

Channel State Acquisition in FDD Massive MIMO: Rate-Distortion Bound and Effectiveness of “Analog” Feedback

Mahdi Barzegar Khalilsarai, Yi Song, Tianyu Yang, and Giuseppe Caire
 {m.barzegarkhalilsarai, yi.song, tianyu.yang, caire}@tu-berlin.de

Abstract—We consider the problem of estimating channel fading coefficients (modeled as a correlated Gaussian vector) via Downlink (DL) training and Uplink (UL) feedback in wideband FDD massive MIMO systems. Using rate-distortion theory, we derive optimal bounds on the achievable channel state estimation error in terms of the number of training pilots in DL (β_{tr}) and feedback dimension in UL (β_{fb}), with random, spatially isotropic pilots. It is shown that when the number of training pilots exceeds the channel covariance rank (r), the optimal rate-distortion feedback strategy achieves an estimation error decay of $\Theta(\text{SNR}^{-\alpha})$ in estimating the channel state, where $\alpha = \min(\beta_{\text{fb}}/r, 1)$ is the so-called quality scaling exponent. We also discuss an “analog” feedback strategy, showing that it can achieve the optimal quality scaling exponent for a wide range of training and feedback dimensions with no channel covariance knowledge and simple signal processing at the user side. Our findings are supported by numerical simulations comparing various strategies in terms of channel state mean squared error and achievable ergodic sum-rate in DL with zero-forcing precoding.

Index Terms—Wideband FDD massive MIMO, channel state estimation error, quality scaling exponent, rate-distortion theory, analog feedback.

I. INTRODUCTION

Massive MIMO consists of employing a large number of antennas (M) at the Base Station (BS) to simultaneously multiplex data over the spatial domain and serve a much smaller number of User Equipment (UEs) (K) at the same time-frequency resource in the Downlink (DL) [1, 2]. Achieving the capacity improvements of massive MIMO requires availability of accurate channel state information at the multi-antenna BS transmitter (CSIT). Therefore the BS needs to learn (or *train*) the CSIT, namely the fading coefficients associated with each of its M antennas and those of the K UEs. In the time division duplexing (TDD) operation mode, the BS learns the CSIT by receiving pilots from the UEs in Uplink (UL) and, relying on UL-DL channel reciprocity, extrapolating the DL channel [3]. In frequency division duplexing (FDD) mode however, channel reciprocity does not hold and the BS needs to broadcast training pilots in the DL to the UEs and receive their estimated channel state via explicit feedback in UL. The error in estimating the CSIT strongly affects the DL spectral efficiency. For example, it is well-known that when the error between true and estimated CSIT decreases as $O(\text{SNR}^{-\alpha})$ in SNR (equivalently, BS transmission power) for some constant $\alpha \in [0, 1]$, then zero-forcing (ZF) precoding can achieve only a fraction α of the optimal degrees of freedom (DoF) per UE

[4, 5]. Therefore the choice of CSIT training and feedback strategies is crucial in achieving faster estimation error decays.

In this paper, we study the CSIT estimation problem for wideband massive MIMO systems, in which the channel is modeled as a spatially correlated, stationary Gaussian random process that evolves in time according to a block-fading model [6]. We consider a generic design of random, spatially isotropic pilots of arbitrary dimension β_{tr} and we derive a lower bound on the achievable CSIT estimation error using rate-distortion theory and the idea of *remote source coding* at the UE. This lower-bound results in an upper-bound on the achievable quality scaling exponent as a function of training and feedback dimension pairs $(\beta_{\text{tr}}, \beta_{\text{fb}})$, showing the fastest rate of error decay among all feedback strategies. If covariance knowledge is available at the UE, we demonstrate how one can approach the optimal performance at the price of small overhead in the number of feedback bits using the entropy-coded scalar quantization (ECSQ) [7]. We then study a variation of the well-known analog feedback (AF) strategy [8–10], in which the training measurements are sent to the BS via unquantized quadrature amplitude modulation (QAM), using which the BS computes an MMSE estimate of the channel (channel covariance knowledge at the BS is assumed). This variation of AF is an attractive strategy, because it requires no channel covariance knowledge and no sophisticated processing (namely, channel estimation and quantization) at the UE side, both of which come at a high price in the case of wideband massive MIMO channels. We emphasize this point by deriving an expression for the achievable quality scaling exponent with AF, and showing its optimality for a wide range of choices of training and feedback dimensions (see Fig. 1).

The effect of channel training and feedback on CSIT estimation error and spectral efficiency in FDD MIMO systems has been studied before in several works [4, 11–14]. In [4] and [12] lower bounds on the achievable DL rate with ZF precoding for analog and digital (quantized) feedback are given, where it is assumed that the number of training pilots exceeds the channel dimension. In [13] achievable rates of an FDD massive MIMO system with optimized training pilots and with channel covariance knowledge at the UE side was studied. In [15] the authors studied sufficient conditions to achieve full DoF by considering channel covariance knowledge at the UE and an error-free channel state feedback to the BS (i.e. an ideal feedback link). The present work provides the following

novelties with respect to the above:

- 1) We consider training and feedback of the channel on the whole bandwidth of N subcarrier, while most previous works assume a narrow-band model, neglecting frequency-domain channel correlation. From a practical standpoint, the recent releases of the 5G new radio heavily emphasize on exploiting this frequency-domain correlation feature of the channel to reduce feedback overhead [16]. This aspect is captured in our model and reflected in the rate-distortion bound as well as the proposed AF strategy.
- 2) We consider training the channel with an arbitrary number of pilots (β_{tr}) that can be larger or smaller than the CSI dimension (MN). Most previous works have assumed training with a number of pilots *larger* than the channel dimension, which is impractical in massive MIMO where the CSI dimension potentially exceeds the dimension of the coherence block.
- 3) We provide optimal information-theoretic bounds on the achievable CSIT estimation error and the quality scaling exponent in terms of training and feedback dimensions. To the best of our knowledge, such analysis for massive MIMO channels with spatial correlation has not been considered before.
- 4) We show the effectiveness of AF without the need for either channel covariance knowledge, or sophisticated estimation and quantization at the UE side.

The rest of the paper is organized as follows. In Section II we describe the model, training method and error metrics. In Section III we derive a lower bound on CSIT estimation error via rate-distortion theory and we explain feedback via ECSQ. In Section IV we discuss AF and derive an expression for the quality scaling exponent it can achieve. Finally, Section V concludes the paper with numerical simulation results.

Notation: We denote scalars, vectors, and matrices by small, small bold-face and capital bold-face letters x , \mathbf{x} , \mathbf{X} , respectively. For a positive integer N , we define $[N] \triangleq \{1, \dots, N\}$. Superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian transpose, respectively. The function $\mathbf{1}_{\mathcal{X}}(x)$ is the indicator function such that $\mathbf{1}_{\mathcal{X}}(x) = 1$ if $x \in \mathcal{X}$ and $\mathbf{1}_{\mathcal{X}}(x) = 0$ if $x \notin \mathcal{X}$. For a real-valued functions $f(\cdot)$ and $g(\cdot)$, defined over all positive reals $x > 0$, we say $f(x) = \Theta(g(x))$ if $\exists A_1, A_2 > 0$ such that $A_1 g(x) \leq f(x) \leq A_2 g(x)$ for all $x \geq x_0$.

II. CHANNEL TRAINING

We consider a broadcast MIMO OFDM system, consisting of a BS with an array of M antennas and $K \leq M$ single-antenna user equipment (UEs). The frequency-domain signal corresponding to subcarrier n received by an arbitrary UE can be expressed as $y[n] = \mathbf{h}^H[n]\mathbf{x}[n] + z[n]$, where $\mathbf{h}[n] \in \mathbb{C}^M$ contains the channel fading coefficients between the BS and the UE at subcarrier n , $\mathbf{x}[n] \in \mathbb{C}^M$ is the transmit signal satisfying the power constraint $\mathbb{E}[\|\mathbf{x}[n]\|^2] \leq \text{snr}_{\text{dl}}$ for all n , where snr_{dl} denotes SNR in the DL, while $z \sim \mathcal{CN}(0, 1)$ is additive

white Gaussian noise (AWGN).¹ We concatenate the channel over all subcarriers in a vector $\mathbf{h} = [\tilde{\mathbf{h}}[1]^T, \dots, \tilde{\mathbf{h}}[N]^T]^T \in \mathbb{C}^{MN}$, which we refer to as the channel state information at the transmitter (CSIT). We assume that \mathbf{h} evolves according to a block-fading model in which it is constant over frames of length T and changes from frame to frame according to an ergodic stationary, spatially correlated, zero-mean Gaussian process, i.e. $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}^h)$ where $\mathbf{\Sigma}^h = \mathbb{E}[\mathbf{h}\mathbf{h}^H]$ is the channel covariance of rank $r = \text{rank}(\mathbf{\Sigma}^h)$.

Throughout this work we assume that UEs have perfect estimates of their channels (perfect CSIR). We then focus on CSIT acquisition by the BS via the following process. Given a set $\mathcal{N}_p \subseteq [N]$ of N_p pilot subcarriers, the BS broadcasts a sequence of $T_p \leq T$ training vectors per pilot subcarrier. The training measurements received at the UE can be written as

$$\tilde{\mathbf{y}}^{\text{tr}}[n] = \tilde{\mathbf{h}}^H[n]\tilde{\mathbf{X}}^{\text{tr}}[n] + \tilde{\mathbf{z}}^{\text{tr}}[n], \quad n \in \mathcal{N}_p, \quad (1)$$

where $\tilde{\mathbf{X}}^{\text{tr}}[n] = [\tilde{\mathbf{x}}_1[n], \dots, \tilde{\mathbf{x}}_{T_p}[n]] \in \mathbb{C}^{M \times T_p}$ is a matrix containing the T_p pilot vectors as its columns. The training dimension $\beta_{\text{tr}} = T_p N_p$ denotes the total number of dimensions dedicated to training in a time-frequency block of dimension TN . The β_{tr} training measurements at the UE can be represented by a single vector $\mathbf{y}^{\text{tr}} = [\tilde{\mathbf{y}}[n_1], \dots, \tilde{\mathbf{y}}[n_{N_p}]] \in \mathbb{C}^{1 \times \beta_{\text{tr}}}$, where using (1) we have

$$\mathbf{y}^{\text{tr}} = \mathbf{h}^H \mathbf{X}^{\text{tr}} + \mathbf{z}^{\text{tr}}, \quad (2)$$

where $\mathbf{X}^{\text{tr}} = (\mathbf{B}_{n,\ell})_{\substack{\ell \in [N_p] \\ n \in [N]}} \in \mathbb{C}^{MN \times \beta_{\text{tr}}}$ is the *training matrix* consisting of $M \times T_p$ blocks $\mathbf{B}_{n,\ell}$, where $\mathbf{B}_{n,\ell} = \tilde{\mathbf{X}}^{\text{tr}}[n_\ell]$ if $n_\ell \in \mathcal{N}_p$ and $\mathbf{B}_{n,\ell} = \mathbf{0}$ if $n_\ell \notin \mathcal{N}_p$. The training matrix can be designed in several ways, for example by optimizing various performance criteria based on channel covariance knowledge [13, 17]. However, we consider a simpler, and therefore practically more available design, in which pilot vectors are randomly and independently generated according to an isotropic Gaussian distribution,

$$\tilde{\mathbf{x}}_i^{\text{tr}}[n] \sim \mathcal{CN}(\mathbf{0}, \frac{\text{snr}_{\text{dl}}}{M} \mathbf{I}), \quad i \in [T_p], \quad n \in \mathcal{N}_p, \quad (3)$$

where we verify that this design satisfies the transmission power constraint $\mathbb{E}[\|\tilde{\mathbf{x}}_i^{\text{tr}}[n]\|^2] \leq \text{snr}_{\text{dl}}$. In other words, the elements of \mathbf{X}^{tr} that are not identically zero, are generated as $\mathcal{CN}(0, \text{snr}_{\text{dl}}/M)$ Gaussian random variables.

After receiving pilot symbols, the UE computes a message containing information about the channel state and sends it to the BS via β_{fb} uses of the UL channel. Given the feedback channel output, the BS computes an estimate $\hat{\mathbf{h}}$ of the CSIT. We consider the mean squared error (MSE)

$$d(\mathbf{h}, \hat{\mathbf{h}}) = \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2] \quad (4)$$

as the error (distortion) metric between true and estimated CSIT. For a fixed tuple $(\beta_{\text{tr}}, \beta_{\text{fb}}, \text{snr}_{\text{dl}})$ and a given realization of \mathbf{X}^{tr} , we say that an error D is achievable if $d(\mathbf{h}, \hat{\mathbf{h}}) \leq D$. Accordingly, we say that a quality scaling exponent of α is achievable if $d(\mathbf{h}, \hat{\mathbf{h}}) = \Theta(\text{snr}_{\text{dl}}^{-\alpha})$, with the exponent indicating how fast the error decays with SNR [5, 18]. Note

¹This definition of transmit power and noise variance simplifies notation, since we only need the ratio of the two, i.e. the SNR.

that an error decay of $\Theta(\text{snr}_{\text{dl}}^{-\alpha})$ in estimating the wideband channel implies an error decay in estimating the channel over each subcarrier that is at least as fast. In other words, if $d(\mathbf{h}, \hat{\mathbf{h}}) = \Theta(\text{snr}_{\text{dl}}^{-\alpha})$, then $d(\tilde{\mathbf{h}}[n], \hat{\tilde{\mathbf{h}}}[n]) = O(\text{snr}_{\text{dl}}^{-\alpha})$, where $\hat{\tilde{\mathbf{h}}}[n]$ is the estimate of $\tilde{\mathbf{h}}[n]$ (the channel over subcarrier n).

The quality scaling exponent is related to the system DoF as follows. It is known that for a multi-user system with $K \leq M$ UEs, if for some $\alpha \in [0, 1]$ the MSE in estimating the CSIT decays as $O(\text{snr}_{\text{dl}}^{-\alpha})$, then ZF precoding achieves a total DoF of $K\alpha$, with $\alpha = 1$ corresponding to the full DoF [19]. One can achieve a slightly improved DoF of $1 + (K-1)\alpha$ with rate-splitting [18], which is coincidentally also an upper-bound, i.e. no scheme can achieve a higher DoF. It is therefore apparent that, when $\alpha = 0$, we can achieve a maximum DoF of 1 with rate-splitting, whereas $\alpha = 1$ yields a full DoF of K .

The Feedback Channel

In what follows we model the UL as a MIMO-MAC channel where all the UEs send their feedback simultaneously to the BS. We assume for simplicity that the BS has perfect knowledge of the UL channel and we use the high-SNR capacity formula $C^{\text{ul}} = \log(1 + M\text{snr}_{\text{ul}})$ with a so-called diversity-multiplexing trade-off factor of one [4]. We assume the SNR in UL to be proportional to the DL SNR as $\text{snr}_{\text{ul}} = \kappa \text{snr}_{\text{dl}}$, with $\kappa > 0$ being a positive constant. The modeling of the feedback link as such was considered in [13] and is a simplifying assumption that allows for a meaningful and elegant development of the theory but is not fundamental, in the sense that one can obtain similar results by considering other feedback channel models.

III. LOWER BOUND ON CSIT ESTIMATION ERROR VIA RATE-DISTORTION THEORY

To derive a lower-bound on the MSE in estimating the channel for given training and feedback dimensions, we think of the UE as an *encoder* that aims at encoding a *source* that produces channel realizations \mathbf{h} , given the training noisy linear measurements of those realizations as in (2). Since the encoder does not have direct access to the source output, this problem is an example of *remote source coding* [20]. Following standard information theoretic notation, we can formulate this problem by modeling the source as a stationary sequence of vector symbols $\mathbb{h}^\Delta = \{\mathbf{h}^{(i)}\}_{i=1}^\Delta$ with distribution $\mathbf{h}^{(i)} \sim \mathcal{CN}(\mathbf{0}, \Sigma^h)$, where $\mathbf{h}^{(i)}$ denotes the channel in frame i and Δ is the total number of frames. The UE observes a sequence of measurements of the source as $\mathbf{y}^{\text{tr}\Delta} = \{\mathbf{y}^{\text{tr}(i)}\}_{i=1}^\Delta$, where from (2) we have $\mathbf{y}^{\text{tr}(i)} = \mathbf{h}^{(i)}\mathbf{H}\mathbf{X}^{\text{tr}} + \mathbf{z}^{\text{tr}(i)}$. Because we seek a lower-bound on the achievable error, we can assume that the UE knows the channel statistics and encodes the source given measurements over infinitely many blocks. Now, a $(2^{\Delta R}, \Delta)$ remote rate-distortion code consists of a sequence of encoding $f_\Delta : \mathbf{y}^{\text{tr}\Delta} \rightarrow [2^{\Delta R}]$, and decoding $g_\Delta : [2^{\Delta R}] \rightarrow \hat{\mathbb{h}}^\Delta$ functions, where $R \geq 0$ denotes the code rate and where $\hat{\mathbb{h}}^\Delta = \{\hat{\mathbf{h}}^{(i)}\}_{i=1}^\Delta$ is the sequence of channel estimates. A remote rate-distortion pair (R, D) is said to be achievable if there exists a sequence of $(2^{\Delta R}, \Delta)$ codes such

that $\lim_{\Delta \rightarrow \infty} \frac{1}{\Delta} \sum_{i=1}^\Delta d(\mathbf{h}^{(i)}, \hat{\mathbf{h}}^{(i)}) \leq D$, where $d(\cdot, \cdot)$ is defined in (4). The closure of all such pairs is the rate-distortion region.

Remark: Note that here the encoder is only required to yield an average error less than D over all frames. This is a weaker condition in comparison to the per-frame achievable error defined in the previous section as the condition $d(\mathbf{h}, \hat{\mathbf{h}}) \leq D$, in the sense that, if a feedback strategy achieves the latter, it also achieves the former. It follows that, for the same training and feedback rates, the achievable error with these assumptions serves as a strict lower-bound for the achievable error of all feedback strategies that operate over finite blocks of the source, in which the UE has no access to channel statistics, and consider the stronger notion of per-frame achievable error as defined in the previous section.

The remote rate-distortion function $R_{\mathbf{h}}^r(D)$ is the infimum of rates R such that (R, D) is in the rate-distortion region for given D , and shows the infimum number of bits needed to quantize a single channel vector to achieve an error of D . We derive an expression for the remote rate-distortion function via the following lemma.

Lemma 1 (remote rate-distortion function): For a fixed realization of the training matrix \mathbf{X}^{tr} , let $\Sigma^u = \Sigma^h \mathbf{X}^{\text{tr}} (\mathbf{X}^{\text{tr}\text{H}} \Sigma^h \mathbf{X}^{\text{tr}} + \mathbf{I})^{-1} \mathbf{X}^{\text{tr}\text{H}} \Sigma^h$ denote the covariance of the posterior mean of the channel given pilot measurements $\mathbf{u}^{(i)} = \mathbb{E}[\mathbf{h}^{(i)} | \mathbf{y}^{\text{tr}(i)}]$ and denote its eigenvalues by $\{\lambda_\ell^u\}_{\ell=1}^{MN}$. The remote rate-distortion function is given by

$$R_{\mathbf{h}}^r(D) = \sum_{\ell=1}^{MN} \left[\log \frac{\lambda_\ell^u}{\gamma} \right]_+, \text{ for } D \geq D_{\text{mmse}} \quad (5)$$

where γ is chosen such that $\sum_{\ell=1}^{MN} \min\{\gamma, \lambda_\ell^u\} = D - D_{\text{mmse}}$, and where $D_{\text{mmse}} = d(\mathbf{h}^{(i)}, \mathbf{u}^{(i)})$.

Proof: See Appendix A. ■

Accordingly, we can define the remote distortion-rate function $D_{\mathbf{h}}^r(R)$ as the infimum of errors D such that (R, D) is in the rate-distortion region for given R and is equivalent to the inverse of $R_{\mathbf{h}}^r(\cdot)$ [21].

With rate-distortion feedback strategy, the UE remotely encodes the channel using $R_{\mathbf{h}}^r(D)$ bits and sends the quantization index in the UL via a channel code. A direct application of the source-channel separation theorem with distortion (see [21], exercise 10.17) yields that an end-to-end error of D between the channel vector and its estimate at the BS via feedback over a channel of capacity C^{ul} is achievable if and only if $\beta_{\text{fb}} C^{\text{ul}} > R_{\mathbf{h}}^r(D)$. Thereby for a given feedback dimension β_{fb} , we can achieve an error of

$$D > D_{\mathbf{h}}^r(\beta_{\text{fb}} C^{\text{ul}}) \quad (6)$$

Therefore, for given β_{fb} and β_{tr} , the lower-bound on the achievable error is $D_{\mathbf{h}}^r(\beta_{\text{fb}} C^{\text{ul}})$, where the dependence on β_{tr} is implicit in the expression for the rate-distortion function $D_{\mathbf{h}}^r$. It seems very difficult to make this relation more explicit in the general case, but we make it explicit for the large SNR regime via the following theorem.

Theorem 1: The rate-distortion feedback strategy achieves a CSIT estimation error of $\Theta(\text{snr}_{\text{dl}}^{-\alpha_{\text{rd}}})$ with probability one

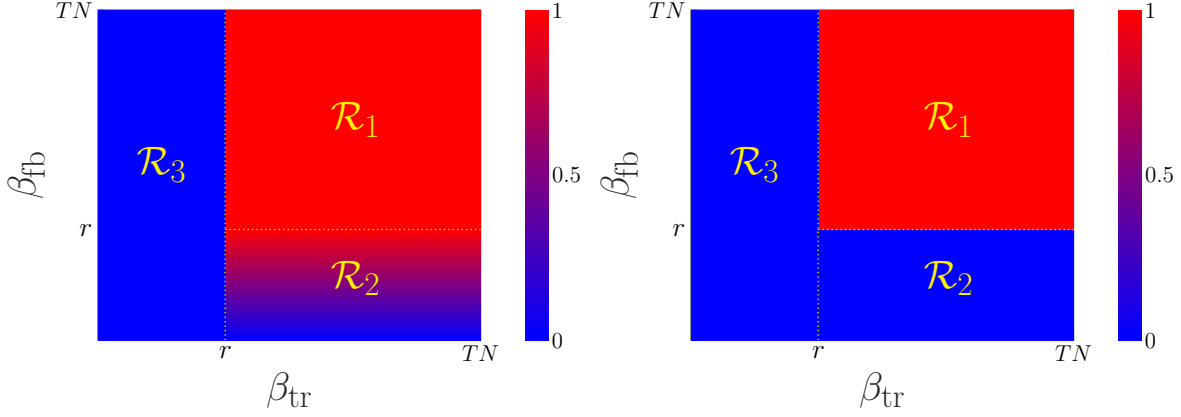


Fig. 1: The quality scaling exponent as a function of $(\beta_{\text{tr}}, \beta_{\text{fb}})$, represented as a heat-map. The left figure corresponds to the optimal rate-distortion feedback α_{rd} , whereas the right figure corresponds to AF α_{af} . Three regions for training and feedback dimension parameters are distinguished by \mathcal{R}_1 , \mathcal{R}_2 and \mathcal{R}_3 . In regions \mathcal{R}_1 and \mathcal{R}_3 the quality scaling exponent of AF coincides with that of the rate-distortion feedback, whereas in region \mathcal{R}_2 it is strictly sub-optimal ($\alpha_{\text{af}} = 0$ vs $\alpha_{\text{rd}} = \beta_{\text{fb}}/r$).

over the realizations of the training matrix \mathbf{X}^{tr} , where

$$\alpha_{\text{rd}} = \min(\beta_{\text{fb}}/r, 1) \mathbf{1}_{[r, \infty)}(\beta_{\text{tr}}). \quad (7)$$

is the quality scaling exponent.

Proof: See Appendix B. ■

Note that in this theorem $r = \text{rank}(\Sigma^h)$ is the channel covariance rank. The quality scaling exponent of the rate-distortion quantizer is an upper-bound on the quality scaling exponent of all feedback strategies and is illustrated as a heat map in Fig. 1 (left). The resulting system DoF with rate-splitting in this case is given by

$$\text{DoF}_{\text{rd}} = 1 + \mathbf{1}_{[r, \infty)}(\beta_{\text{tr}}) \min(\beta_{\text{fb}}/r, 1)(K - 1). \quad (8)$$

Entropy-Coded Scalar Quantization and Feedback

The rate-distortion bound is achieved by vector quantization applied to a large block of MMSE channel estimates $\{\mathbf{u}^{(i)}\}_{i=1}^{\Delta}$. This is impractical, given the high channel and block dimensions and the notorious difficulty of designing optimal vector quantizers. Therefore, given channel covariance knowledge, to produce the feedback message the UE can employ a much simpler *entropy-coded scalar quantizer* (ECSQ) as follows. Given the training vector, the UE computes the MMSE channel estimate $\mathbf{u} = \mathbb{E}[\mathbf{h}|\mathbf{y}^{\text{tr}}]$. The Karhunen-Loève (KL) expansion of \mathbf{u} can be written as $\mathbf{u} = \sum_{\ell=1}^{MN} w_{\ell} \mathbf{g}_{\ell}$, where \mathbf{g}_{ℓ} are eigenvectors of the covariance Σ^u and $w_{\ell} \sim \mathcal{CN}(0, \lambda_{\ell}^u)$ are complex Gaussian coefficients with variance λ_{ℓ}^u , equivalent to the eigenvalues of Σ^u . The idea is to quantize the vector of coefficients $\mathbf{w} = [w_1, \dots, w_{MN}]^T$, component by component, so as to achieve an error $d(\mathbf{u}, \hat{\mathbf{u}}) = \mathbb{E}[\|\mathbf{u} - \hat{\mathbf{u}}\|^2] = \mathbb{E}[\|\mathbf{w} - \hat{\mathbf{w}}\|^2] \leq D - D_{\text{mmse}}$. Inspired by the *reverse water filling* formulation used to derive the remote rate-distortion function in (5) (see proof of Lemma 1), the UE only quantizes those coefficients w_{ℓ} for which $\lambda_{\ell}^u \geq \gamma$, where γ is chosen such that $\sum_{\ell=1}^{MN} \min\{\gamma, \lambda_{\ell}^u\} = D - D_{\text{mmse}}$. With this choice, the error of quantizing each coefficient is given by $E[|w_{\ell} - \hat{w}_{\ell}|^2] = \min(\lambda_{\ell}^u, \gamma)$ and the associated

rate is $\lceil \log \frac{\lambda_{\ell}^u}{\gamma} \rceil_+$ bits. Now, instead of using the optimal rate-distortion vector quantizer, the UE can use a *dithered* scalar quantizer to encode coefficients with variance above γ . Using a classic result from Ziv [7], one can show that this quantizer achieves an error γ with a number of bits $b_{\ell} = \log \frac{\lambda_{\ell}^u}{\gamma} + 1.508$. No bits are assigned to coefficients for which $\lambda_{\ell}^u < \gamma$. Since both the UE and the BS know the channel covariance, there is no need to encode the position of quantized coefficients. Therefore, the total number of feedback bits to achieve an error D with ECSQ can be computed as

$$R_{\text{scalar}}(D) = \sum_{\ell: \lambda_{\ell}^u \geq \gamma} \left(\log \frac{\lambda_{\ell}^u}{\gamma} + 1.508 \right). \quad (9)$$

ECSQ can be seen as a practical quantizer that can have a performance close to optimal (in the sense of achievable error), when the UE has access to the channel covariance knowledge and at the price of higher feedback rate.

IV. ANALOG FEEDBACK

In analog feedback the UE extracts its β_{tr} received DL pilot symbols \mathbf{y}^{tr} from the DL training and feeds them back to the BS via Quadrature Amplitude Modulation (QAM) symbols with unquantized I and Q components via $\beta_{\text{fb}} = \zeta \beta_{\text{tr}}$ channel symbols [8, 9]. In particular, the training vector is modulated by a full-rank matrix Ψ of dimension $\beta_{\text{tr}} \times \beta_{\text{fb}}$, known to both UE and BS. The received feedback at the BS is given by

$$\mathbf{y}^{\text{af}} = \mathbf{y}^{\text{tr}} \Psi + \tilde{\mathbf{z}} = \mathbf{h}^H \mathbf{X}^{\text{tr}} \Psi + \mathbf{z}^{\text{af}}, \quad (10)$$

where $\tilde{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the AWGN over the feedback channel and $\mathbf{z}^{\text{af}} = \mathbf{z}^{\text{tr}} \Psi + \tilde{\mathbf{z}}$. The scalar $\zeta = \beta_{\text{fb}}/\beta_{\text{tr}}$ denotes the number of feedback channel uses per training coefficient. It is customary to choose $\zeta \geq 1$ (so that each training symbol is fed back at least once) and to select Ψ to be a unitary “spreading” matrix ($\Psi \Psi^H = \mathbf{I}$) [4, 10]. However, we allow ζ to be any positive value to keep the generality of the problem. Therefore Ψ is only required to be full-rank (not necessarily unitary), and we nevertheless call it the spreading matrix

for simplicity. Considering the MIMO-MAC capacity formula $C^{\text{ul}} = \log(1 + M\text{snr}_{\text{ul}})$ and from the feedback model (10), the feedback channel input has to satisfy the per-symbol average power constraint $\mathbb{E}[|x|^2] \leq M\text{snr}_{\text{ul}}$. Therefore, the i -th column of the spreading matrix ψ_i is chosen such that

$$\psi_i^H \Sigma_{\mathbf{y}^{\text{tr}}} \psi_i = M\text{snr}_{\text{ul}}, \quad i \in [\beta_{\text{fb}}], \quad (11)$$

where $\Sigma_{\mathbf{y}^{\text{tr}}} = \mathbb{E}[\mathbf{y}^{\text{tr}} \mathbf{y}^{\text{tr}H}] = \mathbf{X}^{\text{tr}H} \Sigma^h \mathbf{X}^{\text{tr}} + \mathbf{I}$ is the covariance of the training vector. Note that selecting a set of β_{fb} vectors that satisfy (11) and which contain a subset of $\min(\beta_{\text{tr}}, \beta_{\text{fb}})$ linearly independent elements is always possible because $\Sigma_{\mathbf{y}^{\text{tr}}}$ is of rank β_{tr} . The BS computes the minimum MSE (MMSE) estimate of the full-dimensional channel given the feedback as

$$\hat{\mathbf{h}} = \mathbb{E}[\mathbf{h}|\mathbf{y}^{\text{af}}] = \Sigma^h \mathbf{X}^{\text{tr}} \Psi \Sigma_{\mathbf{y}^{\text{af}}}^{-1} \mathbf{y}^{\text{af}}, \quad (12)$$

where $\Sigma_{\mathbf{y}^{\text{af}}} = \Psi^H \mathbf{X}^{\text{tr}H} \Sigma^h \mathbf{X}^{\text{tr}} \Psi + \Psi^H \Psi + \mathbf{I}$. Note that unlike the rate-distortion quantizer and the ECSQ, with AF we do *not* assume channel covariance knowledge at the UE. The CSIT estimation error with AF can be computed as

$$\begin{aligned} D &= \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2] \\ &= \text{Tr} \left(\Sigma^h - \Sigma^h \mathbf{X}^{\text{tr}} \Psi \Sigma_{\mathbf{y}^{\text{af}}}^{-1} \Psi^H \mathbf{X}^{\text{tr}H} \Sigma^h \right) \end{aligned} \quad (13)$$

where $\text{Tr}(\cdot)$ denotes the trace. The following theorem shows the scaling law of this error for large SNR.

Theorem 2: The analog feedback strategy achieves a CSIT estimation error of $\Theta(\text{snr}_{\text{dl}}^{-\alpha_{\text{af}}})$ with probability one over the realizations of the training matrix \mathbf{X}^{tr} , where

$$\alpha_{\text{af}} = \mathbf{1}_{[r, \infty)}(\min(\beta_{\text{tr}}, \beta_{\text{fb}})), \quad (14)$$

is the quality scaling exponent.

Proof: See Appendix C. ■

The resulting system DoF with AF is given as

$$\text{DoF}_{\text{af}} = 1 + \mathbf{1}_{[r, \infty)}(\min(\beta_{\text{tr}}, \beta_{\text{fb}}))(K - 1). \quad (15)$$

The quality scaling exponent of AF is illustrated as a heat map in Fig. 1 (right), where a comparison between the right and left figures shows that the exponent achieved by AF is the same as that of the rate-distortion quantizer for all training and feedback dimensions $(\beta_{\text{tr}}, \beta_{\text{fb}})$ belonging to regions \mathcal{R}_1 and \mathcal{R}_3 . In region \mathcal{R}_2 , rate-distortion feedback achieves an exponent of $\beta_{\text{fb}}/r > 0$, whereas AF has exponent zero, and is therefore strictly sub-optimal.

V. NUMERICAL RESULTS

We consider a ULA with $M = 32$ antennas at the BS communicating with $K = 6$ UEs over a total of $N = 24$ OFDM subcarriers. The channel coherence time is assumed to be 5 ms, corresponding to $T = 5 \times 14 = 70$ OFDM symbols in LTE [22].² We consider $N_p = 4$ pilot subcarriers, uniformly placed one per 6 subcarriers. Different training dimensions are considered by varying the number of pilots sent per pilot subcarrier, i.e. by changing the variable T_p . With these parameters, the isotropic pilot vectors are generated according to (3). We produce UE channels according to a

multipath model [23, 24], as

$$h_{m,n} = \sum_{\ell=1}^L c_{\ell} e^{j\pi m \frac{2d}{\lambda} \sin \theta_{\ell}} e^{-j2\pi n \Delta f \tau_{\ell}}, \quad (16)$$

where L denotes the number of paths, $\theta_{\ell} \in [-\pi/2, \pi/2]$ is the angle-of-arrival (AoA) of the ℓ -th signal path, $\tau_{\ell} \in [0, \tau_{\text{max}}]$ is the path delay where τ_{max} is the maximum delay spread of the channel which is bounded by the length of the OFDM cyclic prefix, $c_{\ell} \in \mathbb{C}$ is the complex gain of path ℓ and Δf is the subcarrier spacing (assumed uniform), λ is the wavelength corresponding to the central carrier frequency, and d is the uniform spacing between array elements, taken to be $d = \lambda/2$ for simplicity. The path AoAs and delays are chosen uniformly at random and gains are generated as standard Gaussian random variables. The resulting channel covariance is normalized such that $\text{Tr}(\Sigma_h) = MN$. The number of paths L is the same for all UEs, but the AoAs, delays and gains are generated independently across UEs. Because the AoAs and delays are randomly generated, the channel covariance rank is given, with probability one, by $r = \min(L, MN)$.

We first study the CSIT estimation MSE performance as a function of SNR for all the feedback strategies. We generate 10 realizations of training matrices and K random covariances as above, and for each realization, we generate 100 random instances of each UE's channel. Here the covariance rank is set to $r = 30$. The *average* MSE is computed as

$$\text{MSE}_{\text{avg}} = \frac{1}{K} \sum_k \mathbb{E}[\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2],$$

where the mean $\mathbb{E}[\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2]$ is empirically calculated from the random realizations of training matrices, covariances and channels. We plot the the average MSE against DL SNR for two points in the $\beta_{\text{tr}} - \beta_{\text{fb}}$ plane, namely $(\beta_{\text{tr}}, \beta_{\text{fb}}) = (40, 40)$ and $(40, 10)$. From (7) and (14) we expect that all feedback methods achieve a quality scaling exponent of $\alpha = 1$ since $\beta_{\text{tr}} = \beta_{\text{fb}} > r$. In the second case however, we expect that the rate-distortion and ECSQ achieve a quality scaling exponent of $\alpha = \beta_{\text{fb}}/r = 1/3$, while AF achieves $\alpha = 0$ because $\beta_{\text{tr}} > r > \beta_{\text{fb}}$. These are confirmed by the average MSE vs SNR curves of Fig. 2, where the slope of the curves in high SNR is equivalent to the quality scaling exponent. In the first case, AF achieves the optimal quality scaling exponent because both training and feedback dimensions exceed the covariance rank. In the second case, the rate-distortion and ECSQ feedbacks achieve a non-zero exponent, unlike AF which has a constant error ($\alpha = 0$) even for large SNR. Note also that ECSQ yields an error close to the optimal, with a performance gap between the two that widens when the feedback dimension is lower. This is expected, since when the feedback rate is higher, the gap between the error achieved by scalar quantization and the optimal error is lower.

In the second experiment, we compare the performance of different feedback strategies in terms of achievable Downlink sum-rate in a multi-user system. Here we consider ZF precoding in DL, where the transmit data vector over subcarrier n is given by $\mathbf{x}^{\text{d}}[n] = \sum_{k=1}^K \sqrt{P_k} s_k[n] \mathbf{v}_k[n]$,

²The coherence time can vary to emulate fast-varying (smaller T) and slow-varying (large T) channels.

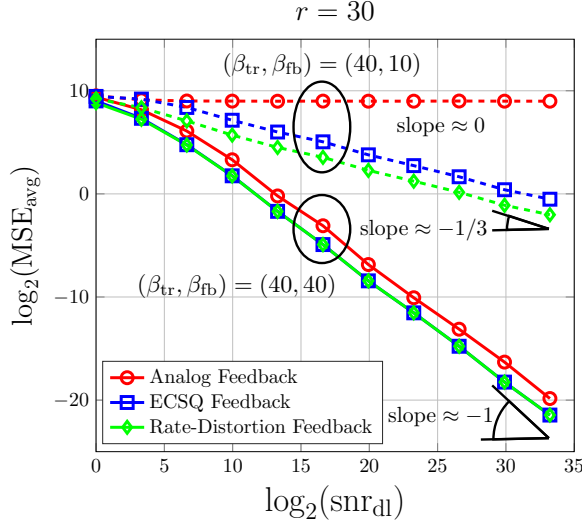


Fig. 2: CSIT estimation MSE vs Downlink SNR for the three feedback strategies, where $M = 32$, $N = 24$ and $K = 6$.

where $P_k = \text{snr}_{\text{dl}}/K$ is the (uniform) transmission power per UE, $s_k[n] \in \mathbb{C}$ is the data symbol intended for UE k such that $\mathbb{E}[|s_k[n]|^2] \leq 1$, and $\{\mathbf{v}_k[n]\}_{k=1}^K$ are the precoding vectors, given by the column-normalized pseudo-inverse of $\tilde{\mathbf{H}}[n] = [\tilde{\mathbf{h}}_1[n], \dots, \tilde{\mathbf{h}}_K[n]]^H$. The choice of ZF precoding (rather than rate-splitting) is for the sake of simplicity and the fact that it is by far the most practical scheme used in real systems. Defining the variables $g_{k,k'}[n] = \sqrt{P_{k'}}(\tilde{\mathbf{h}}_k[n]^H \mathbf{v}_{k'}[n])$, we can write the achievable ergodic rate³ as [25]

$$R_{\text{ub}}[k, n] = \mathbb{E} \left[\log \left(1 + \frac{|g_{k,k}[n]|^2}{N_0 + \sum_{k' \neq k} |g_{k,k'}[n]|^2} \right) \right]. \quad (17)$$

The average sum-rate is computed by averaging (17) over all subcarriers, distinguishing between pilot and data subcarriers, and summing the result over all UEs. Fig. 3 illustrates two sets of curves, comparing the sum-rate vs training dimension for the three feedback strategies. In the first set, we have assumed a high DL SNR value of 50 dBs and we have set $\zeta = \beta_{\text{fb}}/\beta_{\text{tr}} = 1$, so that at each point of the associated curves, the number of training and feedback dimensions are equal. This corresponds to a line trajectory in the $\beta_{\text{tr}} - \beta_{\text{fb}}$ plane. From Fig. 1 we expect that, moving along the line $\beta_{\text{fb}} = \beta_{\text{tr}}$, AF achieves the same quality scaling exponent as the optimal rate-distortion feedback, which implies that in high SNR the two feedback strategies should have close rate performance. This is confirmed by the first set of curves in Fig. 3 where we see that AF achieves a sum-rate that is very close to that of the rate-distortion feedback.

The second set of curves corresponds to a moderate DL SNR of 20 dBs, and $\zeta = 1/4$, which means that for each point of the curves the feedback dimension is taken to be 1/4 the training dimension (rounded up when $\beta_{\text{tr}}/4$ is not an integer). In this case, we have a noticeable gap between the sum-rate

³This rate can be achieved assuming perfect knowledge of the coefficients $\{g_{k,k'}[n]\}$ at UE k , and is a harmless assumption for our purposes.

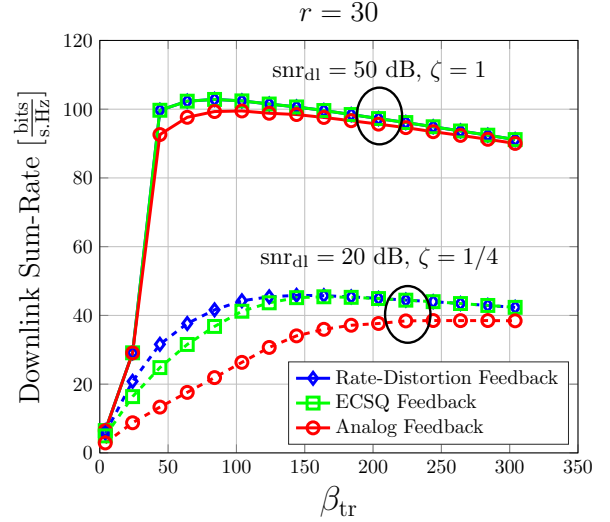


Fig. 3: Achievable sum-rate vs training dimension for the three feedback strategies, where $M = 32$, $N = 24$ and $K = 6$.

with AF and that of the optimal because the SNR is set to a moderate value and more importantly, because of the fact that in this case for training dimension values of $\beta_{\text{tr}} \in (30, 120)$, we have $\beta_{\text{fb}} = \beta_{\text{tr}}/4 \in (7, 30)$, which means that for these points we have $(\beta_{\text{tr}}, \beta_{\text{fb}}) \in \mathcal{R}_2$ (see Fig. 1). In the region \mathcal{R}_2 , AF is *strictly* suboptimal, in the sense that while the optimal quality scaling exponent is $\alpha_{\text{rd}} = \beta_{\text{fb}}/r \in (1/4, 1)$, for AF it is $\alpha_{\text{af}} = 0$.

Note also the gap between the sum-rate with ECSQ and the optimal in the two sets of curves in Fig. 3. In the first set, because we have a large SNR and a large feedback rate ($\zeta = 1$ in comparison to $\zeta = 1/4$), the CSIT error achieved by ECSQ for the same pair of training and feedback dimensions is very close to the optimal. However, in the second case we have a relative shortage of feedback bits, so that the 1.5 bit per quantized coefficient overhead of ECSQ corresponds to a noticeable CSIT error and we see a larger gap between the sum-rates. Nevertheless, ECSQ is still close to the optimal and performs better than AF, which comes at the price of channel covariance knowledge at the UE side. Therefore there exists a decision point where we can choose the ECSQ feedback in case the channel covariance is available at the UE and the AF in case it is not.

APPENDIX

A. Proof of Lemma 1

We start by stating a few standard results regarding the (remote) rate-distortion function. It is well-known that the rate-distortion function of an i.i.d source represented by the random variable \mathbf{u} with distribution $p_{\mathbf{u}}$ can be computed as (see [26] Theorem 10.2.1)

$$R_{\mathbf{u}}(D) = \min_{p_{\bar{\mathbf{u}}|\mathbf{u}}: d(\bar{\mathbf{u}}, \mathbf{u}) \leq D} I(\bar{\mathbf{u}}; \mathbf{u}), \quad (18)$$

where $\bar{\mathbf{u}}$ is the quantization of \mathbf{u} , $I(\bar{\mathbf{u}}; \mathbf{u})$ is the mutual information between \mathbf{u} and $\bar{\mathbf{u}}$, $d(\bar{\mathbf{u}}, \mathbf{u})$ is the distortion between \mathbf{u}

and $\bar{\mathbf{u}}$ (see (4)) and the minimum is taken over all conditional distributions for which the joint distribution $p_{\mathbf{u}, \bar{\mathbf{u}}}$ satisfies the distortion constraint. It is also known that the remote rate distortion function of a source represented by the random variable \mathbf{h} , and encoded given its observations denoted by the random variable \mathbf{y}^{tr} is given by (see [20])

$$R_{\mathbf{h}}^r(D) = \min_{p_{\bar{\mathbf{h}}|\mathbf{y}^{\text{tr}}}: d(\bar{\mathbf{h}}, \mathbf{h}) \leq D} I(\bar{\mathbf{h}}; \mathbf{y}^{\text{tr}}) \quad (19)$$

where $\bar{\mathbf{h}}$ is the quantization of \mathbf{h} , $I(\bar{\mathbf{h}}; \mathbf{y}^{\text{tr}})$ is the mutual information between $\bar{\mathbf{h}}$ and \mathbf{y}^{tr} , and the minimum is taken over all conditional distributions $p_{\bar{\mathbf{h}}|\mathbf{y}^{\text{tr}}}$ for which the joint distribution $p_{\mathbf{h}, \bar{\mathbf{h}}} = p_{\mathbf{h}} p_{\mathbf{y}^{\text{tr}}|\mathbf{h}} p_{\bar{\mathbf{h}}|\mathbf{y}^{\text{tr}}}$ satisfies the distortion constraint. Note that since all sources are i.i.d, we have removed realization index superscripts from the variables (hence \mathbf{h} instead of $\mathbf{h}^{(i)}$). From the premise of the lemma, \mathbf{u} is the MMSE estimate of the channel given the training measurements, i.e. $\mathbf{u} = \mathbb{E}[\mathbf{h}|\mathbf{y}^{\text{tr}}]$. Using the same technique employed to prove inequality (15) of [27] (see Appendix A in [27]), we can verify that the remote rate-distortion function of \mathbf{h} is related to the rate distortion function of \mathbf{u} by

$$R_{\mathbf{h}}^r(D) = R_{\mathbf{u}}(D - D_{\text{mmse}}), \quad (20)$$

for $D \geq D_{\text{mmse}}$, where $D_{\text{mmse}} = \mathbb{E}[\|\mathbf{h} - \mathbf{u}\|^2]$ is the MMSE of estimating the channel at the UE.

On the other hand, the rate-distortion function of a correlated vector Gaussian source is given by reverse water-filling over its covariance eigenvalues [21]. If we denote the eigenvalues of Σ^u by $\{\lambda_{\ell}^u\}_{\ell=1}^{MN}$, then we have $R_{\mathbf{u}}(D) = \sum_{\ell=1}^{MN} \left[\log \frac{\lambda_{\ell}^u}{\gamma} \right]_+$, where γ is chosen such that $\sum_{\ell=1}^{MN} \min\{\gamma, \lambda_{\ell}^u\} = D$. Plugging this in (20) we get

$$R_{\mathbf{h}}^r(D) = \sum_{\ell=1}^{MN} \left[\log \frac{\lambda_{\ell}^u}{\gamma} \right]_+, \quad (21)$$

where γ is chosen such that $\sum_{\ell=1}^{MN} \min\{\gamma, \lambda_{\ell}^u\} = D - D_{\text{mmse}}$. The proof is complete. \blacksquare

B. Proof of Theorem 1

We divide the proof to two parts. In the first part, we show that if $\beta_{\text{tr}} < r$, then the achievable error behaves as $\Theta(1)$ for all realizations of \mathbf{X}^{tr} . In the second part we show that if $\beta_{\text{tr}} \geq r$, then an error decaying as $\Theta(\text{snr}_{\text{dl}}^{-\min(\beta_{\text{tr}}/r, 1)})$ is achievable with probability one over the realizations of \mathbf{X}^{tr} .

Part I. To prove part I, we first bound the minimum mean squared error (MMSE) of estimating the channel given the training measurements at the UE, namely the variable D_{mmse} . From $\mathbf{u} = \mathbb{E}[\mathbf{h}|\mathbf{y}^{\text{tr}}]$ we have

$$D_{\text{mmse}} = \text{Tr}(\mathbb{E}[\mathbf{h}\mathbf{h}^H] - \mathbb{E}[\mathbf{h}\mathbf{y}^{\text{tr}}]\mathbb{E}[\mathbf{y}^{\text{tr}H}\mathbf{y}^{\text{tr}}]^{-1}\mathbb{E}[\mathbf{h}\mathbf{y}^{\text{tr}}]^H), \quad (22)$$

where $\mathbb{E}[\mathbf{h}\mathbf{h}^H] = \Sigma^h$, $\mathbb{E}[\mathbf{h}\mathbf{y}^{\text{tr}}] = \Sigma^h \mathbf{X}^{\text{tr}}$, and $\mathbb{E}[\mathbf{y}^{\text{tr}H}\mathbf{y}^{\text{tr}}] = \mathbf{X}^{\text{tr}H} \Sigma^h \mathbf{X}^{\text{tr}} + \mathbf{I}$. The eigendecomposition of Σ^h can be written as $\Sigma^h = \mathbf{U}_h \Lambda_h \mathbf{U}_h^H$, where $\mathbf{U}_h \in \mathbb{C}^{MN \times r}$ is a tall unitary matrix and $\Lambda_h = \text{diag}(\boldsymbol{\lambda}) \in \mathbb{R}^{r \times r}$ is a diagonal matrix of positive eigenvalues represented by the vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_r]^T$. Using this decomposition and applying the Sherman-Morrison-

Woodbury matrix identity to $(\mathbf{X}^{\text{tr}H} \Sigma^h \mathbf{X}^{\text{tr}} + \mathbf{I})^{-1}$, we have

$$\mathbb{E}[\mathbf{h}\mathbf{y}^{\text{tr}}]\mathbb{E}[\mathbf{y}^{\text{tr}H}\mathbf{y}^{\text{tr}}]^{-1}\mathbb{E}[\mathbf{h}\mathbf{y}^{\text{tr}}]^H = \mathbf{U}_h \Lambda_h^{1/2} \mathbf{G} \Lambda_h^{1/2} \mathbf{U}_h^H - \mathbf{U}_h \Lambda_h^{1/2} \mathbf{G} (\mathbf{I} + \mathbf{G})^{-1} \mathbf{G} \Lambda_h^{1/2} \mathbf{U}_h^H, \quad (23)$$

where we have defined

$$\mathbf{G} = \Lambda_h^{1/2} \mathbf{U}_h^H \mathbf{X}^{\text{tr}} \mathbf{X}^{\text{tr}H} \mathbf{U}_h \Lambda_h^{1/2}. \quad (24)$$

Plugging this into (22) we have

$$D_{\text{mmse}} = \text{Tr}(\Lambda_h (\mathbf{I} - \mathbf{G} + \mathbf{G}(\mathbf{I} + \mathbf{G})^{-1} \mathbf{G})). \quad (25)$$

Using a simple trace inequality, one can show that

$$\lambda_{\min} g(\text{snr}_{\text{dl}}) \leq D_{\text{mmse}} \leq \lambda_{\max} g(\text{snr}_{\text{dl}}), \quad (26)$$

where λ_{\min} and λ_{\max} are minimum and maximum channel covariance eigenvalues, respectively, and we have defined $g(\text{snr}_{\text{dl}}) = \text{Tr}(\mathbf{I} - \mathbf{G} + \mathbf{G}(\mathbf{I} + \mathbf{G})^{-1} \mathbf{G})$, where we have made the dependency of $g(\cdot)$ on snr_{dl} explicit. We now demonstrate this dependency. For a given realization of the training matrix \mathbf{X}^{tr} , denote the eigenvalues of \mathbf{G} by μ_i , $i = 1, \dots, r$. We can write

$$\begin{aligned} g(\text{snr}_{\text{dl}}) &= r - \sum_i \mu_i + \sum_i \frac{\mu_i^2}{\mu_i + 1} \\ &= r - \sum_{i=1}^r \frac{\mu_i}{\mu_i + 1} \end{aligned} \quad (27)$$

Also note that we can represent the training matrix as $\mathbf{X}^{\text{tr}} = \sqrt{\text{snr}_{\text{dl}}} \mathbf{X}_0^{\text{tr}}$, where \mathbf{X}_0^{tr} is randomly generated and independent from snr_{dl} . From this and the definition (24), it follows that $\mu_i = \Theta(\text{snr}_{\text{dl}})$ for all $\mu_i \neq 0$. Using this and (27), we deduce that if \mathbf{G} is full-rank ($\mu_i \neq 0$ for all i) then $g(\text{snr}_{\text{dl}}) = \Theta(\text{snr}_{\text{dl}}^{-1})$ and using (26) we have $D_{\text{mmse}} = \Theta(\text{snr}_{\text{dl}}^{-1})$. Conversely, if \mathbf{G} has at least one zero eigenvalue ($\mu_i = 0$ for some i) then from (27) we have $g(\text{snr}_{\text{dl}}) > 1$ and from (41) we have $D_{\text{mmse}} = \Theta(1)$.

Now, the rank of \mathbf{G} depends on the specific realization of \mathbf{X}^{tr} . When $\beta_{\text{tr}} \geq r$ and \mathbf{X}^{tr} consists of Gaussian isotropic pilot vectors, \mathbf{G} is full-rank with probability one because of the following. The product $\mathbf{U}_h^H \mathbf{X}^{\text{tr}}$ consists of β_{tr} independent Gaussian columns, each of dimension r . The event that these vectors span a space of dimension less than r has probability zero. Therefore, $\mathbf{U}_h^H \mathbf{X}^{\text{tr}} \mathbf{X}^{\text{tr}H} \mathbf{U}_h$ has rank r with probability one, and since $\Lambda_h^{1/2}$ has positive diagonal elements, by definition (24) \mathbf{G} also has rank r with probability one and $\mu_i \neq 0$ for all i . Conversely, if $\beta_{\text{tr}} < r$, \mathbf{G} has rank at most equal to β_{tr} for any realization of the training matrix, leading to $\mu_i = 0$ for some i . This results in $D_{\text{mmse}} = \Theta(1)$. In short, we have proved

$$D_{\text{mmse}} = \begin{cases} \Theta(\text{snr}_{\text{dl}}^{-1}), & \beta_{\text{tr}} \geq r, \\ \Theta(1), & \beta_{\text{tr}} < r. \end{cases} \quad (28)$$

In addition, Lemma 1 states that only errors $D \geq D_{\text{mmse}}$ are achievable. It follows that, if $\beta_{\text{tr}} < r$, then the minimum achievable error in estimating the CSIT behaves as $\Theta(1)$.

Part II. To prove the second part, first note that if $\beta_{\text{tr}} \geq r$, then the covariance of the MMSE channel estimate \mathbf{u} at the

UE, given as

$$\Sigma^u = \Sigma^h \mathbf{X}^{\text{tr}} \left(\mathbf{X}^{\text{trH}} \Sigma^h \mathbf{X}^{\text{tr}} + \mathbf{I} \right)^{-1} \mathbf{X}^{\text{trH}} \Sigma^h \quad (29)$$

has rank r with probability one over the realizations of \mathbf{X}^{tr} . Without loss of generality assume the eigenvalues of Σ^u to be ordered as $\lambda_1^u \geq \dots \geq \lambda_r^u > 0$. Next, consider the remote rate-distortion function in Lemma 1, given as

$$R_{\mathbf{h}}^r(D) = \sum_{\ell=1}^{MN} \left[\log \frac{\lambda_{\ell}^u}{\gamma} \right]_+, \text{ for } D \geq D_{\text{mmse}} \quad (30)$$

where γ is chosen such that $\sum_{\ell=1}^{MN} \min\{\gamma, \lambda_{\ell}^u\} = D - D_{\text{mmse}}$. Consider an interval of error values D for which $D - D_{\text{mmse}} < \varepsilon$ for some ε . If ε is sufficiently small, then $\gamma = (D - D_{\text{mmse}})/r$ and the remote rate-distortion function is given by

$$R_{\mathbf{h}}^r(D) = f(r) - r \log(D - D_{\text{mmse}}), \quad (31)$$

where $f(r) = \sum_{\ell=1}^r \log \lambda_{\ell}^u + r \log r$ is a value independent of D . Therefore, we can write the remote distortion-rate function as

$$D_{\mathbf{h}}^r(R) = 2^{\frac{f(r)-R}{r}} + D_{\text{mmse}}. \quad (32)$$

for all $R > R_{\varepsilon}$, for a sufficiently large $R_{\varepsilon} > 0$. Now let $R = \beta_{\text{fb}} C^{\text{ul}}$. Replacing the MIMO-MAC capacity formula $C^{\text{ul}} = \log(1 + M\kappa \text{snr}_{\text{dl}})$, we notice that there exists some $\text{snr}_{\text{dl}}^{\varepsilon}$ such that $\beta_{\text{fb}} \log(1 + M\kappa \text{snr}_{\text{dl}}) > R_{\varepsilon}$ for all $\text{snr}_{\text{dl}} > \text{snr}_{\text{dl}}^{\varepsilon}$. For these snr_{dl} values we have we have

$$\begin{aligned} \log(D_{\mathbf{h}}^r(\beta_{\text{fb}} C^{\text{ul}}) - D_{\text{mmse}}) &= \log r + \sum_{\ell=1}^r \log \lambda_{\ell}^u / r \\ &\quad - \frac{\beta_{\text{fb}}}{r} \log(1 + M\kappa \text{snr}_{\text{dl}}). \end{aligned} \quad (33)$$

From (29) one can show that the non-zero eigenvalues of Σ^u scale as $\Theta(1)$ for large snr_{dl} , i.e. $\lambda_{\ell}^u = \Theta(1)$, $\ell = 1, \dots, r$. Therefore, the right-hand-side of (33) behaves as $\Theta(\log(\text{snr}_{\text{dl}}^{-\beta_{\text{fb}}/r}))$ in snr_{dl} . It follows that

$$\begin{aligned} D_{\mathbf{h}}^r(\beta_{\text{fb}} C^{\text{ul}}) &= D_{\text{mmse}} + \Theta(\text{snr}_{\text{dl}}^{-\beta_{\text{fb}}/r}) \\ &= \Theta(\text{snr}_{\text{dl}}^{-1}) + \Theta(\text{snr}_{\text{dl}}^{-\beta_{\text{fb}}/r}) \\ &= \Theta(\text{snr}_{\text{dl}}^{-\min(\beta_{\text{fb}}/r, 1)}) \end{aligned} \quad (34)$$

Finally, from the source-channel separation with distortion theorem, we can achieve a CSIT estimation error of $D > D_{\mathbf{h}}^r(\beta_{\text{fb}} C^{\text{ul}})$ if and only if we use the UL channel over β_{fb} feedback dimensions (see Section III), which combined with (34) shows that when $\beta_{\text{tr}} \geq r$, we can achieve an error decay of $\Theta(\text{snr}_{\text{dl}}^{-\min(\beta_{\text{fb}}/r, 1)})$ with probability one over the realizations of \mathbf{X}^{tr} with a feedback dimension of β_{fb} . This completes the proof. ■

C. Proof of Theorem 2

The distortion of MMSE estimation of the channel given the feedback in (10) is given by

$$d(\mathbf{h}, \hat{\mathbf{h}}) = \text{Tr} \left(\mathbb{E}[\mathbf{h}\mathbf{h}^{\text{H}}] - \mathbb{E}[\mathbf{h}\mathbf{y}^{\text{af}}] \mathbb{E}[\mathbf{y}^{\text{afH}} \mathbf{y}^{\text{af}}]^{-1} \mathbb{E}[\mathbf{h}\mathbf{y}^{\text{af}}]^{\text{H}} \right), \quad (35)$$

where $\Sigma^h = \mathbb{E}[\mathbf{h}\mathbf{h}^{\text{H}}]$, $\mathbb{E}[\mathbf{h}\mathbf{y}^{\text{af}}] = \Sigma^h \mathbf{X}^{\text{tr}} \Psi$, and

$$\mathbb{E}[\mathbf{y}^{\text{afH}} \mathbf{y}^{\text{af}}] = \Psi^{\text{H}} \mathbf{X}^{\text{trH}} \Sigma^h \mathbf{X}^{\text{tr}} \Psi + \Psi^{\text{H}} \Psi + \mathbf{I}, \quad (36)$$

The eigendecomposition of Σ^h can be written as $\Sigma^h = \mathbf{U}_h \Lambda_h \mathbf{U}_h^{\text{H}}$, where $\mathbf{U}_h \in \mathbb{C}^{MN \times r}$ is a tall unitary matrix and $\Lambda_h = \text{diag}(\boldsymbol{\lambda}) \in \mathbb{R}^{r \times r}$ is a diagonal matrix of positive eigenvalues represented by the vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_r]^{\text{T}}$. Using this decomposition, the expression in (36) and applying the Sherman-Morrison-Woodbury matrix identity we have

$$\begin{aligned} \mathbb{E}[\mathbf{y}^{\text{afH}} \mathbf{y}^{\text{af}}]^{-1} &= (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1} \\ &\quad - (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1} \Psi^{\text{H}} \mathbf{X}^{\text{trH}} \mathbf{U}_h \Lambda_h^{1/2} \times \\ &\quad \left(\mathbf{I} + \Lambda_h^{1/2} \mathbf{U}_h^{\text{H}} \mathbf{X}^{\text{tr}} \Psi (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1} \Psi^{\text{H}} \mathbf{X}^{\text{trH}} \mathbf{U}_h \Lambda_h^{1/2} \right)^{-1} \times \\ &\quad \Lambda_h^{1/2} \mathbf{U}_h^{\text{H}} \mathbf{X}^{\text{tr}} \Psi (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1}. \end{aligned} \quad (37)$$

Then we can write the second term appearing within the $\text{Tr}(\cdot)$ in (35) as

$$\begin{aligned} \mathbb{E}[\mathbf{h}\mathbf{y}^{\text{af}}] \mathbb{E}[\mathbf{y}^{\text{afH}} \mathbf{y}^{\text{af}}]^{-1} \mathbb{E}[\mathbf{h}\mathbf{y}^{\text{af}}]^{\text{H}} &= \mathbf{U}_h \Lambda_h^{1/2} \mathbf{G} \Lambda_h^{1/2} \mathbf{U}_h^{\text{H}} \\ &\quad - \mathbf{U}_h \Lambda_h^{1/2} \mathbf{G} (\mathbf{I} + \mathbf{G})^{-1} \mathbf{G} \Lambda_h^{1/2} \mathbf{U}_h^{\text{H}}, \end{aligned} \quad (38)$$

where we have defined

$$\mathbf{G} = \Lambda_h^{1/2} \mathbf{U}_h^{\text{H}} \mathbf{X}^{\text{tr}} \Psi (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1} \Psi^{\text{H}} \mathbf{X}^{\text{trH}} \mathbf{U}_h \Lambda_h^{1/2} \quad (39)$$

Plugging (38) into (35) we have

$$d(\mathbf{h}, \hat{\mathbf{h}}) = \text{Tr} \left(\Lambda_h (\mathbf{I} - \mathbf{G} + \mathbf{G}(\mathbf{I} + \mathbf{G})^{-1} \mathbf{G}) \right). \quad (40)$$

This formula is exactly the same as (25) except for the definition of \mathbf{G} . Therefore the same trace inequality as in (26) holds here for the CSI estimation distortion at the BS, i.e. we have

$$\lambda_{\min} g(\text{snr}_{\text{dl}}) \leq d(\mathbf{h}, \hat{\mathbf{h}}) \leq \lambda_{\max} g(\text{snr}_{\text{dl}}), \quad (41)$$

where $g(\text{snr}_{\text{dl}}) = \text{Tr}(\mathbf{I} - \mathbf{G} + \mathbf{G}(\mathbf{I} + \mathbf{G})^{-1} \mathbf{G})$. We now show how $g(\cdot)$ behaves as a function of snr_{dl} . For a given realization of \mathbf{X}^{tr} , denote the eigenvalues of \mathbf{G} by μ_i , $i = 1, \dots, r$. We can write

$$\begin{aligned} g(\text{snr}_{\text{dl}}) &= r - \sum_i \mu_i + \sum_i \frac{\mu_i^2}{\mu_i + 1} \\ &= r - \sum_i \frac{\mu_i}{\mu_i + 1} \end{aligned} \quad (42)$$

From the definition in (39), the constituents of \mathbf{G} depend on snr_{dl} as follows:

- We can represent the training matrix as $\mathbf{X}^{\text{tr}} = \sqrt{\text{snr}_{\text{dl}}} \mathbf{X}_0^{\text{tr}}$, where \mathbf{X}_0^{tr} is generated randomly independent from snr_{dl} . Hence, for a single realization the elements of \mathbf{X}^{tr} scale with snr_{dl} as $\Theta(\sqrt{\text{snr}_{\text{dl}}})$.
- From constraint (11), one can show that each column of Ψ can be written as $\psi_i = \sqrt{a_i} \phi_i$, where $a_i = \frac{M\kappa \text{snr}_{\text{dl}}}{c_i \text{snr}_{\text{dl}} + 1}$ with $c_i = \phi_i^{\text{H}} \mathbf{X}_0^{\text{trH}} \Sigma_h \mathbf{X}_0^{\text{tr}} \phi_i$. Here ϕ_i , $i \in [\beta_{\text{fb}}]$ are a set of unit-norm vectors that contains a subset of $\min(\beta_{\text{tr}}, \beta_{\text{fb}})$ linearly independent vectors, and are independent of snr_{dl} . The existence of this set is guaranteed because $\Sigma_{\mathbf{y}^{\text{tr}}}$ is full-rank. It follows that

$$\Psi (\Psi^{\text{H}} \Psi + \mathbf{I})^{-1} \Psi^{\text{H}} = \Phi (\Phi^{\text{H}} \Phi + \mathbf{S})^{-1} \Phi^{\text{H}}, \quad (43)$$

where $\Phi = [\phi_1, \dots, \phi_{\beta_{\text{fb}}}]$ is independent of snr_{dl} and \mathbf{S} is a diagonal matrix whose (i, i) element is given by $S_{i,i} = \frac{c_i}{M\kappa} + \frac{1}{M\kappa \text{snr}_{\text{dl}}}$. The matrix \mathbf{S} is the only variable dependent on snr_{dl} , and its diagonal elements

are asymptotically scaling as $S = \Theta(1)$.

(c) The matrices \mathbf{U}_h and $\mathbf{\Lambda}_h$ are independent of snr_{dl} .

From these we conclude that the non-zero eigenvalues of \mathbf{G} scale as $\Theta(\text{snr}_{\text{dl}})$, i.e. $\mu_i = \Theta(\text{snr}_{\text{dl}})$ for all $\mu_i \neq 0$.

Now, the rank of \mathbf{G} depends on the specific realization of \mathbf{X}^{tr} . When $\min(\beta_{\text{tr}}, \beta_{\text{fb}}) \geq r$ and \mathbf{X}^{tr} contains isotropic Gaussian pilot vectors, \mathbf{G} is full-rank with probability one because of the same argument as used in Part I of the proof of Theorem 1 and considering the fact that the constituent matrix $\Psi(\Psi^H \Psi + \mathbf{I})^{-1} \Psi^H$ is positive semi-definite with rank $\min(\beta_{\text{tr}}, \beta_{\text{fb}})$. In this case we have $d(\mathbf{h}, \hat{\mathbf{h}}) = \Theta(\text{snr}_{\text{dl}}^{-1})$ and therefore an error of $\Theta(\text{snr}_{\text{dl}}^{-1})$ is achievable. Conversely, if $r > \beta_{\text{tr}}$ or $r > \beta_{\text{fb}}$, \mathbf{G} has rank at most $\min(\beta_{\text{tr}}, \beta_{\text{fb}}) < r$ for any design of pilot matrices, leading to $\mu_i = 0$ for some i and from (41), the error is bounded from below and above by constants, i.e. we have an error of $\Theta(1)$. Therefore AF achieves an error of $\Theta(\text{snr}_{\text{dl}}^{-\alpha})$ with probability one over the realizations of \mathbf{X}^{tr} where $\alpha = \mathbf{1}_{[r, \infty)}(\min(\beta_{\text{tr}}, \beta_{\text{fb}}))$. This completes the proof. ■

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE communications magazine*, vol. 52, no. 2, pp. 74–80, 2014.
- [3] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *2006 fortieth asilomar conference on signals, systems and computers*. IEEE, 2006, pp. 359–363.
- [4] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical: Achievable rates with simple channel state estimation and feedback schemes," *Arxiv preprint cs. IT*, vol. 710, 2007.
- [5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on information theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE transactions on information theory*, vol. 44, no. 6, pp. 2619–2692, 1998.
- [7] J. Ziv, "On universal quantization," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 344–347, 1985.
- [8] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems," *IEEE transactions on communications*, vol. 54, no. 7, pp. 1335–1345, 2006.
- [9] T. A. Thomas, K. L. Baum, and P. Sartori, "Obtaining channel knowledge for closed-loop multi-stream broadband MIMO-OFDM communications using direct channel feedback," in *GLOBECOM'05. IEEE Global Telecommunications Conference, 2005.*, vol. 6. IEEE, 2005, pp. 5–pp.
- [10] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1268–1278, 2006.
- [11] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, 2011.
- [12] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero-forcing based achievable rates," in *2007 IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 2046–2050.
- [13] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, 2015.
- [14] H. Shirani-Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2713–2723, 2009.
- [15] S. Bazzi and W. Xu, "On the amount of downlink training in correlated massive MIMO channels," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2286–2299, 2018.
- [16] "Physical layer procedures for data," *3GPP TS 38.214 version 16.2.0 Release 16*, 2020.
- [17] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, 2004.
- [18] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, 2016.
- [19] A. G. Davoodi and S. A. Jafar, "Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5603–5618, 2016.
- [20] T. Berger, "Rate distortion theory, a mathematical basis for data compression (prentice-hall," *Inc. Englewood Cliffs, New Jersey*, 1971.
- [21] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.
- [22] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.
- [23] A. M. Sayeed, "A virtual representation for time-and frequency-selective correlated MIMO channels," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, vol. 4. IEEE, 2003, pp. IV–648.
- [24] W. U. Bajwa, A. Sayeed, and R. Nowak, "Compressed sensing of wireless channels in time, frequency, and space," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*. IEEE, 2008, pp. 2048–2052.
- [25] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3258–3268, 2018.
- [26] M. Thomas and A. T. Joy, "Elements of information theory," 2006.
- [27] K. Eswaran and M. Gastpar, "Remote source coding under Gaussian noise: Dueling roles of power and entropy power," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4486–4498, 2019.