Pilot Sequence Design for Mitigating Pilot Contamination With Reduced RF Chains

Shahar Stein Ioushua, *Student Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*— Massive multiple-input multiple-output (MIMO) communication is a promising technology for increasing spectral efficiency in wireless networks. Two of the main challenges massive MIMO systems face are degraded channel estimation accuracy due to pilot contamination and increase in computational load and hardware complexity due to the massive number of antennas. In this paper, we focus on the problem of channel estimation in massive MIMO systems, while addressing these two challenges: We *jointly* design the pilot sequences to mitigate the effect of pilot contamination and propose an analog combiner which maps the high number of sensors to a low number of RF chains, thus reducing the computational and hardware cost. We consider a statistical model in which the channel covariance obeys a Kronecker structure, and treat two special cases, corresponding to fully- and partially-separable correlations. We prove that under these models, the analog combiner design can be performed independently of the pilot sequences. Given the resulting combiner, we derive a closed-form expression for the optimal pilot sequences in the fully-separable case and suggest a greedy sum of ratio traces maximization (GSRTM) method for designing sub-optimal pilots in the partially-separable scenario. We demonstrate via simulations that our pilot design framework achieves lower mean squared error than the common pilot allocation framework previously considered for pilot contamination mitigation.**

***Index Terms*— Massive multiple-input multiple-output (MIMO), pilot contamination, multiple access interference, antenna arrays, hybrid beamforming, cellular communication, low-complexity, ratio-trace minimization.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) wireless systems have emerged as a leading candidate for 5G wireless access [1]–[3], offering increased throughput, which is scalable with the number of antennas. However, to fully utilize the possibilities of such large-scale arrays, accurate channel state information is crucial, making the channel estimation task a critical step in the communication process.

In standard MIMO systems, channel estimation can be performed using either time-division duplex (TDD) or frequency-division duplex (FDD) frameworks. In FDD systems, the time required to estimate the downlink channel is proportional to the number of antennas at the base station (BS), which may be extremely large in the massive MIMO regime. To address this challenge, TDD based estimation was suggested in [4], exploiting the TDD channel reciprocity. In this approach, channel estimation is performed by sending known pilot sequences from the user terminals (UTs) to the BS, where different UTs are assigned orthogonal pilots to avoid intra-cell interference (assuming the number of pilot symbols is larger than the number of users in each cell). Since the number of pilot symbols is limited by the coherence duration of the channel and the desire to avoid high training overhead, a limited number of orthogonal pilots can be allocated [5]. Consequently, orthogonal pilots are only assigned to users in the same cell and are typically reused between different cells. Pilot reuse causes inter-cell interference referred to as pilot contamination [5], [6], which is believed to be one of the main performance bottlenecks of massive MIMO systems.

Mitigating pilot contamination has been the focus of considerable research attention, see, e.g., [7]–[18]. The majority of works on this subject can be divided into four main families: blind channel estimation, downlink precoding, pilot allocation, power control and pilot design. Blind channel estimation aims at avoiding the need for pilot information, and instead estimating the channel from the received unknown data. Two such techniques were suggested in [9], [10], based on subspace projection and on eigenvalue decomposition. Both methods exploit the asymptotic orthogonality between the channels of different UTs in massive MIMO systems to cancel interferences from adjacent cells. This approach does not aim at optimizing the pilot sequences.

Downlink precoding methods assume fixed pilot sequences, and optimize the precoding matrix. Specifically, these techniques first carry out standard channel estimation with full pilot reuse, resulting in pilot contamination. Then, using the contaminated estimated channel, they optimize the downlink precoding matrices in all cells to compensate for the pilot contamination effect. In [7], a joint downlink precoding matrix is designed by minimizing the mean-squared error (MSE). The objective consists of two parts: the MSE of each UT in the current cell, and the interference of the UTs in all other cells. These quantities depend on all the estimated channels in the system, thus requiring the BSs to exchange large amounts of information, which results in a large overhead.

An alternative precoding strategy is suggested in [8] which exploits the contamination effect, by letting the BSs serve UTs from different cells. Each BS linearly combines the downlink

Manuscript received July 3, 2019; revised December 16, 2019 and February 16, 2020; accepted February 16, 2020. Date of publication March 6, 2020; date of current version June 16, 2020. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant agreement No. 646804-ERC-COG-BNYQ, and from the Israel Science Foundation under Grant No. 335/14.The associate editor coordinating the review of this article and approving it for publication was E. Björnson. *(Corresponding author: Shahar Stein Ioushua.)*

Shahar Stein Ioushua is with the Department of Electrical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: steinioushua@mail.tau.ac.il).

Yonina C. Eldar is with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.2979142

data designated for all the UTs that share the same pilot sequence, thus leveraging the contaminated channel. In this case, only the downlink data of all users and the channels' long-term statistics need to be shared between different cells. Downlink precoding techniques optimize only the precoding matrices and not the pilot sequence, and can be applied with any given sequences. In particular, they may be applied after a pilot optimization step, which, as we show in this work, contributes significantly to reducing the overall contamination.

In contrast to the previous two approaches, pilot allocation methods optimize the pilot sequences themselves. In this framework, the same predefined set of potential pilot sequences is used in all cells. The optimization is over the allocation of the sequences to the users, such that the cross interference caused by users that share the same sequence is minimized. A common approach for pilot allocation is based on angle-of-arrival (AOA) of the UTs' signal to the BSs. One such example is [11], which aims to exploit the spatial orthogonality between different users and proposes a greedy pilot assignment algorithm, which minimizes the sum of estimation errors in multipath channels by assigning the same pilot sequence to UTs with non-overlapping angle spread. Later results in [18], [19] suggest that this method achieves the same performance as a simple random pilot assignment. Partial pilot reuse is proposed in [13], [14], assigning non-orthogonal pilots only to cell-centered UTs, while UTs located at the edge of the cell are assigned orthogonal ones. The pilot allocation scheme in [15] is based on an iterative algorithm across the different cells, assigning the pilot sequence with the weakest interference to the UT with the most attenuated channel in each iteration. However, this method is only suitable when the number of UTs in each cell equals the number of pilot symbols. In [20], the authors formulated the pilot allocation problem as a potential game and suggested a game-theoretic approach that can achieve the performance of optimal allocation and has the advantage that they do not require a centralized coordinator. In [16] the authors propose a transmission scheme based on time-shifted pilots, in which the cells are partitioned into groups and the scheduling of pilots is shifted in time from one group to the next, canceling interference between adjacent cells.

Another pilot optimization approach is pilot power control, in which all the pilot sequences are pre-determined, and only their power allocation is optimized. It was shown in [16] that the power allocation algorithm proposed in [21] for CDMA can also be used in multi-cell multi-user TDD massive MIMO systems. This algorithm aims at minimizing the outage ratio, defined as the ratio of the number of non-supported users to the total number of users in the system, via a soft removal criterion. That is, users who cannot achieve their target SINRs are allocated with lower power. This method results in gains of over 15dB in the downlink signal to interference ratio in the interference limited scenario. In [18], a pilot power control algorithm was proposed assuming a least-squares (LS) channel estimate, which reduces the power of UTs that are close to their BS. In [22], a low-complexity power control algorithm is suggested to maximize spectral efficiency. The authors assume all the long-term statistics are known across all cells, and aim at suppressing both intra-cell and inter-cell interference. To that aim they utilize a large scale approximation of both uplink and downlink SINR. Additional interesting approach is the joint power-control and pilot allocation in [23]. This work considers a single-cell scenario and using geometric programming tools maximizes the sum- and max-min spectral efficiency. The extension of this work to the multi-cell scenario is not trivial.

Both the works on pilot allocation and pilot control assume *given pilot sequences*, and do not attempt to *design* the sequences. Rather, they limit the optimization to their allocation or transmit power.

Previous pilot sequence design methods typically consider a single-cell perspective [24]–[29]. These works treat a single desired channel with multiple interferers and optimize the pilot sequences in the specific cell to minimize its channel estimation error assuming all the pilots of surrounding cells are given. Generalization of these solutions to multiple cells for jointly designing all pilot matrices in the system jointly is not trivial. A possible straightforward approach for generalizing these methods is to iteratively apply the single-cell solution over each cell independently. However, there is no guarantee that this extension improves the overall channel estimation performance, as the optimization focuses only a single-cell. A joint pilot design method was recently considered in [30], where a convex relaxation was suggested to minimizing the total MSEs of the minimum mean-squared error (MMSE) estimators of all BSs subject to the transmit power constraints of individual users in the network. The optimization is carried successively over the cells. However, this solution results in a sub-optimal MSE due to both its iterative nature and the convex relaxation to the non-convex problem. In contrast to the single-cell solutions, when jointly optimizing over the entire network, we show that under our channel model the solution for the joint-design problem is either obtained in closed-form or is based on a greedy algorithm, hence it is guaranteed to improve the overall channel estimation accuracy.

Previous works on pilot contamination do not consider the problem of *jointly designing the pilot sequences* across the cells. An exception is the recent paper [31] which was submitted with close proximity to this work. Here, the authors express the pilot signals using an orthogonal basis and consider a max-min problem over the UTs' SINR across all cells, where the optimization variables are the associated power coefficients for each basis vector. They then relax the problem and find a local optimum. In all previous papers including [31], it is assumed that there is a dedicated RF chain per antenna. Such architectures are very costly in terms of hardware. In addition, the massive amount of data received from hundreds of antennas at each BS constitutes a huge computational challenge. Therefore, it is desirable to reduce the number of RF chains via analog combining while maintaining high accuracy in estimating the channel. Previous works on analog combiner design [32]–[39] focused on the problems of channel estimation and hybrid beamforming, under various hardware constraints. However, to the best of our knowledge, *joint analog combiner and pilot sequence design* has not been studied to date.

In this work, we treat the joint design of pilot sequences and analog combiners, aimed at minimizing the sum of MSEs across all cells, for a fixed number of RF chains and pilot symbols. We consider a general statistical channel model in which the channel's correlation matrices have a Kronecker structure and focus on two scenarios: fully-separable and partially-separable correlations. For both cases, the only information assumed to be shared between different cells is the long-term statistics of the channels. We prove that, in both scenarios, the analog combiners can be first designed independently of the pilot sequences by applying existing algorithms such as the ones in [39]. Given the resulting combiners, the pilot sequences are designed, where the solution depends on the channel model.

In the fully-separable case [40], the optimal pilot sequences are obtained in closed-form. In particular, when the transmit correlation matrix is diagonal, the received pilots correspond to user selection, i.e., choosing the UTs with strongest average links to their BS. For the partially-separable correlation model, we express the pilot sequence design problem as a maximization of a weighted sum of ratio traces. We then propose a greedy algorithm that generalizes a previously suggested method for maximizing a single ratio trace, the greedy ratio trace maximization (GRTM) [39], to a sum. At each iteration of the greedy sum of ratio traces maximization (GSRTM), one additional pilot symbol is enabled, and the algorithm chooses the optimal symbol to add for each user, given the previous selections.

We demonstrate the advantage of joint pilot design in simulations for both cases. We compare the suggested algorithms with other design methods and state-of-the-art pilot allocation techniques, and show that our algorithms enjoy lower sum-MSE.

The rest of this paper is organized as follows: Section II presents the system model and problem formulation. Section III formulates the MMSE channel estimation problem. In Section IV we study the analog combiner design problem. Pilot sequence optimization is treated in Section V. Section VI illustrates the performance of our approach in simulation examples.

Throughout the paper the following notations are used: We denote column vectors with boldface letters, e.g., $\mathbf{x}$, and matrices with boldface upper-case letters, e.g., $\boldsymbol{X}$. The set of complex numbers is denoted by $\mathbb{C}$, $\mathcal{CN}$ is the complex-normal distribution, $\mathbf{1}_n$ is the $n \times 1$ all ones vector, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Hermitian transpose, transpose, Frobenius norm, and Kronecker product are denoted by $(\cdot)^*$, $(\cdot)^T$, $\|\cdot\|_F$, and $\otimes$, respectively. For two $n \times 1$ real-valued vectors $\mathbf{x}_1, \mathbf{x}_2$, the inequality $\mathbf{x}_1 \leq \mathbf{x}_2$ indicates that all the entries of $\mathbf{x}_1$ are less than or equal to the corresponding entries of $\mathbf{x}_2$. For an $n \times n$ matrix $\boldsymbol{X}$, $\mathrm{tr}(\boldsymbol{X})$ is the trace of $\boldsymbol{X}$, $\lambda_i(\boldsymbol{X})$ is the i-th largest real eigenvalue of $\boldsymbol{X}$ and $\mathrm{vec}(\boldsymbol{X})$ is the $n^2 \times 1$ column vector obtained by stacking the columns of $\boldsymbol{X}$ one below the other. The phase of a variable is denoted as $\angle(\cdot)$, $\mathcal{R}(\boldsymbol{X})$ is the range space of $\boldsymbol{X}$ and $\boldsymbol{P}_{\boldsymbol{X}}$ is the orthogonal projection onto $\mathcal{R}(\boldsymbol{X})$. The $n \times n$ diagonal matrix $\mathrm{diag}(\boldsymbol{x})$ has the vector $\boldsymbol{x}$ on its diagonal, and $\mathrm{diag}^{-1}(\boldsymbol{X})$ is the $n \times 1$ vector whose entries are the diagonal elements of $\boldsymbol{X}$. Finally, for a sequence of $n \times m$ matrices $\{\boldsymbol{X}_i\}_{i=1}^k$, $\mathrm{blkdiag}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is the $kn \times km$ block-diagonal matrix with on-diagonal matrices $\{\boldsymbol{X}_i\}_{i=1}^k$.

II. Problem Formulation and Channel Model

A. Problem Formulation

Consider a network of M time-synchronized cells with full spectrum reuse. In each cell, a BS equipped with N_{BS} antennas and $N_{\mathrm{RF}} \leq N_{\mathrm{BS}}$ RF chains serves K single antenna UTs. At the BS, a network of analog components, typically phase shifters and switches, maps the N_{BS} antennas to the N_{RF} RF chains. In the digital domain, only the reduced number of inputs, N_{RF}, are accessible.

In the uplink channel estimation phase, each UT sends $\tau < MK$ training symbols to the BS, during which the channel is assumed to be constant. Here, the number of pilot symbols τ is pre-determined according to the coherence budget and not subject to optimization. For optimal training length please see for example [4], [41]. Let $\boldsymbol{s}_{ik}$ be the $\tau \times 1$ pilot sequence vector of the k-th user in the i-th cell, assumed to be subject to a per-user power constraint $\boldsymbol{s}_{ik}^* \boldsymbol{s}_{ik} \leq \mathcal{P}$. The pilot sequence matrix of the UTs in the i-th cell is denoted by $\boldsymbol{S}_i = \left[\boldsymbol{s}_{i1}, \ldots, \boldsymbol{s}_{iK}\right]$. Let $\boldsymbol{H}_{ij} \in \mathbb{C}^{N_{\mathrm{BS}} \times K}$ be the channel matrix between the UTs of the j-th cell and the i-th BS, and $\boldsymbol{W}_i \in \mathcal{W}_{\mathrm{RF}}$ be the analog combiner matrix at the i-th BS, where $\mathcal{W}_{\mathrm{RF}} \subseteq \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{BS}}}$ is the feasible set of analog matrices. The properties of $\mathcal{W}_{\mathrm{RF}}$ are discussed below.

We consider an interference limited setting, based on the analysis in [5], which shows that in the limit of an infinite number of BS antennas, the effect of uncorrelated noise vanishes. In this case, the (discrete-time) received signal at the i-th cell BS, $\boldsymbol{Y}_i \in \mathbb{C}^{N_{\mathrm{RF}} \times \tau}$, $1 \leq i \leq M$, is given by

$$\boldsymbol{Y}_i = \boldsymbol{W}_i \boldsymbol{H}_{ii} \boldsymbol{S}_i^T + \boldsymbol{W}_i \sum_{j \neq i} \boldsymbol{H}_{ij} \boldsymbol{S}_j^T. \tag{1}$$

At each BS the channel $\boldsymbol{H}_{ii}$ is estimated from the measurements $\boldsymbol{Y}_i$, using an MMSE estimator, where the expectation is taken over the channels $\boldsymbol{H}_{ij}$, $1 \leq i \leq M$. We note that the ith BS, $1 \leq i \leq M$, needs to estimate only the channel to its corresponding UTs, namely, $\boldsymbol{H}_{ii}$. In this work we assume synchronous communication, i.e. the propagation delays from all users to all BSs are much shorter than the pilot sequence.

The pilot contamination problem implies that the estimation of $\boldsymbol{H}_{ii}$ is degraded by the presence of the interfering channels $\boldsymbol{H}_{ij}$, $i \neq j$, in the received signal $\boldsymbol{Y}_i$. Since the dimension τ of the pilot matrices $\{\boldsymbol{S}_i\}_{i=1}^M$ is smaller than the total number of users in the system MK, these interferences cannot be fully eliminated at the ith BS. We wish to jointly design the pilot sequences and analog combiners $\{\boldsymbol{S}_i, \boldsymbol{W}_i\}_{i=1}^M$ to minimize the sum of the MSEs in estimating the desired channels $\boldsymbol{H}_{ii}$ from the observations $\boldsymbol{Y}_i$, $1 \leq i \leq M$, over the entire network, subject to the per-user power constraint on $\boldsymbol{S}_i$, and to the feasible set of $\boldsymbol{W}_i$. We note that here we assume the pilot and data power budget are independent. A joint-budget system is an interesting topic but out-of-scope here.

The feasible set $\mathcal{W}_{\mathrm{RF}}$ depends on the specific architecture of the analog network, which can vary according to power,

area, and budget constraints. For example, in the case of a fully-connected phase shifters network, $\mathcal{W}_{\mathrm{RF}}$ is the set of $N_{\mathrm{RF}} \times N_{\mathrm{BS}}$ unimodular matrices. For a review of common hardware schemes, the reader is referred to [35], [39].

Designing pilot sequences in the interference-limited setup (1) without RF reduction has been previously considered in [26], [29] under a fully-separable Kronecker model. These works treated a single desired channel, $\boldsymbol{H}_{11}$, with multiple interferers, $\boldsymbol{H}_{1j}$, $j = 2, \cdots, M$, that share the same receive side correlation. In contrast to our approach, they optimized only the pilot sequence of the specific cell, $\boldsymbol{S}_1$, to minimize its channel estimation error assuming all the interferers pilot matrices $\boldsymbol{S}_j$, $j = 2, \cdots, M$, are given. This leads to a water filling solution, assigning more power to directions with larger channel gains and weaker interference. Here, we aim at jointly optimizing the pilot matrices of all cells. Furthermore, we consider both fully- and partially-separable correlation structures for the channel covariance.

B. Channel Model

We focus on statistical channel models obeying the following structure:

$$\boldsymbol{H}_{ij} = \boldsymbol{Q}_{ij}^{\frac{1}{2}} \bar{\boldsymbol{H}}_{ij} \boldsymbol{P}_{ij}^{\frac{1}{2}}. \tag{2}$$

Here $\boldsymbol{Q}_{ij} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{BS}}}, \boldsymbol{P}_{ij} \in \mathbb{C}^{K \times K}$ are deterministic positive semi-defninte (PSD) Hermitian matrices, referred to as the correlation matrices, and $\bar{\boldsymbol{H}}_{ij} \in \mathbb{C}^{N_{\mathrm{BS}} \times K}, 1 \leq i, j \leq M$ are random matrices with i.i.d complex-normal zero-mean unit variance entries, representing the fast-fading channel coefficients. We assume that $\boldsymbol{Q}_{ij}, \boldsymbol{P}_{ij}$, $1 \leq i, j \leq M$, are known at all the BSs. We focus on two special cases of (2):

1) **Fully-Separable Correlations.** In this case,

$$\boldsymbol{Q}_{ij} = \boldsymbol{Q}_i, \ \boldsymbol{P}_{ij} = \boldsymbol{P}_j, \tag{3}$$

i.e., the receive and transmit side correlations are independent of each other. This model is also known as the *doubly correlated Kronecker model* [42]. Here $\boldsymbol{Q}_i = \boldsymbol{R}_{r_i}$ is the receive side correlation matrix of the i-th cell (assumed here for simplicity to be full rank), and $\boldsymbol{P}_j = \boldsymbol{R}_{t_j}$ the transmit side correlation matrix of the users in the j-th cell. It is a reasonable assumption that $\boldsymbol{P}_j$ is an invertible diagonal matrix due to the distance between different UTs. However, in our setup we allow general transmit correlation matrices.
The fully-separable model implies that the transmitters do not affect the spatial properties of the received signal and vice versa. This model was shown to accurately describe systems using space diversity arrays [42]. Nonetheless, we note that it does not account for the distance of different users from the same BS, and it inherently rules out spatial orthogonality between different UTs. However, it is a simple special case of the more general partially-separable model (4), which we show to have a closed form optimal solution that can give some intuition to the behavior of the general problem.

2) **Partially Separable Correlations.** In this model only $\boldsymbol{Q}_{ij}$ is separable, that is

$$\boldsymbol{Q}_{ij} = \boldsymbol{Q}_i. \tag{4}$$

An example of a partially-separable channel is the widely used model of [5], which hereinafter we refer to as the *MU-MIMO channel fading model*. In this case, the columns of $\boldsymbol{H}_{ij}$, denoted $\{\boldsymbol{g}_{ikj}\}_{k=1}^{K}$, are independent random vectors distributed as $\boldsymbol{g}_{ikj} \sim \mathcal{CN}(0, \beta_{ikj} \boldsymbol{I}_{N_{\mathrm{BS}}})$ for some set of decay factors $\{\beta_{ikj}\}$. The channel matrices can then be written as

$$\boldsymbol{H}_{ij} = \bar{\boldsymbol{H}}_{ij} \boldsymbol{D}_{ij}^{\frac{1}{2}}, \tag{5}$$

with $\boldsymbol{D}_{ij} = \mathrm{diag}\,(\beta_{i1j} \cdots \beta_{iKj})$, resulting in

$$\boldsymbol{Q}_i = \boldsymbol{I}_{N_{\mathrm{BS}}}, \ \boldsymbol{P}_{ij} = \boldsymbol{D}_{ij}. \tag{6}$$

In this work, the correlation matrices $\{\boldsymbol{Q}_{ij}, \boldsymbol{P}_{ij}\}$ are assumed to be known a-priori. In practice, they need to be estimated, which can be computationally complex in massive MIMO systems due to the large number of estimated matrix entries. However, under the Kronecker structures considered here, the number of parameters to be estimated is significantly reduced. Moreover, the correlation matrices are typically long-term statistics, varying much slower compared to the fast-fading channel coefficients, so that estimation can be performed once every many coherence intervals, without substantial overhead.

III. MMSE Channel Estimation

We begin by deriving the MMSE channel estimator and its resulting MSE under the model (2).

Vectorizing the received signal in (1), $\boldsymbol{y}_i = \mathrm{vec}\,(\boldsymbol{Y}_i)$, we have

$$\boldsymbol{y}_i = (\boldsymbol{S}_i \otimes \boldsymbol{W}_i)\, \boldsymbol{h}_{ii} + \sum_{j \neq i} (\boldsymbol{S}_j \otimes \boldsymbol{W}_i)\, \boldsymbol{h}_{ij}, \tag{7}$$

where $\boldsymbol{h}_{ij} = \mathrm{vec}\,(\boldsymbol{H}_{ij})$ and $\boldsymbol{h}_{ij} \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{P}_{ij} \otimes \boldsymbol{Q}_{ij}\right)$. The i-th BS estimates its desired channel $\boldsymbol{h}_{ii}$ using the MMSE estimator [43, Chp. 12], given by

$$\hat{\boldsymbol{h}}_{ii} = (\boldsymbol{P}_{ii} \boldsymbol{S}_i^* \otimes \boldsymbol{Q}_{ii} \boldsymbol{W}_i^*) \left[\sum_{j=1}^{M} \left(\boldsymbol{S}_j \boldsymbol{P}_{ij} \boldsymbol{S}_j^* \otimes \boldsymbol{W}_i \boldsymbol{Q}_{ij} \boldsymbol{W}_i^* \right) \right]^{-1} \boldsymbol{y}_i. \tag{8}$$

Here, we assume all inverses exist; similar expressions can be derived using the pseudo-inverse. The corresponding estimation error using (8) is given by

$$\epsilon_i = \mathrm{tr}\,(\boldsymbol{A}_i) - \mathrm{tr}\,(\boldsymbol{B}_i \boldsymbol{C}_i \boldsymbol{B}_i^*), \tag{9}$$

with $\boldsymbol{A}_i \triangleq \boldsymbol{P}_{ii} \otimes \boldsymbol{Q}_{ii}$,

$$\boldsymbol{B}_i \triangleq \boldsymbol{P}_{ii} \boldsymbol{S}_i^* \otimes \boldsymbol{Q}_{ii} \boldsymbol{W}_i^*, \tag{10}$$

and

$$\boldsymbol{C}_i \triangleq \left[\sum_{j=1}^{M} \boldsymbol{S}_j \boldsymbol{P}_{ij} \boldsymbol{S}_j^* \otimes \boldsymbol{W}_i \boldsymbol{Q}_{ij} \boldsymbol{W}_i^* \right]^{-1}. \tag{11}$$

We wish to design the pilot sequence matrices $\boldsymbol{S}_i$ and analog combining matrices $\boldsymbol{W}_i$ for $1 \leq i \leq M$ to minimize the sum of the errors

$$\epsilon = \sum_{i=1}^{M} \epsilon_i = \sum_{i=1}^{M} \left(\mathrm{tr}\left(\boldsymbol{A}_i\right) - \mathrm{tr}\left(\boldsymbol{B}_i \boldsymbol{C}_i \boldsymbol{B}_i^*\right)\right), \tag{12}$$

under the given power and hardware constraints. This approach will result in a different solution than the single target cell design in [24]–[29], as an interfering user in one cell is the desired user in another.

Since the matrices $\boldsymbol{A}_i$ do not depend on the optimization variables $\boldsymbol{W}_i$ and $\boldsymbol{S}_i$, minimizing ϵ is equivalent to maximization of $\sum_i \mathrm{tr}\left(\boldsymbol{B}_i \boldsymbol{C}_i \boldsymbol{B}_i^*\right)$, which yields the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{S}_i, \boldsymbol{W}_i} \; & \sum_i \mathrm{tr}(\boldsymbol{B}_i \boldsymbol{C}_i \boldsymbol{B}_i^*) \\ \text{s.t. } & \boldsymbol{s}_{ik}^* \boldsymbol{s}_{ik} \leq \mathcal{P}, \;\; 1 \leq i \leq M, \; 1 \leq k \leq K, \\ & \boldsymbol{W}_i \in \mathcal{W}_{\mathrm{RF}}, \; 1 \leq i \leq M \end{aligned} \tag{13}$$

where $\boldsymbol{B}_i$ and $\boldsymbol{C}_i$ are given by (10) and (11) respectively.

In the next section, we show that in both scenarios (3) and (4) the combiner design can be performed independently of the pilot sequences. Furthermore, previously suggested methods such as those of [39] may be used. In Section V we solve the remaining pilot design problem for each scenario separately.

IV. Analog Combiner Design

A. Analog Combiner Design Problem Formulation

Note that in both correlation scenarios (3) and (4) we have $\boldsymbol{Q}_{ij} = \boldsymbol{Q}_i$. By defining

$$\alpha_i \triangleq \mathrm{tr}\left(\boldsymbol{Q}_i \boldsymbol{W}_i^* \left(\boldsymbol{W}_i \boldsymbol{Q}_i \boldsymbol{W}_i^* \right)^{-1} \boldsymbol{W}_i \boldsymbol{Q}_i \right), \tag{14}$$

and using standard Kronecker product properties, (13) becomes

$$\begin{aligned} \max_{\boldsymbol{S}_i, \boldsymbol{W}_i} \; & \sum_i \alpha_i \cdot \mathrm{tr}\left(\boldsymbol{S}_i \boldsymbol{P}_{ii}^2 \boldsymbol{S}_i^* \left[\sum_j^M \boldsymbol{S}_j \boldsymbol{P}_{ij} \boldsymbol{S}_j^* \right]^{-1} \right) \\ \text{s.t. } & \boldsymbol{s}_{ik}^* \boldsymbol{s}_{ik} \leq \mathcal{P}, \;\; 1 \leq i \leq M, \; 1 \leq k \leq K, \\ & \boldsymbol{W}_i \in \mathcal{W}_{\mathrm{RF}}, \; 1 \leq i \leq M. \end{aligned} \tag{15}$$

Since the matrices $\boldsymbol{P}_{ij}$ are PSD for all $1 \leq i, j \leq M$, it follows that

$$\mathrm{tr}\left(\boldsymbol{S}_i \boldsymbol{P}_{ii}^2 \boldsymbol{S}_i^* \left[\sum_j^M \boldsymbol{S}_j \boldsymbol{P}_{ij} \boldsymbol{S}_j^* \right]^{-1} \right) \geq 0, \; 1 \leq i \leq M.$$

Consequently, the objective in (15) is a weighted sum of non-negative elements, and hence an increasing function in all the weights α_i, $1 \leq i \leq M$. Since the constraint on $\boldsymbol{W}_i$ is independent of $\{\boldsymbol{S}\}_{i=1}^M$ and of $\boldsymbol{W}_j$ for $j \neq i$, we can maximize the sum in (15) by maximizing each weight α_i separately for each cell and independently of $\{\boldsymbol{S}\}_{i=1}^M$. The resulting α_i, $1 \leq i \leq M$, can then be plugged back into (15) and the resulting problem may be solved with respect to the pilot matrices. Note that since the optimization of the combiner is independent of the pilot sequences, no alternations between the two is required.

The optimization problem for the analog combiner of the i-th BS is therefore

$$\begin{aligned} \max_{\boldsymbol{W}_i} \; & \mathrm{tr}\left(\boldsymbol{W}_i \boldsymbol{Q}_i^2 \boldsymbol{W}_i^* \left(\boldsymbol{W}_i \boldsymbol{Q}_i \boldsymbol{W}_i^* \right)^{-1} \right) \\ \text{s.t. } & \boldsymbol{W}_i \in \mathcal{W}_{\mathrm{RF}}. \end{aligned} \tag{16}$$

Since (16) can be solved for each BS separately, for clarity, we drop the index i. This problem has been previously studied in [39], [44], under various hardware constraints. In [44], a permissive analog scheme that consists of phase shifters, gain shifters and switches was considered. The feasible set $\mathcal{W}_{\mathrm{RF}}$ in this case is the set of all matrices $\boldsymbol{W} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{BS}}}$. For this choice, the family of optimal solutions to (16) is given by

$$\boldsymbol{W}_{\mathrm{FD}} = \boldsymbol{T} \tilde{\boldsymbol{U}}^*, \tag{17}$$

where $\tilde{\boldsymbol{U}}$ is an $N_{\mathrm{RF}} \times N_{\mathrm{BS}}$ matrix whose columns are the eigenvectors of the Hermitian matrix $\boldsymbol{Q}$ corresponding to its N_{RF} largest eigenvalues, and $\boldsymbol{T}$ is some $N_{\mathrm{RF}} \times N_{\mathrm{RF}}$ invertible matrix. The combiner (17) is commonly referred to as the fully-digital solution. The specific solution derived in [44] is (17) with $\boldsymbol{T} = \boldsymbol{I}_{N_{\mathrm{RF}}}$. Problem (16) is also a special case of the framework treated in [39]. There, more restrictive hardware schemes were considered, under which the optimal solution for (16) may be intractable. For these scenarios, two low complexity algorithms were suggested to obtain a sub-optimal solution: Minimal gap iterative quantization (MaGiQ) and GRTM. Both techniques may be used here to solve (16). For completeness, in the following subsection we describe MaGiQ [39].

The problem (16) depends only on the long-term statistics of the channel. Hence, the combiner can be designed once every many coherence intervals. After the analog combiners are designed for each cell separately, the pilot sequences are optimized given the resulting combiners. The impact of the RF reduction on the channel's MSE is illustrated via simulation examples in Section VI.

B. MaGiQ - Minimal Gap Iterative Quantization

One approach for treating (16), is to approximate the fully-digital solution (17) with an analog one. It was shown in [39] that minimizing the approximation gap between the analog combiner $\boldsymbol{W}$ and the fully-digital combiner $\boldsymbol{W}_{\mathrm{FD}}$ minimizes an upper bound on the MSE (16). Motivated by this insight, it was suggested to consider the problem

$$\begin{aligned} \min_{\boldsymbol{W}, \boldsymbol{T}} \; & \| \boldsymbol{T} \tilde{\boldsymbol{U}}^* - \boldsymbol{W} \|_F^2 \\ \text{s.t. } & \boldsymbol{W} \in \mathcal{W}_{\mathrm{RF}}, \quad \boldsymbol{T} \in \mathcal{U}, \end{aligned} \tag{18}$$

where $\mathcal{U}$ is the set of unitary $N_{\mathrm{RF}} \times N_{\mathrm{RF}}$ matrices. Different from previous approaches [35], [36] which considered $\boldsymbol{T} = \boldsymbol{I}_{N_{\mathrm{RF}}}$ and optimized (18) over $\boldsymbol{W}$ alone, here the optimization is carried out with respect to both $\boldsymbol{W}$ and $\boldsymbol{T}$. To solve (18), an alternating minimization approach was suggested, referred to as MaGiQ.

Algorithm 1 MaGiQ - Minimal Gap Iterative Quantization

Input: $\tilde{\boldsymbol{U}}^*$, threshold t
Output: analog combiner $\boldsymbol{W}$
Initialize $\boldsymbol{T}=\boldsymbol{I}_{N_{\mathrm{RF}}}, \boldsymbol{W}=\boldsymbol{0}$
While $\|\boldsymbol{T}\tilde{\boldsymbol{U}}^*-\boldsymbol{W}\|_F^2 \geq t$ do:
1) $\boldsymbol{W}=\boldsymbol{P}_{\mathcal{W}}(\boldsymbol{T}\tilde{\boldsymbol{U}}^*)$
2) Calculate the SVD $\tilde{\boldsymbol{U}}^*\boldsymbol{W}^*=\bar{\boldsymbol{U}}\boldsymbol{\Lambda}\bar{\boldsymbol{V}}^*$
3) $\boldsymbol{T}=\bar{\boldsymbol{V}}\bar{\boldsymbol{U}}^*$

The benefit of MaGiQ is that at each iteration, a closed-form solution for both variables is available. For fixed $\boldsymbol{W}$, the optimal $\boldsymbol{T}$ is given by

$$\boldsymbol{T}_{\mathrm{opt}}=\bar{\boldsymbol{V}}\bar{\boldsymbol{U}}^* \tag{19}$$

with $\tilde{\boldsymbol{U}}^*\boldsymbol{W}^*=\bar{\boldsymbol{U}}\boldsymbol{\Lambda}\bar{\boldsymbol{V}}^*$ the singular value decomposition (SVD) of $\tilde{\boldsymbol{U}}^*\boldsymbol{W}^*$. For a given $\boldsymbol{T}$, the optimal combiner is

$$\boldsymbol{W}_{\mathrm{opt}}=\boldsymbol{P}_{\mathcal{W}_{\mathrm{RF}}}\left(\boldsymbol{T}\tilde{\boldsymbol{U}}^*\right) \tag{20}$$

with $\boldsymbol{P}_{\mathcal{W}}(\boldsymbol{A})$ denoting the orthogonal projection of the matrix $\boldsymbol{A}$ onto the set $\mathcal{W}$. This operator depends on the specific hardware scheme in use and needs to be calculated for each architecture separately. For example, in the case of a fully-connected phase shifter network, the operator reduces to an entry-wise projection onto the unimodular set:

$$\left[\boldsymbol{P}_{\mathcal{W}_{RF}}(\boldsymbol{A})\right]_{il}=e^{j2\pi\angle[\boldsymbol{A}]_{il}}. \tag{21}$$

For further examples see [39]. MaGiQ is summarized in Algorithm 1.

V. Pilot Sequence Design

After the analog combiners are optimized, the resulting weights α_i are plugged back into (15), leading to the following problem with respect to the pilot sequences $\boldsymbol{S}_i$, $1\leq i\leq M$:

$$\begin{aligned}\max_{\boldsymbol{s}_i}\ & \textstyle\sum_i \alpha_i\cdot\mathrm{tr}\left(\boldsymbol{S}_i\boldsymbol{P}_{ii}^2\boldsymbol{S}_i^*\left[\sum_j^M\boldsymbol{S}_j\boldsymbol{P}_{ij}\boldsymbol{S}_j^*\right]^{-1}\right)\\ \text{s.t.}\ & \boldsymbol{s}_{ik}^*\boldsymbol{s}_{ik}\leq\mathcal{P},\ 1\leq i\leq M,\ 1\leq k\leq K.\end{aligned} \tag{22}$$

We now study problem (22) under the channel models (3) and (4). In Section V-A, we consider the fully-separable model in (3), under which we derive a closed-form solution for the pilot sequences $\boldsymbol{S}_i$. Next, in Section V-B, we focus on the partially-separable model (4) and express the pilot design problem as a sum of quadratic ratios. We then suggest a greedy algorithm to solve it, referred to as GSRTM.

A. Fully Separable Correlations

For the fully-separable scenario we show that although this model may lack in description for massive MIMO systems, it holds a key advantage in facilitating the solution of (13), as it results in a closed form expression for the optimal pilot sequences.

Under the model (3), we have $\boldsymbol{P}_{ij}=\boldsymbol{P}_j$. By letting

$$\boldsymbol{S}\triangleq[\boldsymbol{S}_1,\cdots,\boldsymbol{S}_M] \tag{23}$$

be the $\tau\times MK$ pilot matrix,

$$\bar{\boldsymbol{P}}\triangleq\mathrm{blkdiag}\left([\boldsymbol{P}_1,\cdots,\boldsymbol{P}_M]\right), \tag{24}$$

the $MK\times MK$ Hermitian transmit correlation matrix and

$$\bar{\boldsymbol{J}}\triangleq\mathrm{diag}\left([\alpha_1,\cdots,\alpha_M]\right)\otimes\boldsymbol{I}_K \tag{25}$$

the $MK\times MK$ non-negative valued diagonal weights matrix, problem (22) can be written as

$$\max_{\boldsymbol{S}}\ \mathrm{tr}\left(\left(\boldsymbol{S}\bar{\boldsymbol{P}}\boldsymbol{S}^*\right)^{-1}\boldsymbol{S}\bar{\boldsymbol{P}}\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}\boldsymbol{S}^*\right) \tag{26}$$

$$\text{s.t.}\quad \mathrm{diag}^{-1}\left(\boldsymbol{S}^*\boldsymbol{S}\right)\leq\mathcal{P}\cdot\boldsymbol{1}_{M\cdot K}. \tag{27}$$

In the following theorem, we derive an optimal solution $\boldsymbol{S}_{\mathrm{opt}}$ to (26)-(27), which we refer to as the eigen-pilots.

Theorem 1: A pilot symbol matrix which solves (26)-(27) *is given by*

$$\boldsymbol{S}_{\mathrm{opt}}=\sqrt{\mathcal{P}}\boldsymbol{U}_1^*, \tag{28}$$

with $\boldsymbol{U}_1$ the $MK\times\tau$ matrix whose columns are the τ eigenvectors of the Hermitian matrix $\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}$ corresponding to the τ largest eigenvalues.

Proof: Let $\boldsymbol{Q}=\boldsymbol{S}\bar{\boldsymbol{P}}^{\frac{1}{2}}$. Then the objective in (26) becomes $\mathrm{tr}\left(\boldsymbol{P}_{\boldsymbol{Q}}\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}\right)$, where $\boldsymbol{P}_{\boldsymbol{Q}}=\boldsymbol{Q}^*\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-1}\boldsymbol{Q}$ is the orthogonal projection onto $\mathcal{R}(\boldsymbol{Q}^*)$, the range space of $\boldsymbol{Q}^*$. Let $\boldsymbol{V}\in\mathbb{C}^{MK\times\tau}$ be a matrix with orthogonal columns that spans $\mathcal{R}(\boldsymbol{Q}^*)$. Then $\boldsymbol{P}_{\boldsymbol{Q}}=\boldsymbol{V}\boldsymbol{V}^*$ and $\boldsymbol{Q}=\boldsymbol{D}\boldsymbol{V}^*$ for some invertible $\tau\times\tau$ matrix $\boldsymbol{D}$.

If we ignore the constraint (27), then (26) is solved by choosing $\boldsymbol{S}=\boldsymbol{D}\boldsymbol{V}^*\bar{\boldsymbol{P}}^{-\frac{1}{2}}$, where $\boldsymbol{V}$ is the solution of

$$\begin{aligned}\max_{V}\ & \mathrm{tr}\left(\boldsymbol{V}^*\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}\boldsymbol{V}\right)\\ \text{s.t.}\ & \boldsymbol{V}^*\boldsymbol{V}=\boldsymbol{I}_\tau.\end{aligned} \tag{29}$$

This problem has a closed form solution, $\boldsymbol{V}=\boldsymbol{U}_1$ [45, Ch. 20.A.2]. For this choice of $\boldsymbol{V}$, the objective in (29) becomes

$$\mathrm{tr}\left(\boldsymbol{V}^*\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}\boldsymbol{V}\right)=\textstyle\sum_{i=1}^{\tau}\lambda_i\left(\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}\right). \tag{30}$$

The expression (30) is an upper bound on (26) under the power constraint (27). We now show that there is a specific choice of $\boldsymbol{D}$ such that $\boldsymbol{S}=\boldsymbol{D}\boldsymbol{V}^*\bar{\boldsymbol{P}}^{-\frac{1}{2}}$ both satisfies the power constraint and achieves the upper bound, and is therefore optimal. First, we note that

$$\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}=\mathrm{blkdiag}(\alpha_1\boldsymbol{P}_1,\cdots,\alpha_M\boldsymbol{P}_M). \tag{31}$$

Thus, $\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}$ and $\bar{\boldsymbol{P}}$ share the same eigenvectors (possibly in different order). It then follows that $\boldsymbol{U}_1^*\bar{\boldsymbol{P}}^{-\frac{1}{2}}=\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{U}_1^*$, where $\bar{\boldsymbol{\Sigma}}$ is a $\tau\times\tau$ diagonal matrix with the eigenvalues of $\bar{\boldsymbol{P}}$ corresponding to $\boldsymbol{U}_1$ on its diagonal. Hence, $\boldsymbol{S}=\boldsymbol{D}\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{U}_1^*$, and the power constraint (27) can be written in terms of $\boldsymbol{D}$ as

$$\mathrm{diag}^{-1}\left(\boldsymbol{U}_1\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{D}^*\boldsymbol{D}\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{U}_1^*\right)\leq\mathcal{P}\cdot\boldsymbol{1}_{M\cdot K}. \tag{32}$$

Let $\boldsymbol{D}=\sqrt{\mathcal{P}}\bar{\boldsymbol{\Sigma}}^{\frac{1}{2}}$. Then, $\mathrm{diag}^{-1}\left(\boldsymbol{S}^*\boldsymbol{S}\right)=\mathcal{P}\mathrm{diag}^{-1}\left(\boldsymbol{U}_1\boldsymbol{U}_1^*\right)$. Since $\boldsymbol{U}_1$ are columns of a unitary matrix, $\mathrm{diag}\left(\boldsymbol{U}_1\boldsymbol{U}_1^*\right)$ is a vector with positive elements that are smaller or equal to 1,

and the constraint (27) is satisfied. It thus follows that $\boldsymbol{S}_{opt} = \sqrt{\mathcal{P}}\boldsymbol{U}_1^*$ is an optimal solution to (26), proving the theorem. □

Note that the solution $\boldsymbol{S}_{opt}$ is not unique. In fact, if $\boldsymbol{S}_{opt}$ is a solution to (26), then

$$\boldsymbol{S} = \zeta \boldsymbol{T}\boldsymbol{S}_{opt} \tag{33}$$

is also a solution, where $\boldsymbol{T}$ is any $\tau \times \tau$ invertible matrix and ζ is chosen such that $\boldsymbol{S}$ complies with the power constraint. For example, a possible choice is $\zeta = 1$ and any unitary $\boldsymbol{T}$.

Insight into the optimal pilot sequences in Theorem 1 can be obtained when considering the special case in which $\boldsymbol{P}_i$, $1 \leq i \leq M$ are diagonal matrices.

Corollary 1: Let $\boldsymbol{P}_i$, $1 \leq i \leq M$ be diagonal matrices with $p_{i,kk}$ the k-th diagonal entry of $\boldsymbol{P}_i$. Then, the optimal solution $\boldsymbol{S}_{opt}$ corresponds to user selection, where the τ UTs with strongest link $d_{ik} = \alpha_i \cdot p_{i,kk}$ are chosen and assigned orthogonal sequences, while all other UTs are nullified.

Proof: In this case, $\bar{\boldsymbol{J}}\bar{\boldsymbol{P}}$ is a diagonal matrix with diagonal elements $d_{ik} = \alpha_i \cdot p_{i,kk}$, $1 \leq i \leq M$, $1 \leq k \leq K$. Furthermore, the columns of $\boldsymbol{U}_1$ are unit vectors. Therefore, although our problem is not a user scheduling problem but pilot design, the resulting $\boldsymbol{S}_{\rm opt}$ corresponds to a user selection solution. For each UT k in cell i, we calculate its corresponding d_{ik}, and choose the τ UTs with largest d_{ik} values. The remaining $MK - \tau$ UTs are set to zero. □

We emphasize that in contrast to classical user scheduling algorithms, e.g. [46], [47], here the UTs are selected according to their long-term statistics regardless of the specific channel realization.

The specific case of Corollary 1 demonstrates some of the main differences between our design framework and the pilot allocation approach, e.g., [11]–[16]. In allocation solutions, all the MK UTs are necessarily transmitting, and the optimization is over how to choose the users that will be assigned the same pilot sequence. In contrast, in our solution, some UTs' transmissions may be set to zero. The main disadvantage of this approach is its fairness across different UTs. If one wishes to promote fairness in the system, alternative objectives than the sum in (13), for example max-min, may be preferable [48]–[51]. Nevertheless, when aiming at minimizing the overall MSE in the system, the eigen-pilots of (28) are optimal (for the fully-separable case).

In the simulations in Section VI, we compare our eigen-pilots solution with other pilot design and allocation methods and show that it results in lower sum-MSEs.

B. Partially Separable Correlations

We now consider the scenario of partially-separable correlations, in which the channel correlation matrices satisfy (4). In this case, $\boldsymbol{P}_{ij}$ depends both on the transmitter index j and the receiver index i. Hence, instead of (24), we have the $MK \times MK$ block diagonal matrix

$$\bar{\boldsymbol{P}}_i = \text{blkdiag}\left(\boldsymbol{P}_{i1}, \cdots, \boldsymbol{P}_{iM}\right). \tag{34}$$

Let $\boldsymbol{Z}_i$ be an $M \times M$ matrix whose entries are zero except for the i-th diagonal entry, which equals one, and set $\boldsymbol{L}_i \triangleq \boldsymbol{Z}_i \otimes \boldsymbol{I}_K$. Then, $\bar{\boldsymbol{P}}_i\boldsymbol{L}_i$ is a block-diagonal matrix whose entries are zero except for the i-th block, which equals $\boldsymbol{P}_{i,i}$. Using these notations, and recalling the definition of $\boldsymbol{S}$ in (23), we can rewrite (22) as

$$\begin{aligned} \max_{\boldsymbol{S}} \; & \sum_i \alpha_i \cdot \text{tr}\left(\boldsymbol{S}\bar{\boldsymbol{P}}_i^2\boldsymbol{L}_i\boldsymbol{S}^*\left[\boldsymbol{S}\bar{\boldsymbol{P}}_i\boldsymbol{S}^*\right]^{-1}\right) \\ \text{s.t.} \; & \text{diag}^{-1}\left(\boldsymbol{S}^*\boldsymbol{S}\right) \leq \mathcal{P}\cdot\boldsymbol{1}. \end{aligned} \tag{35}$$

Note that in contrast to (26), here we still have the sum over the receiver index i. For $\boldsymbol{P}_{ij} = \boldsymbol{P}_j$ as in the fully-separable case, $\bar{\boldsymbol{P}}_i = \bar{\boldsymbol{P}}$, and (35) coincides with (26).

To the best of our knowledge, there is no known solution for (35) in general. However, there are some special cases for which a solution is available. For a single-cell network, $M = 1$, (35) becomes the ratio-trace problem as in (26), and can be solved using the same method. Another interesting scenario is when the number of pilot symbols is fixed to $\tau = 1$. Although this setting is unlikely in massive MIMO systems, in which τ is typically larger than K, we use this special case to illustrate the complexity of the considered problem. For a single pilot symbol the pilot matrix $\boldsymbol{S}$ is an $1 \times MK$ vector $\boldsymbol{s}$, and (35) becomes

$$\begin{aligned} \max_{\boldsymbol{s}} \; & \sum_i \alpha_i \frac{\boldsymbol{s}\bar{\boldsymbol{P}}_i^2\boldsymbol{L}_i\boldsymbol{s}^*}{\boldsymbol{s}\bar{\boldsymbol{P}}_i\boldsymbol{s}^*} \\ \text{s.t.} \; & s_l^* s_l \leq \mathcal{P}, l = 1 \cdots MK. \end{aligned} \tag{36}$$

Problem (36) is known as the sum of quadratic ratios maximization, and has been addressed in previous works via branch and bound [52], [53] and harmony search [54] algorithms. However, the large dimension of $\boldsymbol{s}$ in (36), which is equal to the total number of users in the system MK, and the singularity for $\boldsymbol{s} = 0$, makes these (and other similar) methods difficult to apply, even when $\tau = 1$.

In the following we propose a greedy method, referred to as GSRTM, for maximizing (35). We begin by considering pilot sequences of length $\tau = 1$, and at each step we expand the sequences with one additional symbol. We choose the optimal symbol to add to each UT's sequence, by solving a vector problem similar to (36). The suggested GSRTM method extends the GRTM algorithm proposed in [39] to the case of sum of ratios trace as in (35).

In the simulation section, we compare the performance of GSRTM with other pilot design and allocation techniques, and demonstrate that it results in lower sum-MSEs than alternative methods in the partially-correlated case.

C. GSRTM - Greedy Sum of Ratio Traces Maximization

In order to formulate the GSRTM algorithm, we first note that due to the quadratic form of the objective in (35), it is not sensitive to a scale in $\boldsymbol{S}$. That is, the objective value corresponding to a given $\boldsymbol{S}$ is identical to the value that corresponds to $\alpha\boldsymbol{S}$, for any scalar $\alpha \neq 0$. Therefore, we can first solve (35) without the power constraint, and then scale the solution to comply with it.

Assume we have a solution $\boldsymbol{S}_{(N)} \in \mathbb{C}^{N \times MK}$ for the N-sized problem, i.e. (35) with $\tau = N$. We now wish to

Algorithm 2 GSRTM

Input correlation matrices $\bar{\boldsymbol{P}}_i$, $1 \le i \le M$
Output pilot matrix $\boldsymbol{S}$
Initialize $\boldsymbol{A}_i = \bar{\boldsymbol{P}}_i^2 \boldsymbol{L}_i$, $\boldsymbol{B}_i = \bar{\boldsymbol{P}}_i$, $1 \le i \le M$
$\boldsymbol{G}_i = \boldsymbol{A}_i$, $\boldsymbol{T}_i = \boldsymbol{B}_i$, $1 \le i \le M$
$\boldsymbol{S}_{(0)} = []$
For $K = 0 : \tau - 1$:

1) **Solve** the base case (38) to obtain $\boldsymbol{s}$
2) **Update** $\boldsymbol{S}_{(K+1)} = \left[\boldsymbol{S}_{(K)}^T \ \boldsymbol{s}^T\right]^T$
3) **Update**

$$\begin{aligned}
\boldsymbol{T}_i &= \boldsymbol{B}_i^{\frac{1}{2}} \left(\boldsymbol{I}_{MK} - \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(K)}^*} \right) \boldsymbol{B}_i^{\frac{1}{2}} \\
\gamma_i &= \mathrm{tr}\left(\boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(K)}^*} \boldsymbol{B}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{B}_i^{-\frac{1}{2}} \right) \\
\boldsymbol{X}_i &= \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(K)}^*} \boldsymbol{B}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{\frac{1}{2}} - \boldsymbol{A}_i^{\frac{1}{2}} \\
\boldsymbol{G}_i &= \gamma_i \boldsymbol{T}_i + \boldsymbol{X}_i \boldsymbol{X}_i^*
\end{aligned}$$

add an additional pilot symbol, and compute the optimal row vector $\boldsymbol{s} \in \mathbb{C}^{1 \times MK}$ to add to the previously selected sequences such that $\boldsymbol{S}_{(N+1)} = [\boldsymbol{S}_{(N)}^T \ \boldsymbol{s}^T]^T \in \mathbb{C}^{N+1 \times MK}$.

First, note that in order for the objective in (35) to be well defined in the $(N+1)$-sized case, we require $\boldsymbol{S}_{(N+1)} \bar{\boldsymbol{P}}_i \boldsymbol{S}_{(N+1)}^*$ to be invertible for all $1 \le i \le M$, namely, that $\boldsymbol{s} \notin \mathcal{R}(\boldsymbol{S}_{(N)})$. This condition is intuitive, since if the new pilot row in the pilot matrix depends linearly on the previous ones, then it will not contribute to the separation between users.

Given this condition, the $(N+1)$-sized optimization problem becomes

$$\begin{aligned}
\max_{\boldsymbol{s}} \quad & \sum_i \alpha_i \mathrm{tr}\Big(\boldsymbol{S}_{(N+1)} \boldsymbol{A}_i \boldsymbol{S}_{(N+1)}^* \left[\boldsymbol{S}_{(N+1)} \boldsymbol{B}_i \boldsymbol{S}_{(N+1)}^* \right]^{-1} \Big) \\
\text{s.t.} \quad & \boldsymbol{s} \notin \mathcal{R}(\boldsymbol{S}_{(N)}),
\end{aligned} \tag{37}$$

with $\boldsymbol{A}_i \triangleq \bar{\boldsymbol{P}}_i^2 \boldsymbol{L}_i$ and $\boldsymbol{B}_i \triangleq \bar{\boldsymbol{P}}_i$.

In the next proposition we show that for a specific choice of $MK \times MK$ PSD matrices $\{\boldsymbol{G}_i\}_{i=1}^M$ and $\{\boldsymbol{T}_i\}_{i=1}^M$, solving (37) is equivalent to solving the following sum of quadratic ratios problem:

$$\begin{aligned}
\max_{\boldsymbol{s}} \ & \sum_i \alpha_i \frac{\boldsymbol{s} \boldsymbol{G}_i \boldsymbol{s}^*}{\boldsymbol{s} \boldsymbol{T}_i \boldsymbol{s}^*} \\
\text{s.t.} \ & \boldsymbol{s} \boldsymbol{T}_i \boldsymbol{s}^* > 0, \quad 1 \le i \le M.
\end{aligned} \tag{38}$$

Proposition 1 leads to the formulation of the GSRTM algorithm for approaching the solution of (35), where in each iteration we choose the best row vector to add to the previously selected pilot matrix by solving (38). Once all τ rows of $\boldsymbol{S}$ are chosen, the entire matrix is scaled to comply with the power constraint. The GSRTM algorithm is summarized in Algorithm 2.

Proposition 1: Problems (38) and (37) are equivalent with

$$\begin{aligned}
\boldsymbol{T}_i &= \boldsymbol{B}_i^{\frac{1}{2}} \left(\boldsymbol{I}_{MK} - \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(k)}^*} \right) \boldsymbol{B}_i^{\frac{1}{2}} \\
\gamma_i &= tr\left(\boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(k)}^*} \boldsymbol{B}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{B}_i^{-\frac{1}{2}} \right) \\
\boldsymbol{X}_i &= \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(k)}^*} \boldsymbol{B}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{\frac{1}{2}} - \boldsymbol{A}_i^{\frac{1}{2}} \\
\boldsymbol{G}_i &= \gamma_i \boldsymbol{T}_i + \boldsymbol{X}_i \boldsymbol{X}_i^*.
\end{aligned}$$

The proof of Proposition 1 below is similar to the proof of [39, Prop. 3]. However, here the proof takes into account a sum of ratio traces, rather than a single ratio trace, and considers a different constraint on $\boldsymbol{s}$.

Proof: To prove the proposition, we rely on the following lemma:

Lemma 1: [55, Ch. 3] Let $\tilde{\boldsymbol{S}} = \left[\boldsymbol{S}^T \ \boldsymbol{s}^T\right]^T$ *and denote* $\boldsymbol{Q} = (\boldsymbol{S}\boldsymbol{S}^*)^{-1}$. *Then*

$$\left(\tilde{\boldsymbol{S}}\tilde{\boldsymbol{S}}^*\right)^{-1} = \begin{bmatrix} \boldsymbol{Q} + \alpha \boldsymbol{Q}\boldsymbol{S}\boldsymbol{s}^*\boldsymbol{s}\boldsymbol{S}^*\boldsymbol{Q}^* & -\alpha \boldsymbol{Q}\boldsymbol{S}\boldsymbol{s}^* \\ -\alpha \boldsymbol{s}\boldsymbol{S}^*\boldsymbol{Q}^* & \alpha \end{bmatrix},$$

with $\alpha = \frac{1}{\boldsymbol{s}\boldsymbol{s}^* - \boldsymbol{s}\boldsymbol{S}^*\boldsymbol{Q}\boldsymbol{S}\boldsymbol{s}^*}$.

Using Lemma 1 and straightforward algebraic operations we get equality between the i-th summand of the sums in (37) and in (38). Consequently, these two objectives are equal, and it remains to show that the constraints are equivalent to prove the proposition.

From the fact that $\boldsymbol{B}_i$, $1 \le i \le M$, are invertible, it follows that $\boldsymbol{s}^* \in \mathcal{R}(\boldsymbol{S}_{(N)}^*)$ if and only if $\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* \in \mathcal{R}(\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(N)}^*)$, $1 \le i \le M$. Therefore, if $\boldsymbol{s}^* \in \mathcal{R}(\boldsymbol{S}_{(N)}^*)$, then

$$\begin{aligned}
\boldsymbol{s}\boldsymbol{T}_i\boldsymbol{s}^* &= \boldsymbol{s}\boldsymbol{B}_i\boldsymbol{s}^* - \boldsymbol{s}\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(N)}^*} \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* \\
&= \boldsymbol{s}\boldsymbol{B}_i\boldsymbol{s}^* - \boldsymbol{s}\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* = 0, \quad \forall 1 \le i \le M.
\end{aligned}$$

In the other direction, note that $\boldsymbol{T}_i$ is a non-negative Hermitian matrix, and hence can be decomposed as $\boldsymbol{T}_i = \boldsymbol{Q}_i \boldsymbol{Q}_i^*$ for some $\boldsymbol{Q}_i$. Thus, $\boldsymbol{s}\boldsymbol{T}_i\boldsymbol{s}^* = 0$ if and only if $\boldsymbol{Q}_i^*\boldsymbol{s}^* = \boldsymbol{0}$. Multiplying both sides of the equation by $\boldsymbol{Q}_i$ yields

$$\boldsymbol{T}_i\boldsymbol{s}^* = \boldsymbol{B}_i^{\frac{1}{2}} \Big(\boldsymbol{I}_{MK} - \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(N)}^*} \Big) \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* = \boldsymbol{0},$$

and for invertible $\boldsymbol{B}_i$ it follows that $(\boldsymbol{I}_{MK} - \boldsymbol{P}_{\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(N)}^*}) \boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* = \boldsymbol{0}$, or, $\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{s}^* \in \mathcal{R}(\boldsymbol{B}_i^{\frac{1}{2}} \boldsymbol{S}_{(N)}^*)$, which is equivalent to $\boldsymbol{s}^* \in \mathcal{R}\big(\boldsymbol{S}_{(N)}^*\big)$. Hence, we proved that

$$\boldsymbol{s}^* \in \mathcal{R}\left(\boldsymbol{S}_{(K)}^*\right) \iff \boldsymbol{s}\boldsymbol{T}_i\boldsymbol{s}^* = 0, \ \forall 1 \le i \le M. \tag{39}$$

Since $\boldsymbol{T}_i$ is non-negative definite, $\boldsymbol{s}\boldsymbol{T}_i\boldsymbol{s}^* \ge 0$, for all $\boldsymbol{s}$. The previous connection then yields

$$\boldsymbol{s}^* \notin \mathcal{R}\left(\boldsymbol{S}_{(K)}^*\right) \iff \boldsymbol{s}\boldsymbol{T}_i\boldsymbol{s}^* > 0, \ \forall 1 \le i \le M.$$

As there is equality between the objectives and feasible sets of (38) and (37), both problems are equivalent. □

Our remaining task is now to solve the sum of quadratic ratios problem in (38). As mentioned before, a possible approach can be to use one of the algorithms [52]–[54], but

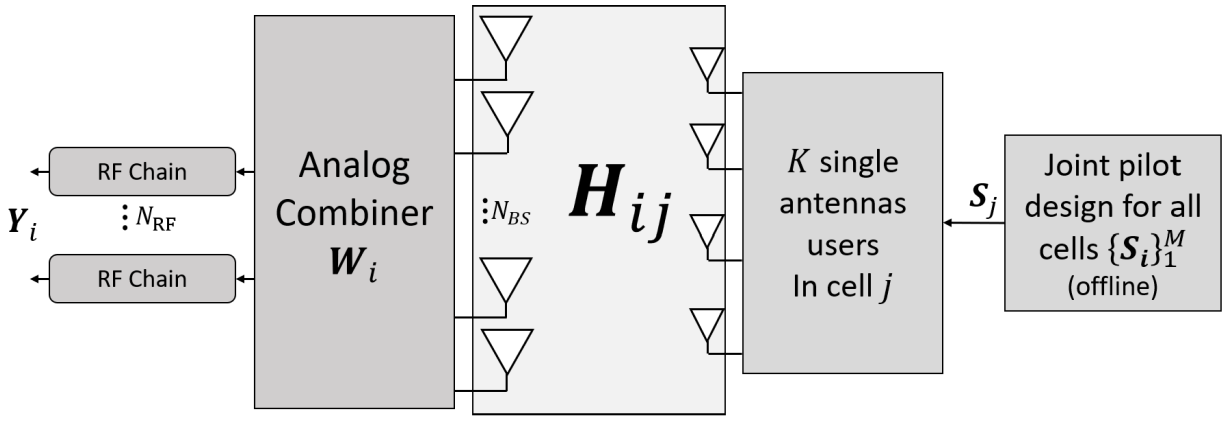

Fig. 1. Reduced RF scheme of the hybrid digital-analog receiver at the base station of the ith cell: the pilot signals from the K users in the jth cell are carried over the channel $\boldsymbol{H}_{ij}$ to the antennas of the ith base station, where they are mapped to the N_{RF} RF chains by the analog combiner $\boldsymbol{W}_i$ to produce the baseband signal $\boldsymbol{Y}_i$.

due to the large dimension of $\boldsymbol{s}$, these will result in very large runtime.

One simple method for obtaining a (sub-optimal) solution for (38) is searching over a selected dictionary, and calculating the objective value in (38) for each vector in the dictionary. The vector that corresponds to the largest objective value is then chosen. The computational load of this approach is reduced both compared to [52]–[54], and to applying similar dictionary solutions directly on (37), since here each element in the sum is a scalar ratio rather than a matrix ratio. The choice of dictionary is flexible. In the simulations detailed in Section VI, we compare different choices.

VI. Numerical Experiments

We now present some numerical results that both demonstrate the benefit of our joint pilot design framework and the effect of the RF reduction on the system's performance. First, in Section VI-A we consider the fully-separable correlations case with different combiner methods and show that the eigen-pilots of Section V-A enjoy lower MSE than other pilot allocation and design techniques. In Section VI-B we consider partially-separable correlations, compare the GSRTM of Algorithm 2 for pilot sequence design with other methods and show that in this case as well the joint design framework yields lower MSE.

Since the combiner design problem is identical for both correlation scenarios, the effect of RF reduction will be tested only under the fully-separable case. In the partially-separable experiments a full combiner, i.e. $N_{\text{RF}} = N_{\text{BS}}$ is considered.

We compare our proposed algorithms with four other approaches: the common reused orthogonal pilots, where at all the cells the same K orthogonal pilots are transmitted without any optimization over the allocation order, the random pilot design, where at each cell i the pilot matrix $\boldsymbol{S}_i$ is drawn randomly from i.i.d. zero-mean unit variance complex-normal distribution, the smart pilot assignment (SPA) method from [15], which is a near-optimal allocation algorithm when $\tau = K$, and the successive optimization approach (SOA) for joint pilot design from [30] which uses a convex relaxation to the design problem and successively minimizes the sum-MSEs over the cells. As mentioned before, pilot design can also be performed from a single-cell perspective as in [24]–[29]. To use the single-cell method for joint pilot sequence design a generalization is needed. One straightforward approach is to perform the design iteratively over the cells. In this case, in each iteration the objective function is different (since it is the MSE of the current cell). In practice we observed that a straightforward implementation did not converge.

Notice that in the reused orthogonal pilots, only K orthogonal sequences are used across all cells, regardless of the number of pilot symbols τ. In practice, when $\tau > K$, there are more than K possible orthogonal sequences, and the reuse ratio can be decreased. This scenario is similar to the random pilot case, where with high probability τ orthogonal pilots will be drawn. In all techniques, the pilots are normalized to comply with the power constraint $\mathcal{P} = 1$.

Throughout this section, the performance measure used is the sum of normalized MSEs, defined as $\bar{\epsilon} = \frac{1}{M}\sum_{i=1}^{M} \frac{\|\boldsymbol{h}_i - \hat{\boldsymbol{h}}_i\|^2}{\|\boldsymbol{h}_i\|^2}$.

A. Fully Separable Model Experiments

We begin by treating the fully-separable channel correlation model (3) and test the performance of the eigen-pilot method of Section V-A.

We test two analog combiners: A fully-digital combiner (17) with $\boldsymbol{T} = \boldsymbol{I}_{N_{\text{RF}}}$, and an analog combiner obtained using the GRTM algorithm proposed in [39]. GRTM is a greedy dictionary-based method, that at each step adds one RF chain to the system and chooses the optimal combiner vector to add to the previously selected ones from a feasible dictionary. We assume a fully-connected phase shifter network, namely, the feasible set is all matrices with unimodular entries. The choice of combiner defines the weights α_i in (14) that will be used for the pilot design.

The first experiment compares between the eigen-pilots, orthogonal pilot reuse, and random pilots methods, with fully-digital and GRTM based analog combiners. Here we use $M = 7$ cells, $N_{\text{BS}} = 20$ BS antennas, $K = 5$ UTs in each cell

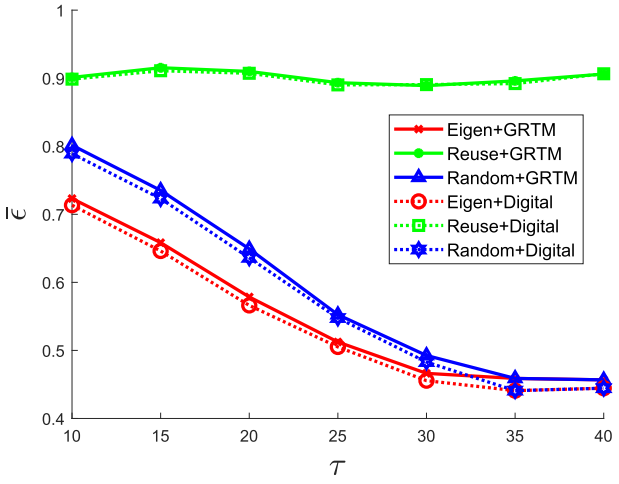

Fig. 2. Estimation error of different pilot design methods vs. number of pilot symbols τ, with GRTM and fully-digital combiners and low rank receive correlation with rank 10, $M = 7$ cells, $K = 5$ users in each cell, $N_{\mathrm{BS}} = 20$ antennas and $N_{\mathrm{RF}} = 5$.

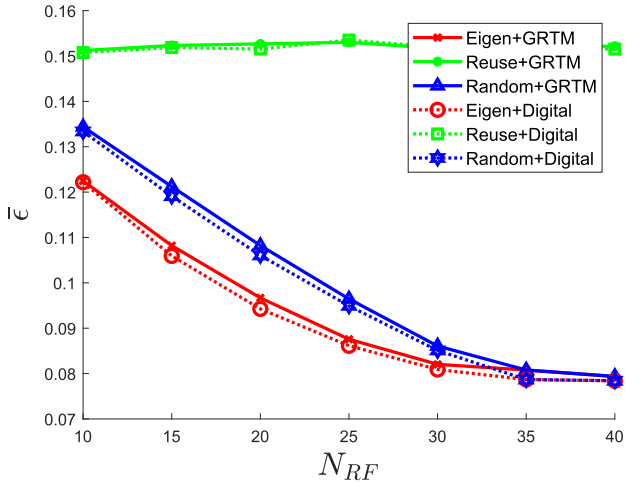

Fig. 3. Estimation error of different pilot design methods vs. number of RF chains N_{RF}, with GRTM and fully-digital combiners, $M = 7$ cells, $K = 10$ users per cell, $N_{\mathrm{BS}} = 128$ antennas, $\tau = 10$ pilot symbols and low-rank receive correlation structure.

and $N_{\mathrm{RF}} = 5$. We set the receive side correlation $\boldsymbol{Q}_i$ to be a random matrix given by $\boldsymbol{Q}_i = \boldsymbol{X}_i \boldsymbol{X}_i^*$, where $\{\boldsymbol{X}_i\}_{i=1}^M$ is a set of independent random matrices with i.i.d zero-mean unit variance complex-Gaussian entries. Then, to achieve low-rank correlation, typical for large arrays that are usually spatially correlated, we nullify all but its first 10 eigenvalues. The transmit side correlation $\boldsymbol{P}_j$ is a diagonal matrix with diagonal elements drawn uniformly from the interval $[0, 1]$. A new realization of $\boldsymbol{Q}_i, \boldsymbol{P}_j$ is generated for each of the 10000 Monte Carlo simulations.

Figure 2 depicts the sum of normalized estimation error $\bar{\epsilon}$ for different number of pilot symbols. While additional symbols improve the eigen-pilot performance, as it enables choosing more users, it does not affect the reused pilot performance. This is because when there is full pilot reuse between cells, additional symbols can only improve intra-cell interference, which in this case does not occur. The random pilot design improves with additional symbols, but falls short in comparison to the eigen-pilots, up to the point when full orthogonality is achieved in both techniques, i.e. $\tau = MK = 35$. As expected, for all methods the fully-digital combiner experiences slightly better performance than the analog GRTM.

We now investigate the effect of the number of RF chains on the estimation error. Here, we consider a more massive scenario, in which $N_{\mathrm{BS}} = 128$, $\tau = 10$ and $K = 10$. Again, we set the receive side correlation matrix to be low rank, by nulling all but its first 20 eigenvalues. This type of correlation is typical for massive arrays. Figure 3 depicts the sum of normalized estimation error $\bar{\epsilon}$ for different number of RF chains. In contrast to the previous experiment, the performance gap between the GRTM and fully-digital combiner is smaller. This is due to the low rank of the receive correlation matrix.

Next, we compare the eigen-pilots method with SPA, which is a near-optimal pilot assignment method when $K = \tau$ [15], and SOA, which is the successive joint pilot design method in [30]. The framework for these methods was developed assuming the MU-MIMO channel fading model (5). To comply with this model, we set

$$\boldsymbol{P}_j = \mathrm{diag}\left(\beta_{1j}, \cdots, \beta_{Mj}\right), \ \boldsymbol{Q}_i = \boldsymbol{I}_{N_{\mathrm{BS}}}$$

where $\beta_{i,j}$ are drawn uniformly from the interval $[0, 1]$. For this choice, the fully-separable model (3) coincides with the correlation model (5), where $\boldsymbol{D}_{ij} = \boldsymbol{P}_j$. The pilot sequences used for SPA are the first τ eigenvectors of $\boldsymbol{X}\boldsymbol{X}^*$, with $\boldsymbol{X}$ a random matrix with i.i.d complex-Normal entries with zero-mean and unit variance, that is drawn once every Monte-Carlo simulation.

Figure 4 depicts the eigen-pilots, SPA, SOA and random-pilots performance versus different numbers of RF chains N_{RF}, for a fully-digital combiner (17) and a full-receiver without RF reduction, i.e. $N_{\mathrm{RF}} = N_{\mathrm{BS}}$. Here we set $\tau = 5$, $K = 4$ and $N_{\mathrm{BS}} = 10$. Since $\boldsymbol{Q}_i = \boldsymbol{I}_{N_{\mathrm{BS}}}$, it follows from (17) that one possible optimal RF reduction solution is antenna selection, where all antennas are equally important, and every additional RF chain enables choosing one more antenna and improves the estimation error linearly. Figure 4 demonstrates the effect of the reduction on the different methods. It can be seen that the performance gap between the techniques grows larger as more RF chains are added to the system. For all N_{RF} values the eigen-pilots outperform the other approaches.

Next we set $N_{\mathrm{BS}} = N_{\mathrm{RF}} = 128$ and $K = 5$. The resulting $\bar{\epsilon}$ versus the number of pilot symbols τ is depicted in Fig. 5. Since SPA optimizes only the allocation of the pilots and not the sequences themselves, and is limited to using a number of sequences equal to the number of users, it does not improve with additional symbols. While the comparison may be unfair, it demonstrates the fact that pilots allocation is a sub-optimal approach in general. In contrast to SPA, the SOA method performs joint pilot design and gains from additional symbols. However, its MSE is higher than the eigen-pilots and random pilots.

Finally, we demonstrate the achievable rate for the different methods. Here, we assumed the joint decoding of all the users

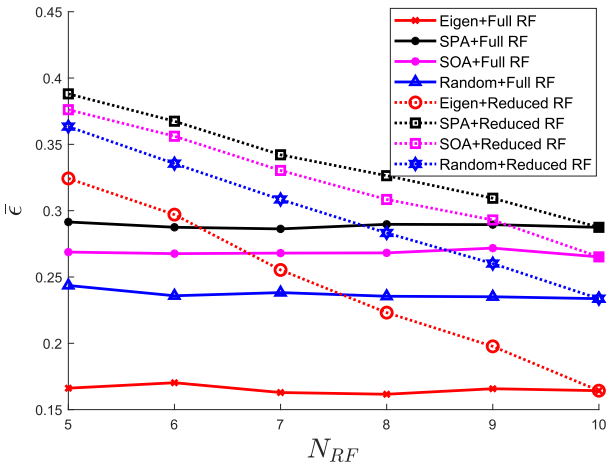

Fig. 4. Estimation error of different pilot design methods vs. number of RF chains N_{RF}, with reduced- and full-RF combiners, $N_{\mathrm{BS}} = 10$ antennas, $M = 3$ cells, $\tau = 5$ pilot symbols and $K = 4$ users in each cell.

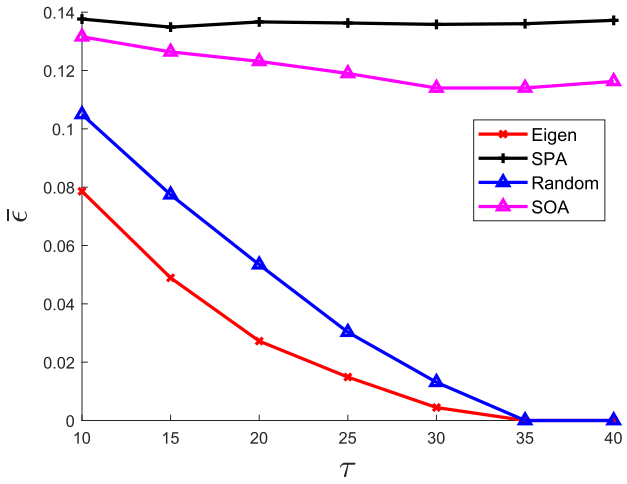

Fig. 5. Estimation error of different pilot design methods vs. number of pilot symbols τ, with fully-RF combiner, $N_{\mathrm{BS}} = 128$ antennas, $M = 7$ cells and $K = 5$ users in each cell.

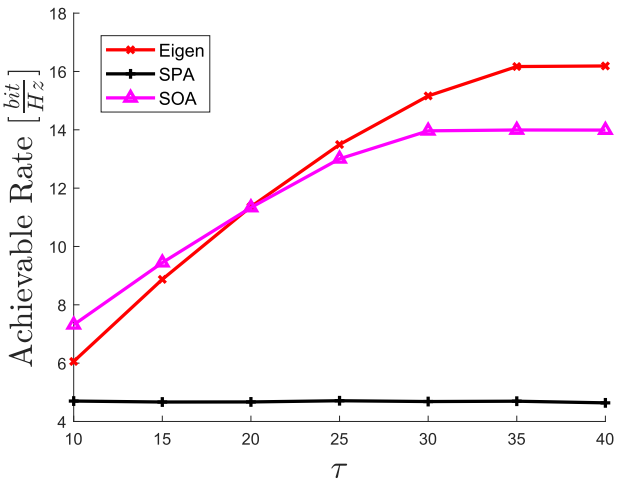

Fig. 6. Average achievable rate of a single user vs. number of pilot symbols τ, with fully-RF combiner, $N_{\mathrm{BS}} = 128$ antennas, $M = 7$ cells and $K = 5$ users in each cell.

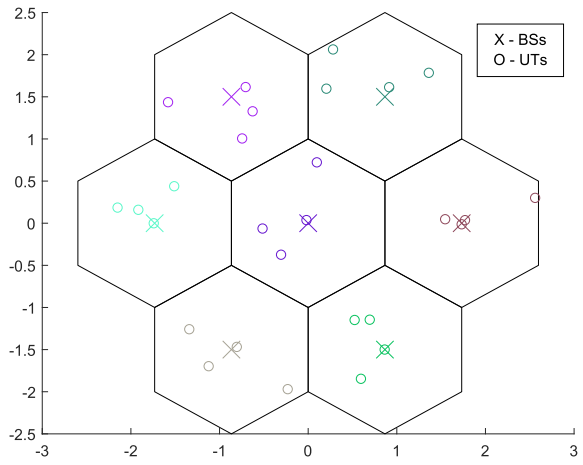

Fig. 7. MU-MIMO system with 7 hexagonal cells.

inside each cell. Figure 6 shows the average achievable rate across all users in the system for the eigen-pilots, SOA and SPA. As expected, the SPA does not improve with additional symbols. The SOA rate is better than the eigen-pilots when the number of pilot symbols is much smaller than the total number of users in the system. This is because the eigen-pilot solution nullifies most of the users. However, as the number of pilot symbols grows, fewer users are nullified and the achievable rate of the eigen-pilots is higher than the SOA's.

B. Partially Separable Model Experiments

For the partially-separable correlations case, we adopt the MU-MIMO channel fading model [5], in which β_{ikj}, $1 \leq i, j \leq M$, $1 \leq k \leq K$ is defined via

$$\beta_{ikj} = \frac{z_{ikj}}{r_{ikj}^{\gamma}} \tag{40}$$

with r_{ikj} representing the distance between the k-th UT in the j-th cell and the i-th BS, γ is the decay exponent, and z_{ikj} is a log-normal random variable, such that $10 \log z_{ikj}$ is zero-mean Gaussian with a standard deviation σ_s. Here we used $\gamma = 3$ and $\sigma_s = 8$[dB]. The diagonal transmit correlation matrices are given by $\boldsymbol{P}_{ij} = \mathrm{diag}\,(\beta_{i1j}, \cdots, \beta_{iMj})$. The MU-MIMO network consists of $M = 7$ hexagonal cells in which the UTs are distributed uniformly. A realization of such a network is illustrated in Fig. 7.

We compare the proposed GSRTM algorithm for pilot sequence design to random pilot design, SPA, SOA, and the sequence design method suggested in [25], which perform optimization from a single cell perspective; the pilots in each cell are designed to minimize the cell's MSE, while treating the other cells as interferers. To adjust this method for our scenario, we begin be setting $\boldsymbol{S}_1$, the pilots of the first cell, to K arbitrary orthogonal sequences. Then, the pilots $\boldsymbol{S}_2$ of the second cell are designed given $\boldsymbol{S}_1$ and $\boldsymbol{P}_{21}$, and so on for all M cells. Here we drop the orthogonal pilot reuse method, as it is inferior to SPA in the sum-MSEs sense. We use $N_{\mathrm{BS}} = 128$ antennas at each BS and $K = 10$ UTs in each cell. Since, as noted in Section V-B, the analog combiner design problem in the partially-separable correlation

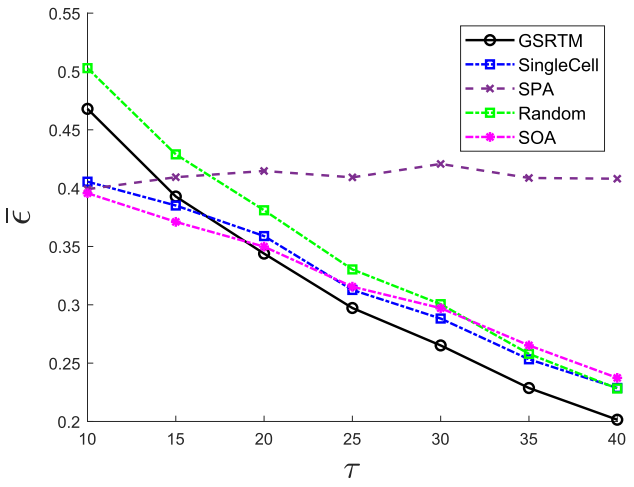

Fig. 8. Estimation error of different pilot design methods vs. number of pilot symbols τ, with $M = 7$ cells, $K = 10$ users per cell and $N_{\mathrm{BS}} = 128$ antennas.

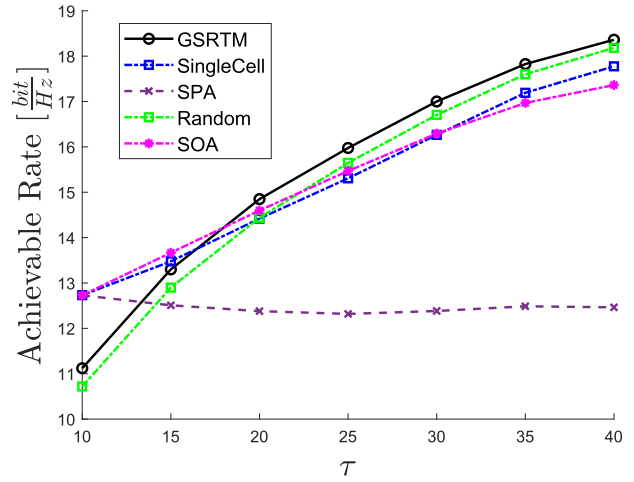

Fig. 9. Average achievable rate of a single user vs. number of pilot symbols τ, with $M = 7$ cells, $K = 10$ users per cell and $N_{\mathrm{BS}} = 128$ antennas.

profile is identical to the fully-separable correlations scenario, in the following we focus only on the pilot sequence design algorithms, and consider a system without RF chain reduction, i.e., $N_{\mathrm{RF}} = N_{\mathrm{BS}} = 128$.

To numerically evaluate the effect of the number of pilot symbols τ on the performance of the considered methods, we depict in Fig. 8 the estimation performance and in Fig 9 the average achievable rate for $\tau \in [10, 40]$. Observing Fig. 8 and 9, we note that while for $K = \tau$, SPA, SOA and the single cell optimization achieve better estimation accuracy and rate, as the number of pilot symbols increases, GSRTM results in lower sum-MSEs and higher average achievable rate. This is a result of; 1) the fact that SPA cannot exploit additional symbols and is limited to the allocation of K sequences, and 2) as the number of symbols increases, the gain in coordination between cells and joint optimization is more dominant than the loose GSRTM suffers from being limited to a dictionary in oppose to SPA and the single-cell optimization.

Next, we study the performance of GSRTM with different dictionaries. Here we considered three options: a general

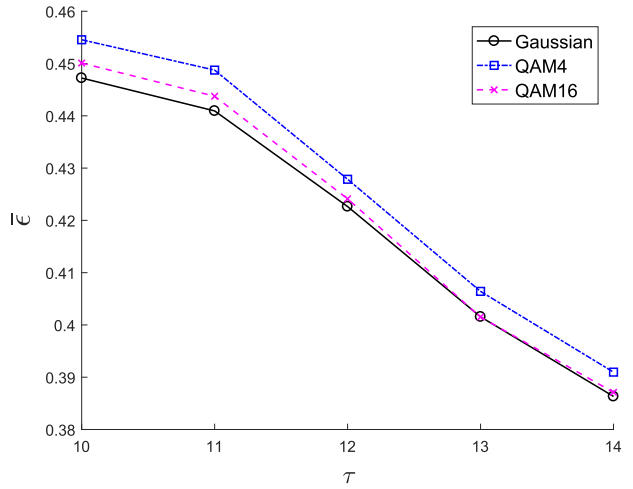

Fig. 10. Estimation error of GSRTM with different dictionaries vs. number of pilot symbols τ, with $M = 7$ cells, $K = 10$ users per cell and $N_{\mathrm{BS}} = 128$ antennas.

Gaussian dictionary, where the dictionary matrix has a i.i.d complex-Normal entries with zero-mean and unit variance, quadratic amplitude modulation (QAM) with 4 constellation points and QAM with 16 constellation points. All dictionaries are composed of $Q = 300$ different possible sequences. That is, the difference between the dictionaries is not the number of possible sequences, but the flexibility of symbol values. As expected, the complex-Normal dictionary results in the lowest MSE, as it offers more flexibility and diversity. The QAM16 dictionary performance is not much lower than the Gaussian one, as its flexibility is quite high due to a large number of constellation points. Yet, it is much simpler to implement in practice. As expected, QAM4 has the highest MSE as it is the most limited. However, it requires fewer representation bits for each symbol.

VII. Conclusion

We treated the problem of joint pilot sequence and analog combiner design in massive MIMO systems with a reduced number of RF chains. We considered two channel correlation profiles: fully- and partially-separable correlations. We showed that for the considered setup, the RF reduction problem can be solved separately from the pilot sequences design and that previously suggested reduction methods fit this framework. For the fully-separable case, we derived an optimal pilot design solution. We demonstrated that in the special case when the transmit correlation matrices are diagonal, the solution corresponds to a user selection method, where the users are chosen according to their long-term statistics. For the partially-separable correlations scenario, we suggested a greedy method for pilot design which solves a sum-of-quadratic-ratios problem. At each step, it allows for a single pilot symbol to be added. Numerical examples demonstrated the substantial benefits of our proposed designs compared to existing methods that do not jointly optimize the pilots.

Acknowledgement

The authors would like to thank Dr. N. Shlezinger for his valuable input.

REFERENCES

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

[2] 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.

[3] E. Bjornson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[4] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. 14th Asilomar Conf. Signals Syst. Comput. (ACSSC)*, 2006, pp. 359–363.

[5] 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.

[6] J. Hoydis, S. Ten 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] 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.

[8] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1137–1141.

[9] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[10] R. R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[11] 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.

[12] Z. Chen and C. Yang, "Pilot decontamination in wideband massive MIMO systems by exploiting channel sparsity," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 5087–5100, Jul. 2016.

[13] L. Su and C. Yang, "Fractional frequency reuse aided pilot decontamination for massive MIMO systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–6.

[14] X. Yan, H. Yin, M. Xia, and G. Wei, "Pilot sequences allocation in TDD massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1488–1493.

[15] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1644–1647, Sep. 2015.

[16] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[17] 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.

[18] V. Saxena, G. Fodor, and E. Karipidis, "Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–6.

[19] M. N. Khormuji, "Pilot-decontamination in massive MIMO systems via network pilot-data alignment," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2016, pp. 93–97.

[20] H. Ahmadi, A. Farhang, N. Marchetti, and A. MacKenzie, "A game theoretic approach for pilot contamination avoidance in massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 12–15, Feb. 2016.

[21] M. Rasti and A. R. Sharafat, "Distributed uplink power control with soft removal for wireless networks," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 833–843, Mar. 2011.

[22] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, Dec. 2017, Art. no. 117.

[23] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," 2016, *arXiv:1612.02574*. [Online]. Available: http://arxiv.org/abs/1612.02574

[24] J. Pang, J. Li, L. Zhao, and Z. Lu, "Optimal training sequences for MIMO channel estimation with spatial correlation," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 651–655.

[25] Y. Liu, T. F. Wong, and W. W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.

[26] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[27] 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.

[28] T. E. Bogale and L. B. Le, "Pilot optimization and channel estimation for multiuser massive MIMO systems," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.

[29] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[30] H. Al-Salihi, T. Van Chien, T. A. Le, and M. R. Nakhai, "A successive optimization approach to pilot design for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1086–1089, May 2018.

[31] T. Van Chien, E. Bjornson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.

[32] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[33] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[34] N. Li, Z. Wei, H. Yang, X. Zhang, and D. Yang, "Hybrid precoding for mmWave massive MIMO systems with partially connected structure," *IEEE Access*, vol. 5, pp. 15142–15151, 2017.

[35] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[36] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[37] X. Yu, J. Zhang, and K. B. Letaief, "Partially-connected hybrid precoding in mm-Wave systems with dynamic phase shifter networks," 2017, *arXiv:1705.00859*. [Online]. Available: http://arxiv.org/abs/1705.00859

[38] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714–2720, Jun. 2015.

[39] S. Stein Ioushua and Y. C. Eldar, "Hybrid analog-digital beamforming for massive MIMO systems," 2017, *arXiv:1712.03485*. [Online]. Available: http://arxiv.org/abs/1712.03485

[40] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[41] H. Quoc Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," 2011, *arXiv:1112.3810*. [Online]. Available: http://arxiv.org/abs/1112.3810

[42] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[44] S. S. Ioushua and Y. C. Eldar, "Pilot contamination mitigation with reduced RF chains," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[45] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Springer, 2011.

[46] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[47] H. Shirani-Mehr, H. Papadopoulos, S. A. Ramprashad, and G. Caire, "Joint scheduling and ARQ for MU-MIMO downlink in the presence of inter-cell interference," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 578–589, Feb. 2011.

[48] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 461–464, Jun. 2009.

[49] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.
[50] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.
[51] A. R. Abhyankar, S. A. Soman, and S. A. Khaparde, "Min-max fairness criteria for transmission fixed cost allocation," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2094–2104, Nov. 2007.
[52] P. Shen, Y. Chen, and Y. Ma, "Solving sum of quadratic ratios fractional programs via monotonic function," *Appl. Math. Comput.*, vol. 212, no. 1, pp. 234–244, Jun. 2009.
[53] P. Shen, W. Li, and X. Bai, "Maximizing for the sum of ratios of two convex functions over a convex set," *Comput. Oper. Res.*, vol. 40, no. 10, pp. 2301–2307, Oct. 2013.
[54] M. Jaberipour and E. Khorram, "Solving the sum-of-ratios problems by a harmony search algorithm," *J. Comput. Appl. Math.*, vol. 234, no. 3, pp. 733–742, Jun. 2010.
[55] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Tech. Univ. Denmark, Lyngby, Denmark, Tech. Rep., Nov. 2012. [Online]. Available: http://localhost/pubdb/p.php?3274

Shahar Stein Ioushua (Student Member, IEEE) received the B.Sc. degree *(cum laude)* in electrical engineering and the M.Sc. degree *(cum laude)* in electrical engineering from the Technion–Israel Institute of Technology, Haifa, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree in electrical engineering with Tel-Aviv University (TAU), Tel-Aviv, Israel.

From 2015 to 2020, she has been a Teaching Assistance with the Viterbi Faculty of Electrical Engineering and the Iby and Aladar Fleischman Faculty of Engineering at Tel Aviv University, and a Project Supervisor with the Signal Acquisition, Modeling and Processing Laboratory (SAMPL), Electrical Engineering Department, Technion. She is a Co-Founder of the WomEng forum of women in engineering at Tel Aviv university and the WomEE forum for women in graduate studies at the EE Department, Technion. Her research interests include information theory, data compression, theoretical aspects of signal processing, compressed sensing, and signal processing for communication signals.

Mrs. Stein Ioushua is a Gutwirth Fellow since 2017. She received the Weinstein for excellent research and the Meyer Foundation Excellence prize.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

She was a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She was a Visiting Professor with Stanford. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology and optics.

Dr. Eldar was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is a member of the Israel Academy of Sciences and Humanities (elected 2017) and a EURASIP Fellow. She has received numerous awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award in 2013, the IEEE/AESS Fred Nathanson Memorial Radar Award in 2014, and the IEEE Kiyo Tomiyasu Award in 2016. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (twice). She received several best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel. She was the co-chair and technical co-chair of several international conferences and workshops. She is the Editor-in-Chief of Foundations and Trends in Signal Processing, a member of the IEEE Sensor Array and Multichannel Technical Committee, and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, a member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing, SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*.