Achievable Rates of FDD Massive MIMO Systems with Spatial Channel Correlation

Zhiyuan Jiang, Andreas F. Molisch, *Fellow, IEEE*, Giuseppe Caire, *Fellow, IEEE*, and Zhisheng Niu, *Fellow, IEEE*

***Abstract*—It is well known that the performance of frequency-division-duplex (FDD) massive MIMO systems with *i.i.d. channels* is disappointing compared with that of time-division-duplex (TDD) systems, due to the prohibitively large overhead for acquiring channel state information at the transmitter (CSIT). In this paper, we investigate the achievable rates of FDD massive MIMO systems with *spatially correlated channels*, considering the CSIT acquisition dimensionality loss, the imperfection of CSIT and the regularized-zero-forcing linear precoder. The achievable rates are optimized by judiciously designing the downlink channel training sequences and user CSIT feedback codebooks, exploiting the multiuser spatial channel correlation. We compare our achievable rates with TDD massive MIMO systems, i.i.d. FDD systems, and the joint spatial division and multiplexing (JSDM) scheme, by deriving the deterministic equivalents of the achievable rates, based on the one-ring model and the Laplacian model. It is shown that, based on the proposed eigenspace channel estimation schemes, the rate-gap between FDD systems and TDD systems is significantly narrowed, even approached under moderate number of base station antennas. Compared to the JSDM scheme, our proposal achieves dimensionality-reduction channel estimation without channel pre-projection, and higher throughput for moderate number of antennas and moderate to large channel coherence block length, though at higher computational complexity.**

***Index Terms*—Massive MIMO systems, Frequency-division-duplex, Spatial channel correlation, Training sequences design, Feedback codebook design.**

I. Introduction

Scaling-up multiple-input-multiple-output (MIMO) systems, thus exploiting the spatial degree-of-freedom (DoF), plays a pivotal role in boosting the capacity of next generation wireless communication systems. In cellular systems, it is desirable to deploy a large number of antennas at base stations (BSs) [1], resulting in what is referred to as the massive MIMO system. Such designs have several advantages, including significant improvements of spectral efficiency and radiated energy efficiency [2], immunity to small-scale channel fading due to the channel hardening effect, simplification of the media-access-control (MAC) layer design, etc.

Striving to reap the dramatic throughput gain of massive MIMO systems, it is found that such capacity improvements rely heavily on the availability of channel state information at the transmitter (CSIT). Without CSIT, e.g., when the user channels are identically distributed and are i.i.d. (independent identically distributed) in time and frequency, the total DoF reduces to one [3].[1] In practice, a pilot-assisted CSIT acquisition approach is widely adopted, where the BS first broadcasts downlink channel training sequences, and then listens to the channel feedback from the users. This is the case for the frequency-division-duplex (FDD) system or the uncalibrated time-division-duplex (TDD) system.[2] For the calibrated TDD system, the channel reciprocity is exploited to allow the BS to obtain the CSIT through uplink channel training [5]. Assuming the channel coefficients are i.i.d. for different users and BS antennas, the CSIT acquisition overhead, which leads to a dimensionality loss of the time-frequency resource, scales with the number of BS antennas for FDD systems, and the number of users for TDD systems, respectively. As we scale up the number of BS antennas, the overhead will become prohibitively large for the FDD system. Therefore, it is commonly considered that the TDD mode is the better, if not the only, choice for massive MIMO systems. Nonetheless, since currently deployed cellular systems are dominantly FDD, and many frequency bands are assigned explicitly for use in FDD, it is of great interest to design schemes that realize the massive MIMO gains with an FDD mode.

Given the fact that the dimensionality loss due to CSIT acquisition overhead is devastating with closed-loop channel estimation in FDD and uncalibrated TDD systems, and that the system performance without CSIT is unacceptably poor, it is natural to pose the question whether there exists other information that can be estimated at a much lower cost, while accomplishing the same task as the CSIT. To this end, it is found that the second order channel statistics, specifically the channel correlation matrices (CCMs) of the channel coefficients, are of tremendous help [6]–[10]. Compared with the instantaneous CSIT realizations, the CCMs, which are determined by user-locations and large-scale fading, vary at a much slower time scale, e.g., seconds to tens of seconds in cellular systems. Therefore, their estimation cost is drastically lower than instantaneous CSIT. In the mean time, recent

Z. Jiang and Z. Niu are with Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China. Emails: jiang-zy10@mails.tsinghua.edu.cn; niuzhs@tsinghua.edu.cn.

A. F. Molisch and G. Caire are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2565, USA. Emails: molisch@usc.edu; caire@usc.edu.

This work is sponsored in part by the National Basic Research Program of China (2012CB316001), and the Nature Science Foundation of China (61201191, 61321061, 61461136004), Hitachi Ltd. and Intel Research under the 5G Program.

[1] In such condition it has been shown that even when the CSIT is known within a mean-square error that does not decrease with SNR, the DoF collapses to one [4].

[2] Since in practice TDD reciprocity is quite difficult to obtain, which requires reciprocity calibration of the transmit and receive radio frequency chains. In fact, the only current system that uses MU-MIMO, which is 802.11ac, uses explicit polling of the users through downlink pilots, and explicit quantized closed-loop feedback from the users, even though it is a TDD system.

work shows the CCMs can be leveraged, in many ways, to facilitate FDD massive MIMO transmission. While the optimal transmission scheme with the aid of CCMs is still unclear, significant rate gain can be expected [7].

A large body of work has been done studying TDD massive MIMO systems. The seminal work in [1] first proposes to deploy BS antennas with a number much larger than the number of users, eliminating the impact of small-scale channel fading and uncorrelated noise due to the *channel hardening* effect, while only the inter-cell interference remains due to *pilot contamination*. In [2], the authors show that in addition to the spectral efficiency improvement, the massive MIMO system increases the radiated energy efficiency by a factor of M, where M is the number of BS antennas, or $\sqrt{M}$ in the presence of imperfect channel estimation. Recent work in [11] further shows the pilot-contamination problem is *not* inherent. Several other issues are also studied extensively, such as downlink precoding, detection, hardware impairment, etc. [12]–[14]

For FDD massive MIMO systems, the research can be categorized into three directions: Compressive-sensing-based, temporal-correlation-based and spatial-correlation-based. In [15] and references therein, the authors exploit the *sparsity* in massive MIMO channel matrix due to limited number of scatterers around the BS, using a compressive sensing approach. Moreover, the time correlation of the channels is leveraged to reduce the CSIT overhead, e.g., [16]–[18] and references therein, where a trellis-code based quantization codebooks are leveraged to decrease the CSIT estimation overhead in [16] [17] and a memory-based channel training sequence design is presented in [18]. The other direction is exploiting the spatial correlation of channel coefficients, pioneered by the work in [7] and extended in [8]–[10], [19], which propose the joint spatial division and multiplexing (JSDM) scheme. Based on the JSDM scheme, the users are divided into groups based on their CCMs, and a two-stage precoding is performed, namely the pre-beamforming and the beamforming, which utilizes the CCMs to counteract the inter-group-interference (IGI) and the instantaneous CSIT to manage the interference inside each group, respectively.

The current work endeavors to optimize the achievable rates of FDD massive MIMO systems. Specifically, we propose eigenspace channel estimation methods to improve the system achievable rates, for the case of spatially correlated channels. The main contributions of this paper include:

- The low-rank covariance matrices of the channels are exploited in order to design efficient channel training and feedback schemes, which enable dimensionality reduced channel estimation, e.g., it may suffice to train the downlink broadcast channel (BC) with pilots less than the number of BS antennas. In fact, the proposed channel training and feedback schemes can be seen as an alternative to the pre-projection and effective channel approach in JSDM. We derive deterministic equivalents of the achievable rates for our schemes with a regularized-zero-forcing (RZF) precoder, considering distinct CCMs of different users, the dimensionality loss due to channel training and feedback processes, and the imperfection of channel estimations. The proposed approach requires minimal modifications of the widely-adopted pilot-assisted scheme, thus making it desirable to implement in practice.
- The optimal channel training sequences with distinct CCMs for different users are studied for the first time. We propose a heuristic iterative algorithm to find the optimized training sequences, within the heuristics of the algorithm, based on maximizing the mutual information between the channel coefficients and the received channel training signals. The training sequences found by the algorithm are shown to improve the system achievable rates substantially, compared with the training sequences optimized for the i.i.d. case.
- The Karhunen-Loeve (KL) transform followed by entropy coded scalar quantization (SQ) with reverse water filling bit-loading for the feedback codebook design (KLSQ) is proposed. We compare its performance with two vector quantization (VQ) methods designed for the spatially correlated channel case. It is shown that the KLSQ is a simple way to approach the optimal VQ performance for correlated Gaussian channel vectors. The simplicity is due to the fact that it is only SQ followed by Huffman entropy coding. Therefore, it is of very low complexity for real time implementation, which justifies and motivates its use.
- Comprehensive numerical results are given to evaluate the performance. We consider the one-ring channel model and the Laplacian angular spectrum channel model, and compare our achievable sum rate with the TDD system and the i.i.d. FDD system under various system parameters. Significant rate gains are obtained by our proposed channel estimation scheme in spatially correlated channels. Furthermore, in comparison with the JSDM scheme, it is shown that the achievable sum rate with our proposal is better in most scenarios, except when the channel coherence block length is very small and the users are well separated in the angular domain.

The remainder of the paper is organized as follows. In Section II, the system model is characterized. In Section III, we specify the proposed eigenspace channel training and feedback schemes, and derive the achievable rates. In Section IV, we derive the deterministic equivalents of the achievable rates. Section V gives the simulation results, including the comparison with TDD and i.i.d. FDD systems, and the JSDM scheme, under various system parameters. Finally, in Section VI, we conclude our work.

Notations : Throughout the paper, we use boldface uppercase letters, boldface lowercase letters and lowercase letters to designate matrices, column vectors and scalars, respectively. $\boldsymbol{X}^\dagger$ denotes the complex conjugate transpose of matrix $\boldsymbol{X}$. $\boldsymbol{X}(:,i)$ denotes the i-th column of $\boldsymbol{X}$. x_i denotes the i-th element of vector $\boldsymbol{x}$. $\text{diag}[x_1, x_2, ..., x_n]$ denotes a diagonal matrix with $x_1, x_2, ..., x_n$ on its diagonal. $\det(\boldsymbol{X})$ and $\text{tr}(\boldsymbol{X})$ denote the determinant and the trace of matrix $\boldsymbol{X}$, respectively. Denote by $\mathbb{E}(\cdot)$ as the expectation operation. Denote by $\boldsymbol{I}_N$ as the N dimensional identity matrix. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes

circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The logarithm $\log(x)$ denotes the binary logarithm. We use Cov$(\cdot)$ to denote the covariance matrix of a random vector.

II. System Model

We consider a downlink BC, where an M-antenna BS serves N single-antenna users. The receive signal of the n-th user is expressed as

$$y_n = \boldsymbol{h}_n^\dagger \boldsymbol{W}\boldsymbol{s} + n_n, \tag{1}$$

where $\boldsymbol{s} \in \mathbb{C}^N$ is the data symbols transmitted to the users, $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s}$ denotes the precoded downlink signals, $\boldsymbol{W} \in \mathbb{C}^{M\times N}$ denotes the precoding matrix, $\boldsymbol{h}_n$ is the channel vector of user-n, and $\boldsymbol{y} \in \mathbb{C}^N$ are the received signals of users. The downlink total transmit power constraint is

$$\text{tr}\left\{\mathbb{E}[\boldsymbol{W}\boldsymbol{s}\boldsymbol{s}^\dagger\boldsymbol{W}^\dagger]\right\} \le P, \tag{2}$$

and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ is the Gaussian distributed uncorrelated noise.

A. Spatially Correlated Channel Matrix

Define the compound downlink channel matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_N]^\dagger$, where $\boldsymbol{h}_n \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_n)$. The CCM of user n is

$$\boldsymbol{R}_n = \mathbb{E}\left[\boldsymbol{h}_n\boldsymbol{h}_n^\dagger\right], \tag{3}$$

where by the Karhunen-Loeve representation,

$$\boldsymbol{h}_n = \boldsymbol{R}_n^{\frac{1}{2}}\boldsymbol{z}_n, \tag{4}$$

where $\boldsymbol{z}_n \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$. It is assumed that the channel vectors of users are mutually *independent*, since users are usually well separated. Denote the singular-value-decomposition (SVD) of the CCM as $\boldsymbol{R}_n = \boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{U}_n^\dagger$, wherein $\boldsymbol{U}_n$ is an $M \times M$ orthogonal matrix and $\boldsymbol{\Sigma}_n = \text{diag}[\lambda_1^{(n)}, \lambda_2^{(n)}, ..., \lambda_M^{(n)}]$.

B. Dominant Eigenspace Representation of CCM

Let us define the order-r_n dominant eigenspace representation of $\boldsymbol{R}_n$ (r_n-DER) as

$$\boldsymbol{R}_n^{(r_n)} = \boldsymbol{U}_n^{(r_n)}\boldsymbol{\Sigma}_n^{(r_n)}(\boldsymbol{U}_n^{(r_n)})^\dagger, \tag{5}$$

where $\boldsymbol{\Sigma}_n^{(r_n)} \in \mathbb{C}^{r_n\times r_n}$ contains the r_n dominant singular values with $1 \le r_n \le M$, and $\boldsymbol{U}_n^{(r_n)} \in \mathbb{C}^{M\times r_n}$ denotes the corresponding r_n eigenvectors of $\boldsymbol{R}_n$. The order-r_n channel vector approximation (r_n-CVA) is

$$\boldsymbol{h}_n^{(r_n)} = \boldsymbol{U}_n^{(r_n)}(\boldsymbol{\Sigma}_n^{(r_n)})^{\frac{1}{2}}\boldsymbol{z}_n^{(r_n)}, \tag{6}$$

where $\boldsymbol{z}_n^{(r_n)} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{r_n})$. And let

$$\boldsymbol{h}_n = \boldsymbol{h}_n^{(r_n)} + \boldsymbol{e}_n^{(r_n)}, \tag{7}$$

where $\boldsymbol{e}_n^{(r_n)}$ denotes the error introduced by only considering the dominant r_n singular values, which, therefore, can be represented as

$$\boldsymbol{e}_n^{(r_n)} = \bar{\boldsymbol{U}}_n^{(r_n)}(\bar{\boldsymbol{\Sigma}}_n^{(r_n)})^{\frac{1}{2}}\bar{\boldsymbol{z}}_n^{(r_n)}, \tag{8}$$

where $\bar{\boldsymbol{U}}_n^{(r_n)} \in \mathbb{C}^{M\times(M-r_n)}$ denotes the remaining $M - r_n$ eigenvectors of $\boldsymbol{R}_n$, and $\bar{\boldsymbol{\Sigma}}_n^{(r_n)} \in \mathbb{C}^{(M-r_n)\times(M-r_n)}$ contains the remaining $M - r_n$ non-dominant singular values. The approximation, namely the r_n-DER, which only accounts for the dominant r_n singular values of $\boldsymbol{R}_n$ is leveraged to improve the CSIT feedback efficiency, which is discussed in details in Section III-B.

Assumptions : Throughout the paper, we assume the BS has perfect knowledge of the CCMs of all users, i.e., $\boldsymbol{R}_n, \forall n$, and the users know their respective CCMs. In practice, the BS can obtain the downlink CCMs by direct transformation from the uplink CCMs without using any training symbols [20]. In addition, we assume the receive and transmit antennas are uncorrelated [21], and we only consider the transmit correlation since we assume single-antenna users [22].

We adopt the block fading channel model, where the channel is constant for T channel uses measured on the time-frequency plane, i.e., in complex dimensions, and evolves independently to another block. The channel coherence block length T is a dimensionality, which is given by the product of channel coherence time and channel coherence bandwidth in an orthogonal frequency-division multiplexing (OFDM) system, on the time-frequency plane. In Long-Term Evolution (LTE) systems, a resource block is a tile of 14 OFDM symbols in time multiplied by 12 subcarriers in frequency, for a total of $T = 168$ complex symbols [23], over which the channel is constant (within the time and frequency selectivity for which the system is designed).

III. FDD Massive MIMO Achievable Rates

In this section, we will specify the rate-achieving transmission scheme proposed in this work. The structure of the transmission strategy is identical with the widely adopted pilot-assisted FDD MU-MIMO system, which encompasses three steps:

- Downlink channel training.
- Uplink CSIT feedback.
- Data transmission.

The rate improvement stems from optimizing the channel training sequences and the CSIT feedback codebooks under the spatially correlated channels, thus requiring minimum modifications to current systems. In what follows, we will investigate the aforementioned steps in order, namely the channel training sequences, feedback codebooks, and derive the achievable rates on account of the dimensionality loss and imperfection of channel estimations with the RZF linear precoder.

A. Optimized Downlink Training with Per-User CCM

The signal model of the channel training phase is expressed as

$$\begin{aligned}&\boldsymbol{Y}_\tau = \boldsymbol{H}\boldsymbol{X}_\tau + \boldsymbol{N}_\tau,\\ &\text{tr}\left[\boldsymbol{X}_\tau\boldsymbol{X}_\tau^\dagger\right] \le \tau P,\end{aligned} \tag{9}$$

where $\boldsymbol{X}_\tau$ is an $M \times \tau$ training signal matrix, containing the training sequences and is known to the BS and the users. τ

is the training length, and $\boldsymbol{Y}_\tau = [\boldsymbol{y}_{\tau,1}, \boldsymbol{y}_{\tau,2}, ..., \boldsymbol{y}_{\tau,N}]^\dagger$ is the corresponding channel output observed by the user, disturbed by Gaussian noise $\boldsymbol{N}_\tau$ with i.i.d. unit-variance entries. The n-th user observes

$$\boldsymbol{y}_{\tau,n}^\dagger = \boldsymbol{h}_n^\dagger \boldsymbol{X}_\tau + \boldsymbol{n}_{\tau,n}^\dagger, \tag{10}$$

and applies the minimum-mean-square-error (MMSE) estimation [24, Section 19.5]

$$\hat{\boldsymbol{h}}_n = \boldsymbol{R}_n \boldsymbol{X}_\tau (\boldsymbol{X}_\tau^\dagger \boldsymbol{R}_n \boldsymbol{X}_\tau + \boldsymbol{I}_\tau)^{-1} \boldsymbol{y}_{\tau,n}. \tag{11}$$

Applying the MMSE decomposition, the user channel $\boldsymbol{h}_n$ and the covariance matrix of the channel estimation error due to imperfect channel training are expressed as [25]

$$\begin{aligned} \boldsymbol{h}_n &= \hat{\boldsymbol{h}}_n + \hat{\boldsymbol{e}}_n, \\ \boldsymbol{C}_{\hat{e}_n} &= \mathrm{Cov}(\boldsymbol{h}_n) - \mathrm{Cov}(\hat{\boldsymbol{h}}_n) \\ &= \boldsymbol{R}_n - \boldsymbol{R}_n \boldsymbol{X}_\tau (\boldsymbol{X}_\tau^\dagger \boldsymbol{R}_n \boldsymbol{X}_\tau + \boldsymbol{I}_\tau)^{-1} \boldsymbol{X}_\tau^\dagger \boldsymbol{R}_n \\ &= (\boldsymbol{R}_n^{-1} + \boldsymbol{X}_\tau \boldsymbol{X}_\tau^\dagger)^{-1}, \end{aligned} \tag{12}$$

respectively. The last equation in (12) follows from the matrix inversion lemma[3]. The total mean-square error (MSE) is

$$\mathrm{MSE} = \sum_{n=1}^{N} \mathrm{tr}\left[\boldsymbol{C}_{\hat{e}_n}\right]. \tag{13}$$

Notice that by assumption $\boldsymbol{R}_n$ is the CCM, thus it may be rank-deficient and not invertible. Nonetheless, let $\bar{\boldsymbol{R}}_n = \boldsymbol{R}_n + \epsilon^* \boldsymbol{I}_M$ such that ϵ^* is small but $\bar{\boldsymbol{R}}_n$ is invertible. Then (12) holds true if we substitute $\bar{\boldsymbol{R}}_n$ for $\boldsymbol{R}_n$. Then we can let $\epsilon^* \to 0$ due to the continuity of the function involved.

In [26], the optimal training sequences where users have *identical* CCMs are given, in the sense of minimizing the MSE or the mutual information between the channel coefficients and received signals conditioned on the transmitted training signals. However, to the best of our knowledge, the optimal training sequences under the per-user CCMs is still unknown, because multiple users share the same downlink training sequences, and thus the training sequences can no longer match one specific CCM, as in the case where user CCMs are identical. In what follows, we develop an iterative algorithm to find the optimized training sequences, within the heuristics of the algorithm, in terms of maximizing the conditional mutual information (CMI) between the channel vector and the received signal. The optimization problem, given the training length τ and total transmit power P is expressed as,[4]

$$\begin{aligned} \text{maximize:} \quad & \sum_{n=1}^{N} \log\det\left(\boldsymbol{I} + \boldsymbol{X}_\tau^\dagger \boldsymbol{R}_n \boldsymbol{X}_\tau\right) \\ \text{s.t.} \quad & \mathrm{tr}\left[\boldsymbol{X}_\tau \boldsymbol{X}_\tau^\dagger\right] \le \tau P, \end{aligned} \tag{14}$$

and we have the following theorem.

[3]The matrix inversion lemma states $(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}$, where $\boldsymbol{A}, \boldsymbol{U}, \boldsymbol{C}, \boldsymbol{V}$ are all matrices with correct sizes.

[4]Notice that (14) is based on the long-term statistics, i.e., CCMs, instead of directly on instantaneous CSIT. We stress that it is impossible to base the optimization of the training sequences on any knowledge of instantaneous CSIT, which varies in time, due to causality.

Theorem 1: The training sequences that maximize the CMI satisfy the following condition

$$\sum_{n=1}^{N} \left[\boldsymbol{R}_n \boldsymbol{X}_{\text{opt}} \left(\boldsymbol{I}_\tau + \boldsymbol{X}_{\text{opt}}^\dagger \boldsymbol{R}_n \boldsymbol{X}_{\text{opt}}\right)^{-1}\right] = \lambda \boldsymbol{X}_{\text{opt}}, \tag{15}$$

where $\lambda \ge 0$ is a constant chosen to satisfy the power constraint.

Proof: The proof is straightforward by deriving the Karush-Kuhn-Tucker (KKT) conditions of the Lagrangian dual problem [27] of (14). ∎

Remark 1: Unfortunately, in general, the problem in (14) is not a convex problem. Consider the special case where $N = 1$ and $\boldsymbol{R}_1$ is rank-deficient, then any $\boldsymbol{X}_\tau$ satisfying

$$\boldsymbol{X}_\tau = [\boldsymbol{x}_0, \boldsymbol{x}_0, , ..., \boldsymbol{x}_0], \tag{16}$$

where $\boldsymbol{x}_0$ is the singular vector of $\boldsymbol{R}_1$ corresponding to the singular value of 0, is a solution of (15) when $\lambda = 0$. Therefore, there are multiple sequences that satisfy the KKT condition in (15), and clearly, none of which satisfying (16) is the optimal solution, since by plugging (16) into (14), the objective is zero. To obtain an improved performance, we develop a heuristic iterative algorithm which is based on the condition in (15) to find the optimized training sequences. Based on the simulation results, the algorithm performs fairly well and converges fast.

Remark 2: Observing the condition in (15), one can immediately infer that when $N = 1$, the optimal training sequences developed in [26] based on identical CCMs, which contain the singular vectors of the CCM with optimal power allocation given by the water-filling solution, satisfy (15), i.e., the one with identical CCMs is a special case for our problem.

Remark 3: The reason that we set the objective to be maximizing the CMI, rather than minimizing the total MSE, is that the algorithm based on minimizing the MSE does *not* converge. This non-convergent behavior is the result of the ill-conditioned matrices involved in computing the KKT conditions in the MSE problem. Consider the derivative of the MSE

$$\frac{\partial \mathrm{MSE}}{\partial \boldsymbol{X}_\tau} = -2 \sum_{n=1}^{N} \left(\boldsymbol{R}_n^{-1} + \boldsymbol{X}_\tau \boldsymbol{X}_\tau^\dagger\right)^{-2} \boldsymbol{X}_\tau. \tag{17}$$

The matrix $\left(\boldsymbol{R}_n^{-1} + \boldsymbol{X}_\tau \boldsymbol{X}_\tau^\dagger\right)$ is often ill-conditioned, when $\boldsymbol{R}_n$ is rank-deficient, whereas in the CMI problem, the matrices involved are all well-conditioned. Moreover, based on [28], the MMSE and the mutual information have very strong relationships, and the numerical results show that the obtained training sequences have very good MSE performance.

The iterative algorithm, which aims to find the optimum training sequences based on the first-order KKT condition in (15) is specified as follows

- Step 1) Initialization:

$$\begin{aligned} & i = 1, \\ & \boldsymbol{X}_\tau^{(1)} = \boldsymbol{X}_\tau^{(0)}; \end{aligned} \tag{18}$$

- Step 2) Iteration:

$$i \geq 2,$$
$$\boldsymbol{X}_\tau^{(i)} = \sum_{n=1}^{N} \boldsymbol{R}_n \boldsymbol{X}_\tau^{(i-1)} \left(\boldsymbol{I}_\tau + \left(\boldsymbol{X}_\tau^{(i-1)} \right)^\dagger \boldsymbol{R}_n \boldsymbol{X}_\tau^{(i-1)} \right)^{-1} \quad (19)$$

Then apply the power normalization, where

$$\boldsymbol{X}_\tau^{(i)} \leftarrow \frac{\tau P}{\mathrm{tr}\left[\boldsymbol{X}_\tau^{(i)} \left(\boldsymbol{X}_\tau^{(i)} \right)^\dagger \right]} \boldsymbol{X}_\tau^{(i)}. \quad (20)$$

If $\|\boldsymbol{X}_\tau^{(i)} - \boldsymbol{X}_\tau^{(i-1)}\|_2 < \epsilon$, where $\|\cdot\|_2$ denotes the spectral norm, the algorithm is finished, and the output is $\boldsymbol{X}_\tau^{(i)}$. Else, go to step 2.

Remark 4: Such a KKT-based iteration is a canonical method to get convergent schemes that yield local optimum of the objective function, which aims at solving a non-convex problem. KKT-based iterations have been proposed in different contexts, e.g., [29].

Remark 5: Notice that $\boldsymbol{X}_\tau^{(0)} \neq 0$, otherwise the algorithm would be stuck at zero. In our simulations, letting $\boldsymbol{X}_\tau^{(0)}$ have orthogonal rows works well. Also notice that in the algorithm, we normalize the power of the training signals to be equal to the power constraint, due to the fact that it is clear that the optimal solution satisfies the power constraint with equality.

B. Uplink CSIT Feedback

After the users estimate their respective channel coefficients based on received channel training signals, they feed back their estimates using predefined codebooks. In this subsection, efficient channel feedback codebooks are designed with spatially correlated channels. We first propose the entropy coded scalar quantization after KL transform, which is a simple way to universally approach the optimal VQ performance. Then we compare its performance with two VQ approaches, which are shown to be near-optimal with spatially correlated channels and also serve as two implementation options.

1) Entropy Coded Scalar Quantization: We consider a scalar quantization method (component by component) of the KL-transformed channel vector, denoted by $\hat{\boldsymbol{h}}_n^{\mathrm{KL}}$. Specifically, denote

$$\hat{\boldsymbol{h}}_n^{\mathrm{KL}} = \boldsymbol{U}_n^\dagger \hat{\boldsymbol{h}}_n = \boldsymbol{\Sigma}_n^{\frac{1}{2}} \boldsymbol{z}_n - \boldsymbol{U}_n^\dagger \hat{\boldsymbol{e}}_n \quad (21)$$

as the KL-transform of the channel vector of user-n, after channel training where $\hat{\boldsymbol{h}}_n$ is the MMSE channel estimation after channel training and $\boldsymbol{U}_n$ is the left singular matrix of $\boldsymbol{R}_n$. Putting aside the channel training error $\hat{\boldsymbol{e}}_n$ for ease of exposition[5], this yields M mutually independent Gaussian variables with non-identical variances. The reverse water-filling approach (RWF) [30] can be implemented to achieve the rate-distortion function (in terms of MSE distortion) in this scenario, i.e., we allocate the quantization bits according to the following conditions

$$\sum_{i=1}^{M} \min\left[\nu_n, \lambda_i^{(n)}\right] = D,$$
$$R_i = \log\left(\frac{\lambda_i^{(n)}}{\nu_n}\right),$$
$$\sum_{i=1}^{M} R_i = B_n, \quad (22)$$

where D is the total MSE distortion, R_i denotes the number of bits allocated to the i-th component of $\hat{\boldsymbol{h}}_n^{\mathrm{KL}}$, B_n is the total number of feedback bits for user-n, and $0 \leq \nu_n \leq \max\{\lambda_i^{(n)}, \forall i\}$ denotes the water level. The MSE distortion for the i-th component is the minimum of $\{\nu_n, \lambda_i^{(n)}\}$, i.e.,

$$D_i^{(n)} = \min\left[\nu_n, \lambda_i^{(n)}\right]. \quad (23)$$

After the BS recovers the KL-transformed channel vector from the user feedback, it can reconstruct the channel vector by left multiplying $\boldsymbol{U}_n$. By this scheme, we obtain the relationship between the channel estimation at the BS side, denoted by $\hat{\hat{\boldsymbol{h}}}_n$, and the real channel,

$$\boldsymbol{h_n} = \hat{\hat{\boldsymbol{h}}}_n + \underbrace{\hat{\boldsymbol{e}}_n + \boldsymbol{U}_n \hat{\hat{\boldsymbol{e}}}_n}_{\boldsymbol{\varepsilon}_n}, \quad (24)$$

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_n) = \underbrace{\boldsymbol{C}_{\hat{e}_n}}_{\mathcal{M}_1} + \underbrace{\boldsymbol{U}_n \boldsymbol{D}_n \boldsymbol{U}_n^\dagger}_{\mathcal{M}_2} \quad (25)$$

$$\hat{\hat{\boldsymbol{R}}}_n \triangleq \mathrm{Cov}(\hat{\hat{\boldsymbol{h}}}_n) = \boldsymbol{R}_n - \mathrm{Cov}(\boldsymbol{\varepsilon}_n), \quad (26)$$

where $\boldsymbol{C}_{\hat{e}_n}$ is defined in (12), $\boldsymbol{D}_n \triangleq \mathrm{diag}\left[D_1^{(n)}, D_2^{(n)}, ..., D_M^{(n)}\right]$, and $\hat{\hat{\boldsymbol{e}}}_n$ is the feedback quantization error. Observing the error covariance matrix in (25), $\mathcal{M}_1$ and $\mathcal{M}_2$ represent channel estimation error covariance due to imperfect channel training and CSI quantization error covariance, respectively.

Remark 6: There are several approaches to mimic such behavior using a scalar quantizer, e.g., apply uniform quantization levels and encode the quantization points with Huffman code for each of the components with $\lambda_i^{(n)} > \nu_n$, based on the fact that the component is Gaussian distributed with variance $\lambda_i^{(n)}$. The advantage of this quantizer is that it does not involve any VQ, thus can be implemented very efficiently in parallel. Notice also that when $\boldsymbol{U}_n$ is a slice of a Discrete Fourier Transform (DFT) matrix (as in large linear antenna arrays), the KL-transform can be well implemented by a Fast Fourier Transform (FFT), therefore the overall quantization can be made extremely computationally efficient. The quantization error performance and comparison with VQ approaches will be shown in Section V-D.

Remark 7: It should be noted that although the SQ-RWF quantizer quantizes the instantaneous channel vector, the parameters describing the SQ-RWF quantizer, including the KL-transformation, the bit allocation information (the number of allocated bits per selected channel entries) and etc., are solely determined by the second-order statistics, i.e., the channel correlations matrices, which are slowly varying. Therefore it

[5]Normally the channel training error is small, therefore we ignore it when designing feedback codebooks.

is not a significant overhead to exchange the description of the SQ-RWF quantizer, since it needs to be communicated only at the rate at which the statistics change significantly .

2) VQ: Isotropical and Skewed Random Codebooks: In the literature, extensive work has been done regarding the VQ feedback codebook design in spatial CCMs. It is well understood that in the asymptotic regime where the number of feedback bits B goes to infinity, the quantization MSE scales down with B as MSE $\sim 2^{\frac{-B}{M-1}}$, regardless whether the channel distribution is i.i.d. or correlated [31] [32]. However, when the number of feedback bits B is limited, which is the case for FDD massive MIMO systems due to scarce channel estimation resources, the exact analysis for the quantization MSE performance is unavailable. In [33], a "skewed codebook" (i.e., a codebook based on skewing an isotropical codebook) that matches the eigenspace of the CCM is shown to be close to optimal by simulation results. The authors of [31] try to derive closed-form expressions for the SNR loss for general skewed codebooks, but the expressions are too complicated to find the optimal skew matrix in closed form. Notwithstanding the difficulty in deriving the optimal codebook in closed form, the Lloyd algorithm can be implemented to find the optimal codebook, however with high computational complexity [34].

Observing that the CCMs of the users are usually rank-deficient, in the sense that a number of singular values of the CCMs are extremely small (see numerical results in Section V for eigenvalue distributions in popular channel models), it is advantageous for the users to compress their feedback overhead by *only* feeding back along the order-r_n dominant eigenspace of the channel, i.e., a r_n-CVA in (6). It will be shown later that this scheme performs better than feeding back all the channel space, when B is finite. Specifically, we consider two kinds of feedback schemes, both of which concentrate the feedback bits in the dominant eigenspace of the channels, while one of them leverages an isotropical random vector to quantize the dominant eigenspace, the other explores the benefit of a skewed codebook design.

a) Isotropical Quantization in Dominant Eigenspace: First, the n-th user decorrelates the channel vector leveraging the r_n-DER of the CCM,

$$\hat{\boldsymbol{z}}_n^{(r_n)} = (\boldsymbol{\Sigma}_n^{(r_n)})^{-\frac{1}{2}} (\boldsymbol{U}_n^{(r_n)})^\dagger \hat{\boldsymbol{h}}_n. \tag{27}$$

Notice that assuming the r_n-CVA is accurate and the channel training is perfect, namely $\hat{\boldsymbol{h}}_n = \boldsymbol{h}_n^{(r_n)}$, then $\hat{\boldsymbol{z}}_n^{(r_n)}$ has r_n independently Gaussian distributed unit-norm entries. Based on this observation, we then use a predefined isotropical codebook to quantize $\hat{\boldsymbol{z}}_n^{(r_n)}$. After the feedback, the BS obtains a quantized version of the channel estimation, after multiplying the channel correlation eigenvectors,

$$\hat{\hat{\boldsymbol{h}}}_n = \boldsymbol{U}_n^{(r_n)} (\boldsymbol{\Sigma}_n^{(r_n)})^{\frac{1}{2}} \hat{\hat{\boldsymbol{z}}}_n^{(r_n)}, \tag{28}$$

where $\hat{\hat{\boldsymbol{z}}}_n^{(r_n)}$ denotes the quantized version of $\hat{\boldsymbol{z}}_n^{(r_n)}$ at the BS side, with quantization error $\hat{\hat{\boldsymbol{e}}}_n$ satisfying

$$\hat{\boldsymbol{z}}_n^{(r_n)} = \hat{\hat{\boldsymbol{z}}}_n^{(r_n)} + \hat{\hat{\boldsymbol{e}}}_n. \tag{29}$$

The quantization error $\hat{\hat{\boldsymbol{e}}}_n$ can be computed based on [32], where random vector quantization (RVQ) is assumed, by which the codebook is obtained by generating 2^{B_n} quantization vectors independently and uniformly distributed on the unit sphere in $\mathbb{C}^{r_n}$. The quantization error $\hat{\hat{\boldsymbol{e}}}_n$ is i.i.d. and independent with $\hat{\hat{\boldsymbol{z}}}_n^{(r_n)}$. It follows that

$$\mathrm{Cov}(\hat{\hat{\boldsymbol{e}}}_n) = \frac{2^{\frac{-B_n}{r_n-1}}}{r_n} \beta \boldsymbol{I}_{r_n}, \tag{30}$$

where

$$\begin{aligned} \beta &= \mathrm{tr}\left[\hat{\boldsymbol{z}}_n^{(r_n)} (\hat{\boldsymbol{z}}_n^{(r_n)})^\dagger\right] \\ &= \mathrm{tr}\left[\boldsymbol{I}_{r_n} - (\boldsymbol{\Sigma}_n^{(r_n)})^{-\frac{1}{2}} (\boldsymbol{U}_n^{(r_n)})^\dagger \boldsymbol{C}_{\hat{e}_n} \boldsymbol{U}_n^{(r_n)} (\boldsymbol{\Sigma}_n^{(r_n)})^{-\frac{1}{2}}\right] \end{aligned} \tag{31}$$

Combining (12), (27), (28), (30) and the r_n-CVA in (7), we obtain the relationship between the channel estimation at the BS side and the real channel, i.e.,

$$\begin{aligned} \boldsymbol{h_n} = \hat{\hat{\boldsymbol{h}}}_n + \\ \underbrace{\boldsymbol{U}_n^{(r_n)} (\boldsymbol{U}_n^{(r_n)})^\dagger \hat{\boldsymbol{e}}_n + \boldsymbol{U}_n^{(r_n)} (\boldsymbol{\Sigma}_n^{(r_n)})^{\frac{1}{2}} \hat{\hat{\boldsymbol{e}}}_n + \boldsymbol{e}_n^{(r_n)}}_{\boldsymbol{\varepsilon}_n}, \end{aligned} \tag{32}$$

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\varepsilon}_n) = &\underbrace{\boldsymbol{U}_n^{(r_n)} (\boldsymbol{U}_n^{(r_n)})^\dagger \boldsymbol{C}_{\hat{e}_n} \boldsymbol{U}_n^{(r_n)} (\boldsymbol{U}_n^{(r_n)})^\dagger}_{\mathcal{M}_1} \\ &+ \underbrace{\frac{2^{\frac{-B_n}{r_n-1}}}{r_n} \beta \boldsymbol{R}_n^{(r_n)}}_{\mathcal{M}_2} + \underbrace{\bar{\boldsymbol{U}}_n \bar{\boldsymbol{\Sigma}}_n \bar{\boldsymbol{U}}_n^\dagger}_{\mathcal{M}_3} \end{aligned} \tag{33}$$

$$\hat{\hat{\boldsymbol{R}}}_n \triangleq \mathrm{Cov}(\hat{\hat{\boldsymbol{h}}}_n) = \boldsymbol{R}_n - \mathrm{Cov}(\boldsymbol{\varepsilon}_n), \tag{34}$$

where $\boldsymbol{C}_{\hat{e}_n}$ is defined in (12), and $\bar{\boldsymbol{U}}_n$, $\bar{\boldsymbol{\Sigma}}_n$ are defined in (8). Observing the error covariance matrix in (33), $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ represent channel estimation error due to imperfect channel training, CSI quantization error, and the error from only feeding back the order-r_n dominant eigenspace of the channel vectors, respectively.

b) Skewed Codebook Quantization in Dominant Eigenspace: Although we concentrate our feedback bits in the dominant eigenspace based on the isotropical dominant codebook design in the preceding subsection, there is still *imbalance* among the singular values of the CCMs, rendering the isotropical RVQ codebook described above not optimal. To this end, we adopt a skewed codebook

$$\mathfrak{C}_{\mathrm{sk}} = \left\{ \frac{\boldsymbol{A}_n^{\frac{1}{2}} \boldsymbol{f}_i}{\left\| \boldsymbol{A}_n^{\frac{1}{2}} \boldsymbol{f}_i \right\|}, \, i = 1, ..., 2^B \right\} \tag{35}$$

where $\boldsymbol{f}_i \in \mathbb{C}^{r_n}$ is isotropically distributed on the unit-sphere, and $\boldsymbol{A}_n = \boldsymbol{U}_n^{(r_n)} (\boldsymbol{\Sigma}_n^{(r_n)})^{\frac{1}{2}}$. It is clear that by design we only feed back the dominant r_n eigenmodes of the channel, i.e., $\boldsymbol{h}_n^{(r_n)}$, neglecting the remaining eigenmodes. The skewed matrix is designed to match the dominant eigenspace of the channel, such that the correlation matrix of the codebook is identical with the r_n-DER. By adopting the codebook design, the total *quantization error*, which is defined as

$$\mathrm{MSE_q} = \mathrm{tr}\left[\mathbb{E}\left(\hat{\hat{\boldsymbol{e}}}_n^\dagger \hat{\hat{\boldsymbol{e}}}_n\right)\right], \tag{36}$$

can be upper bounded based on the following theorem.

Theorem 2: Given a channel vector $\boldsymbol{h}_n$, the quantization error based on the skewed codebook defined in (35) is upper bounded as

$$\text{MSE}_{\text{q}} \leq \frac{\sum_{i=1}^{r_n} (\lambda_i^{(n)})^2}{\lambda_1^{(n)}} 2^{\frac{-B_n}{r_n-1}} + \text{tr}\left[\bar{\boldsymbol{U}}_n \bar{\boldsymbol{\Sigma}}_n \bar{\boldsymbol{U}}_n^{\dagger}\right]. \quad (37)$$

Proof: The proof is based upon the distribution results developed for the i.i.d. channels in [32]. The detail proof is in Appendix A. ∎

Remark 8: It is clear that the first and second terms in (37) represent the error resulting from quantizing the channel and neglecting the subdominant eigenmodes of the channel, respectively. Also notice that the quantization MSE by the skewed codebook is smaller than that by isotropical codebook,

$$\begin{aligned}\text{MSE}_{\text{q,iid}} &= \text{tr}\left[\frac{2^{-B_n}}{r_n-1}\boldsymbol{R}_n^{(r_n)} + \bar{\boldsymbol{U}}_n \bar{\boldsymbol{\Sigma}}_n \bar{\boldsymbol{U}}_n^{\dagger}\right] && (38)\\ &= \sum_{i=1}^{r_n} \lambda_i^{(n)} 2^{\frac{-B_n}{r_n-1}} + \text{tr}\left[\bar{\boldsymbol{U}}_n \bar{\boldsymbol{\Sigma}}_n \bar{\boldsymbol{U}}_n^{\dagger}\right] && (39)\\ &\geq \frac{\sum_{i=1}^{r_n} (\lambda_i^{(n)})^2}{\lambda_1^{(n)}} 2^{\frac{-B_n}{r_n-1}} + \text{tr}\left[\bar{\boldsymbol{U}}_n \bar{\boldsymbol{\Sigma}}_n \bar{\boldsymbol{U}}_n^{\dagger}\right] = \text{MSE}_{\text{q,sk}}, && (40)\end{aligned}$$

where (38) stems from (33) with $\beta = r_n$ for fair comparison since we assume the channel vector to be quantized in the derivation of Theorem 2 has unit entries. The equality holds if and only if $\lambda_1^{(n)} = \lambda_2^{(n)} = ... = \lambda_{r_n}^{(n)}$.

Remark 9: Notice that the quantization error in Theorem 2 does not scale with B to zero. This can be explained that when B is large, it is better to quantize all the channel eigenmodes instead of neglecting the sub-dominant modes, i.e., $r_n = M$. Thus the quantization error with the optimal r_n, which minimizes the quantization error, scales with B to zero, when B goes to infinity. Meanwhile, the bound in Theorem 2 is tighter than the one with r_n fixed to be M, when B is finite. The numerical results in Section V agrees with our analysis.

Remark 10: Notice that the dominant rank r_n, i.e., the order of the CVA we choose to approximate the correlated channels, plays an important role in the feedback scheme. The larger r_n is, the more accuracy we obtain by approximating the correlated channels, whereas the feedback quantization error is also larger due to the increased quantization dimension. Therefore, there exists a tradeoff in terms of the dominant rank, r_n. The optimal r_n can be determined by a simple one-dimensional search over $1{:}M$, performed by the n-th user.

C. Data Transmission

For fair comparisons, also in line with the work in [7] and [35], we consider the RZF linear precoder schemes. The precoder treats the channel estimates as the real channel coefficients. Corresponding achievable rates on account of the imperfect channel estimations are computed in the following section. Define

$$\boldsymbol{K}_{\text{rzf}} = \left(\hat{\bar{\boldsymbol{H}}}^{\dagger}\hat{\bar{\boldsymbol{H}}} + M\alpha \boldsymbol{I}_M\right)^{-1}. \quad (41)$$

The RZF precoding matrix is expressed as

$$\boldsymbol{W}_{\text{rzf}} = \zeta \boldsymbol{K}_{\text{rzf}} \hat{\bar{\boldsymbol{H}}}^{\dagger}, \quad (42)$$

where $\hat{\bar{\boldsymbol{H}}} = \left[\hat{\bar{\boldsymbol{h}}}_1, \hat{\bar{\boldsymbol{h}}}_2, ..., \hat{\bar{\boldsymbol{h}}}_N\right]^{\dagger}$, ζ is a normalization scalar to fulfill the power constraint in (2), and α is the regularization factor [36]. Based on (2), we obtain

$$\zeta^2 = \frac{N}{\text{tr}\left[\hat{\bar{\boldsymbol{H}}}\boldsymbol{K}_{\text{rzf}}^2\hat{\bar{\boldsymbol{H}}}^{\dagger}\right]}, \quad (43)$$

where equal power allocation is assumed, i.e., $\left[\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^{\dagger}]\right]_{i,i} = \frac{P}{N}$. The signal-to-interference-and-noise-ratio (SINR) of user n is

$$\gamma_{n,\text{rzf}} = \frac{\left|\hat{\bar{\boldsymbol{h}}}_n^{\dagger}\boldsymbol{K}_{\text{rzf}}\hat{\bar{\boldsymbol{h}}}_n\right|^2}{\frac{N}{P\zeta^2} + \left|\boldsymbol{\varepsilon}_n^{\dagger}\boldsymbol{K}_{\text{rzf}}\hat{\bar{\boldsymbol{h}}}_n\right|^2 + \boldsymbol{h}_n^{\dagger}\boldsymbol{K}_{\text{rzf}}\hat{\bar{\boldsymbol{H}}}_{[n]}^{\dagger}\hat{\bar{\boldsymbol{H}}}_{[n]}\boldsymbol{K}_{\text{rzf}}\boldsymbol{h}_n}, \quad (44)$$

where $\hat{\bar{\boldsymbol{H}}}_{[n]} = \left[\hat{\bar{\boldsymbol{h}}}_1, ..., \hat{\bar{\boldsymbol{h}}}_{n-1}, \hat{\bar{\boldsymbol{h}}}_{n+1}, ..., \hat{\bar{\boldsymbol{h}}}_N\right]^{\dagger}$. The training dimensionality loss is the length of the training sequence τ. Assuming the feedback bits are transmitted over the uplink MIMO-multiple-access-channel (MIMO-MAC), and based on [25], the total number of feedback channel uses is computed as

$$\delta = \frac{\sum_{n=1}^{N} B_n}{C_{\text{MIMO-MAC}}}. \quad (45)$$

For ease of exposition, we assume $B_n = B$, $\forall n$, and

$$C_{\text{MIMO-MAC}} = \kappa \min[M, N] \log(M\mathsf{SNR}_{\text{ul}}), \quad (46)$$

where $\kappa \in (0, 1)$ is a scalar representing the diversity-multiplexing tradeoff in MIMO-MAC as defined in [25], and SNR_{ul} is the uplink SNR. The achievable sum rate considering imperfect channel training and feedback is expressed as the solution of the following optimization problem

$$\begin{aligned}\text{maximize:} \quad & \left(1 - \frac{\tau+\delta}{T}\right)\sum_{n=1}^{N}\log(1+\gamma_{n,\text{rzf}})\\ \text{s.t.} \quad & 1 \leq \tau + \delta \leq T,\\ & \tau \geq 1, \quad \delta \geq 1, \end{aligned} \quad (47)$$

where the optimization is over the training and feedback length. The fundamental tradeoff is that larger training and feedback length provides a more accurate channel estimation whereas resulting in larger dimensionality loss. Since our focus is on the performance of the downlink BC achievable rates with spatially correlated channels, we use an exhaustive search to find the optimal training and feedback length. The analysis of the optimal training and feedback length for i.i.d. channels can be found in [35] and [37].

IV. Performance Analysis

In this section, we provide expressions for the downlink achievable sum rate under the per-user CCMs, leveraging the deterministic equivalent techniques provided in [35], with necessary modifications.

Following the approach in [35], when M goes to infinity, the SINR of user n, i.e., $\gamma_{n,\text{rzf}}$, satisfies

$$\gamma_{n,\text{rzf}} - \gamma_{n,\text{rzf}}^{o} \overset{M\to\infty}{\longrightarrow} 0 \text{ with probability 1}, \tag{48}$$

where $\gamma_{n,\text{rzf}}^{o}$ is a deterministic quantity that can be computed as

$$\gamma_{n,\text{rzf}}^{o} = \frac{\frac{(\hat{\bar{e}}_n^o)^2}{(1+\hat{\bar{e}}_n^o)^2}}{\frac{\phi^o}{P} + \hat{\bar{E}}_n^o + I_n^o}, \tag{49}$$

where the parameters involved are specified in (50)-(62). The derivation is mostly based upon [35], with generalizations to uncorrelated channel estimation error matrices. The details are omitted for brevity.

$$\phi^o = \frac{1}{M}\sum_{n=1}^{N} \frac{\hat{\bar{e}}_n^{o\prime}}{\left(1+\hat{\bar{e}}_n^o\right)^2}, \tag{50}$$

$$\hat{\bar{e}}_n^o = \frac{1}{M}\text{tr}\left[\hat{\bar{\boldsymbol{R}}}_n \boldsymbol{T}\right], \tag{51}$$

$$\boldsymbol{T} = \left(\frac{1}{M}\sum_{j=1}^{N}\frac{\hat{\bar{\boldsymbol{R}}}_j}{1+\hat{\bar{e}}_j^o} + \alpha \boldsymbol{I}_M\right)^{-1}, \tag{52}$$

$$\hat{\bar{\boldsymbol{e}}}^{o\prime} = \left[\hat{\bar{e}}_1^{o\prime}, \hat{\bar{e}}_2^{o\prime}, ..., \hat{\bar{e}}_N^{o\prime}\right]^T = (\boldsymbol{I}_N - \boldsymbol{J})^{-1}\boldsymbol{v}, \tag{53}$$

$$[\boldsymbol{J}]_{i,j} = \frac{1}{M}\frac{\frac{1}{M}\text{tr}\left[\hat{\bar{\boldsymbol{R}}}_i \boldsymbol{T} \hat{\bar{\boldsymbol{R}}}_j \boldsymbol{T}\right]}{\left(1+\hat{\bar{e}}_j^o\right)^2}, \tag{54}$$

$$\boldsymbol{v} = \frac{1}{M}\left[\text{tr}\left(\hat{\bar{\boldsymbol{R}}}_1 \boldsymbol{T}^2\right), \text{tr}\left(\hat{\bar{\boldsymbol{R}}}_2 \boldsymbol{T}^2\right), ..., \text{tr}\left(\hat{\bar{\boldsymbol{R}}}_N \boldsymbol{T}^2\right)\right]^{\text{T}} \tag{55}$$

$$\hat{\bar{E}}_n^o = \frac{d_{n,n}^o}{M\left(1+\hat{\bar{e}}_n^o\right)^2}, \tag{56}$$

$$\boldsymbol{d}_n^o = \left[d_{n,1}^o, d_{n,2}^o, ..., d_{n,N}^o\right]^T = (\boldsymbol{I}_N - \boldsymbol{J})^{-1}\boldsymbol{b}_n, \tag{57}$$

$$\boldsymbol{b}_n = \frac{1}{M}\left[\text{tr}\left(\hat{\bar{\boldsymbol{R}}}_1 \boldsymbol{T}\left(\boldsymbol{R}_n - \hat{\bar{\boldsymbol{R}}}_n\right)\boldsymbol{T}\right), ... , \text{tr}\left(\hat{\bar{\boldsymbol{R}}}_N \boldsymbol{T}\left(\boldsymbol{R}_n - \hat{\bar{\boldsymbol{R}}}_n\right)\boldsymbol{T}\right)\right]^{\text{T}}, \tag{58}$$

$$I_n^o = \frac{u_n}{\left(1+\hat{\bar{e}}_n^o\right)^2} + \sum_{j\neq n}^{N}\frac{d_{n,j}^o}{M\left(1+\hat{\bar{e}}_j^o\right)^2}, \tag{59}$$

$$u_n = \frac{1}{M}\sum_{j\neq n}^{N}\frac{f_{n,j}^o}{\left(1+\hat{\bar{e}}_j^o\right)^2}, \tag{60}$$

$$\boldsymbol{f}_n^o = \left[f_{n,1}^o, ..., f_{n,N}^o\right]^T = (\boldsymbol{I}_N - \boldsymbol{J})^{-1}\boldsymbol{c}_n, \tag{61}$$

$$\boldsymbol{c}_n = \frac{1}{M}\left[\text{tr}\left(\hat{\bar{\boldsymbol{R}}}_1 \boldsymbol{T}\hat{\bar{\boldsymbol{R}}}_n \boldsymbol{T}\right), ..., \text{tr}\left(\hat{\bar{\boldsymbol{R}}}_N \boldsymbol{T}\hat{\bar{\boldsymbol{R}}}_n \boldsymbol{T}\right)\right]^{\text{T}}. \tag{62}$$

V. Numerical Results

In our simulations, we evaluate the FDD massive MIMO achievable rates with various spatially correlated channel models, and compare those with the TDD system, the FDD system with i.i.d. channels, and the JSDM scheme. The azimuth angles of users, i.e., θ_n, are uniformly distributed in $[-60°, 60°]$, unless stated otherwise. We set the tolerance in the iterative algorithm in Section III-A, i.e., ϵ, to be 10^{-6}. The regularization factor in the RZF precoder is $\alpha = 0.01$.

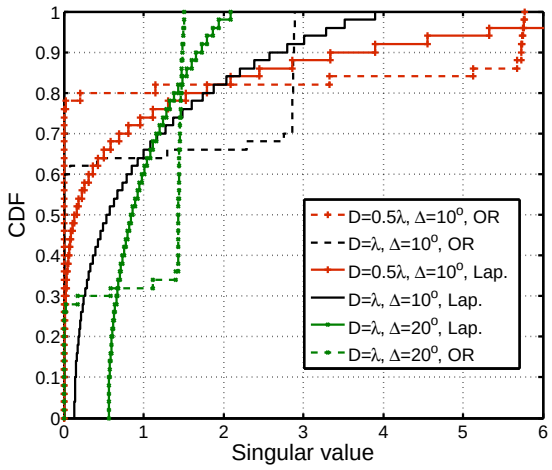

Fig. 1. The CDF of the singular values of user CCMs for various parameters. $M = 50$. $\theta = \frac{\pi}{2}$.

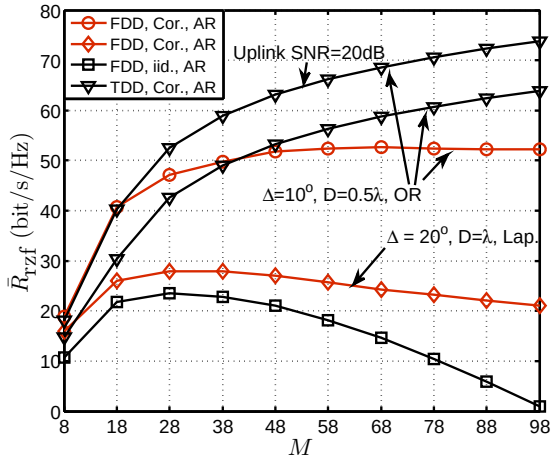

Fig. 2. Achievable sum rates (AR) in massive MIMO systems with i.i.d. channels, per-user correlation channels (Cor.), TDD mode and FDD mode respectively. The downlink and uplink SNR are set to 20 dB and 10 dB, respectively, unless labeled otherwise. The channel coherence block length is $T = 200$. The number of users in the cell is $N = 8$. The per-user channel correlation matrices are calculated according to (63) and (64).

A. CCMs: One-Ring Model and Laplacian Model

First, we evaluate the singular value distribution of CCMs. In Fig. 1, the cumulative probability function (CDF) of the singular values of the user CCMs is shown. We adopt two models to calculate the CCM of a uniform linear antenna array. The first one is the one-ring model (OR) [38], based on which

$$[\boldsymbol{R}]_{i,j} = \frac{1}{2\Delta}\int_{-\Delta+\theta}^{\Delta+\theta} e^{-j2\pi D(i-j)\sin(\alpha)}d\alpha, \tag{63}$$

where Δ denotes the angular spread, θ denotes the mean user azimuth angle seen from the BS and D is the antenna spacing. Alternatively, the Laplacian angular spectrum model (Lap.) is also considered [24, Section 7.4.2], where

$$[\boldsymbol{R}]_{i,j} = \frac{1}{\sqrt{2}\Delta}\int_{\theta-\pi}^{\theta+\pi} e^{-\frac{\sqrt{2}}{\Delta}|\alpha-\theta|-j2\pi D(i-j)\sin(\alpha)}d\alpha. \tag{64}$$

From Fig. 1, it is observed that the singular values of the CCMs are generally distributed with large deviations under

various parameters, i.e., some singular values are large while some others are effectively close to zeros, thus we define the *effective rank (ER)* of the CCM. Generally, as the antenna spacing is smaller, or the angular spread is smaller, the ER will be smaller. Note that usually the ER calculated by the Laplacian model is larger than that by the one-ring model, due to the one-ring model restricting the direction-of-arrival (DoA) to a finite support. Also note that the number of BS antennas is relevant, which is shown in [7] that the ratio of ER and M approaches a constant asymptotically with M going to infinity. In the following simulations, we evaluate the FDD massive MIMO achievable rates under different parameters and both models depicted in Fig. 1.

B. Comparisons between the proposed scheme in correlated FDD systems with TDD and i.i.d. FDD systems

In presence of spatially correlated channels, the achievable rates under the proposed scheme are shown in Fig. 2, in comparison with i.i.d. FDD systems and also correlated TDD systems. The achievable rates of FDD systems with correlated channels are obtained using the training sequences obtained by the iterative algorithm in Section III-A, and the SQ-RWF feedback codebook design in Section III-B. First, it is noteworthy that in FDD systems, in general, the achievable sum rate is not monotonously increasing with the number of BS antennas, as it does so in the TDD system, due to the fact that when the number of BS antennas grows large with FDD mode, the channel estimation dimensionality loss will become non-negligible. Therefore, there is a large rate gap between the i.i.d. FDD system and the TDD system, rendering the FDD mode unfavorable for massive MIMO transmission.

Nevertheless, when the channel is spatially correlated, the FDD system achievable sum rate under per-user CCMs is significantly larger than that in i.i.d. channels, especially when the number of BS antennas is large, thanks to the judiciously designed dominant channel estimation schemes. The rate gap between the TDD mode and the FDD mode is narrowed significantly, especially when M is moderate, which suggests that it is promising to exploit the large-system gain even with FDD mode. Note that, while the throughput of TDD systems generally increases with M, the throughput of FDD systems would eventually go down if we further increase M, even with our proposed schemes.

It is shown that the achievable rates of FDD systems are even larger than TDD systems (with uplink SNR=10 dB) under some parameters shown in Fig. 2. The phenomenon is explained by the fact that the uplink SNR is set 10 dB lower than the downlink SNR in the corresponding simulation results, which is typical for a cellular system due to the smaller transmit power of user-terminals, rendering the TDD system performance inferior due to the imperfect *uplink* channel training. Observe that when M becomes larger, the TDD system sum rate will go up unbounded, eventually surpassing the FDD system. Moreover, when the uplink SNR is set to be the same as the downlink SNR, see corresponding curves, the TDD system performs better, which is as expected.

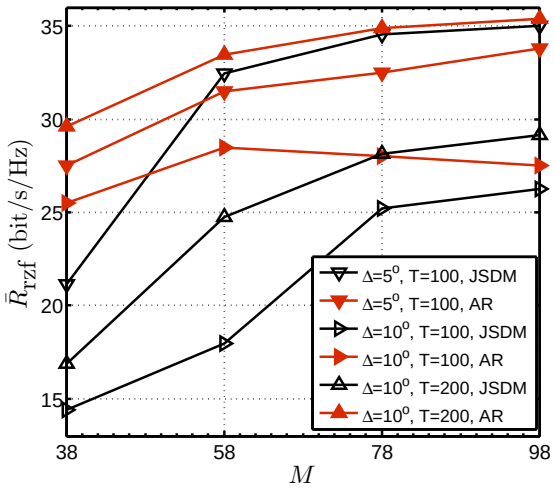

Fig. 3. The achievable sum rates (AR) obtained by the eigenspace channel estimation, compared to the JSDM scheme. The downlink and uplink SNR are both 10 dB. The number of simultaneous users is $N = 8$.

C. Comparisons with JSDM

In Fig. 3, we compare the achievable sum rates obtained by the proposed eigenspace channel estimation to the JSDM scheme [7], [8], which was the first to exploit the spatial correlation to benefit the FDD massive MIMO system. Note that the uplink CSIT feedback is not treated in the previous JSDM papers [7], [9]. To make fair comparison, we assume that the JSDM scheme uses an isotropical VQ feedback codebook, since it is unknown whether the JSDM scheme can also benefit from a better-designed codebook for correlated channels after the pre-projection of channel vectors. To get more insights and understand the simulation results better, it is important to first illustrate the merits and demerits of the JSDM scheme compared to our scheme.

The JSDM scheme has the advantage to better suppress the channel estimation overhead. Specifically, by grouping the users based on their respective CCMs and performing the pre-beamforming, the equivalent number of BS antennas in each *virtual sector*, i.e., b_g in [7], can be optimized to strike a good balance between the power gain, which scales with b_g, and the channel estimation overhead. In an extreme case, b_g can be made as small as the number of users in each virtual sector, thus, the overall channel estimation overhead scales with the number of users in each virtual sector, which drastically decreases the dimensionality loss. However, on the downside, while the JSDM scheme adopts a *divide-and-multiplex* approach, the division is imperfect, in the sense that the JSDM scheme suffers from the inherent residual IGI, especially when the CCMs of the users in each group are different, rendering that the pre-beamforming cannot counteract the IGI completely. Notice that in our framework, the proposed dominant channel estimations incorporate all the user CCMs into the scheme design, which significantly mitigates the IGI. Moreover, it is noteworthy that the *computational complexity* of the JSDM scheme is smaller compared with our proposed scheme, since our scheme deals with a higher dimensional

channel matrix[6].

Specifically, we follow the parameters used in the simulation in [8, Section IV-C]. The *fixed angular quantization* method is adopted to divide users into $G = 8$ user-groups, where each group performs the per-group-processing. The quantization points are

$$\theta \in \{-57.5^\circ, -41.5^\circ, -23^\circ, -7.5^\circ, 7.5^\circ, 23.5^\circ, 41.5^\circ, 57.5^\circ\}, \tag{65}$$

and the angular spread for the quantization matrices are identical with the user-angular-spread, which is specified in Fig. 3. To keep the IGI under control, similarly with [8], we further divide the user-groups into 2 patterns, where only the users in the same pattern are scheduled simultaneously[7]. The ER in each virtual sector, i.e. $r^\star$ in [7], is chosen while neglecting extremely small singular values of the CCM, and b_g is chosen to optimize the sum rate by exhaustive search. The training sequences of each virtual sector are unitary sequences as in [7, Section VI].

The deterministic equivalents for the JSDM scheme are computed based on [7, Appendix A], with generalizations to distinct CCMs within each user-group. The details are again omitted for brevity.

It is observed from Fig. 3 that the JSDM scheme achieves better sum rate when the channel coherence block length is small, e.g., $T = 100$, and the number of BS antennas M is large. Qualitatively, this is as expected since the small channel coherence block length and large M both put more weight in the need to suppress the channel estimation overhead, and based on [7], a large M also leads to the fact that the singular vectors of the CCMs can be well approximated by the columns of a DFT matrix, which ensures orthogonality as long as the DoA intervals of different users are disjoint. On the other hand, the achievable sum rate of our proposed eigenspace channel estimation shows evidently better achievable rate when the channel coherence block length is larger, which elevates the urgency to suppress the channel estimation overhead, or when the angular spread of users is larger, which causes larger residual IGI in the JSDM scheme. Notice that large angular spread also decreases the achievable rates of our scheme, due to the increased channel estimation dimensionality loss, however our scheme turns out to be more resilient in this regard. Although there are several parameters in the JSDM scheme, that may be properly tuned to achieve better rate than the eigenspace channel estimation scheme, the eigenspace channel estimation scheme still has the advantage of low complexity, and the optimization for the JSDM scheme goes out of the scope of this paper.

D. Performance Gain Leveraging Dominant Eigenspace Channel Estimation

Furthermore, we demonstrate how much gain we can get from leveraging the training and feedback schemes designed for the multi-user CCMs, by comparing with the unitary training and feedback schemes used in the i.i.d. channel case. The rate gain is depicted in Fig. 4, showing leveraging dominant eigenspace channel estimation can indeed improve the sum rate of FDD massive MIMO system with spatially correlated channels. The detailed performance analyses of the proposed channel training and feedback schemes are shown in Fig. 5 and Fig. 6, respectively.

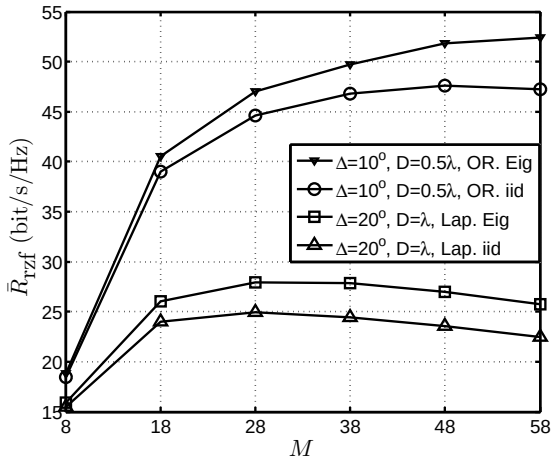

Fig. 4. Achievable sum rates (AR) in massive MIMO systems with eigenspace training and feedback schemes (Eig), compared with unitary channel estimation schemes commonly used for the i.i.d. channels. The downlink and uplink SNR are set to 20 dB and 10 dB, respectively. The channel coherence block length is $T = 200$. The number of users in the cell is $N = 8$.

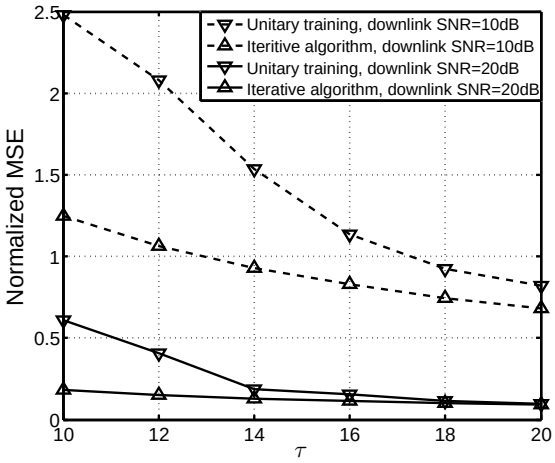

Fig. 5. The MSE caused by only the channel training process normalized by the number of users, versus the number of training symbols of the optimized training signals given by the iterative algorithm, compared with random orthogonal training sequences. $N = 8$, $M = 20$. The per-user channel correlation matrices are calculated according to the one-ring model, with $D = 0.5\lambda$ and $\Delta = 10^\circ$.

For the training process, the MSE performance of the iterative algorithm we developed in Section III-A, which finds the optimized training sequences with per-user CCMs, is shown in Fig. 5. For comparison purposes, the simulation also considers the unitary training sequences, which are shown to be optimal with i.i.d. channels [39]. We assume, for the unitary training

[6]Possible operations on the channel matrix include inversion and SVD, depending on the precoding algorithm.

[7]For fair comparison, we set the number of users in the achievable sum rate of the proposed scheme to be half of the total users in the JSDM, since there are 2 patterns.

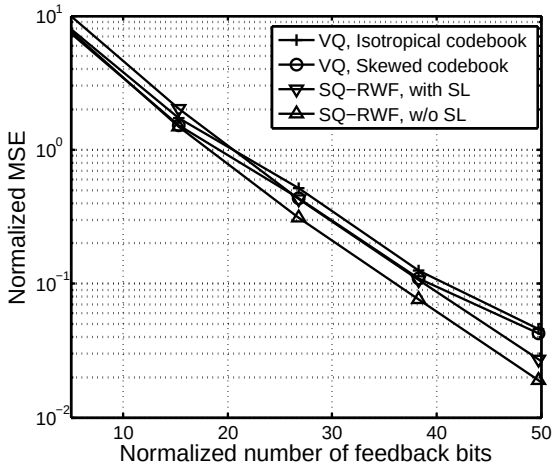

Fig. 6. The quantization MSE, normalized by the number of users, with and without the shaping loss (SL). $N = 8$, $M = 20$. The per-user CCMs are calculated according to the one-ring model, with $D = 0.5\lambda$ and $\Delta = 10^\circ$.

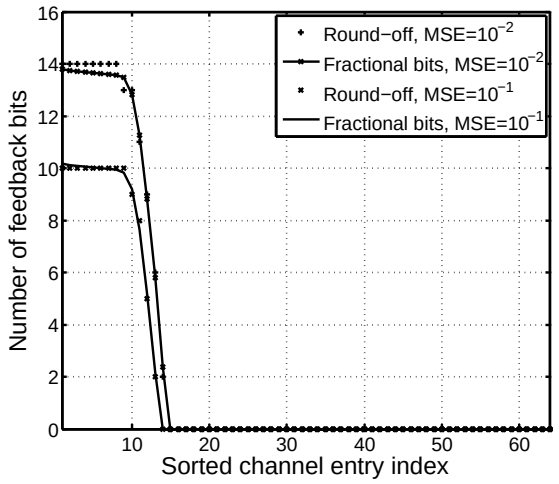

Fig. 7. Number of feedback bits for each channel entry in the SQ-RWF scheme with SL, versus the sorted channel entry index based on the variations. The mean DoA $\theta = \pi/6$, angular spread $\Delta = 10^\circ$, antenna spacing $D = 0.5\lambda$ and the CCM is calculated based on the one-ring model. The number of BS antennas $M = 64$.

sequences[8],

$$\begin{cases} \boldsymbol{X}_\tau \boldsymbol{X}_\tau^\dagger = \dfrac{\tau P}{M} \boldsymbol{I}_M & \text{if } \tau \geq M, \\ \boldsymbol{X}_\tau^\dagger \boldsymbol{X}_\tau = P \boldsymbol{I}_\tau & \text{if } \tau < M. \end{cases} \tag{66}$$

.

When the number of training symbols is small, the MSE achieved by the iterative algorithm is much lower than the unitary training sequences, due to the fact that in presence of channel correlation, the training sequences obtained by our algorithm can find the eigenspace that needs to be estimated more accurately and concentrate the power to that subspace. Note that when the downlink SNR is large and the number of training symbols is large enough[9] to train all the subspaces, the unitary training sequences are asymptotically optimal. Such observations are further evaluated by setting the downlink SNR to 10 dB, which shows a certain MSE gap between the optimized training sequences and the unitary training sequences, even when $\tau = M$.

Fig. 6 shows the quantization MSE performances of SQ-RWF and VQ with isotropical and skewed codebooks. It is observed that SQ-RWF achieves better MSE performance, even with the shaping loss (SL)[10] when the number of feedback bits is large. Notice that in general VQ is more efficient than SQ, especially when the vector is correlated. However, after the KL-transform, the channel vector is decorrelated into independent Gaussian variables with non-identical variances, in which case the RWF is the optimal bit allocation in terms of MSE distortion.

Moreover, to compare the SQ-RWF scheme with conventional VQ approaches, according to the well-known results on VQ methods in the literature [17], to achieve the perfect-CSIT DoF, the total number of feedback bits for an M-dimensional channel vector is approximately

$$B = \frac{M-1}{3} \rho_{\text{dB}}, \tag{67}$$

where ρ_{dB} is the downlink SNR (in dB). Given the parameters in Fig. 7 and $\rho_{\text{dB}} = 20$ (for fair comparison with MSE $= 10^{-2}$ since it is found that the MSE should scale inversely with the SNR to achieve the perfect-CSIT DoF), the number of feedback bits in (67) is 420, whereas the number of total feedback bits with SQ-RWF is 164 by simulation. However, it should be noted that the such a comparison is not quite fair since the feedback codebook design in [17] is not optimized for the spatially correlated channels.

Fig. 7 gives a concrete example to specify the bit allocation of the SQ-RWF scheme. Note that we perform the proposed quantization method on the user-channel after the KL-transform, i.e., $\hat{\boldsymbol{h}}_n^{\text{KL}}$ in (21). The entries of $\hat{\boldsymbol{h}}_n^{\text{KL}}$ are sorted based on the variations in descending order and the indices are shown as the x-axis. To achieve a target quantization MSE, it is shown that the round-off SQ-RWF, which rounds off the number of feedback bits given by the RWF approach to its nearest integer, uses a slightly larger number of bits than the SQ-RWF scheme which allows fractional number of feedback bits for each channel entry. It is shown that the SQ-RWF scheme "throws away" a number of channel entries due to their small variations, i.e., allocating zero bits to them, while concentrates its feedback bits to a few dominant channel entries. In addition, the SQ-RWF scheme uses a larger number of feedback bits to quantize the dominant entries, as well as

[8]Notice that the unitary training sequences for the case $\tau < M$ is not well defined in [39], since it does not suffice to have $\tau < M$ pilots in i.i.d. channels. Here we assume X_τ has orthogonal rows when $\tau < M$.

[9]In this case we have $\tau \geq M$, then there are enough channel observations to recover the channel coefficients perfectly when the downlink SNR goes to infinity.

[10]The shaping loss is defined as the loss due to the fact that scalar quantization operates on a hypercube, while optimal vector quantization operates on a hypersphere. It is equal to $\frac{1}{2}\log(\pi e/6)$ bits per real dimension, and corresponds to the difference of the differential entropies of a Gaussian and a uniform distribution with the same (unit) variance. A full thorough treatment of entropy-coded scalar quantization in an information theoretic sense can be found in [40].

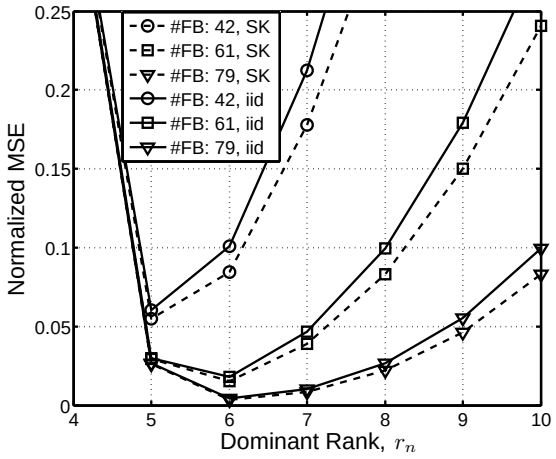

Fig. 8. The quantization MSE, normalized by the number of users, versus the dominant rank we choose to feedback the CSIT, with various number of feedback bits. $N = 8$, $M = 20$. The per-user channel correlation matrices are calculated according to the one-ring model, with $D = 0.5\lambda$ and $\Delta = 10^\circ$.

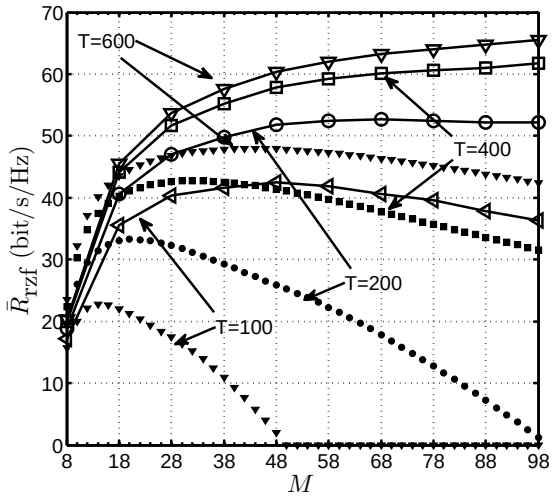

Fig. 9. Performance of FDD massive MIMO systems with CCMs and various channel coherence block length. The dotted and solid lines represent the system with i.i.d. channels and spatially correlated channels, respectively. The downlink and uplink SNR are set to 20 dB and 10 dB, respectively. The number of users in the cell is $N = 8$. The per-user channel correlation matrices are calculated according to the one-ring model, with $D = 0.5\lambda$ and $\Delta = 10^\circ$.

quantizing more channel entries, when a better quantization accuracy is required.

The impact of the dominant rank, i.e., r_n, we choose in the VQ feedback process on the quantization MSE is shown in Fig. 8, with different number of feedback bits. The numbers of feedback bits (42, 61, 79) shown in the figure correspond approximately to 2, 3, 4 bits per channel entry. The tradeoff between the quantization accuracy of the effective channel and the estimation error resulting from the neglected eigenspace of the CCM is shown. It is observed that there exists an optimal r_n in terms of minimizing the total feedback error. The optimal r_n is increasing with the number of feedback bits, for the reason that when we have more feedback bits, we can afford to estimate a higher-dimensional eigenspace, rendering a better accuracy of the CSIT feedback estimation. The performance of the skewed feedback codebook is also shown in the figure. The gain in terms of MSE is fairly small, when the optimal dominant rank is chosen, because the error mainly stems from neglecting the non-dominant eigenspace. Note that when r_n is large, the performance gain of the skewed codebook is more evident since the MSE in this regime is dominated by the channel quantization. It is worth mentioning that the absolute values of the quantization error are fairly small, compared with the training error. We find that the channel estimation error mainly comes from the downlink channel training process which is analog in nature, rather than the digitalized CSIT feedback process, for the reason that the MSE scales down linearly with the number of training symbols (13), but exponentially with the feedback bits (30).

In Fig. 9, the impact of channel coherence block length on the sum rate improvements is shown. Significant rate improvement, which is up to two-fold, is expected for all channel coherence block length. The results suggest that under the spatially correlated channels, which is especially common with mm-wave channels [9], along with a well-designed transmission strategy, namely the training and feedback schemes, the FDD system is capable of realizing significant massive MIMO gain.

VI. Conclusions

By computing the achievable rates with a RZF precoder of FDD massive MIMO systems, on account of the downlink channel training and uplink CSIT feedback dimensionality loss and corresponding channel estimation error, we showed that spatial channel correlation at the BS side is beneficial to the FDD massive MIMO system. The benefit is especially prominent if the channels are strongly correlated, namely the CCMs are effectively rank-deficient. In particular, we proposed an iterative algorithm to find the optimized channel training sequences in presence of multiuser spatial correlation, and a KL-transform followed by SQ with RWF bit-loading feedback codebook design, which is extremely computationally efficient and thus easy to implement in practice while achieving near-optimal performance. Our proposed approach, which achieves dimensionality-reduction channel estimation, can be seen as an alternative to the pre-projection and effective channel approach in the JSDM scheme. Moreover, it is noteworthy that while achieving a significant performance gain, our approach only requires minimum modifications of the widely-used training-based transmission scheme, and thus it is easy to implement.

Numerical results show significant rate improvements when leveraging our proposed eigenspace channel estimation approaches under spatially correlated channels, in comparison with i.i.d. FDD massive MIMO systems. In fact, when the channel correlation is strong and the number of BS antennas is not very large, the achievable sum rate of FDD massive MIMO systems can even outperforms TDD systems. Comparisons with the JSDM scheme reveal both schemes have advantages under different channel conditions, such as coherence time and angular spread. In particular, our proposed schemes display better performance when channel coherence block length is large, or the angular spread of the users is large, while

requiring a higher computational complexity due to operating on a higher-dimensional matrix.

These results suggest that in FDD massive MIMO systems, increasing spatial channel correlation, e.g., by decreasing the antenna spacing, more line-of-sight transmission, etc., can be beneficial. Although this differs from the favorable propagation conditions in TDD systems, which prefer i.i.d. channels to maximize the total DoF, the FDD system benefits significantly from correlation, which enables dimensionality loss reduction as far as channel estimation is concerned. The tradeoff between the DoF of the downlink BC and the spatial correlation in FDD massive MIMO is an interesting problem for future work.

APPENDIX A
PROOF OF THEOREM 2

Proof: The quantization MSE of the skewed codebook is expressed as

$$\mathrm{MSE_q} = \mathbb{E}_{\boldsymbol{z}_n}\left\{\mathbb{E}_{\mathfrak{C}_{\mathrm{sk}}}\left[\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n - \max_i\left[\frac{\boldsymbol{f}_i^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{f}_i}{\boldsymbol{f}_i^\dagger\boldsymbol{\Lambda}_n\boldsymbol{f}_i}\right]\right]\right\} + \mathrm{tr}\left[\bar{\boldsymbol{U}}_n\bar{\boldsymbol{\Sigma}}_n\bar{\boldsymbol{U}}_n^\dagger\right], \tag{68}$$

where $\boldsymbol{h}_n = \boldsymbol{Q}_n\boldsymbol{\Lambda}_n^{\frac{1}{2}}\boldsymbol{z}_n$, $\boldsymbol{Q}_n$ is an orthogonal matrix, $\boldsymbol{z}_n \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{r_n})$. Define the first term in (68) as Δ_1. We obtain

$$\Delta_1 = \mathbb{E}_{\boldsymbol{z}_n}\int_0^{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n}\left[\Pr\left(\frac{\boldsymbol{f}_i^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{f}_i}{\boldsymbol{f}_i^\dagger\boldsymbol{\Lambda}_n\boldsymbol{f}_i} \le x|\boldsymbol{f}_i^\dagger\boldsymbol{f}_i = 1\right)\right]^N dx \tag{69}$$

$$\le \mathbb{E}_{\boldsymbol{z}_n}\int_0^{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n}\left[\Pr\left(\frac{\boldsymbol{f}_i^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{f}_i}{\lambda_1} \le x|\boldsymbol{f}_i^\dagger\boldsymbol{f}_i = 1\right)\right]^N dx \tag{70}$$

$$= \mathbb{E}_{\boldsymbol{z}_n}\int_0^{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n\boldsymbol{z}_n}\left[\Pr\left(\left(\boldsymbol{f}_i^\dagger\frac{\boldsymbol{\Lambda}_n\boldsymbol{z}_n}{\sqrt{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}}\right)^2 \le \frac{\lambda_1 x}{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}|\boldsymbol{f}_i^\dagger\boldsymbol{f}_i = 1\right)\right]^N dx \tag{71}$$

$$\approx \mathbb{E}_{\boldsymbol{z}_n}\int_0^{\frac{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}{\lambda_1}}\left[\Pr\left(\left(\boldsymbol{f}_i^\dagger\frac{\boldsymbol{\Lambda}_n\boldsymbol{z}_n}{\sqrt{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}}\right)^2 \le \frac{\lambda_1 x}{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}|\boldsymbol{f}_i^\dagger\boldsymbol{f}_i = 1\right)\right]^N dx \tag{72}$$

$$= \mathbb{E}_{\boldsymbol{z}_n}\frac{\boldsymbol{z}_n^\dagger\boldsymbol{\Lambda}_n^2\boldsymbol{z}_n}{\lambda_1^{(n)}}\int_0^1\left[1-(1-x)^{r_n-1}\right]^N dx \tag{73}$$

$$\approx \frac{\sum_{i=1}^{r_n}(\lambda_i^{(n)})^2}{\lambda_1^{(n)}}\ 2^{\frac{-B_n}{r_n-1}}, \tag{74}$$

wherein the equality (69) follows from integrating (68) by parts, the approximation in (72) follows from the work in [31, Appendix J], which shows the dominant term of the integral in (71) is (72), then by [32, Corollary 1], (73) and (74) follows. ■

REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov 2010.

[2] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans Commun.*, vol. 61, pp. 1436–1449, Apr 2013.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, Jul 2003.

[4] A. G. Davoodi and S. A. Jafar, "Aligned image sets under channel uncertainty: Settling a conjecture by Lapidoth, Shamai and Wigger on the collapse of degrees of freedom under finite precision CSIT," *arXiv preprint arXiv:1403.1541*.

[5] G. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," *IEEE Trans. Antennas Propag.*, vol. 52, pp. 1568–1577, Jun 2004.

[6] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast MIMO channels: Curse or blessing?," in *Proc. IEEE Global Communications Conference (GLOBECOM' 08)*, pp. 1–5, 2008.

[7] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inform. Theory*, vol. 59, pp. 6441–6463, Oct 2013.

[8] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Top. Signal Process.*, vol. 8, pp. 876–890, Oct. 2014.

[9] A. Adhikary, E. Al Safadi, M. Samimi, R. Wang, G. Caire, T. Rappaport, and A. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Select. Areas Commun.*, vol. 32, pp. 1239–1255, Jun 2014.

[10] J. Nam, "Fundamental limits in correlated fading MIMO broadcast channels: Benefits of transmit correlation diversity," *arXiv preprint arXiv:1401.7114*, 2014.

[11] R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Top. Signal Process.*, vol. 8, pp. 773–786, Oct 2014.

[12] C. Studer and E. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 303–313, Feb 2013.

[13] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inform. Theory*, Sep 2014.

[14] C. Shepard, H. Yu, and L. Zhong, "ArgosV2: A flexible many-antenna research platform," in *Proc. Annual International Conference on Mobile Computing & Networking*, pp. 163–166, 2013.

[15] X. Rao and V. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Processing*, vol. 62, pp. 3261–3271, Jun 2014.

[16] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-Advanced to FDD massive MIMO systems," *arXiv preprint arXiv:1402.6794*, 2014.

[17] J. Choi, Z. Chance, D. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, pp. 5016–5029, Dec 2013.

[18] J. Choi, D. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Top. Signal Process.*, vol. 8, pp. 802–814, Oct 2014.

[19] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "On the achievable rates of FDD massive MIMO systems with spatial channel correlation," in *Proc. IEEE International Conference on Communications in China (ICCC' 2014)*, 2014.

[20] K. Hugl, J. Laurila, and E. Bonek, "Transformation based downlink beamforming techniques for frequency division duplex systems," in *Proc. Interim Symposium on Antennas and Propagation*, pp. 1529–1532, 2000.

[21] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 90–100, Jan 2006.

[22] S. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE International Conference on Communications (ICC' 2001)*, vol. 7, pp. 2266–2270, 2001.

[23] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach*. Cambridge University Press, 2013.

[24] A. F. Molisch, *Wireless Communications*. 2nd edition, IEEE Press–Wiley, 2011.

[25] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2845–2866, Jun 2010.

[26] J. H. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 546–557, Feb 2004.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2009.

[28] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1261–1282, Apr 2005.

[29] S. Serbetli and A. Yener, "Transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Processing*, vol. 52, pp. 214–226, Jan 2004.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[31] V. Raghavan and V. Veeravalli, "Ensemble properties of RVQ-based limited-feedback beamforming codebooks," *IEEE Trans. Inform. Theory*, vol. 59, pp. 8224–8249, Dec 2013.

[32] C. K. Au-Yeung and D. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 458–462, Feb 2007.

[33] P. Xia and G. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Processing*, vol. 54, pp. 1853–1863, May 2006.

[34] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, pp. 129–137, Mar 1982.

[35] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inform. Theory*, vol. 58, pp. 4509–4537, Jul 2012.

[36] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans Commun.*, vol. 53, pp. 195–202, Jan 2005.

[37] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans Commun.*, vol. 59, pp. 2228–2240, Aug 2011.

[38] D. shan Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar 2000.

[39] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr 2003.

[40] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. 31, pp. 344–347, May 1985.

Zhiyuan Jiang (S'12) was born at Beipiao, Liaoning Province, China in 1987. He received his B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China in 2010. Since September 2010, he has been a graduate student in Niulab of Electronic Engineering Department of Tsinghua University, where he is currently pursuing his doctoral studies. His main research interests include multiuser MIMO systems, green wireless networks, queuing theory and Lyapunov optimization.

Andreas F. Molisch (S'89–M'95–SM'00–F'05) received the Dipl. Ing., Ph.D., and habilitation degrees from the Technical University of Vienna, Vienna, Austria, in 1990, 1994, and 1999, respectively. He subsequently was with AT&T (Bell) Laboratories Research (USA); Lund University, Lund, Sweden, and Mitsubishi Electric Research Labs (USA). He is now a Professor of Electrical Engineering and Director of the Communication Sciences Institute at the University of Southern California, Los Angeles.

His current research interests are the measurement and modeling of mobile radio channels, ultra-wideband communications and localization, cooperative communications, multiple-input?multiple-output systems, wireless systems for healthcare, and novel cellular architectures. He has authored, coauthored, or edited four books (among them the textbook Wireless Communications, Wiley-IEEE Press), 16 book chapters, some 170 journal papers, and numerous conference contributions, as well as more than 70 patents and 60 standards contributions. Dr. Molisch has been an Editor of a number of journals and special issues, General Chair, Tecnical Program Committee Chair, or Symposium Chair of multiple international conferences, as well as Chairman of various international standardization groups. He is a Fellow of the IEEE, Fellow of the AAAS, Fellow of the IET, an IEEE Distinguished Lecturer, and a member of the Austrian Academy of Sciences. He has received numerous awards, among them the Donald Fink Prize of the IEEE, and the Eric Sumner Award of the IEEE.

Giuseppe Caire (S'92 – M'94 – SM'03 – F'05) was born in Torino, Italy, in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino (Italy), in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992 and the Ph.D. from Politecnico di Torino in 1994. He was a recipient of the AEI G.Someda Scholarship in 1991 and has been post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994-1995. He has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and he is currently a professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles and an Alexander von Humboldt Professor with the Electrical Engineering and Computer Science Department of the Technical University of Berlin, Germany.

He served as Associate Editor for the IEEE Transactions on Communications in 1998-2001 and as Associate Editor for the IEEE Transactions on Information Theory in 2001-2003. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the IEEE Communications Society & Information Theory Society Joint Paper Award in 2004 and in 2011. Giuseppe Caire is a Fellow of IEEE since 2005. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.

Zhisheng Niu (M'98–SM'99–F'12) graduated from Beijing Jiaotong University, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. During 1992-94, he worked for Fujitsu Laboratories Ltd., Japan, and in 1994 joined with Tsinghua University, Beijing, China, where he is now a professor at the Department of Electronic Engineering and deputy dean of the School of Information Science and Technology. He is also a guest chair professor of Shandong University, China. His major research interests include queueing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communication and networks.

Dr. Niu has been an active volunteer for various academic societies, including Director for Conference Publications (2010-11) and Director for Asia-Pacific Board (2008-09) of IEEE Communication Society, Membership Development Coordinator (2009-10) of IEEE Region 10, Councilor of IEICE-Japan (2009-11), and council member of Chinese Institute of Electronics (2006-11). He is now a distinguished lecturer (2012-15) and Chair of Emerging Technology Committee (2014-15) of IEEE Communication Society, a distinguished lecturer (2014-16) of IEEE Vehicular Technologies Society, a member of the Fellow Nomination Committee of IEICE Communication Society (2013-14), standing committee member of Chinese Institute of Communications (CIC, 2012-16), and associate editor-in-chief of IEEE/CIC joint publication China Communications.

Dr. Niu received the Outstanding Young Researcher Award from Natural Science Foundation of China in 2009 and the Best Paper Award from IEEE Communication Society Asia-Pacific Board in 2013. He also co-received the Best Paper Awards from the 13th, 15th and 19th Asia-Pacific Conference on Communication (APCC) in 2007, 2009, and 2013, respectively, International Conference on Wireless Communications and Signal Processing (WCSP'13), and the Best Student Paper Award from the 25th International Teletraffic Congress (ITC25). He is now the Chief Scientist of the National Basic Research Program (so called "973 Project") of China on "Fundamental Research on the Energy and Resource Optimized Hyper-Cellular Mobile Communication System" (2012-2016), which is the first national project on green communications in China. He is a fellow of both IEEE and IEICE.