phi(\cdot)$ with respect to the vector U. To calculate the matrix \mathbf{A} one can use the multivariate extension of Stein's lemma provided in [34] i.e.,

$$\mathbf{PA}^{\mathsf{H}} = \mathbb{E}\left[U\Phi(U,\Xi)^{\mathsf{H}}\right].$$
 (18)

Proposition 2. Given the relation in (15) and the conditions (14) and (16), the rows of the matrix pair (\mathbf{Y}, \mathbf{X}) converge empirically as,

$$\lim_{N \to \infty} \{ (\mathbf{y}_{n:}, \mathbf{x}_{n:}) \} \stackrel{PL(2)}{=} (Y, X), \tag{19}$$

and

$$Y = \mathbf{A}X + \Theta, \quad \Theta \sim C\mathcal{N}(\mathbf{0}, \mathbf{T}), \tag{20}$$

with Θ independent of X.

Proof. See Appendix B.

The model (20) shows that transformation of **X** to produce Y results in a linearly scaled X plus Gaussian noise. The scaling matrix A and Gaussian noise covariance matrix T can be computed from the distributions of the components as shown in (17). By substituting the quantization function $Q_{tx}(\cdot)$ in the place of the function $\Phi(\cdot)$, we observe that the timedomain quantization effectively scales the frequency signal z and adds an independent Gaussian noise.

IV. ACHIEVABLE SPECTRAL ENERGY AND RATE

A. Spectral Covariance

We first compute the asymptotic per sample spectral covariance matrix shown in (5). Given the assumptions in Sec. II-D, calculate \mathbf{A}_{tx} and \mathbf{T}_{tx} using (17) with $\Phi(U) = \mathbf{W}_{tx}Q_{tx}(U)$, where $U \sim C\mathcal{N}(\mathbf{0}, \mathbf{P}), \ \mathbf{P} = \sum_m \delta_m \mathbf{P}_m.$

Theorem 1. Let $\mathbf{R} = \mathbf{V}\mathbf{X}$ be the frequency-domain representation of the transmitted signal **X**. Then the covariance matrix in each sub-band converges almost surely to

$$\mathbf{S}_m := \lim_{N \to \infty} \mathbf{S}_m(N) = \delta_m \left[\mathbf{A}_{\mathrm{tx}} \mathbf{P}_m \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}} \right].$$
(21)

In particular, the total covariance matrix per sample converges almost surely as,

$$\mathbf{S} := \lim_{N \to \infty} \frac{1}{N} \mathbf{X}^{\mathsf{T}} \mathbf{X}^{\star} = \mathbf{A}_{\mathrm{tx}} \mathbf{P} \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}}$$
(22)

Proof. See Appendix C.

Based on Thm. 1, spectrospatial power distribution of the transmitter can be computed using (6) and (7). Based on (6), the fraction of power in sub-band m is

$$\nu_m := \frac{\operatorname{Tr}(\mathbf{S}_m)}{\operatorname{Tr}(\mathbf{S})} = \frac{\delta_m \operatorname{Tr} \left(\mathbf{A}_{\mathrm{tx}} \mathbf{P}_m \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}} \right)}{\operatorname{Tr} \left(\mathbf{A}_{\mathrm{tx}} \mathbf{P} \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}} \right)}.$$
 (23)

As a result, we always have

$$\nu_m \ge \frac{\delta_m \operatorname{Tr} \left(\mathbf{T}_{\mathrm{tx}} \right)}{\operatorname{Tr} \left(\mathbf{A}_{\mathrm{tx}} \mathbf{P} \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}} \right)}.$$
 (24)

In fact, the converse is also true. More precisely, in the next proposition we show that for a given pair $(\mathbf{A}_{tx}, \mathbf{T}_{tx})$ and input covariance matrix P, there exist per sub-band covariance matrices \mathbf{P}_m resulting in a power fraction vector $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)$ if and only if $\nu_m \ge 0$, $\sum_m \nu_m = 1$, and (24) is satisfied.

Proposition 3. Let $\mathbf{A}_{tx} \in \mathbb{C}^{N_t \times N_{st}}$, $Tr(\mathbf{T}_{tx}) > 0$, **P** with $Diag(\mathbf{P}) > 0$ and $\delta_m \ge 0$ with $\sum_m \delta_m = 1$ be given. For any $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)$, the following are equivalent:

(a) For a set of $\nu_m s$ given by (23), there exists \mathbf{P}_m with Diag $(\mathbf{P}_m) \ge 0$ such that $\mathbf{P} = \sum_m \delta_m \mathbf{P}_m$. (b) ν_m satisfies (24) for all m and $\sum_m \nu_m = 1$.

Proof. See Appendix D.
$$\Box$$

Note that (24) shows that the power in a sub-band cannot be reduced less than a threshold. This arises from the fact that the quantization noise of a quantizer is white and places power across the spectrum. Similar observation was made for the case of SISO systems in [31]. This suggests that, depending on the system parameters such as DAC resolution, there is a limit on how much OOB emission can be reduced. This is of great importance in practical systems with regulations on the OOB emission levels.

B. Achievable Rate

We next lower bound the achievable rate of the system

$$R_{\rm lin} := \liminf_{N \to \infty} \frac{1}{N} I(\mathbf{Z}; \widehat{\mathbf{Z}}), \tag{25}$$

between the transmitted symbols Z and received symbols Z. We refer to this as the *linear rate* since it is achieved using linear modulation as in Fig. 1. To bound (25), we use the following result which is a non-trivial extension of [31, Lemma 1] derived for the vector case.

Lemma 1. Suppose that $\mathbf{Z} \in \mathbb{C}^{N_{st} \times d}$ is a random matrix with *i.i.d.* columns $Z \sim C\mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \mathbf{P})$. Let **Y** be another random *matrix and define*,

$$\mathbf{P}_C := \mathbf{P}^{-1} \frac{1}{d} \sum_{i=1}^{a} \mathbf{C}_i, \tag{26a}$$

$$\mathbf{C}_{i} = \mathbb{E}\left[\overline{\mathbf{z}}_{:i}\overline{\mathbf{y}}_{:i}^{\mathsf{H}}\right] \mathbb{E}\left[\overline{\mathbf{y}}_{:i}\overline{\mathbf{y}}_{:i}^{\mathsf{H}}\right]^{-1} \mathbb{E}\left[\overline{\mathbf{y}}_{:i}\overline{\mathbf{z}}_{:i}^{\mathsf{H}}\right], \qquad (26b)$$

where $\overline{\mathbf{Z}} = \mathbf{Z} - \mathbb{E}[\mathbf{Z}]$ and $\overline{\mathbf{Y}} = \mathbf{Y} - \mathbb{E}[\mathbf{Y}]$. Then, the mutual information between **Z** and **Y** is bounded below by,

$$I(\mathbf{Z}; \mathbf{Y}) \ge -d\ln(|\mathbf{I} - \mathbf{P}_C|).$$

Proof. See Appendix E.

Assume that the rows of the noise matrix $\boldsymbol{\xi}_{n:}$ are i.i.d. with some distribution Ξ where $\mathbb{E}|\Xi|^2 < \infty.$ Then, using (17) calculate \mathbf{A}_{rx} and \mathbf{T}_{rx} with $\Phi(U) = \mathbf{W}_{\mathrm{rx}}Q_{\mathrm{rx}}(F(\mathbf{W}_{\mathrm{tx}}Q_{\mathrm{tx}}(U), \Xi))$, where $U \sim C\mathcal{N}(\mathbf{0}, \mathbf{P}), \ \mathbf{P} = \sum_{m} \delta_m \mathbf{P}_m.$

Theorem 2. Under the assumptions in Sec. II-D and given (8) and (9), the linear rate is almost surely bounded below by,

$$R_{\rm lin} \ge \sum_{m=1}^{M} \delta_m \log \left(\left| \mathbf{I} + \mathbf{A}_{\rm rx} \mathbf{P}_m \mathbf{A}_{\rm rx}^{\sf H} \mathbf{T}_{\rm rx}^{-1} \right| \right).$$
(27)

Proof. See Appendix E.

The lower bound can be interpreted as follows. For Gaussian inputs, the system in Fig. 1 can be equivalently modeled with a MIMO system with channel matrix gain \mathbf{A}_{rx} where the



Fig. 2: (a) Spatial transmit power, $BF_m(\cdot)$, for the transmit sub-band considering different DAC resolutions based on (7) for $\phi = \pi/4$. (b) ACLR based on (30) for different DAC resolutions. (c) Achievable rate of the system based on Thm. 2 for different values of average received SNR per antenna and DACs and ADCs resolutions.

channel outputs are added with a Gaussian noise vector of covariance $T_{\rm rx}$. The lower bound is basically the Shannon capacity of this equivalent linear system when the per subband covariance matrices are fixed. More precisely, for the case of MIMO channel with one sub-band and no quantization the lower bound leads to

$$R_{\rm lin} \ge \log\left(\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{P}\mathbf{H}^{\mathsf{H}}\right|\right)$$
 (28)

which is the Shannon capacity of MIMO channel \mathbf{H} with the transmit covariance matrix \mathbf{P} .

V. SIMULATIONS AND NUMERICAL RESULTS

For our simulations, we consider a MIMO system, where both transmitter and receiver use ULAs with half wavelength spacing and 16 transmit and eight receive antennas equipped with low resolution uniform DACs and ADCs, respectively. Moreover, for each quantizer, its spacing is optimized to minimize the mean square error distortion between its input and output according to [35]. We assume that the analog precoder and combiner are identity matrices. Also, we use a Rayleigh fading channel whose elements are generated using i.i.d. symmetric complex Gaussian distribution with variance one and mean zero (i.e., $C\mathcal{N}(0,1)$). We consider that there are two sub-bands, m = 2, and the transmitter is transmitting in the first sub-band only (i.e., $P_2 = 0$). For evaluating matrices A and T for Thm. 1 and Thm. 2, we use 1024 point FFT (i.e., N = 1024). More precisely, we assume that the first 512 samples correspond to the desired transmit band and the rest correspond to the adjacent band. Furthermore, we average the simulations over 100 channel realizations.

Spectrospatial power: First, we validate the accuracy of the per sample power formula in (21) from Thm. 1. Let us consider that the transmitter performs digital beamforming to direct its transmit stream towards angle ϕ . More precisely, to generate $\mathbf{z}_{n:}, n \in [512]$, it uses the covariance matrix

$$\mathbf{P}_1 = \frac{1}{N_t} \mathbf{e}(\phi) \mathbf{e}(\phi)^{\mathsf{H}},\tag{29}$$

where $\mathbf{e}(\phi) = [1, e^{i\pi \cos(\phi)}, \dots, e^{(N_t - 1)i\pi \cos(\phi)}]^{\mathsf{T}}$ is the array response of the transmitter's ULA. In Fig. 2a, we have plotted

the spatial transmit power, $BF_m(\psi)$ for $\psi \in (0, \pi]$ using (7) for $\phi = \pi/4$ considering different DAC resolutions when S_1 is calculated based on Thm. 1 and when it is evaluated through simulation. We observe that the theorem accurately predicts the simulations. Furthermore, as the DAC resolution decreases the main lobe gain is reduced while side lobes' gains are increased, leaking more power into the side-lobes.

To evaluate our model's accuracy for the spectral power distribution, let us define the ACLR of this transmitter as ratio of the transmit power in the first sub-band over the leaked power in the second sub-band. Based on (23), we have

$$ACLR = \frac{\nu_1}{\nu_2}.$$
 (30)

In Fig. 2b, we have plotted the ACLR for the covariance matrix in (29) with $\phi = \pi/4$ considering different DAC resolutions when ν_1 and ν_2 are calculated using Thm. 1 and when they are evaluated through simulation. As observed, the theorem accurately predicts the simulations and as expected the ACLR is an increasing function of the DAC resolution.

Achievable rate of the system: Next, we use Thm. 2 to plot the achievable rate of the system. For this case we use the identity matrix for \mathbf{P}_1 . The rate lower bound considering DAC and ADC resolutions of one, three, and five bits are plotted based on (27) from Thm. 2 in Fig. 2c. As a performance benchmark and upper bound, we have also plotted the Shannon capacity of the system corresponding to infinite resolution DACs and ADCs. As expected, increasing the resolution of DACs and ADCs improves the rate. We also observe that the system performs very close to the upper bound at low SNRs.

VI. CONCLUSION

We have studied quantized MIMO systems with linear modulation and provided a rigorous, yet simple equivalent linear model in the limit of large random transformations for modulation. The provided model accurately captures the effect of quantization on the spectrospatial power distribution and achievable rate of the system. We have also validated the spectrospatial predictions through simulations and numerical evaluations.

References

- [1] T. S. Rappaport *et al.*, *Millimeter wave wireless communications*. Pearson Education, 2014.
- [2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.
- [3] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, 2017.
- [4] J. Zhang *et al.*, "On low-resolution ADCs in practical 5G millimeterwave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, 2018.
- [5] M. Abdelghany et al., "Towards All-digital mmWave Massive MIMO: Designing around Nonlinearities," in Proc. IEEE Asilomar Conf. Signals, Syst., Comput., 2018, pp. 1552–1557.
- [6] A. Khalili et al., "On MIMO channel capacity with output quantization constraints," Proc. IEEE Int. Symp. Inf. Theory, pp. 1355–1359, June 2018.
- [7] A. Khalili *et al.*, "Tradeoff between delay and high SNR capacity in quantized MIMO systems," *in Proc. IEEE Int. Symp. Inf. Theory*, pp. 597–601, Jul. 2019.
- [8] —, "On multiterminal communication over MIMO channels with onebit ADCs at the receivers," *Proc. IEEE Int. Symp. Inf. Theory*, pp. 602– 606, July 2019.
- [9] —, "On throughput of millimeterwave MIMO systems with low resolution ADCs," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 5255–5259, 2020.
- [10] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, 2009.
- [11] T. Koch and A. Lapidoth, "At low SNR, asymmetric quantizers are better," *IEEE Trans. Inf. Information Theory*, vol. 59, no. 9, pp. 5421– 5445, 2013.
- [12] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proc. Intl. Conf. Wireless Commun. and Mobile Comput.* ACM, 2006, pp. 1387–1392.
- [13] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. IEEE Inf. Theory Appl. Wkshp. (ITA)*. IEEE, 2015, pp. 191–198.
- [14] S. Dutta *et al.*, "A case for digital beamforming at mmWave," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2019.
- [15] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, 2015.
- [16] S. Rini *et al.*, "A general framework for low-resolution receivers for MIMO channels," *arXiv preprint arXiv:1702.08133*, 2017.
- [17] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 1–5.
- [18] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.

- [19] S. Jacobsson *et al.*, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, 2017.
- [20] C. Mollen *et al.*, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, 2016.
- [21] J. Mo et al., "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, 2017.
- [22] A. Khalili *et al.*, "MIMO networks with one-bit ADCs: Receiver design and communication strategies," *IEEE Transactions on Communications*, 2021.
- [23] P. Skrimponis *et al.*, "Understanding energy efficiency and interference tolerance in millimeter wave receivers," *arXiv preprint arXiv:2201.00229*, 2022.
- [24] ——, "Towards energy efficient mobile wireless receivers above 100 GHz," *IEEE Access*, vol. 9, pp. 20704–20716, 2020.
- [25] A. Gersho and R. M. Gray, Vector quantization and signal compression. Springer Science & Business Media, 2012, vol. 159.
 [26] A. K. Fletcher et al., "Robust predictive quantization: Analysis and
- [26] A. K. Fletcher *et al.*, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 618–632, 2007.
- [27] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1739–1755, 2005.
- [28] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, 1993.
- [29] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, July 1996.
- [30] M. S. Derpich, J. Østergaard, and G. C. Goodwin, "The quadratic Gaussian rate-distortion function for source uncorrelated distortions," in *Data Compression Conf.* IEEE, 2008, pp. 73–82.
- [31] S. Dutta *et al.*, "Capacity bounds for communication systems with quantization and spectral constraints," *arXiv preprint arXiv:2001.03870*, 2020.
- [32] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [33] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, 2019.
- [34] Z. Landsman and J. Nešlehová, "Stein's lemma for elliptical random vectors," *Journal of Multivariate Analysis*, vol. 99, no. 5, pp. 912–927, 2008.
- [35] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.
- [36] K. W. Fang, Symmetric multivariate and related distributions. CRC Press, 2018.
- [37] J. F. Kingman, "On random sequences with spherical symmetry," *Biometrika*, vol. 59, no. 2, pp. 492–494, 1972.
- [38] E. Fischer, "Über den hadamardschen determinantensatz," Arch. Math.(Basel), vol. 13, pp. 32–40, 1908.
- [39] Y. Li and L. Feng, "Extensions of brunn-minkowski's inequality to multiple matrices," *Linear Algebra and its Applications*, vol. 603, pp. 91–100, 2020.

Appendix A

PROOF OF PROPOSITION 1

For the proof of this proposition, we first prove the following results on left circularly symmetric matrices and functional equations.

Lemma 2. Let $\mathbf{V} \in \mathbb{C}^{N \times N}$ be a left spherical symmetric matrix. Then, its characteristic function $\Phi(\mathbf{T})$, satisfies the followings

- 1) $\Phi(\mathbf{T}) = \Phi(\mathbf{V}^{\mathsf{H}}\mathbf{T})$ for $\mathbf{V} \in O(N)$, where O(N) denotes the set of $N \times N$ orthogonal matrices.
- 2) $\Phi(\mathbf{T})$ is a function of $\mathbf{T}^{\mathsf{H}}\mathbf{T}$.

The proof is a straightforward extension of the proof of [36, Theorem 2.1].

Lemma 3. Let $\mathbf{F}(\mathbf{a}) = \mathbf{F}(\mathbf{b})\mathbf{F}(\mathbf{c})$, where $\mathbf{F}(\mathbf{t})$ is a mapping from \mathbb{C}^d to \mathbb{C} and a function of \mathbf{tt}^{H} and $\mathbf{aa}^{\mathsf{H}} = \mathbf{bb}^{\mathsf{H}} + \mathbf{cc}^{\mathsf{H}}$. Therefore,

$$\mathbf{F}(\mathbf{a}) = e^{-\frac{1}{4}\mathbf{a}^{\mathsf{H}}\mathbf{P}\mathbf{a}},\tag{31}$$

where $\mathbf{P} \in \mathbb{C}^{d \times d}$.

Proof. Since $\mathbf{F}(\mathbf{a})$ is a function of \mathbf{aa}^{H} , we can write it as $\mathbf{F}(\mathbf{t}) = \mathbf{G}(\mathbf{t}^{\mathsf{H}}\mathbf{t})$. Set $\mathbf{A} = \mathbf{aa}^{\mathsf{H}}, \mathbf{B} = \mathbf{bb}^{\mathsf{H}}$, and $\mathbf{C} = \mathbf{cc}^{\mathsf{H}}$ we have:

$$\mathbf{G}(\mathbf{A}) = \mathbf{G}(\mathbf{B})\mathbf{G}(\mathbf{C}) \tag{32}$$

where $\mathbf{G}(\mathbf{T})$ is a mapping from $\mathbb{C}^{d \times d}$ to \mathbb{C} . Therefore,

$$\mathbf{G}(\mathbf{T}) = e^{\mathbf{J}(\mathbf{T})},\tag{33}$$

where $\mathbf{J}(\mathbf{T})$ is mapping from $\mathbb{C}^{d \times d}$ to \mathbb{C} and $\mathbf{J}(\mathbf{A}) = \mathbf{J}(\mathbf{B}) + \mathbf{J}(\mathbf{C})$. Therefore,

$$\mathbf{J}(\mathbf{T}) = \sum_{j} \mathbf{p}_{lj}^{\mathsf{T}} \mathbf{T} \mathbf{p}_{rj} \sum_{j} \operatorname{Tr} \left(\mathbf{p}_{lj}^{\mathsf{T}} \mathbf{T} \mathbf{p}_{rj} \right) = \sum_{j} \operatorname{Tr} \left(\mathbf{T} \mathbf{p}_{rj} \mathbf{p}_{lj}^{\mathsf{T}} \right)$$
$$= \operatorname{Tr} \left(\sum_{j} \mathbf{T} \mathbf{p}_{rj} \mathbf{p}_{lj}^{\mathsf{T}} \right) = \operatorname{Tr} \left(\mathbf{T} \left\{ \sum_{j} \mathbf{p}_{rj} \mathbf{p}_{lj}^{\mathsf{T}} \right\} \right).$$
(34)

As a result, we can write

$$\mathbf{J}(\mathbf{T}) = -\frac{1}{4} \operatorname{Tr}\left(\mathbf{PT}\right),\tag{35}$$

Where $\mathbf{P} \in \mathbb{C}^{d \times d}$. Using $\mathbf{T} = \mathbf{t}\mathbf{t}^{\mathsf{H}}$ and substituting (35) in (33), we have

$$\mathbf{F}(\mathbf{t}) = e^{-\frac{1}{4}\mathrm{Tr}\left(\mathbf{Ptt}^{\mathsf{H}}\right)} = e^{-\frac{1}{4}\mathrm{Tr}\left(\mathbf{t}^{\mathsf{H}}\mathbf{Pt}\right)} = e^{-\frac{1}{4}\mathbf{t}^{\mathsf{H}}\mathbf{Pt}}.$$
 (36)

Which concludes the proof.

To prove the proposition, we use a similar proof as of the paper [37]. Note that the matrix U is left-spherically symmetric. Therefore, any rotation of the rows of matrix U would preserve its distribution. Hence, using Finetti's theorem, there is a $\sigma - field, \Psi$ of events conditional upon which the $\mathbf{u}_{n:}, n \in [N]$ are independent and have the same distribution function $F(\mathbf{u})$. Define,

$$\Phi(\mathbf{t}) = \int e^{i\operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{t}^{\mathsf{H}}\mathbf{u}\right)\right)} dF(\mathbf{u}) = \mathbb{E}\left[e^{i\operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{t}^{\mathsf{H}}\mathbf{u}\right)\right)} \middle| \Psi\right],$$
(37)

where $\mathbf{t} \in \mathbb{C}^d$.

The conditional independence means that, for matrix $\mathbf{T} \in \mathbb{C}^{N \times d}$,

$$\mathbb{E}\left[e^{i\operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{T}^{\mathsf{H}}\mathbf{U}\right)\right)}\middle|\Psi\right] = \prod_{n=1}^{N}\Phi(\mathbf{t}_{r})$$
(38)

$$\mathbb{E}\left[e^{i\operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{T}^{\mathsf{H}}\mathbf{U}\right)\right)}\right] = \mathbb{E}\left[\prod_{n=1}^{N}\Phi(\mathbf{t}_{r})\right]$$
(39)

From Lemma 2, we know that the left and therefore the right side of (39) are function of $\mathbf{T}^{\mathsf{H}}\mathbf{T} = \sum_{r=1}^{N} \mathbf{t}_{r}^{\mathsf{H}}\mathbf{t}_{r}$. Let $\mathbf{t}^{\mathsf{H}}\mathbf{t} = \mathbf{h}^{\mathsf{H}}\mathbf{h} + \mathbf{g}^{\mathsf{H}}\mathbf{g}$. Then,

$$\mathbb{E}\left[|\Phi(\mathbf{t}) - \Phi(\mathbf{h})\Phi(\mathbf{g})|^2\right] = \mathbb{E}[\Phi(\mathbf{t})\Phi(-\mathbf{t})] \\ + \mathbb{E}[\Phi(\mathbf{h})\Phi(\mathbf{g})\Phi(-\mathbf{h})\Phi(-\mathbf{g})] - \mathbb{E}[\Phi(\mathbf{t})\Phi(\mathbf{h})\Phi(\mathbf{g})] \quad (40) \\ - \mathbb{E}[\Phi(-\mathbf{t})\Phi(-\mathbf{h})\Phi(-\mathbf{g})].$$

All four elements in the right hand side of (40) are equal since each element is in the form of (39) and is function of $2t^{H}t$. Therefore,

$$\Phi(\mathbf{t}) \stackrel{a.s}{=} \Phi(\mathbf{h})\Phi(\mathbf{g}) \tag{41}$$

Based on Lemma 3, the solution to this functional equation is

$$\Phi(\mathbf{t}^{\mathsf{H}}) = e^{-\frac{1}{4}\mathbf{t}^{\mathsf{H}}\mathbf{P}\mathbf{t}}.$$
(42)

Using (38),

$$\mathbb{E}\left[e^{i\operatorname{Re}\left(\operatorname{Tr}\left(\mathbf{T}^{\mathsf{H}}\mathbf{U}\right)\right)}\Big|\mathbf{P}\right]$$

$$=\mathbb{E}\left[\prod_{n=1}^{N}e^{-\frac{1}{4}\mathbf{t}_{n}^{\mathsf{H}}\mathbf{P}\mathbf{t}_{n}}\Big|\mathbf{P}\right]=\prod_{n=1}^{N}e^{-\frac{1}{4}\mathbf{t}_{n}^{\mathsf{H}}\mathbf{P}\mathbf{t}_{n}}$$
(43)

Therefore, conditioned on P, $\mathbf{u}_{n:}$ are i.i.d. $\sim C\mathcal{N}(\mathbf{0}, \mathbf{P})$. We can find P as follows

$$\mathbf{P} \stackrel{(a)}{=} \frac{1}{N} \sum_{n=1}^{N} \mathbf{u}_{n:} \mathbf{u}_{n:}^{\mathsf{H}} = \frac{1}{N} \mathbf{U}^{\mathsf{T}} \mathbf{U}^{*}$$

$$\stackrel{(b)}{=} \frac{1}{N} \mathbf{X}^{\mathsf{T}} \mathbf{X}^{*} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_{n:} \mathbf{x}_{n:}^{\mathsf{H}} \stackrel{(c)}{=} \mathbb{E}[XX^{\mathsf{H}}],$$
(44)

where (a) follows from strong law of large number, (b) is due to equality $\mathbf{U} = \mathbf{V}\mathbf{X}$, and (c) comes from empirical convergence of rows of **X** to random vector X^{T} .

APPENDIX B PROOF OF PROPOSITION 2

For the proof of this proposition, we use an approach similar to the proof of [33, Theorem 4]. Consider the following steps:

$$\mathbf{U} = \mathbf{V}^{\mathsf{H}}\mathbf{Z},\tag{45a}$$

$$\mathbf{A} = \left\langle \frac{\partial \Phi(\mathbf{U}, \boldsymbol{\Xi})}{\partial U} \right\rangle,\tag{45b}$$

$$\mathbf{B} = \boldsymbol{\Phi}(\mathbf{U}, \boldsymbol{\Xi}) - \mathbf{U}\mathbf{A}^{\mathsf{T}}, \qquad (45c)$$

$$\Theta = \mathbf{VB},\tag{45d}$$

where $\left\langle \frac{\partial \Phi(\mathbf{U}, \Xi)}{\partial U} \right\rangle$ is the Jacobian matrix of $\Phi(\cdot)$ evaluated for each row pair of $(\mathbf{u}_{n:}, \boldsymbol{\xi}_{n:})$ and then averaged over *n*. The transpose of rows of matrix Ξ (i.e., $\boldsymbol{\xi}_{n:}$) are i.i.d. $\mathcal{CN}(0, \mathbf{I}_{N_r})$. Therefore,

$$\lim_{n \to \infty} \{ \boldsymbol{\xi}_{n:} \} \stackrel{PL(2)}{=} \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim \mathcal{CN}(\boldsymbol{0}, I_{N_r})$$
(46)

From Prop. 1, we know that the rows of matrix U converge empirically to random vector $U \sim C\mathcal{N}(\mathbf{0}, \mathbf{P})$. Additionally, since V and Z are independent of Ξ and the function $\Phi(\cdot)$ is Lipschitz continuous, the sequence $(\boldsymbol{\xi}_{n:}, \mathbf{U}_{n:}, \mathbf{b}_{n:})$ converges empirically to (Ξ, U, B) . Following same steps of [33, Theorem 4], we have

$$\Theta = \Theta^{det} + \Theta^{rand}, \tag{47}$$

where $\Theta^{det} = \mathbf{0}$ and the rows of the matrix Θ^{rand} , converge empirically to $\Theta \sim \mathcal{CN}(\mathbf{0}, \mathbf{T})$ where $\mathbf{T} = \mathbb{E} \left[(\Phi(U, \Xi) - \mathbf{A}U) (\Phi(U, \Xi) - \mathbf{A}U)^{\mathsf{H}} \right].$

APPENDIX C Proof of Theorem 1

The theorem is a direct application of the linear model in Prop. 2. To use the proposition, first observe that, due to (10) and the Gaussian distribution on \mathbf{z}_{n} : in (8), we have that the elements of sub-band selection ℓ and the rows of frequencydomain inputs \mathbf{Z} converge empirically as,

$$\lim_{n \to \infty} \{ (\mathbf{z}_{n:}, \boldsymbol{\ell}_n) \} \stackrel{PL(2)}{=} (Z, \ell), \tag{48}$$

where $\ell \in [M]$ is a discrete random variable with $\Pr(\ell = m) = \delta_m$ and Z is the conditional complex Gaussian,

$$Z \sim C\mathcal{N}(\mathbf{0}, \mathbf{P}_m)$$
 when $\ell = m$.

In particular, the covariance matrix of Z is,

$$\mathbb{E}[ZZ^{\mathsf{H}}] = \sum_{m=1}^{M} \delta_m \mathbf{P}_m =: \mathbf{P}.$$
 (49)

Now, the frequency domain components of the transmitted vectors $\mathbf{x}_{:j}, j \in [N_t]$ are given by,

$$\mathbf{R} = \mathbf{V}\mathbf{X} = \mathbf{V}\mathbf{Q}(\mathbf{V}^{\mathsf{H}}\mathbf{Z}).$$

Proposition 2 then shows that the components of $(\mathbf{r}, \mathbf{z}, \boldsymbol{\ell})$ converge empirically as,

$$\lim_{N \to \infty} \{ (\mathbf{r}_{n:}, \mathbf{z}_{n:}, \ell_n) \} \stackrel{PL(2)}{=} (R, Z, \ell)$$

$$R = \mathbf{A}_{\mathrm{tx}} Z + W_{\mathrm{tx}}, \quad W_{\mathrm{tx}} \sim C \mathcal{N}(\mathbf{0}, \mathbf{T}_{\mathrm{tx}}),$$

where W_{tx} is independent of Z. The sub-band energies,

$$\mathbf{S}_{m} := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{r}_{n:} \mathbf{r}_{n:}^{\mathsf{H}} \mathbb{1}_{\{\ell_{n}=m\}} = \mathbb{E} \left[RR^{\mathsf{H}} \mathbb{1}_{\{\ell=m\}} \right]$$
$$= \mathbb{E} \left[(\mathbf{A}_{\mathrm{tx}} Z + W_{\mathrm{tx}}) (\mathbf{A}_{\mathrm{tx}} Z + W_{\mathrm{tx}})^{\mathsf{H}} | \ell = m \right] \Pr(\ell = m)$$
$$= \left[\mathbf{A}_{\mathrm{tx}} \mathbf{P}_{m} \mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}} \right] \delta_{m}.$$

This proves (21). To prove (22),

$$\mathbf{S} = \sum_{m=1}^{M} \mathbf{S}_{m} = \sum_{m=1}^{M} \left[\mathbf{A}_{tx} \mathbf{P}_{m} \mathbf{A}_{tx}^{\mathsf{H}} + \mathbf{T}_{tx} \right] \delta_{m}$$
(50)
= $\mathbf{A}_{tx} \mathbf{P} \mathbf{A}_{tx}^{\mathsf{H}} + \mathbf{T}_{tx},$

where the last step used (49) and the fact that $\sum_{m} \delta_m = 1$.

APPENDIX D The Linear Rate Region

 $(a) \Rightarrow (b)$: Suppose there exists \mathbf{P}_m as in (a) and let ν_m be given by (23). Since $\mathbf{P} = \sum_m \delta_m \mathbf{P}_m$, we have $\sum_m \nu_m = 1$. Also, since $\text{Diag}(\mathbf{P}_m) \ge 0$, we have ν_m in (23) satisfies the lower bound (24).

(b) \Rightarrow (a): Conversely, suppose we are given ν satisfying (24) with $\sum_{m} \nu_{m} = 1$. Set,

$$\mathbf{P}_{m} = \frac{\mathbf{P}}{\mathrm{Tr}\left(\mathbf{A}_{\mathrm{tx}}\mathbf{P}\mathbf{A}_{\mathrm{tx}}^{\mathsf{H}}\right)} \left[\frac{\nu_{m}}{\delta_{m}}\mathrm{Tr}\left(\mathbf{A}_{\mathrm{tx}}\mathbf{P}\mathbf{A}_{\mathrm{tx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{tx}}\right) - \mathrm{Tr}(\mathbf{T}_{\mathrm{tx}})\right]$$
(51)

Therefore, ν_m satisfies (23). Since ν_m satisfies (24), P_m in (51) satisfies $P_m \ge 0$. Also,

$$\sum_{m} \delta_{m} P_{m} = \frac{\mathbf{P}}{\operatorname{Tr} \left(\mathbf{A}_{tx} \mathbf{P} \mathbf{A}_{tx}^{\mathsf{H}} \right)} \left| \sum_{m} \nu_{m} \operatorname{Tr} \left(\mathbf{A}_{tx} \mathbf{P} \mathbf{A}_{tx}^{\mathsf{H}} + \mathbf{T}_{tx} \right) - \sum_{m} \delta_{m} \operatorname{Tr} (\mathbf{T}_{tx}) \right]$$
$$= \frac{\mathbf{P}}{\operatorname{Tr} \left(\mathbf{A}_{tx} \mathbf{P} \mathbf{A}_{tx}^{\mathsf{H}} \right)} \left[\operatorname{Tr} \left(\mathbf{A}_{tx} \mathbf{P} \mathbf{A}_{tx}^{\mathsf{H}} + \mathbf{T}_{tx} - \mathbf{T}_{tx} \right) \right] = \mathbf{P},$$

where we have used the fact that $\sum_{m} \nu_m = \sum_{m} \delta_m = 1$.

APPENDIX E Proof of Theorem 2

We need two basic mutual information lemmas. For $m \in [M]$, let $\mathbb{Z}[m]$ and $\widehat{\mathbb{Z}}[m]$ denote the sub-matrices of \mathbb{Z} and $\widehat{\mathbb{Z}}$ with components in sub-band m.

Lemma 4. The mutual information is bounded below by,

$$I(\mathbf{Z}; \widehat{\mathbf{Z}}) \ge \sum_{m=1}^{M} I(\mathbf{Z}[m]; \widehat{\mathbf{Z}}[m]).$$
(52)

The proof follows using the same steps as of the proof of [31, Lemma 1].

Lemma 5. Suppose that $\mathbf{Z} \in \mathbb{C}^{N_t \times d}$ is a random matrix with *i.i.d.* columns $Z \sim C\mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \mathbf{P})$. Let \mathbf{Y} be another random matrix and define,

$$\mathbf{P}_C := \mathbf{P}^{-1} \frac{1}{d} \sum_{i=1}^d \mathbf{C}_i, \tag{53a}$$

$$\mathbf{C}_{i} = \mathbb{E}\left[\overline{\mathbf{z}}_{:i}\overline{\mathbf{y}}_{:i}^{\mathsf{H}}\right] \mathbb{E}\left[\overline{\mathbf{y}}_{:i}\overline{\mathbf{y}}_{:i}^{\mathsf{H}}\right]^{-1} \mathbb{E}\left[\overline{\mathbf{y}}_{:i}\overline{\mathbf{z}}_{:i}^{\mathsf{H}}\right], \qquad (53b)$$

where $\overline{\mathbf{Z}} = \mathbf{Z} - \mathbb{E}[\mathbf{Z}]$ and $\overline{\mathbf{Y}} = \mathbf{Y} - \mathbb{E}[\mathbf{Y}]$. Then, the mutual information between \mathbf{Z} and \mathbf{Y} is bounded below by,

$$I(\mathbf{Z};\mathbf{Y}) \ge -d\ln(|\mathbf{I} - \mathbf{P}_C|).$$

Proof. The mutual information is,

$$I(\mathbf{Z}; \mathbf{Y}) = H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{Y}).$$
(54)

Since $\widetilde{\mathbf{z}} = \operatorname{vec}(\mathbf{Z}) = [\mathbf{z}_{:1}^{\mathsf{T}}, \mathbf{z}_{:2}^{\mathsf{T}}, ..., \mathbf{z}_{:d}^{\mathsf{T}}]^{\mathsf{T}}$ is distributed as $C\mathcal{N}(\operatorname{vec}(\boldsymbol{\mu}_{\mathbf{z}}), \mathbf{I}_{d} \otimes \mathbf{P}),$

$$H(\mathbf{Z}) = d\ln(|2\pi e\mathbf{P}|).$$
(55)

Define $\tilde{\mathbf{y}} = \text{vec}(\mathbf{Y}) = [\mathbf{y}_{:1}^{\mathsf{T}}, \mathbf{y}_{:2}^{\mathsf{T}}, ..., \mathbf{y}_{:d}^{\mathsf{T}}]^{\mathsf{T}}$. Now, given \mathbf{Y} , we can get a estimate of $\tilde{\mathbf{z}}$, $\tilde{\tilde{\mathbf{z}}}$ which leads to the error covariance matrix,

$$\Sigma_e := \mathbb{E}\left[(\widetilde{\mathbf{z}} - \widehat{\widetilde{\mathbf{z}}}(\widetilde{\mathbf{y}})) (\widetilde{\mathbf{z}} - \widehat{\widetilde{\mathbf{z}}}(\widetilde{\mathbf{y}}))^{\mathsf{H}} \right], \tag{56}$$

where $\tilde{\mathbf{z}}(\tilde{\mathbf{y}})$ is the estimate of $\tilde{\mathbf{z}}$ given $\tilde{\mathbf{y}}$. So, the conditional entropy $H(\mathbf{Z}|\mathbf{Y})$ is bounded below by the entropy of a Gaussian random vector with the covariance Σ_e . Consider linear minimum mean square estimation and we have

$$H(\mathbf{Z}|\mathbf{Y}) \le \ln\left((2\pi e)^{dN_t} |\Sigma_e|\right)$$
(57)

$$\stackrel{(a)}{\leq} \ln \left((2\pi e)^{dN_t} \prod_{i=1}^{a} |\mathbf{C}_{ei}| \right) \tag{58}$$

$$= d \ln \left((2\pi e)^{N_t} \right) + N_t \ln \left(\prod_{i=1}^a |\mathbf{C}_{ei}|^{\frac{1}{N_t}} \right)$$
(59)

$$\stackrel{(b)}{\leq} d\ln\left((2\pi e)^{N_t}\right) + dN_t \ln\left(\frac{1}{d}\sum_{i=1}^d |\mathbf{C}_{ei}|^{\frac{1}{N_t}}\right) \tag{60}$$

$$\stackrel{(c)}{\leq} d\ln\left((2\pi e)^{N_t}\right) + dN_t \ln\left(\frac{1}{d}\left|\sum_{i=1}^d \mathbf{C}_{ei}\right|^{\overrightarrow{N_t}}\right) \tag{61}$$

$$= d \ln \left((2\pi e)^{N_t} \left| \frac{1}{d} \sum_{i=1}^d \mathbf{C}_{ei} \right| \right)$$
(62)

where $\mathbf{C}_{ei} = \mathbb{E}\left[\bar{\mathbf{z}}_{:i}\bar{\mathbf{z}}_{:i}^{\mathsf{H}}\right] - \mathbb{E}\left[\bar{\mathbf{z}}_{:i}\bar{\mathbf{y}}_{:i}^{\mathsf{H}}\right] \mathbb{E}\left[\bar{\mathbf{y}}_{:i}\bar{\mathbf{y}}_{:i}^{\mathsf{H}}\right]^{-1} \mathbb{E}\left[\bar{\mathbf{y}}_{:i}\bar{\mathbf{z}}_{:i}^{\mathsf{H}}\right]$, (a) uses Fischer's inequality [38], (b) uses Jensen's inequality, and (c) uses an extended version of the Hadamard's inequality [39, Prop. 2.7]. Therefore, we have

$$H(\mathbf{Z}|\mathbf{Y}) \le d\ln\left(\left|2\pi e\left(\mathbf{P} - \frac{1}{d}\sum_{i=1}^{d}\mathbf{C}_{i}\right)\right|\right).$$
(63)

Substituting (55) and (63) into (54), we get

$$I(\mathbf{Z}; \mathbf{Y}) \ge d\ln(|2\pi e\mathbf{P}|) - d\ln\left(\left|2\pi e\left(\mathbf{P} - \frac{1}{d}\sum_{i=1}^{d}\mathbf{C}_{i}\right)\right|\right)$$
(64)

$$= -d\ln(|\mathbf{I} - \mathbf{P}_C|),\tag{65}$$

which concludes the proof. \Box

To prove the theorem, we use these lemmas as follows. In each sub-band m, the rows of $\mathbb{Z}[m]$ are i.i.d. complex Gaussians with zero mean and covariance matrix \mathbb{P}_m . So, by Lemma 5,

$$I(\mathbf{Z}[m]; \widehat{\mathbf{Z}}[m]) \ge -N_m \ln(|\mathbf{I} - \mathbf{P}_{Cm}|), \tag{66}$$

where N_m is the number of coefficients in sub-band m and \mathbf{P}_{Cm} is the correlation coefficient matrix,

$$\mathbf{P}_{Cm} := \mathbf{P}_{m}^{-1} \frac{1}{N_{m}} \sum_{i=1}^{N_{m}} \mathbb{E}\left[\mathbf{z}[m]_{:i} \widehat{\mathbf{z}}[m]_{:i}^{\mathsf{H}}\right] \times \\ \mathbb{E}\left[\widehat{\mathbf{z}}[m]_{:i} \widehat{\mathbf{z}}[m]_{:i}^{\mathsf{H}}\right]^{-1} \mathbb{E}\left[\widehat{\mathbf{z}}[m]_{:i} \mathbf{z}[m]_{:i}^{\mathsf{H}}\right].$$
(67)

Now, (10) shows that $N_m/N \to \delta_m$. So, if we divide (66) by N and take the limit we get,

$$\liminf_{N \to \infty} \frac{1}{N} I(\mathbf{Z}[m]; \widehat{\mathbf{Z}}[m]) \ge -\delta_m \ln(|\mathbf{I} - \overline{\mathbf{P}}_{Cm}|).$$
(68)

where $\overline{\mathbf{P}}_{Cm}$ is the limiting correlation,

$$\overline{\mathbf{P}}_{Cm} := \lim_{N \to \infty} \mathbf{P}_{Cm} \tag{69}$$

To compute the limit in (69), we use a similar calculation to the proof of Theorem 1. Specifically, the received symbols are given by,

$$\widehat{\mathbf{Z}} = \mathbf{V}\mathbf{G}(\mathbf{V}^{\mathsf{H}}\mathbf{Z}, \mathbf{W}).$$
(70)

Proposition 2 then shows that the rows of $(\widehat{\mathbf{Z}}, \mathbf{Z}, \boldsymbol{\ell})$ converge empirically as,

$$\lim_{N \to \infty} \{ (\widehat{Z}_{n:}, Z_{n:}, \ell_n) \} \stackrel{PL(2)}{=} (\widehat{Z}, Z, L),$$
(71)

and

$$\widehat{Z} = \mathbf{A}_{\mathrm{rx}} Z + W_{\mathrm{rx}}, \quad W_{\mathrm{rx}} \sim \mathcal{C}N(\mathbf{0}, \mathbf{T}_{\mathrm{rx}}), \tag{72}$$

where $W_{\rm rx}$ is independent of Z. Now, we have that,

$$\lim_{N \to \infty} \mathbf{P}_{Cm} = \frac{\lim_{N_m \to \infty} \mathbf{P}_m^{-1} \frac{1}{N_m} \sum_{i=1}^{N_m} \mathbb{E}\left[\mathbf{z}[m]_{:i} \widehat{\mathbf{z}}[m]_{:i}^{\mathsf{H}}\right] \times}{\mathbb{E}\left[\widehat{\mathbf{z}}[m]_{:i} \widehat{\mathbf{z}}[m]_{:i}^{\mathsf{H}}\right]^{-1} \mathbb{E}\left[\widehat{\mathbf{z}}[m]_{:i} \mathbf{z}[m]_{:i}^{\mathsf{H}}\right]}$$
(73)
$$= \frac{\mathbf{P}_m^{-1} \mathbb{E}\left[Z\widehat{Z}^{\mathsf{H}}|L=m\right] \mathbb{E}\left[\widehat{Z}\widehat{Z}^{\mathsf{H}}|L=m\right]^{-1} \times}{\mathbb{E}\left[\widehat{Z}Z|L=m\right]}$$
(74)

$$= \left[\mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} (\mathbf{A}_{\mathrm{rx}} \mathbf{P}_{m} \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{rx}})^{-1} \mathbf{A}_{\mathrm{rx}} \right] \mathbf{P}_{m}^{\mathsf{H}}$$
(75)

Hence,

$$(\mathbf{I} - \overline{\mathbf{P}}_{C}^{(m)})^{-1}$$

$$\stackrel{(a)}{=} \mathbf{I} + \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} \left(\mathbf{A}_{\mathrm{rx}} (\mathbf{P}_{m} - \mathbf{P}_{m}^{\mathsf{H}}) \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} + \mathbf{T}_{\mathrm{rx}} \right)^{-1} \mathbf{A}_{\mathrm{rx}} \mathbf{P}_{m}^{\mathsf{H}} \quad (76)$$

$$\stackrel{(b)}{=} \mathbf{I} + \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} \mathbf{T}_{-1}^{-1} \mathbf{A}_{\mathrm{rx}} \mathbf{P}_{m} \quad (77)$$

$$\stackrel{(b)}{=} \mathbf{I} + \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} \mathbf{T}_{\mathrm{rx}}^{-1} \mathbf{A}_{\mathrm{rx}} \mathbf{P}_{m}, \tag{77}$$

where (a) follows from Woodbury inversion lemma and (b) uses the fact that \mathbf{P}_m is a covariance matrix. As a result, from (68), we obtain

$$\liminf_{N \to \infty} \frac{1}{N} I(\mathbf{Z}[m]; \widehat{\mathbf{Z}}[m]) \geq \delta_m \log \left(|\mathbf{I} + \mathbf{A}_{\mathrm{rx}}^{\mathsf{H}} \mathbf{T}_{\mathrm{rx}}^{-1} \mathbf{A}_{\mathrm{rx}} \mathbf{P}_m| \right).$$
(78)

Substituting (78) into the sum (52) and using Sylvester's determinant identity proves (27).