



This is a repository copy of *Non-iterative downlink training sequence design based on sum rate maximization in FDD massive MIMO systems*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/165916/>

Version: Published Version

Article:

Alsabah, M.Q., Vehkaperä, M. and O'Farrell, T. orcid.org/0000-0002-7870-4097 (2020) Non-iterative downlink training sequence design based on sum rate maximization in FDD massive MIMO systems. *IEEE Access*, 8. pp. 108731-108747. ISSN 2169-3536

<https://doi.org/10.1109/access.2020.3000302>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Received April 9, 2020, accepted May 1, 2020, date of publication June 5, 2020, date of current version June 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000302

Non-Iterative Downlink Training Sequence Design Based on Sum Rate Maximization in FDD Massive MIMO Systems

MUNTADHER ALSABAH¹, MIKKO VEHKAPERÄ², AND
TIMOTHY O'FARRELL¹, (Senior Member, IEEE)

¹Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 4ET, U.K.

²VTT Technical Research Centre of Finland Ltd., 90571 Oulu, Finland

Corresponding author: Timothy O'Farrell (t.ofarrell@sheffield.ac.uk)

This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Project EP/L026147/1.

ABSTRACT This paper considers the problem of downlink (DL) training sequence design with limited coherence time for frequency division duplex (FDD) massive MIMO systems in a general scenario of single-stage precoding and distinct spatial correlations between users. To this end, a computationally feasible solution for designing the DL training sequences is proposed using the principle of linear superposition of sequences constructed from the users' channel covariance matrices. Based on the non-iterative superposition training structure and the P -degrees of freedom (P -DoF) channel model, a novel closed-form solution for the optimum training sequence length that maximizes the DL achievable sum rate is provided for the eigenbeamforming (BF) precoder. Additionally, a simplified analysis that characterizes the sum rate performance of the BF and regularized zero forcing (RZF) precoders in closed-form is developed based on the method of random matrix theory and the P -DoF channel model. The results show that the superposition training sequences achieve almost the same rate performances as state-of-the-art training sequence designs. The analysis of the complexity results demonstrates that more than four orders-of-magnitude reduction in the computational complexity is achieved using the superposition training design, which signifies the feasibility of this approach for practical implementations compared with state-of-the-art iterative algorithms for DL training designs. Importantly, the results indicate that the analytical solution for the optimum training sequence length with the P -DoF channel model can be effectively used with high accuracy to predict the sum rate performance in the more realistic one ring (OR) channel model, and thus, near optimal solutions can be readily obtained without resorting to computationally intensive optimization techniques.

INDEX TERMS Massive MIMO, achievable sum rate, training sequence design, channel estimation, time division duplex, frequency division duplex, spatial channel correlation, random matrix theory.

I. INTRODUCTION

Next generation cellular systems require to maximize spectral efficiency to satisfy the rapidly increasing demand for wireless data services [1], whilst reducing both the cost and energy consumed [2], [3]. Massive multiple-input multiple-output (massive MIMO), proposed in [4], is introduced as one of the most promising technologies to achieve this goal. In particular, massive MIMO transmission has several advantages such as: (a) allowing the use of linear precoding schemes with low complexity signal processing; (b) achieving a uniform quality of service across the entire cell; (c) providing immunity

against fading; and (d) reducing the base station (BS) energy consumption [5]–[10].

To achieve the full potential of massive MIMO, sufficiently accurate and timely estimates of the channel state information (CSI) at the BS are required [11], [12]. Early research on massive MIMO systems focused on the time division duplex (TDD) operation, where the required CSI is obtained by sending a superposition of orthogonal sequences over a length of T_p symbols in the uplink (UL) direction during each coherence interval [5], [6], [8], [13], [14]. The authors in [15], [16] found that the optimum number of UL training symbols is proportional to the number of user terminals (UTs) K and independent of the number of BS antennas N that can be made as large as required. Under UL and downlink (DL)

The associate editor coordinating the review of this manuscript and approving it for publication was Jie Tang.

channel reciprocity that holds in TDD systems, the UL CSI estimates are used for designing the DL precoder without the requirement for DL CSI estimation. However, in FDD systems, the DL and UL channels occupy different frequency bands [17], and hence, estimation of the DL CSI using UL training sequences is not possible. As such, the framework analysis for the optimum training design developed in [15], [16] cannot be used to predict the performance of an FDD massive MIMO system. To obtain the CSI in FDD systems, the UTs would need to estimate the DL channels of each of the N BS antennas and send the quantized channel estimates back to the BS to design the precoder [18]. This is generally deemed unfeasible for the FDD massive MIMO systems with large N since the overhead for the DL CSI estimation is proportional to the number of BS antennas N [11], [12], [19]. As such, the available coherence time would be largely occupied by the channel training, leaving insufficient time for transmitting useful data to the UTs [19], [20].

To address the challenge of FDD operation in massive MIMO systems, several studies have investigated the design of DL training sequences using different channel models and design criteria, see e.g., [21]–[28]. In particular, the research in [21]–[23] explores the joint use of the spatial and temporal channel correlations in which the training sequences are designed based on the criterion of minimizing the minimum mean squared error (MSE) of the channel estimate in a scenario where all users exhibit a common spatial correlation. However, in practice, users could exhibit distinct spatial correlation patterns; therefore, the optimization framework of the DL training sequences developed in [21]–[23] does not hold in the general scenario with heterogeneous user channels. Another line of research studies have focused on the design of training sequences for FDD massive MIMO systems by utilising a two-stage-precoding technique, termed as joint spatial division and multiplexing (JSDM) [24], [29]. Specifically, the research in [24], [29] exploits correlations in the spatial domain, where the users within each group exhibit the same spatial correlation, and a linear superposition of each group correlation matrix is used to perform the first of two stages of precoding, thus forming a beam for each group. As such, the training sequence length in the DL can be scaled linearly with the number of user groups, which can be less than N , resulting in a feasible pilot overhead requirement for FDD operation. While the two-stage precoding technique helps to constrain the training sequence length and maximizes the sum rate criterion, sophisticated scheduling and clustering algorithms of the user groups, and of the users inside each group, are essential, thus constraining the approach. Furthermore, the research in [24], [29] does not address the challenge of designing the DL training sequence in single-stage precoding with K distinct covariance matrices, and thus, cannot predict the optimum training length that maximizes the sum rate performance in this preferred scenario. The research in [25]–[28] considered a general scenario of single-stage precoding with distinct spatial correlations, where the training sequences

in [25], [26] and [27], [28] are designed iteratively by using different iterative algorithms as a solution to a sum conditional mutual information (SCMI) maximization criterion and a sum mean square error (SMSE) minimization criterion, respectively. While advanced iterative algorithms have been developed in the aforementioned research, they provide no closed-form solution for the optimum DL training sequence that maximizes the sum rate with limited coherence time. Furthermore, the limited coherence time interval implies that the CSI estimation should be determined more frequently, and thus, iterative-based solutions for the DL training sequence design may be infeasible.

A. CONTRIBUTIONS AND PAPER FINDINGS

This paper addresses the challenge of DL channel estimation in an FDD massive MIMO communication system with single-stage precoding and limited coherence time using a non-iterative approach for the DL training sequence design. To this end, the principle of linear superposition, in which the DL training sequences are constructed from the eigenvectors of the K distinct correlation matrices, is proposed, which allows a feasible solution for DL channel estimation to be achieved with a reduced design complexity, thereby avoiding the design of existing training sequences that require computationally demanding iterative algorithms. Based on the superposition training approach and the P -degrees of freedom (P -DoF), a new analytical closed-form solution for the optimum training sequence length that maximizes the DL achievable sum rate in an FDD massive MIMO system is provided for the eigenbeamforming (BF) precoder. In addition, an asymptotic random matrix theory along with the P -DoF channel model are adopted in this paper to provide a straightforward analysis of the sum rate for the BF and regularized zero forcing (RZF) precoders. Comparisons between the sum rates of the superposition training design and the state-of-the-art sequences designed based on iterative algorithms [25], [28], are conducted based on the P -DoF and the one ring (OR) [30] channel models. Furthermore, the computational complexity of the superposition sequence design is analyzed and compared with the state-of-the-art iterative algorithms.

We found that the diversity of spatial correlations between multiple users significantly improves the sum rate performance in comparison to uncorrelated channels with identical covariance matrices. The numerical results demonstrate that the superposition training sequences achieve almost the same rate performances as state-of-the-art training designs while reducing the computational complexity. Importantly, the results show that the pilot length that is optimized for the BF precoder is also sufficient to predict the rate performance of the BF and RZF precoders in the more realistic OR channel model. Overall, the proposed design paradigm allows a pragmatic DL training design for an FDD massive MIMO system to be achieved with a significant computational complexity reduction.

B. PAPER ORGANIZATION AND NOTATION

The paper is organized as follows. In Section II, the system model is introduced. In Section III, we explain the channel estimation process based on the DL training sequence together with the problem formulation. In Section IV, the SINR analyses of the BF and RZF precoders based on the random matrix theory are developed, which are then used in Section V with P -DoF channel model to provide a closed-form solution of the optimum training sequence length for the BF precoder and an explicit mathematical analyses of the DL sum rates for the BF and RZF precoders. In Section VI, numerical results are provided in order to characterize the system performance and validate the analyses. Finally, the paper is concluded in Section VII.

Notation: In the present paper, an upper boldface symbol stands for a matrix whereas a lower boldface symbol stands for a vector. $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ denotes the circularly symmetric complex Gaussian (CSCG) probability distribution with mean $\mathbf{0}$ and covariance matrix \mathbf{R} . The term $\mathbb{E}[\cdot]$ refers to the expectation operator. \mathbf{I}_N denotes the $N \times N$ identity matrix. The operators trace, transpose, Hermitian transpose, inverse and absolute value are denoted by $\text{tr}(\cdot)$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, and $|\cdot|$, respectively. $[\mathbf{A}]_{:,j:m}$ denotes a submatrix containing columns j through m of matrix \mathbf{A} . We use $[\mathbf{a}]_k$ and $[\mathbf{A}]_{k,l}$ to denote the element in the k th row of vector \mathbf{a} and the element in the k th row and l th column of matrix \mathbf{A} , respectively.

II. SYSTEM MODEL

The present model in this paper considers a single-cell, DL mobile wireless communications system where the BS is equipped with a uniform linear array (ULA) of N antennas and employs single-stage precoding, which serves K single antenna user terminals (UTs) with $N \gg K$. Non-line-of-sight (NLOS) Rayleigh fading channels over a single-frequency band are considered with an overall coherence time denoted by $T_c \in \mathbb{Z}^+$ and enumerated in symbols per transmission block.

As depicted in Fig. 1, the available coherence time T_c is divided into the training duration T_p , the feedback time T_f and the data transmission duration T_d . As the purpose of the present paper is to concentrate on the DL training sequence design, minimizing the training sequence length over a limited coherence time, the UL feedback time T_f and associated error rate are assumed to be zero, as considered in [15], [22]–[24], [28], [29], [31].

The DL received signal during the data-phase at the k -th UT may be written as [5], [6]

$$y_k = \sqrt{\rho_d} \mathbf{h}_k^H \mathbf{s} + n_k, \quad (1)$$

where $\mathbf{h}_k \in \mathbb{C}^N$ is the complex channel vector between the BS and the k -th UT and additive noise n_k is modeled as a zero mean unit variance CSCG random variable. The DL transmit vector $\mathbf{s} \in \mathbb{C}^N$ is given as

$$\mathbf{s} = \sqrt{\xi} \mathbf{W} \mathbf{x}, \quad (2)$$

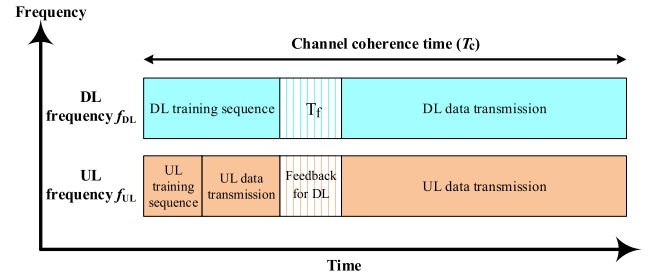


FIGURE 1. Separate DL and UL training and data payload phases of the block fading model in FDD mode.

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K] \in \mathbb{C}^{N \times K}$ is the precoding matrix at the BS and $\mathbf{x} = [x_1, \dots, x_K]^T \in \mathbb{C}^K$ is a zero mean CSCG vector of data symbols satisfying $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_K$. When the matrix power normalization technique is used [6], [32], the normalization constant ξ can be written as

$$\xi = \frac{K}{\mathbb{E}[\text{tr}(\mathbf{W}\mathbf{W}^H)]}, \quad (3)$$

which ensures that $\mathbb{E}[\|\mathbf{s}\|^2] = K$ and the average per-user BS transmit power is ρ_d during the data-phase. To this end, a DL achievable sum rate, C , for the massive MIMO system under consideration can be expressed as

$$C = \left(1 - \frac{T_p}{T_c}\right) \sum_{k=1}^K \log_2(1 + \text{SINR}_k), \quad (4)$$

where the associated signal-to-interference-plus-noise ratio SINR term at the k -th UT is given by (5) [5], [6]. The term $1/\rho_d$ is the inverse of the per user signal-to-noise-ratio (SNR) during the data-phase.

SINR_k

$$= \frac{\xi |\mathbb{E}[\mathbf{h}_k^H \mathbf{w}_k]|^2}{\frac{1}{\rho_d} + \xi \mathbb{E}[|\mathbf{h}_k^H \mathbf{w}_k - \mathbb{E}[\mathbf{h}_k^H \mathbf{w}_k]|^2] + \xi \sum_{l \neq k}^K \mathbb{E}[|\mathbf{h}_k^H \mathbf{w}_l|^2]} \quad (5)$$

The sum rate lower bound in (4) depends on the channel statistics, the channel estimates and the linear precoding technique used at the BS. We consider specifically two commonly prevailing types of linear precoders, the eigenbeamforming (BF) or maximum ratio transmitter (MRT) precoder and the regularized zero forcing (RZF) precoder as defined in (6) [6],

$$\mathbf{W} = \begin{cases} \hat{\mathbf{H}}^H & \text{for BF/MRT,} \\ \hat{\mathbf{H}}^H (\hat{\mathbf{H}} \hat{\mathbf{H}}^H + N \zeta \mathbf{I}_K)^{-1} & \text{for RZF,} \end{cases} \quad (6)$$

where ζ is the regularization parameter, which is considered to be the inverse of the per user SNR, $1/\rho_d$ [6]. The term $\hat{\mathbf{H}}$ is the estimate of the DL channels $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]^H \in \mathbb{C}^{K \times N}$, where each channel \mathbf{h}_k , $k = 1, \dots, K$, is modeled as a zero mean, independent CSCG random vector. The following section explains how $\hat{\mathbf{H}}$, which is required at the BS for precoding, is estimated. The k -th user's correlation matrix,

$\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] \in \mathbb{C}^{N \times N}$, is normalized as $\mathbb{E}[\text{tr}(\mathbf{H}^H \mathbf{H})] = \sum_{k=1}^K \text{tr}(\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H]) = \sum_{k=1}^K \text{tr}(\mathbf{R}_k) = NK$. The covariance matrix \mathbf{R}_k is considered to be locally stationary, varying more slowly than the instantaneous channel of the coherence time [22], [24], [29], [33], [34], and thus, may be accurately estimated by either the FDD or TDD schemes considered [35]–[40].

III. CHANNEL ESTIMATION AND PROBLEM FORMULATION

This section addresses the problem of channel estimation using DL training sequences in an FDD massive MIMO communication system.

A. CHANNEL ESTIMATION USING DOWNLINK TRAINING SEQUENCES

To estimate the DL channel, the BS transmits predetermined pilot sequences of duration or length T_p during the training-phase. The received training-signal, $\mathbf{y}_k \in \mathbb{C}^{T_p}$, at the k -th user is given by

$$\mathbf{y}_k = \sqrt{\rho_p} \mathbf{S}_p^H \mathbf{h}_k + \mathbf{n}_k, \quad (7)$$

where $\mathbf{S}_p \in \mathbb{C}^{N \times T_p}$ is the spatio-temporal common pilot matrix, which is normalized as $\text{tr}(\mathbf{S}_p^H \mathbf{S}_p) = T_p$ so that the average transmitted power during the training-phase is equal to ρ_p . The receiver noise $\mathbf{n}_k \in \mathbb{C}^{T_p}$ exhibits a CSCG distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{T_p})$. Since the channel vector \mathbf{h}_k follows a CSCG distribution with known statistics at the BS, linear filters, which exploit the channel statistics to optimize channel estimation performance, can be used. To this end, an optimized channel estimation performance in the DL with $T_p < N$ is achieved by utilizing Bayesian estimation, i.e. employing minimum-mean square-error (MMSE) filter, which makes use of channel and noise statistics. Accordingly, the k -th user's covariance matrix with MMSE channel estimation $\forall k$ is expressed as [41]

$$\Phi_k = \sqrt{\rho_p} \mathbf{R}_k \mathbf{S}_p (\rho_p \mathbf{S}_p^H \mathbf{R}_k \mathbf{S}_p + \mathbf{I}_{T_p})^{-1} \sqrt{\rho_p} \mathbf{S}_p^H \mathbf{R}_k. \quad (8)$$

In this paper, the structure of a common pilot matrix \mathbf{S}_p is designed by jointly considering all the K distinct channel covariance matrices where the effective eigenvectors of each of the matrices are combined using the principle of linear superposition. In particular, a unique pilot sequence for the training-phase is developed based on the eigenvalue decomposition (EVD) of the channel covariance matrices

$$\mathbf{R}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^H, \quad (9)$$

where $\mathbf{U}_k = [\mathbf{u}_{k,1}, \dots, \mathbf{u}_{k,N}] \in \mathbb{C}^{N \times N}$ is a unitary matrix of the eigenvectors and $\mathbf{\Lambda}_k \in \mathbb{R}^{N \times N}$ is a diagonal matrix of the eigenvalues of \mathbf{R}_k arranged in descending order $\lambda_{k,1} \geq \lambda_{k,2} \geq \dots \geq \lambda_{k,N}$. Specifically, the pilot matrix $\mathbf{S}_p \in \mathbb{C}^{N \times T_p}$ is constructed from the superposition of the first T_p eigenvectors of \mathbf{R}_k , corresponding to the largest eigenvalues,

as expressed in (10).

$$\mathbf{S}_p = \frac{\sum_{k=1}^K [\mathbf{U}_k]_{:,1:T_p}}{\left\| \sum_{k=1}^K [\mathbf{U}_k]_{:,1:T_p} \right\|_F} \quad (10)$$

The pilot matrix in (10) is normalized by the Frobenius norm to satisfy the power constraint $\text{tr}(\mathbf{S}_p^H \mathbf{S}_p) = T_p$. In principle, increasing T_p allows for more pilot signal energy to be received, but it comes at the cost of reduced spectral efficiency due to a shorter data transmission phase. Therefore, the energy in the channel that is related to the last eigenvector columns $[\mathbf{U}_k]_{:,T_p+1:N}$ is not used in precoding.

B. FORMULATION OF THE OPTIMIZATION PROBLEM

Maximizing the downlink sum rate over the training-phase duration in the massive MIMO system under consideration equates to the optimization problem defined in (11).

$$\begin{aligned} & \underset{T_p}{\text{maximize}} \quad \left(1 - \frac{T_p}{T_c}\right) \sum_{k=1}^K \log_2(1 + \text{SINR}_k) \\ & \text{subject to} \quad 1 \leq T_p \leq \min(N, T_c) \end{aligned} \quad (11)$$

Though the problem setup may seem straightforward, solving (11) for arbitrary correlation matrices \mathbf{R}_k and finite values of N is computationally demanding since the expectations in (5) need to be evaluated for different choices of T_p using extensive Monte Carlo simulations. A computationally feasible solution to finding an optimum training sequence length T_p^* may be obtained by invoking asymptotic random matrix theory (RMT) methods [6], [15], [32], [42].

To date, the optimum pilot length aiming to maximize the achievable sum rate performance of FDD massive MIMO systems is obtained by exhaustively searching from all possible combinations of $1 \leq T_p \leq \min(N, T_c)$. However, with the scaling up of the BS array and increasing of the coherence time, such an exhaustive search might be time and resource consuming, and hence, practically infeasible. To the best of our knowledge, finding a closed-form solution for the optimum training sequence length T_p^* that maximizes $(1 - T_p/T_c) \sum_{k=1}^K \log_2(1 + \text{SINR}_k)$ in DL single-stage precoding of an FDD massive MIMO system, has not been solved due to the technical challenge of deriving a closed-form solution.

IV. ACHIEVABLE SUM RATE ANALYSIS

In this section, we provide the expressions that accurately approximate the SINR_k and downlink sum rate for the massive MIMO system under consideration based on the asymptotic random matrix theory approach in [6]. In particular, an asymptotically tight approximation of the SINR_k , denoted $\overline{\text{SINR}}_k$, for the SINR_k equation defined in (5) is obtained as indicated in (12) when N and K grow without bound while the ratio $K/N > 0$ is kept constant.

$$\text{SINR}_k - \overline{\text{SINR}}_k \xrightarrow[N \rightarrow \infty]{} 0 \quad (12)$$

Consistent with previous literature of large system analysis [6], [15], [20], [32], [43], our numerical results show that these approximations are also highly valid for practical, finite values of N and K . This allows us to replace the SINR_k term in (5) for the BF and RZF precoders with the approximations given in this section, so that the optimization problem may be rewritten as in (13).

$$\begin{aligned} & \underset{T_p}{\text{maximize}} \left(1 - \frac{T_p}{T_c} \right) \sum_{k=1}^K \log_2(1 + \overline{\text{SINR}}_k) \\ & \text{subject to } 1 \leq T_p \leq \min(N, T_c) \end{aligned} \quad (13)$$

The following propositions for the BF and RZF precoders are modified and repurposed versions of Theorem 4 and Theorem 6 in [6], respectively. Particularly, the precoders under consideration are designed based on an imperfect channel estimation in the downlink of a single-cell FDD massive MIMO system. As such, the MMSE channel estimation is obtained based on the downlink training sequence that is given in (10). Further details on the asymptotic random matrix theory methods can be found in [6], [42].

Proposition 1: Let $\text{SINR}_k^{\text{BF}}$ denote the SINR_k for BF precoding as given in (5). An asymptotically tight approximation of $\text{SINR}_k^{\text{BF}}$ for the regime where N and K go to infinity with a given ratio reads

$$\overline{\text{SINR}}_k^{\text{BF}} = \frac{\bar{\xi} (\text{tr}(\Phi_k))^2}{\frac{1}{\rho_d} + \sum_{i=1}^K \bar{\xi} \text{tr}(\mathbf{R}_k \Phi_i)}, \quad (14)$$

where $\bar{\xi} = \left(\frac{1}{K} \sum_{k=1}^K \text{tr}(\Phi_k) \right)^{-1}$, which denotes the precoder normalization, and Φ_k is the k -th user's covariance matrix with MMSE channel estimation, which is provided in (8).

The expression (14) is generally valid for any channel correlation model and training sequence type. Unlike the expression for BF precoding, the SINR approximation for the RZF precoder is given in terms of several auxiliary variables. These variables arise from the asymptotic random matrix theory analysis and, in general, need to be numerically solved before the SINR approximation is obtained. A simplified analysis where the performance of the RZF precoder can be characterized in closed-form is provided in Section V based on the P -DoF channel model.

Proposition 2: Let $\text{SINR}_k^{\text{RZF}}$ denote the SINR_k for RZF precoding as given in (5). An asymptotically tight approximation of $\text{SINR}_k^{\text{RZF}}$ for a general correlation model is given as

$$\overline{\text{SINR}}_k^{\text{RZF}} = \frac{N \bar{\xi} \delta_k^2}{\frac{(1+\delta_k)^2}{\rho_d K} + \bar{\xi} \sum_{i=1}^K \left(\frac{1+\delta_k}{1+\delta_i} \right)^2 \mu_{k,i}}, \quad (15)$$

where the term $\bar{\xi} \in \mathbb{R}$ is obtained later in (23). Parameter $\delta_k \in \mathbb{R}$ is a unique solution to a fixed-point equation that arises from the asymptotic random matrix theory analysis and needs to be numerically solved. An analytical closed-form solution for δ_k is provided in subsection V-C based on the P -DoF channel model. Recalling the covariance matrix of

the MMSE channel estimate Φ_k given in (8) and defining a recursion on integer t , where $t = 1, 2, \dots$,

$$\delta_k^{(t)} = \frac{1}{N} \text{tr} \left(\Phi_k \left(\frac{1}{N} \sum_{i=1}^K \frac{\Phi_i}{1 + \delta_i^{(t-1)}} + \zeta \mathbf{I}_N \right)^{-1} \right), \quad (16)$$

with an initial value $\delta_k^{(0)} = 1/\zeta$ for all k where $\zeta > 0$ is the regularization parameter in (6), the variable $\delta_k \in \mathbb{R}$ is found numerically by the standard fixed-point algorithm as the limit

$$\delta_k = \lim_{t \rightarrow \infty} \delta_k^{(t)}, \quad (17)$$

After the solution of the fixed-point equations in (16) and (17) is numerically obtained, it is substituted into

$$\mathbf{T} = \left(\zeta \mathbf{I}_N + \frac{1}{N} \sum_{k=1}^K \frac{\Phi_k}{1 + \delta_k} \right)^{-1}, \quad (18)$$

to obtain random matrix $\mathbf{T} \in \mathbb{C}^{N \times N}$. Auxiliary matrix $\bar{\mathbf{T}} \in \mathbb{C}^{N \times N}$ is given by

$$\bar{\mathbf{T}} = \mathbf{T} \left(\mathbf{I}_N + \frac{1}{N} \sum_{k=1}^K \frac{\Phi_k \bar{\delta}_k}{(1 + \delta_k)^2} \right) \mathbf{T}, \quad (19)$$

and $\bar{\delta} \triangleq [\bar{\delta}_1 \dots \bar{\delta}_K]^T$ is given as

$$\bar{\delta} = (\mathbf{I}_K - \mathbf{J})^{-1} \bar{\mathbf{v}}, \quad (20)$$

where $\mathbf{J} \in \mathbb{C}^{K \times K}$ and $\bar{\mathbf{v}} \in \mathbb{C}^K$ are obtained from the expressions given in (21) and (22).

$$[\mathbf{J}]_{k,l} = \frac{\text{tr}(\Phi_k \mathbf{T} \Phi_l \mathbf{T})}{(N(1 + \delta_k))^2}, \quad 1 \leq k, l \leq K, \quad (21)$$

$$[\bar{\mathbf{v}}]_k = \frac{1}{N} \text{tr}(\mathbf{T} \Phi_k \mathbf{T}), \quad 1 \leq k \leq K. \quad (22)$$

Parameter $\bar{\xi} \in \mathbb{R}$ in (15) corresponds to the precoder normalization and is obtained by substituting the matrices \mathbf{T} and $\bar{\mathbf{T}}$ into

$$\bar{\xi} = (\text{tr}(\mathbf{T}) - \zeta \text{tr}(\bar{\mathbf{T}}))^{-1}. \quad (23)$$

The auxiliary variable $\mu_{k,i} \in \mathbb{R}$ in (15) is obtained using the dominated convergence and the continuous mapping theorems as developed in [6], which is modified here for the massive MIMO system under consideration, and thus, it is provided from the expressions given in (24), (25) and (26),

$$\begin{aligned} \bar{\mu}_{k,i} &= \frac{1}{N} \text{tr}(\mathbf{R}_k \mathbf{T}'_i) \\ &= \frac{2 \text{Re}(\text{tr}(\Phi_k \mathbf{T}) \text{tr}(\Phi_k \mathbf{T}'_i)) (1 + \delta_k) - \text{tr}(\Phi_k \mathbf{T})^2 \delta'_i}{(N(1 + \delta_k))^2}, \end{aligned} \quad (24)$$

$$\mathbf{T}'_i = \mathbf{T} \left(\Phi_i + \frac{1}{N} \sum_{k=1}^K \frac{\Phi_k \delta'_k}{(1 + \delta_k)^2} \right) \mathbf{T}, \quad (25)$$

$$\delta' = (\mathbf{I}_K - \mathbf{J})^{-1} \mathbf{v}', \quad (26)$$

where $\mathbf{v}' \in \mathbb{C}^K$ denotes

$$[\mathbf{v}']_k = \frac{1}{N} \text{tr}(\mathbf{T}\Phi_k \mathbf{T}\Phi_k), \quad 1 \leq k \leq K. \quad (27)$$

The SINR approximation given in (15) is generally valid for any channel correlation model and training sequence type. Determining the sum rate for the RZF precoder in a form that is useful for the optimization problem defined in (13) is still challenging. In order to solve this and gain further insight into the problem, we utilize the analytical P -DoF channel model considered in [6] in the next section. This supports the formulation of both $\overline{\text{SINR}}_k^{\text{BF}}$ and $\overline{\text{SINR}}_k^{\text{RZF}}$, leading to the successful computation of their respective sum rates with very low computational complexity.

V. CLOSED-FORM SUM RATE ANALYSIS AND TRAINING SEQUENCE OPTIMIZATION BASED ON THE P -DoF CHANNEL MODEL

In practice, field measurements have shown that the MIMO channel coefficients are spatially correlated in outdoor [44] and indoor propagation environments [45], [46]. The correlation model depends on the number of degrees of freedom offered by the physical channel, which can be much smaller than the number of BS antennas N . Therefore, the correlation model can be modelled as a P -dimensional subspace, where P is the number of angular spatial directions or bins. These angular bins correspond to the number of significant multipaths in the angular domain. In this section, an analytical physical channel model [6], [10], [47], [48] is considered, where the angular domain is separated into a finite P number of directions i.e. P -DoF. The P -DoF channel model is used to obtain a closed-form solution for the achievable sum rate in the BF and RZF precoders. The model also allows the SINR to be expressed in closed-form for both the BF and RZF precoders and the optimum training sequence length to be mathematically derived and expressed in an analytical form for the BF precoder but not the RZF precoder, the latter requiring numerical computation.

From [6], [10], [47], [48], the analytical P -DoF channel model for the system under consideration is written as

$$\mathbf{h}_k = \sqrt{\frac{N}{P}} \mathbf{A}_k \mathbf{b}_k, \quad (28)$$

where $P/N \in (0, 1]$, the elements of $\mathbf{b}_k \in \mathbb{C}^P$ are i.i.d. $\mathcal{CN}(0, 1)$ and $\mathbf{A}_k \in \mathbb{C}^{N \times P}$ is constructed from $P \leq N$ columns of an arbitrary $N \times N$ unitary matrix so that $\mathbf{A}_k^H \mathbf{A}_k = \mathbf{I}_P$. Clearly, \mathbf{R}_k has rank P and the channel is stochastically rank deficient if $P < N$. In general, the ratio P/N controls the degrees of freedom of the channel and represents the extent of correlation or amount of scattering in the channel, and thus, models the radio environment [6], [11], [47]–[50]. The smaller P/N is, the more concentrated the channel energy is in the non-zero eigenmodes and the more correlated the channel is. The normalization factor in (28) guarantees that the total average power of the channel is $\mathbb{E}[\text{tr}(\mathbf{H}^H \mathbf{H})] = \sum_{k=1}^K \text{tr}(\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H]) = \sum_{k=1}^K \text{tr}(\mathbf{R}_k) = NK$, as desired (see Section II). The covariance matrices of the users for the

channel defined in (28) are of the form $\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] = \frac{N}{P} \mathbf{A}_k \mathbf{A}_k^H$, where the expectation is taken over \mathbf{b}_k .

A. CHANNEL COVARIANCE MATRICES AND ANGLE OF ARRIVAL (AoA) SUPPORTS

In this subsection, we present special cases of the channel covariance matrices namely; non-overlapping and overlapping angle of arrival (AoA) supports.

1) NON-OVERLAPPING AoA SUPPORTS

While the channel of different users associated with the BS can be random, they might exist in mutually orthogonal subspaces. This behaviour of the channel is known as a non-overlapping AoA support [24], [29], [51]–[53], in which the AoA of desired and interfering users are disjoint. In practice, the scenario of the non-overlapping signal subspaces between the desired and the interfering users can be realized when the ULA at the BS has a large number of antenna elements and when the users are well separated in the angular domain [51]. In the multiuser scenario with non-overlapping AoA supports, the covariance matrices of the users are all statistically different and independent of each other.

Recalling the EVD of \mathbf{R}_k given in (9) and the definition of non-overlapping AoA supports [51], in which the channel covariance matrices can be asymptotically orthogonal and satisfy

$$\mathbf{U}_k^H \mathbf{U}_l = 0, \quad \forall l \neq k, \text{ as } N \rightarrow \infty. \quad (29)$$

In this special condition, the transmitted pilot sequences between users do not interfere. The unitary symmetric structure of the discrete Fourier transform (DFT) matrix can be used to obtain perfectly non-overlapping AoA supports with the P -DoF channel model (28). For a ULA with a large number of antenna elements, we expect to have similar performances for the channel model defined in (28) and that discussed in Section VII [29], due to their asymptotic equivalence. For the particular case of the DFT based channel model, the pilot matrix in (10) was constructed based on the principle of linear superposition of subsets formed from the columns of the DFT matrix. To this end, the SINR_k expression for BF precoding, defined by (14), can be further simplified for the particular system under consideration. Specifically, for the P -DoF channel model and based on non-overlapping AoA supports, it is straightforward to obtain simplified expressions for $\text{tr}(\Phi_k)$ and $\text{tr}(\mathbf{R}_k \Phi_k)$ in (14) as

$$\text{tr}(\Phi_k) = \sum_{n=1}^{T_p} \frac{\lambda_{k,n}^2}{\lambda_{k,n} + K/\rho_p}, \quad (30)$$

$$\text{tr}(\mathbf{R}_k \Phi_k) = \sum_{n=1}^{T_p} \frac{\lambda_{k,n}^3}{\lambda_{k,n} + K/\rho_p}, \quad (31)$$

where $\lambda_{k,n} \geq 0 \forall k = 1, \dots, K$ and $\forall n = 1, \dots, T_p$ are the ordered eigenvalues for the channel covariance matrices \mathbf{R}_k namely, $\lambda_{k,1} = \lambda_{k,2} = \dots = \lambda_{k,P} = N/P$ and $\lambda_{k,P+1} = \dots = \lambda_{k,N} = 0$ for each user. Substituting \mathbf{R}_k of the

P -DoF channel model into (8) yields, after some algebraic manipulations, the channel estimate covariance matrix as

$$\Phi_k = \frac{1}{\left(1 + \frac{K}{\rho_p} \frac{P}{N}\right)^{\frac{P}{N}}} \mathbf{U}_{k,m} \mathbf{U}_{k,m}^H, \quad m = \begin{cases} T_p & \text{if } T_p \leq P; \\ P & \text{otherwise,} \end{cases} \quad (32)$$

where $\mathbf{U}_{k,m} \in \mathbb{C}^{N \times m}$ denotes a matrix constructed from m eigenvectors of \mathbf{R}_k .

2) OVERLAPPING AoA SUPPORTS

When the users exhibit completely overlapping AoA supports, the covariance matrices of the users are all statistically the same and of the form $\mathbf{R}_k = \mathbf{R} \forall k = 1, \dots, K$. This implies that the EVD of \mathbf{R} is given by (9) where the pilot matrix $\mathbf{S}_p \in \mathbb{C}^{N \times T_p}$ in (10) is constructed from the first T_p eigenvectors of $\mathbf{U}_k = \mathbf{U} \forall k = 1, \dots, K$, corresponding to the largest eigenvalues of \mathbf{R} . Specifically, for equal channel covariance matrices, the channel estimates of all the users are statistically equal. Straightforward algebra provides a simplified expression for the covariance matrix of the MMSE channel estimate with overlapping AoA supports, which is obtained in a manner similar to the non-overlapping AoA supports, discussed earlier.

Since the channel has $P \leq N$ degrees of freedom and the energy consumed by the training-phase is $\rho_p T_p$, choosing $T_p > P$ while keeping ρ_p constant leads to the same channel estimation performance as $T_p = P$ but unnecessarily consumes more energy in the training-phase. Therefore, $T_p \leq P$ is always assumed in the following. Though the transmitted power per user during the training-phase decreases by a factor of K due to the channel estimates of K users, the numerical results in Section VI show that the achievable sum rate is maximized when the users have orthogonal covariance matrices. This implies that the channel is divided into K interference free single users, and thus, suggests that the achievable sum rate performance benefits from the diversity of spatial channel correlations across multiple users during the data-phase. Thus, the more distinct the spatial correlations exhibited by the users, the less the residual interference is and, thus, enhanced system performance in terms of data rate can be achieved.

While the P -DoF channel model introduced in this section may seem highly idealistic, the numerical results in Section VI show that the optimal training sequence length selection is very similar also in more realistic channel models, such as the OR model [30], which justifies its use. The use of an analytical P -DoF channel model provides a straightforward system design methodology.

B. TRAINING SEQUENCE LENGTH OPTIMIZATION AND CLOSED-FORM SUM RATE ANALYSIS FOR BF PRECODING IN FDD SYSTEMS

In this subsection, we show how the optimum DL training sequence length is obtained analytically in a closed-form for the BF precoder at high SNR in the P -DoF channel.

We start by combining (32) with (30)–(31). Straightforward algebra provides asymptotic tight approximations for the SINR for the BF precoder with overlapping $\overline{\text{SINR}}^{\text{BF}}$ and non-overlapping $\overline{\text{SINR}}_{\perp}^{\text{BF}}$ AoA supports, respectively, as

$$\overline{\text{SINR}}^{\text{BF}} = \frac{\rho_d \rho_p T_p}{(P/N + \rho_p)(P/N + K \rho_d)}, \quad T_p \leq P. \quad (33)$$

$$\overline{\text{SINR}}_{\perp}^{\text{BF}} = \frac{\rho_d \rho_p T_p}{(P/N + \rho_p)(KP/N + \rho_d)}, \quad T_p \leq P. \quad (34)$$

Remark 1: Comparing (33) and (34) to Eq. (28) in [6], which provides an SINR approximation for the BF precoder in TDD massive MIMO when length K uplink pilot sequences are transmitted, shows that the effect of DL channel estimation on SINR is to replace the DoF of the channel P by the training sequence length $T_p \leq P$. Clearly, the choice $T_p = P$ maximizes (34) and yields the same SINR as with UL channel estimation. With this choice, however, the loss in achievable rate due to training, $(1 - T_p/T_c)$, dominates the increase in SINR when P becomes sufficiently large. This intuitive statement is formalized for the case of the high SNR region described in Propositions 3 & 4 below. An interference-free scenario is obtained in the SINR expression (34) with the non-overlapping AoA supports due to the independent correlation matrices between users. In contrast, the interference is maximized in the SINR expression (33) with the fully overlapping AoA supports due to the condition of common correlation for all users $\mathbf{R}_k = \mathbf{R}$ and $\Phi_k = \Phi \forall k = 1, \dots, K$. Both SINR terms (33) and (34) are equal when the number of users $K = 1$ or $\rho_d = P/N$. The necessary condition for the SINR with the non-overlapping AoA supports (34) to be higher than the SINR with the overlapping AoA supports (33) is that $\rho_d > P/N$ and $K > 1$. In practice, when $\rho_d > 0$ dB, $\overline{\text{SINR}}_{\perp}^{\text{BF}} > \overline{\text{SINR}}^{\text{BF}}$.

Substituting first (33) into (13) and then (34) provides a fast numerical optimization method for the BF precoder. Importantly, at high SNR, i.e. when $\rho_d = \rho_p = \rho \rightarrow \infty$, analytical solutions emerge. In this special case the average achievable sum rate with overlapping AoA supports $\bar{C}^{\text{BF},\infty}$ and non-overlapping $\bar{C}_{\perp}^{\text{BF},\infty}$, respectively, further simplify to

$$\bar{C}^{\text{BF},\infty} = \left(1 - \frac{T_p}{T_c}\right) \cdot K \cdot \log_2 \left(1 + \frac{T_p}{K}\right), \quad T_p \leq P, \quad (35)$$

$$\bar{C}_{\perp}^{\text{BF},\infty} = \left(1 - \frac{T_p}{T_c}\right) \cdot K \cdot \log_2 (1 + T_p), \quad T_p \leq P, \quad (36)$$

which leads to Propositions 3 and 4.

Proposition 3: For K users, P -DoF channel model with overlapping AoA supports and channel coherence time T_c , the downlink training sequence length T_p^* that maximizes the average achievable sum rate, in a massive MIMO system using BF precoding, at high SNR and with uniform power allocation $\rho_d = \rho_p$, is

$$T_p^* = \begin{cases} P & \text{if } P < \tau, \\ \arg \max_{T_p \in \lceil \tau \rceil \lceil \tau \rceil} \bar{C}^{\text{BF},\infty} & \text{otherwise,} \end{cases} \quad (37)$$

where

$$\tau = \frac{T_c + K}{W\left(\frac{T_c + K}{K}e\right)} - K. \quad (38)$$

The Lambert W -function $W(\cdot)$ is defined in [54] and e is Euler's number. In (37), $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling and floor functions, respectively, which accommodates the necessary integer value of T_p given that $\tau \in \mathbb{R}$.

Proof: A proof of Proposition 3 is presented in Appendix A. ■

Proposition 4: For K users, P -DoF channel mode with non-overlapping AoA supports and channel coherence time T_c , the novel downlink training sequence length T_p^* that maximizes the average achievable sum rate, in a massive MIMO system using BF precoding, at high SNR and with uniform power allocation is

$$T_p^* = \begin{cases} P & \text{if } P < \tau, \\ \arg \max_{T_p \in \{\lceil \tau \rceil, \lfloor \tau \rfloor\}} \bar{C}_{\perp}^{\text{BF}, \infty} & \text{otherwise,} \end{cases} \quad (39)$$

where

$$\tau = \frac{T_c + 1}{W((T_c + 1)e)} - 1. \quad (40)$$

The proof of Proposition 4 is similar to the proof of Proposition 3 and, thus, we have omitted for brevity. Clearly, Proposition 4 is the same as Proposition 3 with $K = 1$. From Propositions 3 & 4, the optimal DL training sequence lengths for the BF precoder in the P -DoF channel at high SNR are characterized as follows. T_p^* equals the degrees of freedom of the channel when P is less than τ and saturates at, or below, $\lceil \tau \rceil$, when the DoF exceeds (38), (40) with the overlapping and non-overlapping AoA supports, respectively. The numerical results in the next section confirm that the same behavior is observed also at moderate SNRs, as desired.

The achievable sum rate with DL channel estimation and BF precoding can also be upper bounded using Propositions 3 & 4. Specifically, given K and T_c , the achievable sum rates for BF precoding with overlapping and non-overlapping AoA supports and $\rho_d = \rho_p$ are upper bounded respectively by

$$\bar{C}^{\text{BF}, \infty} \leq \left(1 - \frac{\tau}{T_c}\right) \cdot K \cdot \log_2 \left(1 + \frac{\tau}{K}\right), \quad (41)$$

$$\bar{C}_{\perp}^{\text{BF}, \infty} \leq \left(1 - \frac{\tau}{T_c}\right) \cdot K \cdot \log_2 (1 + \tau). \quad (42)$$

Since τ does not depend on P or N , the rate saturates at a constant level below (41), (42) when the DoF exceeds (38), (40), i.e. $P > \tau$. This also means that the asymptotic sum rate of the P -DoF channel with BF precoding, as a function of N , is independent of the ratio $P/N \in (0, 1]$, although the more rank deficient the correlation matrix is the more BS antennas are needed to approach (41), (42). It should be pointed out that this behavior is in contrast to BF precoding with UL training where the SINR, and hence the sum rate, grows without bound as a function of P when K and T_c are fixed.

Importantly, the numerical results in Section VI show that the optimum pilot sequence length obtained from Proposition 4 for the BF precoder based on the P -DoF channel model is correspondingly similar to those based on a practical OR channel model. The observation means near optimal solutions can be readily found without resorting to computationally intensive optimization techniques.

C. CLOSED-FORM SUM RATE ANALYSIS FOR THE RZF PRECODER IN FDD SYSTEMS

This subsection provides an analytical closed-form expression for the achievable sum rate when the BS employs RZF precoding and the P -DoF channel model, as defined in (28), is used. Though the derivation is protracted, it is algebraically straightforward, therefore, the details have been omitted for brevity. The following Propositions summarize the key results.

Proposition 5: An asymptotically tight approximation for the SINR for RZF precoding based on an analytical P -DoF channel model with overlapping AoA supports is given in terms of auxiliary variables δ , ξ and $\tilde{\mu} \in \mathbb{R}$ as

$$\overline{\text{SINR}}^{\text{RZF}} = \frac{N \xi \tilde{\delta}^2}{\frac{(1+\tilde{\delta})^2}{\rho_d} + K \xi \tilde{\mu}}, \quad T_p \leq P, \quad (43)$$

where $\tilde{\delta}$ is a closed-form solution to the fixed-point equation in (16)–(17) when $\mathbf{R}_k = \mathbf{R}$ and $\Phi_k = \Phi \forall k = 1, \dots, K$ and reads

$$\tilde{\delta} = \frac{T_p - K - \tilde{Z} + \sqrt{(K - T_p)^2 + 2(K + T_p)\tilde{Z} + \tilde{Z}^2}}{2\tilde{Z}}, \quad (44)$$

where $\tilde{Z} = P\zeta(1 + 1/\rho_p)$ with training power scaled by a factor of K (i.e. $\rho_p K$). Parameters ξ and $\tilde{\mu}$ are simplified versions of (23) and (24) with overlapping AoA supports, which are obtained from the expressions given in (45) and (46) as shown at the bottom of the next page, with $\tilde{Q} = K + \tilde{Z} + \delta\tilde{Z}$.

$$\tilde{\xi} = \frac{\tilde{Q}\zeta \left((1 + \tilde{\delta})^2 \tilde{Z}^2 + K(2\tilde{Q} - K - T_p) \right)}{(1 + \tilde{\delta}) T_p Z (\tilde{Q} - T_p)} \quad (45)$$

Proposition 6: An asymptotically tight approximation for the SINR for RZF precoding based on the P -DoF channel model with non-overlapping AoA supports is given in terms of auxiliary variables δ , ξ and $\tilde{\mu} \in \mathbb{R}$ as

$$\overline{\text{SINR}}_{\perp}^{\text{RZF}} = \frac{N \xi \tilde{\delta}^2}{\frac{(1+\tilde{\delta})^2}{\rho_d} + \xi \tilde{\mu}}, \quad T_p \leq P. \quad (47)$$

Parameter $\tilde{\delta}$ is a closed-form solution to the fixed-point equations (16)–(17) when Φ_k is of the form (32) and reads

$$\tilde{\delta} = \frac{T_p - 1 - Z + \sqrt{Z^2 + (2 + 2T_p)Z + 1 - 2T_p + T_p^2}}{2Z}, \quad (48)$$

where $Z = P\zeta \left(1 + \frac{PK}{N\rho_p}\right)$, while the variables $\check{\xi}$ and $\check{\mu}$ are simplified forms of (23) and (24), respectively, and given as

$$\check{\xi} = \frac{Q\zeta(Q^2 - T_p)}{(1 + \check{\delta})T_p Z(Q - T_p)}, \quad (49)$$

$$\check{\mu} = \frac{(1 + \check{\delta})^2 N T_p \left(T_p^2 + \frac{Q^2 Z}{P\zeta} - 2QT_p\right)}{Q^2(Q^2 - T_p)}, \quad (50)$$

where we have denoted $Q = 1 + Z + \check{\delta}Z$ for notational convenience.

From Propositions 5 & 6, an approximation for the achievable rate for RZF precoding can be readily calculated using (13) when the relevant system parameters in terms of the P , N , K , ζ , and ρ_d are known. Due to the complexity of the SINR expression, obtaining an analytical solution for the optimum training sequence length, as in the case of BF precoding, is still very challenging. Nonetheless, the expressions of the SINR for the RZF precoder in (43) and (47) are now given in simplified closed-form, which makes evaluation of the achievable sum rate numerically straightforward. As such, the optimization problem in (13) is now feasible even for a brute-force search. The numerical experiments also show that the BF and RZF precoders have very similar optimal training sequence lengths and, thus, T_p^* can be reliably chosen also for the RZF precoder using Proposition 4.

In the following section, numerical results from analysis and simulation are presented for the BF and RZF precoders when the P -DoF and OR channel models are used. The salient system parameters explored are P , N , T_c , SNR and K . For the P -DoF channel model, strong and weak channel correlation are indicated by the ratios $P/N = 0.1$ and $P/N = 1$, respectively. The parameterisation of the OR channel model follows [25], [55] with the angular spread $\omega = \{5^\circ, 20^\circ\}$, normalized antenna spacing $D = \{1/2, 1\}$ and AoAs $\theta_k = \{-57.5^\circ, -45^\circ, -41.5^\circ, -23.5^\circ, -7.5^\circ, 7.5^\circ, 23.5^\circ, 41.5^\circ, 45^\circ, 57.5^\circ\}$. Values of $\omega = 5^\circ$, $D = 1/2$ indicate strong channel correlation whereas $\omega = 20^\circ$, $D = 1$ indicate weak correlation. The system performance is characterized with $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users, assuming fully digital transceivers operating in sub-6GHz frequency bands. For Monte Carlo simulations, 10^4 incidences of instantaneous channel realization are used, where the channels fade independently from symbol block to block. A summary of the simulation parameters is given in Table 1.

VI. NUMERICAL RESULTS AND DISCUSSION

This section presents several simulation and theoretical results, which characterize the system performance of the BF and RZF precoders, in correlated and uncorrelated channels.

TABLE 1. Simulation parameters.

Parameters	Symbol	Value
Number of BS antennas	N	10-500
Number of UTs	K	10
Correlation in the P -DoF model	P/N	{1, 0.1}
Angular spread in the OR model	ω	{ 5° , 20° }
Antenna element separation	D	{1/2, 1} λ
Coherence time	T_c	100 symbols
SNR	ρ	10dB
BS configuration	BS	ULA

The impact of increasing the number of BS antennas while keeping the coherence time fixed on the achievable sum rate of a massive MIMO system is investigated. Comparison between the sum rate of the UL channel estimation as used in a conventional TDD massive MIMO system and the DL channel estimation in an FDD system is also provided. Results that characterize the achievable sum rate performance of the proposed superposition sequence design are presented and compared with the state-of-the-art sequence designs. Furthermore, the computational complexity of the proposed superposition training design is analysed and compared with the state-of-the-art sequence designs.

A. CHARACTERIZATION OF SUM RATE AND OPTIMUM TRAINING SEQUENCE LENGTH FOR FDD SYSTEMS

Fig. 2 and Fig. 3 show plots of the achievable sum rate and the optimum training sequence length T_p^* , respectively, versus the number of BS antennas N , comparing precoder performance in correlated and uncorrelated channels when the P -DoF channel model is used with $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users. Curves for the BF and RZF precoders are plotted for three computational methods as follows: numerical (BF & RZF) based on equations (13), (34), (33), (47), (43); analytical (BF only) based on equations (36), (35) using the pilot sequence lengths that are chosen according to Propositions 4 & 3, respectively; and simulated (BF & RZF) based on equation (11). These computational methods provide validation between the theoretical and simulated performances, which show excellent agreement throughout.

In Fig. 2, the achievable sum rate of the BF precoder, for the uncorrelated channel (i.e. $P/N = 1$), increases steeply with N before saturating at about 14 bit/s/Hz for values of $N > 30$. In contrast, the sum rate for the BF precoder in the correlated channel with $P/N = 0.1$ increases more gradually before saturating at ~ 35 bit/s/Hz, for values of $N > 200$. The saturation of the sum rate for the BF precoder, regardless

$$\check{\mu} = \frac{(1 + \check{\delta})^2 N T_p \left((1 + \check{\delta})^2 \check{Z}^3 - 2KP\zeta T_p - K(K - 2\check{Q})\check{Z} + P\zeta T_p (T_p - 2(1 + \check{\delta})\check{Z}) \right)}{\check{Q}^2 P\zeta \left((1 + \check{\delta})^2 \check{Z}^2 + K(2\check{Q} - K - T_p) \right)} \quad (46)$$

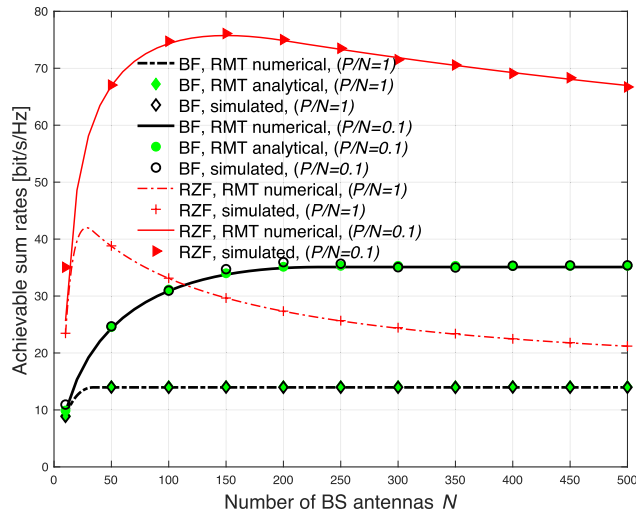


FIGURE 2. Achievable sum rate versus number of BS antennas N for DL channel estimation using the P -DoF channel model with $P/N = \{0.1, 1\}$, $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users.

of the level of correlation in the channel, follows the behavior predicted by (42) & (41) for correlated and uncorrelated channels, respectively. For the RZF precoder in the uncorrelated channel, the sum rate again increases rapidly reaching a peak value of ~ 41 bit/s/Hz at $N = 29$. For values of $N > 29$, the sum rate slowly decreases monotonically, plateauing at ~ 21 bit/s/Hz for $N = 500$. However, the sum rate for the RZF precoder in the correlated channel with $P/N = 0.1$ increases more rapidly before reaching a larger peak value of ~ 75 bit/s/Hz at $N = 150$. For values of $N > 150$, the sum rate slightly decreases monotonically reaching ~ 67 bit/s/Hz at $N = 500$. As the number of BS antennas N increases, the sum rate decreases due to the residual interference caused by imperfect channel estimation. In particular, as N gets large, the residual interference increases because the capability of the RZF precoder to cancel the interference decreases. The results show that the BF and RZF precoders achieve a significant improvement in the maximum value of achievable sum rate when the channel is strongly correlated.

In Fig. 3, for the uncorrelated channel, when $P/N = 1$, the optimum training sequence length T_p^* initially increases linearly with N for both the BF and RZF precoders. This region of the plots corresponds to when $P < \tau$ in (39). Conversely, when $P > \tau$ defines a region where T_p^* saturates. For the BF precoder, T_p^* saturates at 34 symbols when $N = 34$ whereas for the RZF precoder T_p^* continues to increase gradually before saturating at 40 symbols when $N = 400$. For the correlated channel, where $P/N = 0.1$, a similar linear characteristic is observed for N up to 230 for the BF precoder and 200 for the RZF precoder. After these regions, T_p^* saturates at 23 symbols for the BF precoder while T_p^* continues to increase slightly for the RZF precoder. The results in Fig. 2 & Fig. 3 confirm that maximizing the sum rate leads to a feasible optimum training sequence length when DL channel estimation is used in an FDD massive MIMO system.

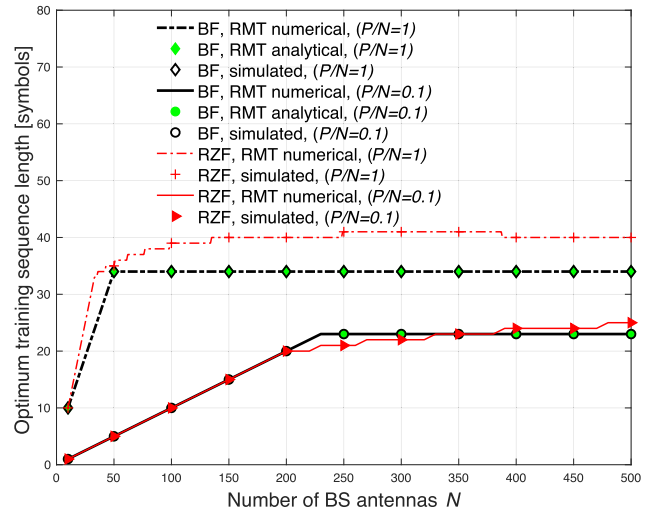


FIGURE 3. Optimum training sequence length T_p^* versus number of BS antennas N for DL channel estimation using the P -DoF channel model with $P/N = \{0.1, 1\}$, $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users.

Importantly, the results show that a feasible sum rate can be realized even with the uncorrelated channels, i.e., $P/N = 1$. This is justified by the fact that the achievable sum rate is obtained by optimizing the training sequence length through the maximization of $(1 - T_p/T_c) \sum_{k=1}^K \log_2(1 + \text{SINR}_k)$, instead of only minimizing the mean square error of the channel estimate, as typically considered in the conventional analyses of FDD and TDD systems. Furthermore, the results confirm that excellent agreement between the numerical, analytical and simulated results was obtained, which underpins the contributions of this research.

B. COMPARING SYSTEM PERFORMANCE OF THE PROPOSED TRAINING SEQUENCE DESIGN AND THE STATE-OF-THE-ART DESIGNS

Having demonstrated the feasibility of a DL training sequence based on superposition, we compare the achievable sum rate performance of the superposition sequence design with the SMSE/SCMI designs. We compare the system performances based on both the P -DoF and OR channel models.

1) SYSTEM PERFORMANCE FOR THE P -DoF CHANNEL MODEL

Results in Fig. 4 were obtained for the correlated P -DoF channel model with $P/N = 0.1$ whereas the other salient system parameters remain unchanged at $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users. Note that the curves for the superposition sequence design correspond to those already presented in Fig. 2. Fig. 4 demonstrates that for both types of precoders, all the training sequence designs exhibit essentially the same sum rate performances. For example, with a BF precoder, the rates saturate at 35 bit/s/Hz, whereas in the RZF precoder the rates peak at ~ 75 bit/s/Hz for $N = 150$. The dotted lines for SMSE, and the dash-dotted lines for SCMI are indistinguishable from the solid line of the

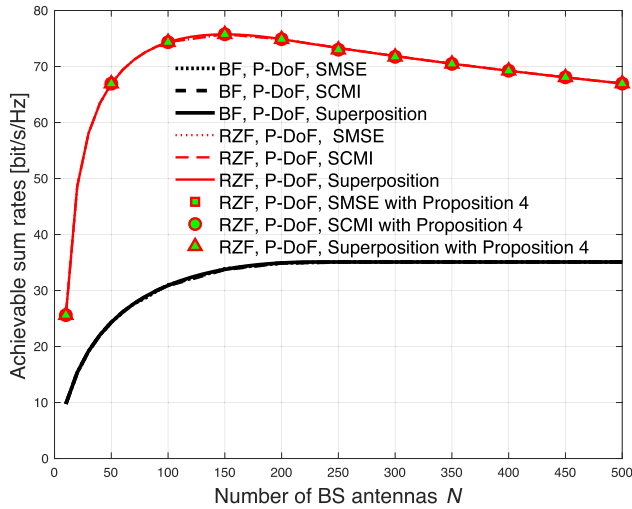


FIGURE 4. Achievable sum rate versus number of BS antennas N comparing different training sequence designs using the P -DoF channel model with $P/N = 0.1$, $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users.

proposed superposition sequence design for both BF and RZF precoders. Importantly, Fig. 4 also shows that the analytical closed-form solution for the optimum pilot sequence length mathematically derived in Proposition 4 for the DL BF precoder with a correlated channel can be reliably selected for the DL RZF precoder (see green markers). Though not plotted here, this observation remains valid for the case of uncorrelated channels, i.e., $P/N = 1$, where the optimum pilot sequence length of the BF precoder in Proposition 3 can be effectively used to predict the rate performance of the RZF precoder for all the training sequence designs.

2) SYSTEM PERFORMANCE FOR THE OR CHANNEL MODEL

So far, the results presented are based on the analytical P -DoF channel model. Being analytical, this model facilitates the simulation, numerical calculation and theoretical analysis of the system performance, giving excellent agreement between the three types of results. It also allows the channel correlation factor to be straightforwardly set in computations. In order to validate the applicability of the P -DoF channel model, this section considers the application of an alternative channel model called the OR scattering channel model [30], which is a more practical channel model frequently encountered in the open literature on MIMO evaluation. The OR scattering channel model represents an environment where all scatterers are located on a ring around the UT and there is no local scattering around the BS.

The system geometric parameters of the user's channel covariance matrices \mathbf{R}_k in the OR model are determined by the angular spread ω , angles of arrival θ_k , and normalized antenna spacing D in wavelengths. Specifically, the (m, n) th element of \mathbf{R}_k is given in Toeplitz form [25], [30] as

$$[\mathbf{R}_k]_{m,n} = \frac{1}{2\omega} \int_{-\omega+\theta_k}^{\omega+\theta_k} e^{-j2\pi D(m-n)\sin(\Delta)} d\Delta. \quad (51)$$

where Δ represents the intervals/ranges of the AoAs distribution (i.e. $\Delta \in [-60^\circ, 60^\circ]$) since a 120° sector is considered in this paper. The integration in (51) is computed numerically whereas the DL instantaneous channel realization is given by $\mathbf{h}_k = \mathbf{R}_k^{1/2} \tilde{\mathbf{h}}_k$, where the elements of $\tilde{\mathbf{h}}_k \in \mathbb{C}^N$ are independent and identically distributed with zero mean and unit variance [25]. Due to the different scattering geometries associated with each user's position in a geographic area, the angular support of each user's channel appears random. The randomness in the user locations captures the fact that the angular supports of the users may partially overlap. In addition, when the BS antennas are closely spaced and the amount of scattering around the UT is limited, as indicated by both D and ω being small, some of the eigenvalues of \mathbf{R}_k are close to zero, making \mathbf{R}_k effectively rank deficient. In contrast to the P -DoF channel model, in the OR channel model the non-zero eigenvalues are not usually equal.

An approximation of the actual rank $r_k \in \mathbb{Z}^+$ for large but finite N that contain the effective non-zero eigenvalues of the channel covariance matrix in (51) is given by [29]

$$r_k = N\beta_k, \quad (52)$$

where $\beta_k = \min\{1, f(D, \omega, \theta_k)\}$ is the asymptotic normalized rank of the channel covariance matrix \mathbf{R}_k with $f(D, \omega, \theta_k) = \lfloor D \sin(\theta_k - \omega) - D \sin(\theta_k + \omega) \rfloor$.

While (52) can accurately predict the rank of the OR channel model that is related to the number of non-zero eigenvalues of \mathbf{R}_k , this number may differ between users due to different θ_k . Hence, the maximum number of the effective non-zero eigenvalues across all of the users is selected, i.e. $r = \max_{k=1, \dots, K} \{r_k\}$, to ensure that all the relevant eigenvectors of each \mathbf{R}_k , corresponding to the largest eigenvalues over all the users, are accounted for. When the rank of the OR channel model is obtained, the training sequence length can be reliably selected based on the results derived in Section V-B for the analytical P -DoF channel model.

Fig. 5 and Fig. 6 plot the achievable sum rate versus the number of BS antennas N , comparing the proposed superposition training sequence design (10) with the SMSE/SCMI designs based on the OR scattering channel model for the BF and RZF precoders, respectively. Fig. 5 and Fig. 6 are obtained with $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users. The solid lines depict numerical analysis based on random matrix theory, while the colored markers denote simulation. Fig. 5 shows almost the same rate performances are obtained in all the training sequence designs for the BF precoder. In Fig. 6, marginal loss in the rate performance is obtained with the proposed superposition training design in comparison with the state-of-the-art SMSE/SCMI designs.

As expected, results in Fig. 6 shows that the RZF precoder achieves greater sum rate than the BF precoder in Fig. 5, both in the correlated and the uncorrelated channels. Training sequences based on the superposition design achieve a slightly lower sum rate than either the SMSE or SCMI designs when RZF precoding is used, which is attributed to

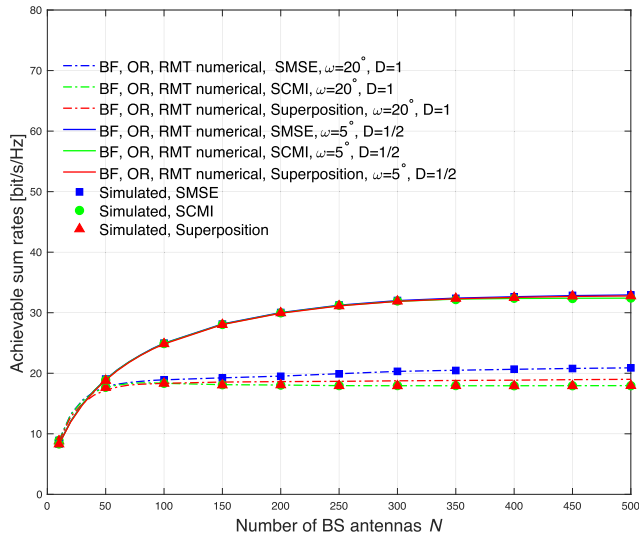


FIGURE 5. Achievable sum rate versus number of BS antennas N , $T_c = 100$ symbols, SNR = 10 dB, and $K = 10$ users, comparing different training sequence designs based on BF precoder using the OR channel model with $\omega = \{5^\circ, 20^\circ\}$, $D = \{1/2, 1\}$.

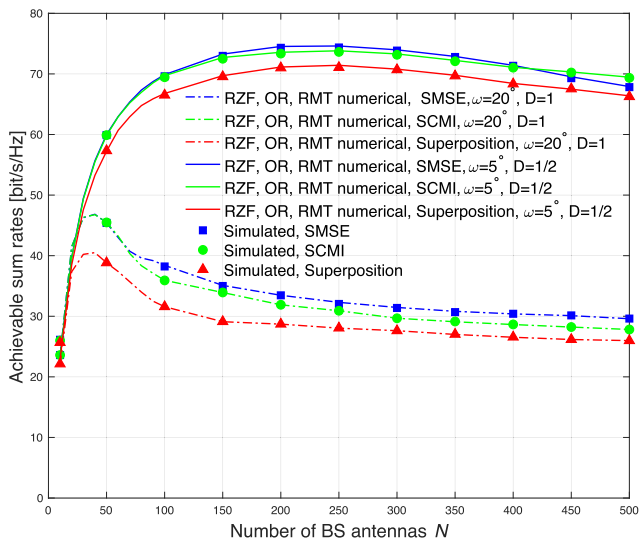


FIGURE 6. Achievable sum rate versus number of BS antennas N , $T_c = 100$ symbols, SNR = 10 dB, and $K = 10$ users, comparing different training sequence designs based on RZF precoder using the OR channel model with $\omega = \{5^\circ, 20^\circ\}$, $D = \{1/2, 1\}$.

the nonoptimal cancellation of interference in the superposition design. Also, the results in Fig. 5 and Fig. 6 demonstrate that large improvements in the sum rate performances are obtained for both the BF and RZF precoders when the channels are correlated i.e. $\omega = 5^\circ$, $D = 1/2$. The results also confirm that excellent agreement between numerical and simulated modelling is obtained with the more realistic OR model.

Fig. 7 plots achievable sum rate versus N for the BF and RZF precoders showing the impact of using the optimum pilot sequence length obtained by Proposition 4 for the BF precoder with the P -DoF channel model over the OR scattering

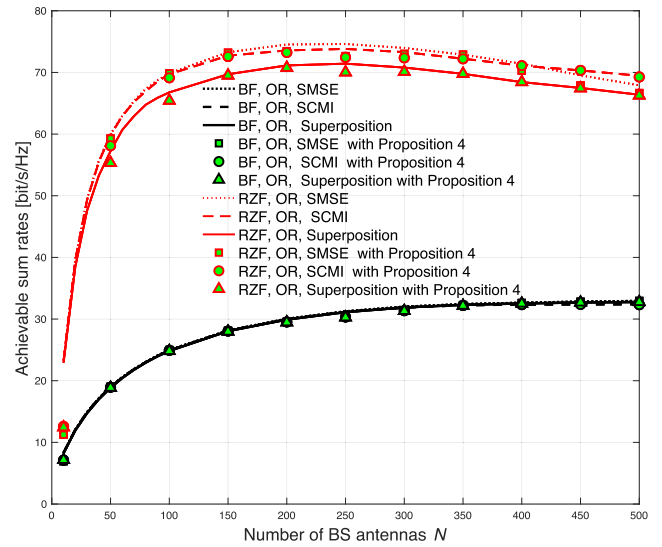


FIGURE 7. Achievable sum rate versus number of BS antennas N , $T_c = 100$ symbols, SNR = 10 dB, and $K = 10$ users, comparing different training sequence designs using the OR model with $\omega = 5^\circ$, $D = 1/2$.

channel model. The curves for the BF and RZF precoders in Fig. 7 are obtained numerically based on equations (13), (14), (15), which were developed using the method of RMT. The parameter values for the OR channel model, are $\omega = 5^\circ$ and $D = 1/2$. The other salient system parameters remain unchanged at $T_c = 100$ symbols, SNR = 10 dB and $K = 10$. The solid lines depict the superposition training sequence design (10) and dotted lines and dash-dotted lines represent the SMSE/SCMI designs, respectively, while the colored markers denote the optimum pilot sequence length obtained by Proposition 4 for the BF precoder using the P -DoF channel model. The result in Fig. 7 confirms that the optimum training sequence length for the BF precoder with the P -DoF channel model provides close agreement also in the more realistic OR channel model for both the BF and RZF precoders. In particular, it is possible to apply Proposition 4 to the three training sequence designs in order to obviate the need to search for an optimum training sequence length for all training methods in the more realistic OR channel model. Though not plotted here, similar behaviour is observed when the channels are relatively uncorrelated, i.e., $\omega = 20^\circ$ and $D = 1$, where the optimum pilot length obtained from Proposition 3 for the BF precoder based on the P -DoF channel model can also be used with high accuracy to predict the sum rate performance in the more realistic OR channel model for both the BF and RZF precoders. Overall, Fig. 7 demonstrates the effectiveness of using the P -DoF channel model to provide a practical system design approach, which accurately predicts the performance in more realistic channel models.

C. COMPARING SYSTEM PERFORMANCE FOR DOWNLINK FDD AND UPLINK TDD CHANNEL ESTIMATION

Having demonstrated the performance of DL training sequence and channel estimation of an FDD multiuser

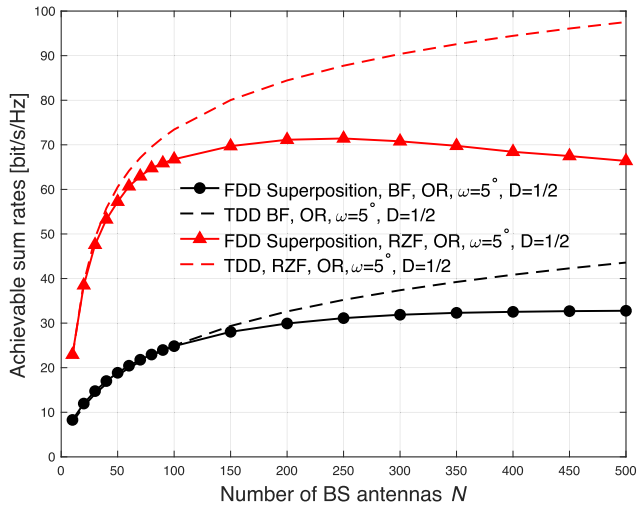


FIGURE 8. Achievable sum rate versus number of BS antennas N based on the OR channel model with $\omega = 5^\circ$, $D = 1/2$, $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users, comparing DL (i.e. FDD) and UL (i.e. TDD) channel estimation.

massive MIMO system, it is pertinent to compare the achievable sum rate performance with that obtained for UL channel estimation as conventionally used in a TDD system whereby the uplink-downlink channel with perfect reciprocity is considered. In a TDD system, a superposition of orthogonal UL training sequences is transmitted by the users to the BS, and the BS estimates the UL channel by using an MMSE channel estimator. For reciprocal channels, the number of pilot symbols required for the UL channel estimation is proportional to the number of users K , which reflects the DoF on the UL.

Fig. 8 plots the achievable sum rate versus the number of BS antennas N , comparing the performance of the proposed superposition training design in (10) for the DL channel estimation of an FDD system with the UL channel estimation of a TDD system in the OR channel model. In particular, Fig. 8 is obtained with $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users, where the parameter values for the OR channel model are chosen with angular spread $\omega = 5^\circ$, and normalized antenna spacing $D = 1/2$. These parameter values imply relatively strong correlation. The results in Fig. 8 show that for the system parameters considered, over a practical number of BS array sizes of $N < 250$ antennas, the DL and UL sum rate performances are comparable in the more realistic OR channel model with correlated channels under both the BF and RZF precoders considered. Specifically, Fig. 8 demonstrates that DL channel estimation with the proposed superposition training design is effective in strongly correlated channels.

D. COMPUTATIONAL COMPLEXITY ANALYSIS AND PERFORMANCE EVALUATION OF DIFFERENT TRAINING DESIGNS IN FDD

In this subsection, we present the computational complexity analysis of the superposition training sequence design and

the SMSE/SCMI sequence designs. We compare the overall computational complexity, which is obtained by multiplying the number of iterations each algorithm needs to converge by the number of complex floating point operations (flops) involved per iteration. The superposition training design requires only *one iteration*. For a fair comparison, we use the same error tolerance $\epsilon = 0.001$ for the SMSE and SCMI sequence designs.

Table 2 summarizes the complexity analysis in flops per iteration [56] for the three training sequence designs. Parameters t_d and t_h represent the number of iteration required to optimize the step size in the SMSE algorithm. The variable X is given as $X = T_p^3 + T_p^2 + N^2(6T_p + 2) + 3NT_p(2T_p - 1) - 2$. Also, r denotes the maximum number of effective non-zero eigenvalues (i.e. the rank) of the channel covariance matrices across the users. Below are the details to explain how the computational complexity analysis of the proposed superposition sequence design was developed in Table 2. Particularly, the analysis is obtained by counting the number of multiplications and additions [56] of each step in the superposition pilot scheme. Appendix B explains further how the computational complexity of state-of-the-art iterative algorithms were obtained in Table 2.

- Obtaining the EVD of an $N \times N$ rank deficient matrix in (9) for K users needs KN^2r flops.
- In order to obtain the term $\sum_{k=1}^K \mathbf{U}_k$ with an $N \times T_p$ matrix in (10), $(K - 1)(NT_p)$ flops are required.
- Obtaining the squared Frobenius norm $\|\cdot\|_F$ of an $N \times T_p$ matrix in (10) requires $2NT_p - 1$ flops.
- The scalar matrix multiplication with an $N \times T_p$ matrix needs NT_p flops.
- Combining all the flops calculated above, leads to the complexity analysis of the superposition pilot scheme in Table 2.

Fig. 9 shows plots of the computational complexity versus the number of BS antennas N , comparing the superimposed training sequence design with the SMSE and SCMI training sequence designs. The results in Fig. 9 were obtained for the OR model with $\omega = 5^\circ$, $D = 1/2$ using the pilot sequence length for the BF precoder.¹ The other salient system parameters remain unchanged at $T_c = 100$ symbols, SNR = 10 dB and $K = 10$ users. The results in Fig. 9 indicates the complexity of the superposition sequence design remains significantly lower than the sequence designs based on iterative algorithms. In particular, the results demonstrate that more than a four orders-of-magnitude reduction in computational complexity is obtained using the proposed superposition approach. Hence, signifying the feasibility of the superposition training sequence design for practical implementations compared with state-of-the-art iterative algorithms. This result is a significant outcome from the research. Table 3 shows the relative increase in complexity

¹For the sake of brevity, we present the BF precoder result only. We observe similar performance trends in all the training sequence designs for the RZF precoder.

TABLE 2. Computational complexity analysis of different training designs.

Algorithms	Complexity in flops per iteration
SMSE [28]	$4N^2T_p - 1 - 1/3T_p^3 + 2NT_p(1+T_p) + (K-1)T_p(16N^2 + T_p(5T_p - 2) + 2N(7T_p - 5)) + (K-1)(t_d+1)X + (K-1)(t_h+2)X$
SCMI [25]	$T_p^3K + T_p^2(N(4K-2) - 1) + 4N^2T_p(K-1) + NT_p(7-3K) - 1$
Superposition	$N^2Kr + NT_pK + 2NT_p - 1$

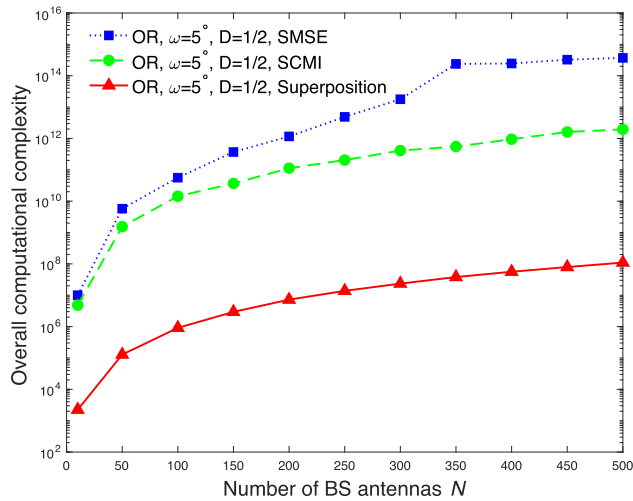


FIGURE 9. Overall computational complexity versus the number of BS antennas N comparing different training sequence methods corresponding to the optimal pilot length in BF precoding in the OR model with $\omega = 5^\circ$, $D = 1/2$.

TABLE 3. The relative complexities of SCMI and SMSE over superposition.

N	Superposition	SCMI [25]	SMSE [28]
N=50	1.00	12047.15	44930.10
N=100	1.00	15797.30	61560.96
N=150	1.00	12441.68	125330.89

of SCMI and SMSE over superposition for representative antenna arrays of $N = 50, 100$ and 150 elements. The table highlights the considerable reduction in complexity achieved by superposition.

VII. CONCLUSIONS

In this paper, the principle of linear superposition of sequences constructed from the users' channel correlation matrices is proposed to provide a feasible solution for DL channel estimation in an FDD multiuser massive MIMO communications systems without resorting to computationally intensive iterative algorithms. Based on the superimposed training structure, we have provided a novel analytical closed-form solution for the optimum training sequence length T_p^* that maximizes the DL achievable sum

rate $(1 - T_p/T_c) \sum_{k=1}^K \log_2(1 + \text{SINR}_k)$, defined by (11) in the BF precoder. Additionally, an analytical approximation for the achievable sum rate of the BF and RZF precoders has been provided using asymptotic random matrix theory and the P -DoF channel model, which is used to avoid executing extensive Monte Carlo simulations and allowed an analytical solution for the optimization problem considered to be obtained with low-complexity. The numerical results showed that these approximations are accurate for practical, finite systems parameters in terms of N and K . Results characterizing the system performance for the BF and RZF precoders are presented for an analytical P -DoF and the OR channel models. The numerical results showed that the proposed training sequences offer comparable sum rate performances to the state-of-the-art sequences while reducing the computational complexity substantially. Furthermore, comparison between the correlated channels with K independent channel covariance matrices and uncorrelated channels with identical channel covariance matrices $\mathbf{R}_k = \mathbf{I}_N \forall k = 1, \dots, K$ was also provided. The analyses of the results have shown that the diversity of spatial correlations between multiple users significantly enhances the achievable sum rate of an FDD massive MIMO system using DL channel estimation. This paper also provided comparisons between the achievable sum rates of the UL CSI estimation as conventionally used in a TDD system and the DL CSI estimation in an FDD system. The results showed that for practical BS array sizes of $N < 250$ antennas and limited coherence time, the sum rate of an FDD system using DL channel estimation is comparable to the performance of a TDD system in relatively strongly correlated channels. Our findings are supported by a rigorous mathematical analysis, which tightly agrees with our simulated results, which underpin the key contribution of this research. Importantly, using the framework analysis developed in this paper, we found that the optimum T_p^* that is analytically optimized for the BF precoder is sufficient to predict the achievable sum rate performance of the RZF precoder, which remains near optimal. This observation also remains highly valid in the more realistic OR channel model. This result leads to large reductions in the computational complexity of the proposed approach. Overall, the proposed design paradigm opens up the possibility for FDD massive MIMO systems operating in a general scenario of single-stage precoding and distinct spatial correlations with limited coherence time. Future work will investigate the effects of other system parameters and configurations including removing the assumption of an error and delay free feedback channel, taking into account multi-cell multi-user operation, and considering a uniform planar array of antenna elements. The authors are also interested in applying the paper's findings to millimetre wave bands.

APPENDIX A
PROOF OF PROPOSITION 3

To find the optimum training sequence length for BF precoding in a P -DoF channel at high SNR, as given in Proposition 3,

we first relax the requirement for the training sequence length to be a positive integer and replace T_p with a real valued variable $\tau \geq 0$ as shown in (38). Treating the high SNR sum rate $\bar{C}^{\text{BF},\infty}$ given in (35) as a smooth continuous function of training sequence length τ allows the partial derivative

$$\frac{\partial \bar{C}^{\text{BF},\infty}}{\partial \tau} = -\frac{1}{T_c \ln(2)} \ln\left(1 + \frac{\tau}{K}\right) + \frac{1}{T_c \ln(2)} \left(\frac{T_c - \tau}{\tau + K}\right), \quad (53)$$

to be obtained. Setting $\frac{\partial \bar{C}^{\text{BF},\infty}}{\partial \tau} = 0$ yields after some algebraic manipulations the equality

$$\ln\left(1 + \frac{\tau}{K}\right) = \frac{T_c + K}{\tau + K} - 1 = \frac{T_c + K}{K} \left(1 + \frac{\tau}{K}\right)^{-1} - 1. \quad (54)$$

Moving the terms containing τ on the left hand side allows (54) to be written as

$$\left(\ln\left(1 + \frac{\tau}{K}\right) + 1\right) \exp\left(\ln\left(1 + \frac{\tau}{K}\right) + 1\right) = \frac{T_c + K}{K} e. \quad (55)$$

Clearly the left hand side of (55) is of the general form $f(\tau)e^{f(\tau)}$ and can be solved using the Lambert W function [54]. In particular, the Lambert W-function, $W()$, satisfies $f(\tau) = W(f(\tau)e^{f(\tau)})$ and as a result, the necessary condition for τ at the stationary point becomes

$$\begin{aligned} \tau &= K \exp\left(W\left(\frac{T_c + K}{K} e\right) - 1\right) - K \\ &= \frac{K}{e} \exp\frac{\frac{T_c + K}{K} e}{W\left(\frac{T_c + K}{K} e\right)} - K, \end{aligned} \quad (56)$$

which leads directly to (40). The second equality in (56) follows from the definition of the Lambert W-function, namely, $W(f(\tau))e^{W(f(\tau))} = f(\tau)$.

Finally, since τ is a real valued relaxation of the integer valued training sequence length T_p , the true optimum value is either $\lceil \tau \rceil$ or $\lfloor \tau \rfloor$, i.e., one of the two integers nearest to τ . Combining this with the constraint $T_p \leq P$ leads to (37), completing the proof. The same approach applied to $\bar{C}_{\perp}^{\text{BF},\infty}$ yields (40).

APPENDIX B FLOPS PER ITERATION CALCULATIONS

This Appendix briefly explains the methodology for calculating the flops per iteration for the SMSE and SCMI algorithms, which are iterative.

- The following explains how to calculate the total number of multiplications and additions in flops of a matrix-matrix product of an $N \times N$ matrix with an $N \times N$ matrix. Let $\mathbf{A} = \mathbf{BC}$, where \mathbf{B} and \mathbf{C} are two matrices each with a size of $N \times N$. Each element in the obtained matrix \mathbf{A} involves an N multiplications and $N - 1$ additions. This is valid for any matrix, which depends on the number of rows and columns of the matrix. Hence, the first element of the obtained matrix \mathbf{A} involves an (N) multiplication + $(N - 1)$ addition = $2N - 1$ flops. The whole matrix-matrix multiplication operation then requires $(2N - 1)(N)(N) = 2N^3 - N^2$

flops. Hence, using the same procedure discussed above for the matrix-matrix products, we have the following calculations.

- Multiplying an $N \times N$ matrix with an $N \times T_p$ matrix entails $2N^2T_p - NT_p$ flops.
- Multiplying a $T_p \times N$ matrix with an $N \times T_p$ matrix, entails $2NT_p^2 - T_p^2$ flops.
- Multiplying a $T_p \times N$ matrix with an $N \times N$ matrix entails $2N^2T_p - NT_p$ flops.
- Multiplying a $T_p \times T_p$ matrix with a $T_p \times T_p$ matrix requires $2T_p^3 - T_p^2$ flops.
- Multiplying an $N \times N$ matrix with an $N \times N$ matrix entails $2N^3 - N^2$ flops.
- Multiplying an $N \times T_p$ matrix with a $T_p \times T_p$ matrix entails $2NT_p^2 - NT_p$ flops.
- Multiplying an $N \times T_p$ matrix with a $T_p \times N$ matrix entails $2N^2T_p - N^2$ flops.
- The scalar matrix multiplication with an $N \times T_p$ matrix needs NT_p flops.
- Adding an $N \times T_p$ matrix to an $N \times T_p$ matrix requires NT_p flops.
- Inverting of a $T_p \times T_p$ matrix requires T^3 flops.
- In order to calculate the squared Frobenius norm, also known as the squared Hilbert Schmidt norm [56], which is denoted by $\|\cdot\|_F$, of an $N \times T_p$ matrix requires the following summations and multiplications. It follows from summing up $N \times T_p$ entries and $N \times T_p$ multiplications. Combining these two operations together, yields the total number of flops of $2NT_p - 1$.
- Calculating the term $\sum_{k=1}^K \mathbf{R}_k \mathbf{S}_p (\mathbf{S}_p^H \mathbf{R}_k \mathbf{S}_p + \mathbf{I}_{T_p})^{-1}$ in the SCMI algorithm requires $(K - 1)(T_p^3 + 4NT_p^2 + 4N^2T_p - 3NT_p)$ flops.
- Calculating the spectral norm in the SCMI algorithm of an $N \times T_p$ matrix requires $T_p^3 - T_p^2 + 2NT_p^2$ flops.
- Calculating the term $\sum_{k=1}^K \mathbf{R}_k^2 \mathbf{S}_p (\mathbf{S}_p^H \mathbf{R}_k \mathbf{S}_p + \mathbf{I}_{T_p})^{-1} + \mathbf{R}_k \mathbf{S}_p \times (\mathbf{S}_p^H \mathbf{R}_k \mathbf{S}_p + \mathbf{I}_{T_p})^{-1} \mathbf{S}_p^H \mathbf{R}_k^2 \mathbf{S}_p (\mathbf{S}_p^H \mathbf{R}_k \mathbf{S}_p + \mathbf{I}_{T_p})^{-1}$ in the SMSE algorithm requires $(K - 1)(5T_p^3 + 16N^2T_p + 14NT_p^2 - 2T_p^2 - 10NT_p)$ flops.

With the above complexity calculations in flops per iteration, the results in Table 2 are obtained for the different training sequence designs.

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [2] W. Guo and T. O'Farrell, "Capacity-energy-cost tradeoff in small cell networks," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–5.
- [3] L. Zhao, G. Zhao, and T. O'Farrell, "Efficiency metrics for wireless communications," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 2825–2829.
- [4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [5] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

- [6] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [8] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
- [9] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [10] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy-efficient downlink transmission for multicell massive DAS with pilot contamination," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1209–1221, Feb. 2017.
- [11] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [12] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, Nov. 2016.
- [13] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Multicell multiuser massive MIMO transmission with downlink training and pilot contamination precoding," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6301–6314, Aug. 2016.
- [14] J. Li, C. Yuen, D. Li, X. Wu, and H. Zhang, "On hybrid pilot for channel estimation in massive MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6670–6685, Jul. 2019.
- [15] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
- [16] L. Fan, D. Qiao, S. Jin, C.-K. Wen, and M. Matthaiou, "Optimal pilot length for uplink massive MIMO systems with low-resolution ADC," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
- [17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [18] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [19] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
- [20] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
- [21] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.
- [22] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.
- [23] J. So, D. Kim, Y. Lee, and Y. Sung, "Pilot signal design for massive MIMO systems: A received signal-to-noise-ratio-based approach," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 549–553, May 2015.
- [24] J. Nam, G. Caire, and J. Ha, "On the role of transmit correlation diversity in multiuser MIMO systems," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 336–354, Jan. 2017.
- [25] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.
- [26] Y. Gu and Y. D. Zhang, "Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2334–2346, May 2019.
- [27] B. Tomasi, A. Decurninge, and M. Guillaud, "SNOPS: Short non-orthogonal pilot sequences for downlink channel state estimation in FDD massive MIMO," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, USA, Dec. 2016, pp. 1–6.
- [28] S. Bazzi and W. Xu, "Downlink training sequence design for FDD multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4732–4744, Sep. 2017.
- [29] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [30] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [31] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, Apr. 2018.
- [32] M. Sadeghi, L. Sanguinetti, R. Couillet, and C. Yuen, "Large system analysis of power normalization techniques in massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9005–9017, Oct. 2017.
- [33] S. Stein, "Fading channel issues in system engineering," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 68–89, Feb. 1987.
- [34] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.
- [35] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.
- [36] B. M. Hochwald and T. L. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 642–653, Mar. 2001.
- [37] A. Decurninge, M. Guillaud, and D. T. M. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.
- [38] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 974–978.
- [39] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.
- [40] K. Upadhyaya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, Apr. 2018.
- [41] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [42] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, Oct. 2011.
- [43] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprasad, "Achieving 'massive MIMO' spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.
- [44] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.
- [45] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.
- [46] J. W. Wallace and M. A. Jensen, "Measured characteristics of the MIMO wireless channel," in *Proc. IEEE 54th Veh. Technol. Conf. (VTC Fall)*, vol. 4, Oct. 2001, pp. 2038–2042.
- [47] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.
- [48] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 3464–3467.
- [49] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, San Francisco, CA, USA, Sep. 2011, pp. 1–5.
- [50] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. Eur. Conf. Antennas Propag. (EUCAP)*, Prague, Czech, Mar. 2012, pp. 433–437.

- [51] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.
- [52] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.
- [53] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.
- [54] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the LambertW function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, Dec. 1996.
- [55] 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. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
- [56] R. Hunger, *Floating Point Operations in Matrix-Vector Calculus*. Munich, Germany: Associate Institute Signal Processing, Technische Univ. München, Oct. 2005.



MIKKO VEHKAPERÄ received the Ph.D. degree from the Norwegian University of Science and Technology, Trondheim, Norway, in 2010. He was a Postdoctoral Researcher with the KTH Royal Institute of Technology, Sweden, an Academy of Finland Postdoctoral Researcher with Aalto University, Finland, an Assistant Professor (Lecturer) with The University of Sheffield, U.K., and a Research Fellow with Aalto University. He is currently a Senior Scientist with the VTT Technical Research Centre of Finland. He is also an Adjunct Professor with Aalto University and an Advanced Visiting Fellow with The University of Sheffield. He held visiting appointments with the Massachusetts Institute of Technology, Cambridge, MA, USA; Kyoto University, Kyoto, Japan; the Tokyo Institute of Technology, Tokyo, Japan; and Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. His research interests include wireless communications, information theory, and signal processing. He was a co-recipient of the Best Student Paper Award at the IEEE International Conference on Networks, in 2011, and the IEEE Sweden Joint VT-COM-IT Chapter Best Student Conference Paper Award, in 2015.



TIMOTHY O'FARRELL (Senior Member, IEEE) is currently the Chair Professor in wireless communications with The University of Sheffield, U.K. He is also an Expert in wireless communication systems specializing in physical layer signal processing, radio resource management, and wireless network planning. He has pioneered research on energy efficient mobile cellular communications, the mathematical modeling of CSMA-based MAC protocols for WiFi, iterative block coding for wireless communication systems, and spreading sequence design for CDMA wireless networks. He is an Entrepreneur, being the Cofounder and the CTO of Supergold Communication Ltd., from 1997 to 2007, a start-up that participated in the standardization of IEEE 802.11g with the MBCK proposal. In the framework of Mobile VCE (mVCE), he was the Academic Coordinator of the Core 5 Green Radio Project, from 2009 to 2012, and the Leader in establishing energy efficiency as a global research field in wireless communications systems. He has published 328 journals and conference papers, book chapters, patents, and technical reports; and has participated in standards, consultancy, and expert witness activities within the wireless sector. He has managed 24 major research projects as a Principal Investigator. His current EPSRC project portfolio includes the SERAN, FARAD, and DDmmMaMi projects. He leads the U.K. Research Strategy Community Organisation in Communications, Mobile Computing, and Networking within the EPSRC portfolio (CommNet2, www.commnet.ac.uk). He is also the Director of the mVCE, a Chartered Engineer, and a member of IET.



MUNTADHER ALSABAH received the B.S. degree in electrical engineering from the University of Baghdad, Iraq, in 2007, and the M.S. degree in wireless communication engineering from Universiti Putra Malaysia (UPM), Malaysia, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, U.K. His current research interests include wireless communications, multiple-input-multiple-output (MIMO) systems, and signal processing.

• • •