IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 23, 2018, accepted February 15, 2018, date of publication February 27, 2018, date of current version March 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810227

Near-Optimal Pilot Signal Design for FDD Massive MIMO System: An Energy-Efficient Perspective

YI WANG[1,2,3], PENGGE MA[1,2], RUI ZHAO[4], (Member, IEEE), YONGMING HUANG[5], (Senior Member, IEEE), CHUNGUO LI[5], (Senior Member, IEEE), JUN ZHU[6], (Student Member, IEEE), KAIZHI HUANG[3], AND BING WANG[3]

[1]School of Electronics and Communication Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450046, China
[2]Collaborative Innovation Center for Aviation Economy Development of Henan Province, Zhengzhou 450046, China
[3]National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China
[4]Xiamen Key Laboratory of Mobile Multimedia Communications, Huaqiao University, Xiamen 361021, China
[5]School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[6]Qualcomm Inc., San Diego, CA 92121, USA

Corresponding author: Yi Wang (yiwang@zua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61720106003, Grant 61401165, Grant 61379006, Grant 61671144, and Grant 61701538, in part by the Natural Science Foundation of Fujian Province under Grants 2015J01262, in part by Science and Technology Innovation Teams of Henan Province for Colleges and Universities (17IRTSTHN014), in part by the Scientific and Technological Key Project of Henan Province under Grant 172102210080 and Grant 182102210449, in part by Natural Science Research Project of Jiangsu Province for Colleges and Universities under Grant 16KJB51000, and in part by the Collaborative Innovation Center for Aviation Economy Development of Henan Province.

ABSTRACT This paper investigates the pilot signal design for a massive multiple-input multiple-out (MIMO) frequency division duplexing downlink system by taking the channel spatial correlation into account. Our objective is to optimize the pilot pattern, including the pilot power and structure, from the perspective of energy-efficient communication, which can evaluate the channel estimation accuracy, spectral efficiency, and power consumption simultaneously. The original problem is established based on maximizing the energy efficiency (EE) with a predefined quality-of-service requirement and the total power budget, where the involved cost function is in a nonanalytic form with the non-convex nature. To solve it, an analytical expression is derived first by the virtue of the deterministic equivalent approximation technology, which is shown to be a tight approximation. Based on this, the structure of the EE maximization-based pilot signal matrix is proved to be column orthogonal, where the column corresponds to the dominant eigen-directions of channel spatial correlation matrix. By use of the derived pilot structure, the primal non-convex fractional optimization problem is recast to an equivalent optimization problem with the objective function in a subtractive form, which can be solved by deliberately manipulating the Lagrangian function. Finally, an iterative algorithm is proposed to obtain the closed-form solution and further insights are extracted. Numerical results validate the performance gain of our proposed near-optimal pilot scheme in terms of the energy efficiency and spectral efficiency compared with the classical mean squared error-based pilot scheme.

INDEX TERMS Massive MIMO, FDD, pilot design, energy efficiency.

I. INTRODUCTION

Massive MIMO has been widely considered as one of the most promising candidate technologies for the 5th mobile communication systems (5G), which has drawn extensive research attention in the wireless communication fields [1]–[3]. By equipping hundreds or more antennas at base station (BS) to provide service for a much smaller number of users simultaneously, massive MIMO exhibits numerous noteworthy merits in the sense of spectral efficiency, energy efficiency, spatial resolution, inter-user interference elimination as well as secure transmission and so on [4]–[7]. What's more, the potential benefits can be realized with low-complexity coherent processing [8].

Note that to fully reap the remarkable gains of massive MIMO, the knowledge of channel state information (CSI) at the BS is a prerequisite [8]. The CSI acquisition has also been recognized as a very challenging task for massive MIMO systems, due to the high dimensionality of channel matrices as well as the training and computation overhead. In practice, the pilot-aided channel estimation approach is commonly

utilized for CSI acquisition in massive MIMO systems [3]. Moreover, in view of the two main duplexing modes in the existing cellular systems, i.e., time division duplexing (TDD) and frequency division duplexing (FDD), the corresponding pilot-based CSI acquisition methods work in different mechanisms, which lead to different training overhead and estimation accuracy.

A large body of the studies on massive MIMO focus on TDD mode [9]–[14], where the downlink CSI can be obtained at the BS by exploiting the channel reciprocity and using uplink pilots sent from the users. As a sequence, the required length of orthogonal pilot sequence is proportional to the number of users, which results in lower training overhead than that in FDD mode. On the other hand, the problem of CSI acquisition in FDD mode becomes more intractable because the channel reciprocity is unavailable. In this case, the BS must first broadcast downlink pilot signals, and then the users send the estimated CSI to the BS through the feedback channel, which means that the downlink training and uplink feedback overhead scale up with the number of BS antennas. Although TDD system has the superiority of training overhead, the inherent imperfections in TDD mode, e.g., pilot contamination [9], calibration error [15], and hardware impairments [16], seriously limit the system performance. Besides these, since FDD mode is still dominant in the currently deployed cellular system [17] and more applicable to the symmetric traffic and delay-sensitive applications [18], it is therefore of great meaning to explore the challenging problems for CSI acquisition in massive MIMO FDD systems. Thus, we pay attention to the potential channel training issues under the FDD massive MIMO framework in this paper.

Even though the CSI acquisition is faced with more difficulties in massive MIMO FDD system, some explorative studies have been carried out in the ares of downlink precoding design, channel estimation and training, as well as the feedback scheme [19]–[25], which all aimed to lower down the training and/or feedback overhead. Nam *et al.* [19] proposed a two-tier precoding scheme by grouping the users with approximately the same channel covariance matrix, which reduced the dimensionality of the original physical channel and accordingly cut down the training overhead. In [20], two novel types of angle-of-arrival (AoA) based beamforming schemes were offered, wherein the required training overhead for CSI acquisition only depends on the number of served users. Duly *et al.* [21] introduced a new misalignment cost function to optimize the training sequences without orthogonality constraint. Choi *et al.* [22] developed a successive channel prediction/estimation strategy at the user side with only small length of training signals by leveraging time correlation of the channels. In [23], the achievable rates were optimized by heuristically designing the downlink training sequences and feedback codebooks. Xie *et al.* [24] summarized the current low-rank channel estimation approaches with the sparse properties of channel environments and then, they built a spatial basis expansion channel model (SBEM) with far fewer parameter dimensions such that both the uplink (for TDD) and downlink (for FDD) channel estimation can be carried out with a small amount of training resource [25]. Apart from these, Dai *et al.* provided a class of compressive sensing (CS) based algorithms to reduce the overhead of channel estimation estimation and feedback by exploiting the sparsity of wireless channels, such as the sparsity of time-domain channel impulse response [26], the spatially common sparsity [27], and the joint spatio-temporally sparsity [28].

It is worth mentioning that the vast majority of aforementioned works only concentrated on the channel estimation accuracy with low overhead and considered the conventional criterion, such as mean squared error (MSE), to design the pilot signal for channel estimation. However, the designed pilot signal not only affects the channel estimation accuracy as well as the achievable rate but also consumes the system energy load. Thus, from the perspective of Green communications [29], [30], the energy efficiency metric is more comprehensive and suitable for the pilot signal designs, which has attracted more and more interest for the system analysis and design. Unfortunately, the energy efficiency based pilot signal design, including the pilot power and structure, is still an open topic, as the cost function capturing energy efficiency is in general non-convex with respect to (w.r.t.) pilot signal matrix and thus is hard to tackle.

To the best of our knowledge, by far few literatures put forth the energy-efficient pilot signal design for massive MIMO FDD systems, among which our previous work [31] proposed an energy-efficient resource allocation scheme between the training phase and the downlink data transmission phase when considering a limited channel coherent interval and a total power budget. However, it only focused on the joint optimization of system parameters, i.e., the total pilot signal power, the data power and the training duration, wherein the conventional MSE-based orthogonal pilot sequence was adopted for simplicity without an in-depth discussion for the pilot signal structure optimization. It is worth noting that although the energy-efficient pilot transmission scheme has been studied in other cellular configurations, like in multiuser TDD MIMO system [32] and multiuser OFDM system [33], such results are not directly applicable to massive MIMO FDD systems. In addition, the orthogonal pilot sequence is a pre-assumption in [32] and [33]. However, in this work we relax the orthogonal restriction on the pilot signal and try to investigate the optimal pilot signal design from the EE point of view, which means a more general case.

Inspired by this, in this paper we aim to use the energy efficiency as a performance metric for the design of pilot pattern in massive MIMO FDD downlink system while meets the target quality-of-service constraint and total power budget. Since the complicated objective function and the non-convex nature of the original problem, obtaining the optimal solution is intractable. Thus, we first derive an analytical expression to approximate the energy efficiency function, based on which, the structure of the near-optimal pilot signal is proved. Then, by utilizing the fractional programming, the fractional non-convex optimization problem is transformed into an

equivalent optimization problem with a subtractive-form objective function, which can be handled by the Lagrangian multiplier method. Accordingly, an iterative algorithm is developed and a closed-form solution for the near-optimal pilot signal is deduced. Finally, numerical results are given to validate the effectiveness and superiority of our proposed pilot pattern.

The rest of this paper is organized as follows. In Section II, the involved massive MIMO FDD system is introduced and the EE-based pilot signal design problem is put forward. Section III derives the EE-based pilot signal pattern by using the majorization theory and the deterministic equivalent approximation technology, based on which, an iterative algorithm is proposed to achieve the near-optimal pilot solution. In Section IV, the numerical simulations are provided to verify the performances of our proposed pilot pattern, including the comparison with the traditional MSE-based pilot stratagem. Section V concludes our work.

Notations—We use the boldface uppercase, lowercase letters and italic letters to represent matrices, vectors and scalars, respectively. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$, and $\mathrm{tr}(\cdot)$ denote the transpose, conjugate transpose, inverse, and trace of a matrix, respectively. $\mathbf{A}(i,j)$ refers to the i-th row and j-th column element of matrix $\mathbf{A}$. $\mathbb{E}\{\cdot\}$ is the statical expected value of random variable x and $\mathbb{E}\{x|\,y\}$ means the conditional expected value of random variable x on a realization of the random variable y. $\mathrm{diag}\{\mathbf{x}\}$ denotes a diagonal matrix with the vector $\mathbf{x}$ on its principal diagonal. The notation $|\cdot|$ and $\|\cdot\|$ denote the absolute value and the Frobenius norm, respectively. The operator $[\cdot]^+$ means $\max\{0,\cdot\}$. $\mathcal{CN}(\mathbf{m},\mathbf{\Omega})$ stands for the circular symmetric complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Omega}$. $\mathbf{I}_N$ is the identity matrix of size $N\times N$. We also use $\mathrm{Span}(\mathbf{U})$ to denote the linear subspace spanned by the columns of $\mathbf{U}$ and $\mathrm{Span}^{\perp}(\mathbf{U})$ for its orthogonal complement. $\xrightarrow{\text{a.s.}}$ means almost sure convergence.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a massive MIMO FDD downlink system over flat Rayleigh fading channels, which consists of a BS with N antennas (N is large) and a single-antenna user as illustrated in Fig. 1. The downlink transmission includes three stages, i.e. pilot-based donwlink channel estimation, CSI feedback and data transmission. Assume that the block-fading channel has a coherence interval of T_c (in symbols), which implies that channel coefficients keep quasi static during a coherence time block.

In each coherent interval, we assume that the BS employs the first L ($<T_c$) symbols to send the pilot signal on its N antennas, which can be denoted as a $N\times L$ matrix $\mathbf{X}$. Furthermore, the total power assigned to the pilot sequence satisfies the constraint $\mathrm{Tr}\left(\mathbf{X}\mathbf{X}^{\mathrm{H}}\right)\le P$. Thus, the received $L\times 1$ pilot vector at the user can be expressed as

$$\mathbf{y}_{\mathrm{p}}=\mathbf{X}^{\mathrm{H}}\mathbf{h}+\mathbf{n}_{\mathrm{p}} \tag{1}$$

where $\mathbf{n}_{\mathrm{p}}\in\mathbb{C}^{L\times 1}\sim\mathcal{CN}(\mathbf{0},\sigma_{\mathrm{p}}^2\mathbf{I}_L)$ is the additive white Gaussian noise (AWGN) vector and $\mathbf{h}\in\mathbb{C}^{N\times 1}\sim\mathcal{CN}(\mathbf{0},\mathbf{R})$ denotes the downlink channel with $\mathbf{R}=\mathbb{E}\{\mathbf{h}\mathbf{h}^{\mathrm{H}}\}$ being the positive semidefinite spatial correlation matrix [34]. Herein, the transmit-side antenna correlation at the BS usually results from the little antenna spacing and insufficient near-field scattering conditions. We assume that $\mathrm{tr}\mathbf{R}=N$ denotes the normalized average channel gain and $\mathbf{R}$ is known to the system [34].[1] To proceed, we introduce the eigenvalue decomposition of $\mathbf{R}$ as follows, which is necessary for the subsequent analysis in this paper.

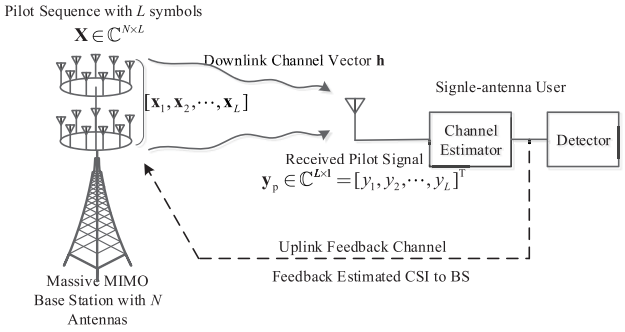


FIGURE 1. Massive MIMO FDD Downlink System.

$$\mathbf{R}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}} \tag{2}$$

where $\mathbf{U}$ is a $N\times K$ unitary matrix with orthonormal columns, which correspond to the eigenvectors of $\mathbf{R}$, and $\mathbf{\Lambda}=\mathrm{diag}\{[\lambda_1,\lambda_2,\ldots\lambda_K]\}$ is a $K\times K$ diagonal matrix with all the non-zero eigenvalues of $\mathbf{R}$ on its principal diagonal, where the eigenvalues is in descending order. It is clear that the rank of $\mathbf{R}$ is K ($\le N$), which implies that $\mathbf{R}$ might be rank-deficient. As $\mathbf{h}$ is included in the subspace $\mathrm{Span}(\mathbf{U})$, the channel vector can be modeled as an equivalent form

$$\mathbf{h}=\mathbf{U}\mathbf{g} \tag{3}$$

where $\mathbf{g}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Lambda})$ can be viewed as the equivalent and effective channel vector with reduced dimensionality.

Thanks to $\mathbf{h}\in\mathrm{Span}(\mathbf{U})$, we still make use of the following substitution

$$\mathbf{X}=\mathbf{U}\widetilde{\mathbf{X}} \tag{4}$$

which will not incur any performance degradation since the pilot signal power assigned to $\mathrm{Span}^{\perp}(\mathbf{U})$ will be directly filtered without affecting the received signal. Thus, we use $\widetilde{\mathbf{X}}\in\mathbb{C}^{K\times L}$ as the equivalent pilot signal instead of $\mathbf{X}$ hereafter, unless otherwise notified. Kindly note that the column orthogonality of the pilot matrix $\mathbf{X}$ is not emphasized so far, which is different from the traditional assumption on the unitary training with equal power allocation as in [22] and [31]. By substituting (3) and (4) into (1), the received pilot signal can be redescribed as

$$\mathbf{y}_{\mathrm{p}}=\left(\mathbf{U}\widetilde{\mathbf{X}}\right)^{\mathrm{H}}\mathbf{U}\mathbf{g}+\mathbf{n}_{\mathrm{p}}=\widetilde{\mathbf{X}}^{\mathrm{H}}\mathbf{g}+\mathbf{n}_{\mathrm{p}} \tag{5}$$

[1] In realistic situations, the channel covariance matrix changes slowly w.r.t. the coherence time of the instantaneous channel matrix, which can be estimated and updated with low training overhead. Thus, it is reasonable to assume the channel covariance matrix known at the BS and the user.

where $\widetilde{\mathbf{X}}$ meets the power constraint $\mathrm{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right) = \mathrm{Tr}\left(\mathbf{X}\mathbf{X}^{\mathrm{H}}\right) \leq P$. From (5), it can be seen that at most K unknowns in $\mathbf{g}$ are needed to be estimated. Thus, for ease of analysis, we assume that the length of pilot sequence is not more than the channel rank as in typical MIMO systems [35], namely, $L \leq K$.[2]

With the observed signal $\mathbf{y}_{\mathrm{p}}$ in (5), the minimum mean squared error (MMSE) estimator is adopted at the user for channel estimation. According to the statistical signal processing methodology in [10], [11], and [36], the MMSE estimate $\widehat{\mathbf{g}}$ of $\mathbf{g}$ can be obtained

$$\widehat{\mathbf{g}} = \mathbb{E}\left\{\mathbf{g}|\,\mathbf{y}_{\mathrm{p}}\right\} = \mathbf{\Lambda}\widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}^{\mathrm{H}}\mathbf{\Lambda}\widetilde{\mathbf{X}} + \sigma_{\mathrm{p}}^2\mathbf{I}_L\right)^{-1}\mathbf{y}_{\mathrm{p}} \tag{6}$$

which is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{\Phi})$, where

$$\Phi = \mathbb{E}\left\{\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}\right\} = \Lambda\widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}^{\mathrm{H}}\Lambda\widetilde{\mathbf{X}} + \sigma_{\mathrm{p}}^2\mathbf{I}_L\right)^{-1}\widetilde{\mathbf{X}}^{\mathrm{H}}\Lambda \tag{7}$$

By virtue of the orthogonally property of MMSE estimation [10], [11], [36], $\mathbf{g}$ has the following decomposition

$$\mathbf{g} = \widehat{\mathbf{g}} + \widetilde{\mathbf{g}} \tag{8}$$

where $\widetilde{\mathbf{g}} \sim \mathcal{CN}(\mathbf{0}, \Omega)$ is the channel estimation error with Ω given by

$$\Omega = \mathbb{E}\left\{\widetilde{\mathbf{g}}\widetilde{\mathbf{g}}^{\mathrm{H}}\right\} = \Lambda - \mathbf{\Phi} \tag{9}$$

What's more, $\widetilde{\mathbf{g}}$ and $\widehat{\mathbf{g}}$ are mutually independent [10]. Therefore, the normalized mean squared error (MSE) of the channel estimation is given by

$$\begin{aligned}\mathrm{MSE}\left(\widetilde{\mathbf{X}}\right) &= \frac{\mathbb{E}\left\{\|\mathbf{g} - \widehat{\mathbf{g}}\|^2\right\}}{\mathbb{E}\left\{\|\mathbf{g}\|^2\right\}} \\ &= \frac{\mathrm{tr}\left\{\Lambda - \Lambda\widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}^{\mathrm{H}}\Lambda\widetilde{\mathbf{X}} + \sigma_{\mathrm{p}}^2\mathbf{I}_L\right)^{-1}\widetilde{\mathbf{X}}^{\mathrm{H}}\Lambda\right\}}{N}\end{aligned} \tag{10}$$

which is a function w.r.t. the pilot matrix $\widetilde{\mathbf{X}}$.

After the downlink channel vector is estimated at the user, the estimated CSI $\widehat{\mathbf{g}}$ will be sent to the BS through the feedback channel. For the analytic simplicity, we assume that the SNR of the feedback channel is high enough such that the feedback error is negligible compared with the channel estimation error [34], [37]. Thus, the ideal no-delay and error-free feedback channel condition is considered here as in [31], [34], and [37] such that we concentrate on the influence of the downlink training and data beamforming.[3]

When the BS obtains the downlink CSI, it performs beamforming scheme for the data transmission during the remainder $T_{\mathrm{c}} - L$ symbols in a coherent interval. Then, the beamformed data signal can be expressed as

$$\mathbf{d} = \mathbf{v}s = \mathbf{U}\widetilde{\mathbf{v}}s \tag{11}$$

[2]Kindly note that there has no effect on the subsequent analytical conclusions when $L > K$ is assumed.

[3]The study of feedback scheme with low overhead is another interesting and nontrivial topic in massive MIMO FDD system, which is beyond the scope of our research. Interested readers can refer to [38] and references therein for in-depth discussion.

where s is useful data symbol with power normalization $\mathbb{E}\left\{|s|^2\right\} = 1$, $\mathbf{v}$ is the beamforming vector with $\|\mathbf{v}\|^2 = 1$ to normalize the transmit power. Once again, we adopt the substitution $\mathbf{v} = \mathbf{U}\widetilde{\mathbf{v}}$ without any loss of performance owing to $\mathbf{h} \in \mathrm{Span}\,(\mathbf{U})$.

Finally, the received data signal at the user can be described as

$$y_{\mathrm{d}} = \sqrt{\rho_{\mathrm{d}}}\mathbf{h}^{\mathrm{H}}\mathbf{d} + n_{\mathrm{d}} \overset{(a)}{=} \sqrt{\rho_{\mathrm{d}}}\widehat{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}s + \sqrt{\rho_{\mathrm{d}}}\widetilde{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}s + n_{\mathrm{d}} \tag{12}$$

where ρ_{d} is the transmit power for the useful data, $n_{\mathrm{d}} \sim \mathcal{CN}(0, \sigma_{\mathrm{d}}^2)$ is the AWGN, (a) comes from the substitution of (3), (8), and (11).

Using a standard bound based on the worst-case uncorrelated additive noise theory [10], the average spectral efficiency (SE) can be obtained as follows

$$R = \mathbb{E}\left\{\log_2\left(1 + \gamma\right)\right\} \tag{13}$$

where γ is the equivalent received signal-to-noise ratio (SNR) and given by

$$\gamma = \frac{\rho_{\mathrm{d}}\widehat{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{H}}\widehat{\mathbf{g}}}{\rho_{\mathrm{d}}\widetilde{\mathbf{v}}^{\mathrm{H}}\Omega\widetilde{\mathbf{v}} + \sigma_{\mathrm{d}}^2} = \frac{\widetilde{\mathbf{v}}^{\mathrm{H}}\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}}{\widetilde{\mathbf{v}}^{\mathrm{H}}\left(\Omega + \delta^{-1}\mathbf{I}_K\right)\widetilde{\mathbf{v}}} \tag{14}$$

where $\delta = \frac{\rho_{\mathrm{d}}}{\sigma_{\mathrm{d}}^2}$ denotes the transmit SNR for data symbol.

Different from the commonly-used beamforming scheme in typical massive MIMO system, i.e., maximum-ratio transmission (MRT) and zero-forcing (ZF), the optimal beamforming vector that maximizes the received SNR is employed here, i.e., $\widetilde{\mathbf{v}}_{\mathrm{opt}} = \arg\max_{\widetilde{\mathbf{v}}} \frac{\widetilde{\mathbf{v}}^{\mathrm{H}}\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}}{\widetilde{\mathbf{v}}^{\mathrm{H}}(\Omega+\delta^{-1}\mathbf{I}_K)\widetilde{\mathbf{v}}}$. Since $\left(\Omega + \delta^{-1}\mathbf{I}_K\right)$ is a positive definite matrix and $\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}$ is a Hermitian matrix, the involved optimal beamforming design is a generalized eigenvalue problem, also known as generalized Rayleigh quotient.

Generally speaking, it is quite difficult to achieve a closed-form solution to a generalized eigenvalue problem. Fortunately, the rank of $\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}$ in the numerator of the right-hand side (RHS) of (14) is equal to 1, which makes us easily solve the problem in this case. By using the substitution $\widetilde{\mathbf{a}} \triangleq \left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{1/2}\widetilde{\mathbf{v}}$, an equivalent problem can be obtained in the form of standard Rayleigh-Ritz ratio, i.e., $\max_{\widetilde{\mathbf{v}}} \frac{\widetilde{\mathbf{v}}^{\mathrm{H}}\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}\widetilde{\mathbf{v}}}{\widetilde{\mathbf{v}}^{\mathrm{H}}(\Omega+\delta^{-1}\mathbf{I}_K)\widetilde{\mathbf{v}}} = \max_{\widetilde{\mathbf{a}}} \frac{\widetilde{\mathbf{a}}^{\mathrm{H}}\mathbf{W}\widetilde{\mathbf{a}}}{\widetilde{\mathbf{a}}^{\mathrm{H}}\widetilde{\mathbf{a}}}$, where $\mathbf{W} = \left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1/2}\widehat{\mathbf{g}}\widehat{\mathbf{g}}^{\mathrm{H}}\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1/2}$. Therefore, we can obtain the optimal beamforming vector and the corresponding maximum received SNR $\widetilde{\mathbf{v}}_{\mathrm{opt}}$ by means of the similar methodology as [39], which are given by

$$\widetilde{\mathbf{v}}_{\mathrm{opt}} = \frac{\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1}\widehat{\mathbf{g}}}{\left\|\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1}\widehat{\mathbf{g}}\right\|} \tag{15}$$

$$\gamma_{\max} = \widehat{\mathbf{g}}^{\mathrm{H}}\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1}\widehat{\mathbf{g}} \tag{16}$$

Consequently, the SE can be obtained as

$$R = \mathbb{E}\left\{\log_2\left(1 + \widehat{\mathbf{g}}^{\mathrm{H}}\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1}\widehat{\mathbf{g}}\right)\right\} \tag{17}$$

Based on the analysis above, it can be seen that the pilot signal not only determines the channel estimation accuracy and further the SE but also affects the system overall power consumption. Hence, there is a tradeoff between the SE and the power consumption, which can be evaluated by a new measurement, i.e., energy efficiency (EE), which is defined as the SE divided by the total power consumption [40]–[43]. To handle the EE problem, a practical power consumption model is necessary. Based on [41]–[45], the power consumed by the pilot sequence and the data signal is given by

$$P_{\text{tot}} = \frac{\text{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)}{\zeta T_{\text{c}}} + \frac{T_{\text{c}} - L}{T_{\text{c}}}\frac{\rho_{\text{d}}}{\zeta} + NP_{\text{ant}} + P_{\text{sta}} \tag{18}$$

where the first two terms of the RHS of (18) denote the pilot power and practical data transmit power, respectively, $\zeta \leq 1$ denotes the power amplifier efficiency, P_{sta} is the static circuit power consumed at the BS, which is irrelevant to the BS antenna number and P_{ant} is the constant circuit power consumption per antenna, which accounts for the power dissipations in the transmit filter, mixer, frequency synthesizer, and digital-to-analog converter which is independent of the actual transmitted power (Please refer to [42] and references therein for more details). For notation ease and analysis simplicity, we let $\zeta = 1$ without loss of generality and define $P_{\text{fix}} \triangleq \frac{T_{\text{c}}-L}{T_{\text{c}}}\frac{\rho_{\text{d}}}{\zeta} + NP_{\text{ant}} + P_{\text{sta}}$.

With the consideration above, we establish the pilot signal design problem for maximizing the EE while guaranteeing the required QoS and stratifying the total power budget

$$\begin{aligned}\mathcal{P}_1: \max_{\widetilde{\mathbf{X}}\in\mathbb{C}^{K\times L}} \ & \text{EE} = \frac{(1-\vartheta)\cdot R}{\text{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)/T_{\text{c}} + P_{\text{fix}}} \\ \text{s.t. C1: } & \text{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right) \leq P, \quad \text{C2: } R \geq R_0\end{aligned} \tag{19}$$

where $\vartheta = \frac{L}{T_{\text{c}}}$ accounts for the training overhead proportion in a coherent interval. It is worth mentioning that the constraints on the QoS can be presented in many different ways and herein, the minimum SE is selected as the constraint.

The EE-based pilot scheme in (19) evaluates the EE practically, yet leads to some obstacle for solving the optimization problem. The main difficulties rest with two-fold: (1) The derivation of the analytical expression of EE function in (19) is extremely challenging; (2) The objective function is in a non-convex fractional form w.r.t. the pilot matrix.

In next section, we will try our best to solve the pilot design problem for the massive MIMO FDD system.

III. ENERGY-EFFICACY BASED PILOT SIGNAL DESIGN ALGORITHM

In this section, we will propose a two-layer iterative algorithm to deal with the optimization problem $\mathcal{P}_1$ in (19). As a precondition, a closed-form expression of the EE is derived with the help of the random matrix theory. Based on the analytical approximation expression, the structure of the near-optimal pilot matrix is deduced by utilizing the majorization theory. Then, by means of the fractional programming technology, the EE-based optimization problem is reformulated into an equivalent problem in subtractive form, wherein the standard Lagrangian duality theory can be employed and a closed-form solution is obtained.

On account of the complicated expression of the SE in (17), exact evaluation for it appears to be intractable in mathematics. Alternatively, by using the deterministic equivalent approximation method in [10] and [46] we can obtain a closed-form expression of the SE, which is quite applicable for the large-dimensional system. Then, we have the following lemma.

Lemma 1: When the BS adopts the vector $\widetilde{\mathbf{v}}_{\text{opt}}$ for downlink beamforming, a tight approximation of the average SE in (17) is given by

$$\overline{R} = \log_2\left(1 + \delta\text{Tr}\left(\Lambda\right) - \text{Tr}\left(\delta\left(\mathbf{S} + \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)^{-1}\mathbf{Z}\right)\right) \tag{20}$$

where $\mathbf{S} = \sigma_{\text{p}}^2\left(\delta\mathbf{I}_K + \Lambda^{-1}\right)$ and $\mathbf{Z} = \sigma_{\text{p}}^2\left(\delta\Lambda + \mathbf{I}_K\right)$. Furthermore, it is guaranteed that $R - \overline{R} \xrightarrow[N\to\infty]{\text{a.s.}} 0$.

Proof 1: Please refer to Appendix B. ■

As pointed out in [10] and [34], although the asymptotic result in Lemma 1 is derived when $N \to \infty$, the tight approximation of the SE derived from the asymptotic analysis is suitable for large but finite system dimensionality.

By using (20), the problem $\mathcal{P}_1$ can be redescribed as

$$\begin{aligned}\mathcal{P}_2: \max_{\widetilde{\mathbf{X}}\in\mathbb{C}^{K\times L}} \ & \frac{(1-\vartheta)\cdot\overline{R}}{\text{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)/T_{\text{c}} + P_{\text{fix}}} \\ \text{s.t. C1}, \quad & \text{C2: } \overline{R} \geq R_0\end{aligned} \tag{21}$$

Although $\mathcal{P}_2$ cannot be solved directly, the structure of the EE maximization pilot sequence to $\mathcal{P}_2$ can be uncovered, which is summarized in Theorem 1.

Theorem 1: The optimal solution $\widetilde{\mathbf{X}}$ to $\mathcal{P}_2$ is a $K \times L$ quasi-diagonal matrix in the form of

$$\widetilde{\mathbf{X}} = \left[\text{diag}\left\{[x_1, x_2, \underbrace{\ldots}, x_L]\right\}, \mathbf{0}\right]^{\text{T}}, \quad x_l \geq 0, \ \forall l. \tag{22}$$

Proof 2: Assume that $\widetilde{\mathbf{X}}$ is in the set of feasible solution of the problem $\mathcal{P}_2$. For notation ease, let us define $\mathbf{V} = \left(\mathbf{S} + \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)^{-1}$. According to [47, Th. 20.A.4], we have

$$\text{Tr}\left(\mathbf{VZ}\right) \geq \sum_{i=1}^{K} \lambda_{\mathbf{V},i} \cdot \lambda_{\mathbf{Z},K-i+1} \tag{23}$$

where $\lambda_{\mathbf{V},i}$ is the i-th largest eigenvalue of $\mathbf{V}$. Since $\mathbf{Z}$ is a diagonal matrix with positive real diagonal entries in decreasing order, the equality in (23) is achieved if and only if $\mathbf{V}$ is diagonal with its elements in the opposite order of $\mathbf{Z}$. Then, the EE in $\mathcal{P}_2$ is upper bounded by

$$\text{EE}\left(\widetilde{\mathbf{X}}\right) \leq \frac{(1-\vartheta)\cdot\log_2\left(1 + \delta N - \delta\sum_{i=1}^{K}\lambda_{\mathbf{V},i}\cdot\lambda_{\mathbf{Z},K-i+1}\right)}{\text{Tr}\left(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}\right)/T_{\text{c}} + P_{\text{fix}}} \tag{24}$$

The upper bound in (24) can be fulfilled when $\mathbf{V}$ is a diagonal matrix with its diagonal entries in increasing order. On the one hand, to make $\mathbf{V}$ a diagonal matrix, $\mathbf{Q} = \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\text{H}}$ should be a

diagonal matrix, as $\mathbf{S}$ is a diagonal matrix. On the other hand, the diagonal elements of $\mathbf{Q}$ should be in decreasing order, which is because the diagonal elements of $\mathbf{S}$ are in increasing order. Furthermore, due to $\text{rank}(\mathbf{Q}) \leq L$, the optimal $\mathbf{Q}$ is in the form of $\mathbf{Q} = \text{diag}\{[q_1, q_2, \ldots, q_L, 0, \ldots, 0]\}$ with $q_1 \geq q_2 \geq \cdots \geq q_L \geq 0$. Therefore, we have a solution to $\mathcal{P}_2$ in the form of $\widetilde{\mathbf{X}}$ given in (22), where $x_l = \sqrt{q_l}$.

This completes the proof. ■

Remark 1: From (4) and (22), one can find that in the EE maximization based pilot scheme, the dominant channel eigen-directions are selected for pilot sequences as that in the MSE minimization based pilot scheme [35], [48], wherein the variable x_l in (22) indicates the assigned power to the corresponding channel eigen-directions for channel estimation. Since $x_1 \geq x_2 \geq \ldots \geq x_L$, it tells us that more power is allocated to the strong channel eigen-direction in EE-base pilot scheme. In addition, substituting (22) into (10) yields

$$\text{MSE}\left(\widetilde{\mathbf{X}}\right) = 1 - \frac{1}{N}\sum_{l=1}^{L}\frac{x_l^2\lambda_l^2}{x_l^2\lambda_l + \sigma_{\text{p}}^2} \tag{25}$$

which confirms that both the pilot power and the pilot sequence length play important roles in the channel estimation accuracy.

Remark 2: One can also find that the EE-based pilot signal structure still has the column orthogonality as the MSE-based pilot pattern, although we have relaxed this requirement at the beginning of the system model establishment.

By using Theorem 1 and the substitution $x_l = \sqrt{q_l}$, $\mathcal{P}_2$ can be rearranged to an equivalent form w.r.t. $\mathbf{m} \triangleq [q_1, q_2, \ldots, q_L]$

$$\begin{aligned}\mathcal{P}_3 : \max_{\mathbf{q}} \ & \frac{(1-\vartheta)\cdot\overline{R}(\mathbf{q})}{P_{\text{tot}}(\mathbf{q})} \\ \text{s.t. C1:} & \sum_{l=1}^{L} q_l \leq P, \quad \text{C2:}\overline{R}(\mathbf{q}) \geq R_0\end{aligned} \tag{26}$$

where $\overline{R}(\mathbf{q})$ and $P_{\text{tot}}(\mathbf{q})$ are given by

$$\begin{aligned}\overline{R}(\mathbf{q}) &= \log_2\left(1 + \delta\left(N - \sum_{l=1}^{L}\frac{\mathbf{Z}(l,l)}{q_l + \mathbf{S}(l,l)} - \sum_{l=L+1}^{K}\lambda_l\right)\right) \\ P_{\text{tot}}(\mathbf{q}) &= \sum_{l=1}^{L} q_l/T_{\text{c}} + P_{\text{fix}}\end{aligned} \tag{27}$$

However, $\mathcal{P}_3$ is still a non-convex optimization problem due to the fractional form of the objective function. To tackle this dilemma, the nonlinear fractional programming methodology is exploited. According to [42, Th. 1], the maximum EE for $\mathcal{P}_3$ can be achieved if and only if

$$\begin{aligned}&\max_{\mathbf{q}\in\mathbb{D}} (1-\vartheta)\cdot\overline{R}\left(\mathbf{q}\right) - \eta^\star P_{\text{tot}}\left(\mathbf{q}\right) \\ &\qquad = (1-\vartheta)\cdot\overline{R}\left(\mathbf{q}^\star\right) - \eta^\star P_{\text{tot}}\left(\mathbf{q}^\star\right) = 0\end{aligned} \tag{28}$$

where $\mathbb{D} = \left\{\mathbf{q}\,|\text{C1, C2}\right\}$ is the feasible set for $\mathcal{P}_3$, $\eta^\star = \frac{(1-\vartheta)\cdot\overline{R}\left(\mathbf{q}^\star\right)}{P_{\text{tot}}\left(\mathbf{q}^\star\right)} = \max_{\mathbf{q}\in\mathbb{D}}\frac{(1-\vartheta)\cdot\overline{R}\left(\mathbf{q}\right)}{P_{\text{tot}}\left(\mathbf{q}\right)}$ is the maximum EE to $\mathcal{P}_3$, and $\mathbf{q}^\star$ is the associated optimal solution.

Eq. (28) tells us that for a fractional optimization problem, there exists an equivalent problem with the objective function in subtractive form, e.g., $(1-\vartheta)\cdot\overline{R}\left(\widetilde{\mathbf{X}}\right) - \eta^\star P_{\text{tot}}\left(\mathbf{q}\right)$ in the considered case. Moreover, one can see that the optimal solution to $\mathcal{P}_3$ can be obtained by solving $\max_{\mathbf{q}\in\mathbb{D}}(1-\vartheta)\cdot\overline{R}\left(\mathbf{q}\right) - \eta^\star P_{\text{tot}}\left(\mathbf{q}\right)$ if $\eta^\star$ is known in advance. Though $\eta^\star$ is unknown at first, an iterative algorithm (also known as Dinkelbach method [42]) can be used to solve $\mathcal{P}_3$ with an equivalent objective function. Therefore, we will focus on the following equivalent objective function for a given parameter η in the rest of the paper.

$$\mathcal{P}_4 : \max \ (1-\vartheta)\cdot\overline{R}\left(\mathbf{q}\right) - \eta P_{\text{tot}}\left(\mathbf{q}\right) \quad \text{s.t. C1, C2} \tag{29}$$

By using the properties of the compound function [49], it can be proved that the objective function in $\mathcal{P}_4$ is concave. Thus, we can conduct $\mathcal{P}_4$ by solving its equivalent Lagrangian dual problem, namely,

$$\min_{\mu,\nu\geq 0}\max_{\mathbf{q}\geq 0}\mathcal{L}(\mu,\nu,\mathbf{q}) \tag{30}$$

where μ and ν are the Lagrangian multipliers corresponding to the constraints C1 and C2, respectively. $\mathcal{L}(\mu,\nu,\mathbf{q})$ is the Lagrange dual function as follows

$$\begin{aligned}\mathcal{L}(\mu,\nu,\mathbf{q}) &= (1-\vartheta)\,\overline{R}(\mathbf{q}) - \eta P_{\text{tot}}(\mathbf{q}) \\ &\quad + \mu\left(P - \sum_{l=1}^{L} q_l\right) + \nu\left(\overline{R}(\mathbf{q}) - R_0\right)\end{aligned} \tag{31}$$

The optimization problem in (30) can be further decomposed into two subproblems, i.e., the inner maximization problem and outer minimization problem, which can be solved by alternate iteration. For the outer minimization problem, the multipliers μ and ν can be updated by means of the gradient method

$$\mu^{(n+1)} = \left[\mu^{(n)} - t_1\left(P - \sum_{l=1}^{L} q_l^\star\right)\right]^+ \tag{32}$$

$$\nu^{(n+1)} = \left[\nu^{(n)} - t_2\left(\overline{R}\left(\mathbf{q}^\star\right) - R_0\right)\right]^+ \tag{33}$$

where n means the iteration number, t_1 and t_2 are the positive step sizes, $\mathbf{q}^\star$ is the optimal solution of the inner maximization problem in (30) by using $\mu^{(n)}$ and $\nu^{(n)}$. The gradient updates of (32) and (33) can always converge to the optimal ν and μ with $t_1^{(n)}$ and $t_2^{(n)}$ being sufficiently small [31].

In the following, we devote to obtaining $\mathbf{q}^\star$ with given μ and ν, i.e.,

$$\max_{\mathbf{q}\geq 0}\mathcal{L}(\mu,\nu,\mathbf{q}) \tag{34}$$

Theorem 2: For the given μ and ν, the optimal solution $\mathbf{q}^\star$ of (34) can be obtained by

$$q_l^\star = \left[\frac{\sqrt{\mathbf{Z}(l,l)}}{\varpi} - \mathbf{S}(l,l)\right]^+, \quad \forall l \tag{35}$$

where ϖ, a, b, and c are defined as

$$\varpi \triangleq \frac{\sqrt{((\eta/T_{\mathrm{c}}+\mu)\,b)^2+4ac\,(\eta/T_{\mathrm{c}}+\mu)}-(\eta/T_{\mathrm{c}}+\mu)\,b}{2a}$$
$$a \triangleq \frac{1-\vartheta+\nu}{\ln 2}, b \triangleq \sum_{i=1}^{L}\sqrt{\mathbf{Z}\,(i,i)}, c \triangleq \frac{1}{\delta}+N-\sum_{i=L+1}^{K}\lambda_i \tag{36}$$

Proof 3: Please refer to Appendix C.

Obviously, the power allocation in (35) is in the form of multilevel water-filling type solution. Apart from this, we can conclude from Theorem 2 that:

- The actual pilot sequence length is equal to the number of the nonzero elements in $\mathbf{q}^{\star}$, which can only be determined numerically through the proposed algorithm. That is to say, we might assign zero power to some of the channel eigen-directions such that the ultimate pilot sequence length might be shortened.
- The water level $\frac{\sqrt{\mathbf{Z}(l,l)}}{\varpi}$ is different from the conventional water level, which is not solely determined by the Lagrange multipliers.

One can also find from (35) that the power variable q_l is updated by the Lagrange multiplier $\mu^{(n)}$ until convergence. However, through an in-depth study of the structure of q_l in (35) w.r.t. μ, it is discovered that q_l has a lower bound, which has an impact on the convergence. Then, we have the following proposition.

Proposition 1: For a given pilot sequence length L and a total power constraint P, the Lagrangian dual problem (30) can converge if and only if

$$P \geq \sum_{l=1}^{L}\left[\frac{\sigma_{\mathrm{p}}^2\sum_{i=1}^{L}\sqrt{(\delta\lambda_i+1)\,(\delta\lambda_l+1)}}{\delta^{-1}+\sum_{i=1}^{L}\lambda_i}-\sigma_{\mathrm{p}}^2\left(\delta+\frac{1}{\lambda_l}\right)\right]^{+} \tag{37}$$

and there will be no solution otherwise. In other words, the required minimum power to support the pilot sequence length L for EE maximization is given by the RHS of (37).

Proof 4: Please refer to Appendix D.

The physical meaning of Proposition 1 is that the maximum pilot length is restricted by the given power budget P. Moreover, we can find that the water level $\frac{\sqrt{\mathbf{Z}(l,l)}}{\varpi}$ is deceasing with L, whose proof is straightforward and omitted here. Thus, when (37) is not satisfied, we should reduce the pilot length gradually until (37) is fulfilled. Furthermore, it can be guaranteed that the minimum pilot length, i.e. $L = 1$, can always satisfy (37), since the RHS of (37) is equal to zero in this case.

Remark 3: Different from our EE-based pilot design, the MSE-based algorithm [35] can always converge for a given power budget P and predefined training length L without any condition such as (37). This is because the involved water-filling levels in [35] does not have the similar bound as in our proposed algorithm.

Based on aforesaid analysis, a near-optimal two-layer iterative algorithm is developed to solve $\mathcal{P}_1$. The detailed steps are summarized in Algorithm 1. Regarding the convergence of the proposed algorithm, it can be easily proved according to [42], which is omitted here.

Algorithm 1 Iterative Algorithm for EE-Based Pilot Design

1: Initialize $m = 1$, $n = 1$, $t_1 > 0$, $t_2 > 0$, $\epsilon_1 > 0$, $\epsilon_2 > 0$, $\mu^{(1)} \geq 0$, $\nu^{(1)} \geq 0$, $\eta^{(1)} \geq 0$;
2: **while** (37) is not satisfied **do**
3: $L = L - 1$;
4: **end while**
5: **repeat** {Layer 1}
6: **repeat** {Layer 2}
7: Compute $\mathbf{q}^{\star}$ by using (35) with $\eta^{(m)}$, $\mu^{(n)}$, and $\nu^{(n)}$;
8: Update $\mu^{(n+1)}$ and $\nu^{(n+1)}$ according to (32) and (33) with $\mathbf{q}^{\star}$;
9: $n = n + 1$;
10: **until** $|\mu^{(n)} - \mu^{(n-1)}| \leq \epsilon_2$ and $|\nu^{(n)} - \nu^{(n-1)}| \leq \epsilon_2$.
11: $\eta^{(m+1)} = \frac{(1-\vartheta)\cdot\overline{R}(\mathbf{q}^{\star})}{P_{\mathrm{tot}}(\mathbf{q}^{\star})}$;
12: $m = m + 1$;
13: **until** $\left|(1-\vartheta)\cdot\overline{R}\,(\mathbf{q}) - \eta^{(m-1)}P_{\mathrm{tot}}\,(\mathbf{q})\right| \leq \epsilon_1$.

IV. NUMERICAL RESULTS

In this section, we evaluate the performance of our proposed EE-based pilot scheme from different aspects. For simplicity and without loss of generality, the path-loss and shadow fading coefficient is set to be 1, and the variance of AWGN is set to be 1W. Herein, the channel correlation matrix $\mathbf{R}$ is characterized via the commonly-used exponential correlation model, where $\mathbf{R}\,(i,j) = r^{|i-j|}$ $(i,j = 1,\cdots,N)$ with $r \in [0,1]$ denoting the correlation coefficient between adjacent transmit antennas. This model basically approximates the correlation in a uniform linear array (ULA) under rich scattering conditions [50], [51]. In our simulations, a correlation coefficient of $r = 0.8$ is considered, which describes a modest correlation in the sense of typical ULA behaving with half-wavelength antenna spacings. The circuit power per antenna is $P_{\mathrm{ant}} = 0.01$W, and the static power consumption is $P_{\mathrm{sta}} = 1$W. With regard to the channel coherence interval, a typical value $T_{\mathrm{c}} = 50$ is chosen, which corresponds to a low or moderate mobility environment with coherence bandwidth of 10KHz and coherence time of 5ms [32]. The convergence thresholds ε_1 and ε_2 are set to be 10^{-5}. For comparison, the pilot sequence design based on the MSE minimization [35] is provided for benchmark, i.e., $\min_{\widetilde{\mathbf{X}}} \mathrm{MSE}\left(\widetilde{\mathbf{X}}\right)$ s.t. C1, C2, where $\mathrm{MSE}\left(\widetilde{\mathbf{X}}\right)$ is defined in (10), C1 and C2 correspond to the total power budget and the minimum SE requirement, respectively, as in problem $\mathcal{P}_1$.

Firstly, we examine the accuracy of the closed-form expression of the SE in Lemma 1 through Monte Carlo simulations. In each simulation, 10^4 independent channel realizations are generated and averaged to produce the numerical results. One can find from Fig. 2 that the analytical

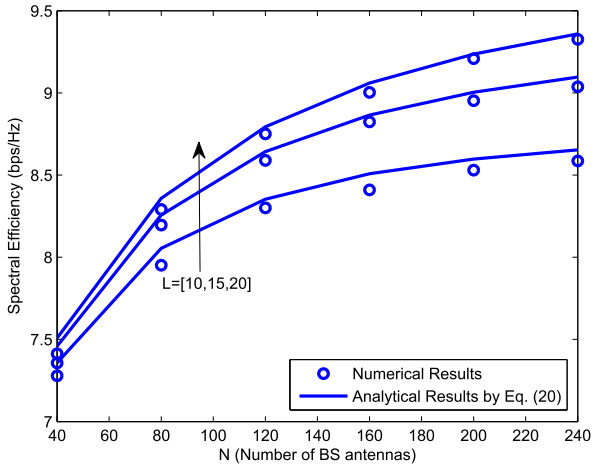


FIGURE 2. The spectral efficiency v.s. the number of BS antennas under different pilot length L, where $K = N$, $r = 0.8$, $\rho_p = \rho_d = 10$dB, $\tilde{\mathbf{X}} = \sqrt{\rho_p}[\text{diag}\{[1, 1, \ldots, 1]\}, \mathbf{0}]^T$.

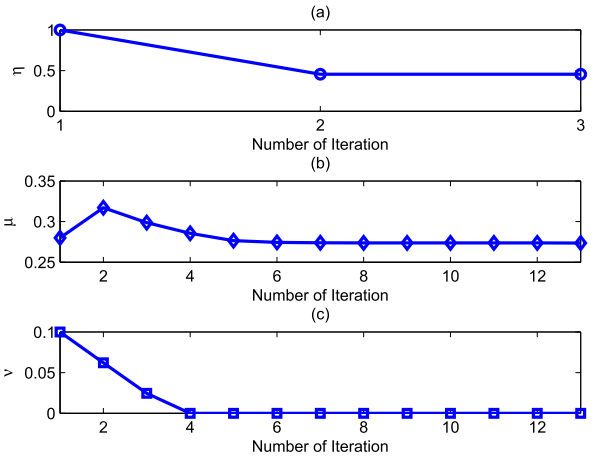


FIGURE 3. The convergence trajectory of the proposed EE-based algorithm, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $P = 9$dB, $R_0 = 2$bps/Hz.

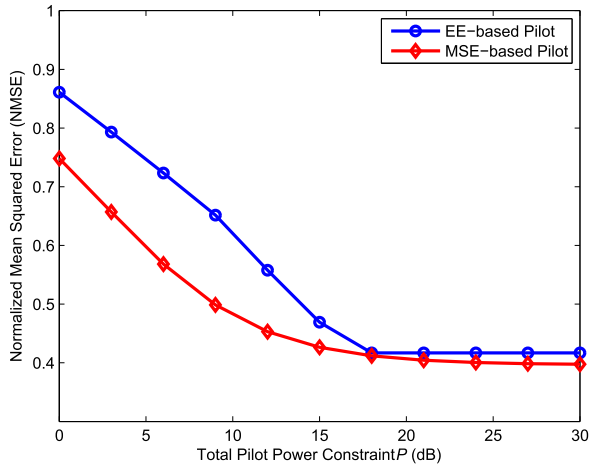


FIGURE 4. The normalized MSE v.s. the total pilot power constraint, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $R_0 = 2$bps/Hz.

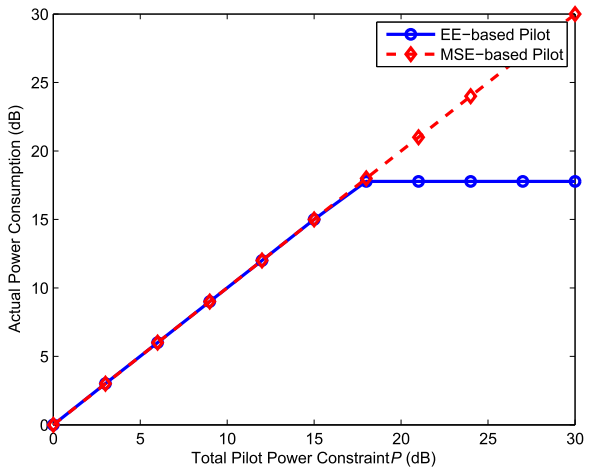


FIGURE 5. The actual power consumption v.s. the total pilot power constraint, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $R_0 = 2$bps/Hz.

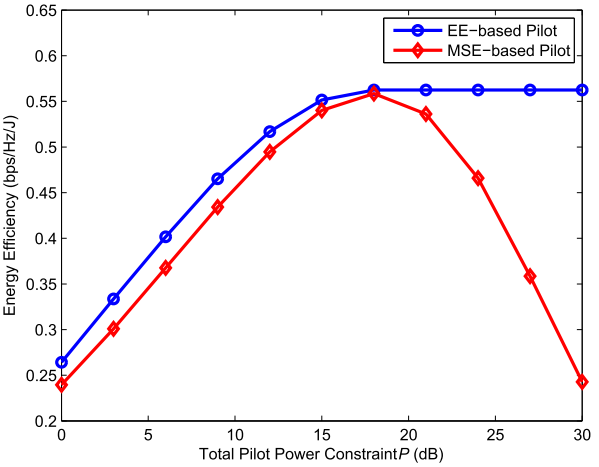


FIGURE 6. The energy efficiency v.s. the total pilot power constraint, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $R_0 = 2$bps/Hz.

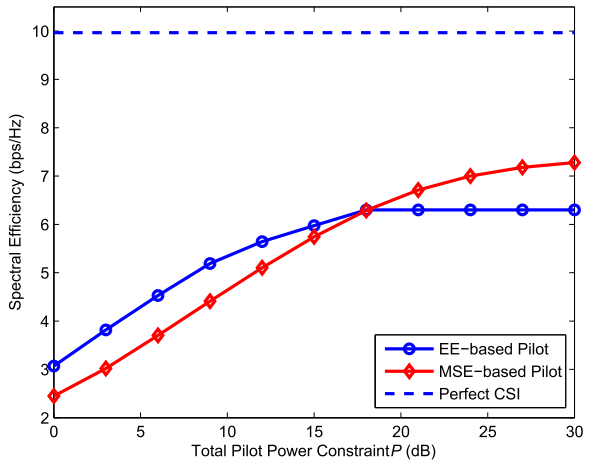


FIGURE 7. The spectral efficiency v.s. the total pilot power constraint, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $R_0 = 2$bps/Hz.

expression of SE in (20) becomes increasingly tight as the antenna number grows large. That is, the concise analytical expression in Lemma 1 is effective for the consequent EE-based algorithm evaluation.

Fig. 3 illustrates the convergence trajectory of Algorithm 1. The sub-figure (a) shows the convergence speed of the variable η in Layer 1 (line 5 in Algorithm 1). The convergence

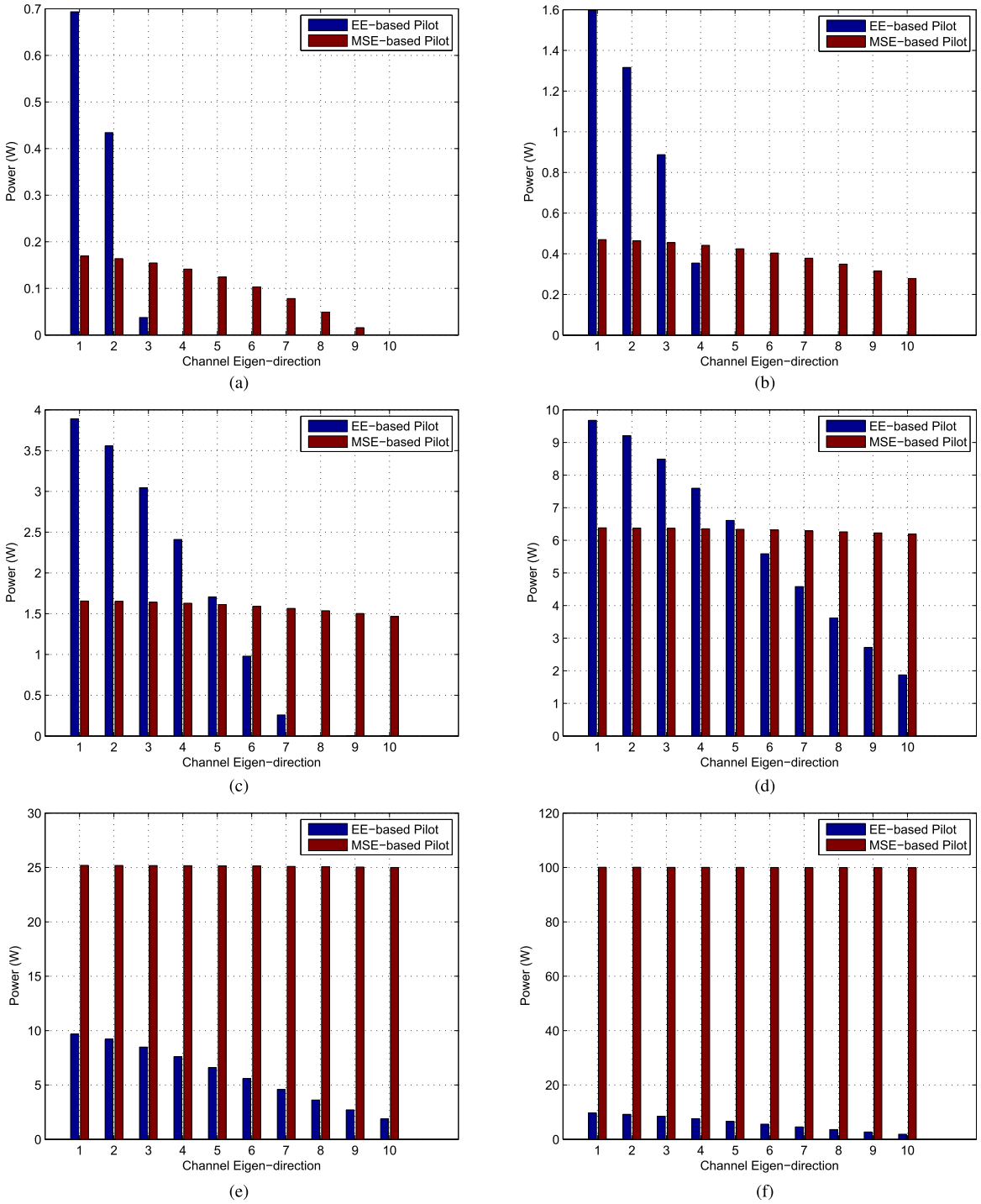


FIGURE 8. Power assignment for the first L strongest channel eigen-directions under different total pilot power constraint, where $N = 100$, $T_c = 50$, $L = 10$, $r = 0.8$, $R_0 = 2$bps/Hz. (a) $P = 0$dB. (b) $P = 6$dB. (c) $P = 12$dB. (d) $P = 18$dB. (e) $P = 24$dB. (f) $P = 30$dB.

behaviors of the variables μ and ν in Layer 2 (line 6 in Algorithm 1) are presented in sub-figures (b) and (c), respectively. It is obvious that the proposed algorithm can always converge to the optimal energy efficiency within a limited number of iterations.

Fig. 4 depicts the performance of the normalized MSE, which is defined as $\mathbb{E}\{\|\mathbf{h}-\widehat{\mathbf{h}}\|^2\}/\mathbb{E}\left\{\|\mathbf{h}\|^2\right\}$. Obviously, the MSE-based pilot scheme outperforms the EE-based pilot scheme. However, as the total power grows large, the performance gap between them becomes small. Especially, when the total power exceeds 18dB, the performance of both schemes saturate to a certain level. This is because both the pilot power and pilot sequence length impact the channel estimation accuracy. Thus, for the fixed pilot sequence

length L, the performance gain will eventually enter into the saturation regime when only relying on the pilot power increment.

Fig. 5 illustrates the actual power consumption for the pilot transmission. One can see that the EE-based pilot scheme allocates the same pilot power as the MSE-based pilot scheme in the low transmit power region, which implies that both schemes transmit the pilot signals with full power in this region. However, when the transmit power constraint is larger than 18dB, i.e., the moderate to high transmit power region, the latter still expends full power for pilot signal transmission to pursuit MSE minimization, whereas the former would rather reduce the pilot signal power so as to guarantee energy efficiency performance. This is also another key reason why the EE-based pilot scheme quickly saturates to a certain level in terms of the MSE performance in Fig. 4.

Fig. 6 compares the energy efficiency performance for both pilot schemes. It is observed that the EE-based pilot scheme is always superior to the MSE-based pilot scheme in the whole transmit power range. More precisely, the performance gain is more remarkable at middle to high transmit power region, namely $P \gg 18$dB. The reasons lie in two-fold: (1). In low transmit power region, the MSE-based pilot scheme provides more accurate CSI at the expense of pilot sequence length, which can be confirmed from Fig. 8(a) - Fig. 8(c). In fact, the EE-based pilot scheme does not take up the full pilot length, i.e., the actual pilot length is smaller than L, which implies that more symbols are remained for effective data transmission in a channel coherent interval. Taking Fig. 8(a) for example, when the total pilot power $P = 6$dB, the actual pilot length is 4 in EE-based pilot scheme, whereas the MSE-based pilot scheme occupies full training length, i.e., $L = 10$. (2). In middle-high transmit power region, the SE gain resulting from the improvement of the channel estimation accuracy cannot make up for the adverse influence of the total power increment, which incurs performance degradation in the MSE-based pilot scheme.

Fig. 7 provides the spectral efficiency performance of the two pilot schemes. Surprisingly, although the MSE-base pilot scheme dominates in term of the channel estimation accuracy at low transmit power region, the EE-based pilot scheme still exhibits better performance than the MSE-based pilot scheme in the sense of the SE performance owing to the shorter pilot sequence length. While at middle to high transmit power region, the performance of EE-based pilot scheme falls short of the MSE-base pilot scheme. Furthermore, the SE performance of the EE-based scheme enters into a saturation regime, which is because the EE-based pilot scheme tends to lower down the transmit power consumption so as to keep the overall energy efficiency maximizing.

From Fig. 4, Fig. 6, and Fig. 7, one can conclude that we should not persistently pursue the channel estimation accuracy but neglect the significant goal of data communication, which may leads to severe waste of the transmission resource, including the time, frequency as well as the energy. In essence, the resource allocation should be carefully managed in order to balance the training phase and data transmission phase.

Fig. 8 shows the pilot power assignment for the first L channel eigen-directions. It can be found that in the EE-based pilot scheme, the strong eigen-directions receive more power assignment than weak in all transmit power region. To be more exact, at low transmit power region, this scheme would rather discard the weaker eigen-directions and make the power concentrate on the stronger eigen-directions. Moreover, as the total pilot power increases, longer pilot length will be utilized. However, in MSE-based pilot scheme, the power is also allocated to strong channel eigen-directions. But with the total pilot power increasing, the equal power is distributed to all the eigen-directions which is different from the EE-based scheme.

V. CONCLUSION

In this paper, we studied the energy-efficient pilot signal design for a single-user massive MIMO FDD downlink system. With considering the total pilot energy budget and the system minimum spectral efficiency requirement, an energy-efficiency based pilot signal optimization problem was established. Since the objective function was complicated even without a analytical expression, the deterministic equivalent approximation technology was utilized to derive a closed-form expression of the cost function, based on which, the pattern of the near-optimal pilot signal was deduced in the light of the theory of majorization. In an effort to solve the non-convex and fractional optimization problem, we resorted to the fractional programming methodology such that the original problem was rearranged into an equivalent subtractive-form problem, where the standard convex theory was adopted and the closed-form solution was obtained. The performance gain of our proposed pilot scheme is justified in comparison with the traditional MSE-based pilot signal by the numerical results. Furthermore, it is shown that the developed iterative algorithm can converge during a few number of iterations.

APPENDIX A
DETERMINISTIC EQUIVALENT APPROXIMATION LEMMA

Lemma 2 ([10, Lemma 4], [46, Lemma 1]): Assume $\mathbf{W}$ and $\mathbf{\Xi} \in \mathbb{C}^{N\times N}$ with uniformly bounded spectral norms w.r.t. N. Consider $\mathbf{x} \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Xi})$ and we have

$$\frac{\mathbf{x}^H\mathbf{W}\mathbf{x}}{N} - \frac{\mathrm{tr}\mathbf{W}\mathbf{\Xi}}{N} \xrightarrow[N\to\infty]{a.s.} 0 \tag{38}$$

APPENDIX B
PROOF OF LEMMA 1

Since $\widehat{\mathbf{g}} \sim \mathcal{CN}(0, \mathbf{\Phi})$, the received SNR in (16) has the corresponding deterministic equivalent approximation as $N \to \infty$ by using Lemma 2 in Appendix A, i.e.,

$$\frac{\gamma_{\max}}{N} \xrightarrow{a.s.} \frac{\overline{\gamma}}{N} = \frac{\mathrm{Tr}\left(\left(\Omega + \delta^{-1}\mathbf{I}_K\right)^{-1}\Phi\right)}{N} \tag{39}$$

$$\begin{aligned}\overline{\gamma} &= \mathrm{Tr}\left(\left(\Lambda-\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}+\delta^{-1}\mathbf{I}_{K}\right)^{-1}\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}\right)\\ &= \mathrm{Tr}\left(\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}\left(\Lambda-\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}+\delta^{-1}\mathbf{I}_{K}\right)^{-1}\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)\\ &= \mathrm{Tr}\left(\left(\Lambda\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)-\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}+\delta^{-1}\left(\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)\right)^{-1}\Lambda\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)\\ &= \mathrm{Tr}\left(\delta\left(\sigma_{\mathrm{p}}^{2}\delta\mathbf{I}_{K}+\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}\left(\Lambda\left(\sigma_{\mathrm{p}}^{2}\delta\mathbf{I}_{r}+\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)-\Lambda\left(\sigma_{\mathrm{p}}^{2}\delta\mathbf{I}_{r}+\sigma_{\mathrm{p}}^{2}\Lambda^{-1}\right)\right)\right)\\ &= \delta\mathrm{Tr}\left(\Lambda\right)-\mathrm{Tr}\left(\delta\left(\sigma_{\mathrm{p}}^{2}\delta\mathbf{I}_{K}+\sigma_{\mathrm{p}}^{2}\Lambda^{-1}+\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{H}}\right)^{-1}\left(\sigma_{\mathrm{p}}^{2}\delta\Lambda+\sigma_{\mathrm{p}}^{2}\mathbf{I}_{K}\right)\right)\end{aligned} \tag{40}$$

$$\frac{1-\vartheta+\nu}{\ln 2}\left(\frac{\sqrt{\mathbf{Z}\left(l,l\right)}}{q_l+\mathbf{S}\left(l,l\right)}\right)^{2}+\left(\frac{\eta}{T_{\mathrm{c}}}+\mu\right)\sum_{i=1}^{L}\sqrt{\mathbf{Z}\left(i,i\right)}\frac{\sqrt{\mathbf{Z}\left(l,l\right)}}{q_l+\mathbf{S}\left(l,l\right)}-\left(\frac{\eta}{T_{\mathrm{c}}}+\mu\right)\left(\frac{1}{\delta}+N-\sum_{i=L+1}^{K}\lambda_i\right)=0 \tag{46}$$

Substituting (7) and (9) into the RHS of (39) yields (40), as shown at the top of this page, where the properties of matrix computation, namely, $(\mathbf{I}+\mathbf{AB})^{-1}\mathbf{A}=\mathbf{A}(\mathbf{I}+\mathbf{BA})^{-1}$, $(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$, and $\mathrm{Tr}\left(\mathbf{AB}\right)=\mathrm{Tr}\left(\mathbf{BA}\right)$, are employed.

Then, on the basis of the dominated convergence and the continuous mapping theorem [10], we have the deterministic equivalent approximation of the average SE as follows

$$R\xrightarrow[N\to\infty]{\text{a.s.}}\overline{R}=\log_2\left(1+\overline{\gamma}\right) \tag{41}$$

This completes the proof. ■

APPENDIX C
PROOF OF THEOREM 2

In order to solve (34), the KKT conditions will be utilized [49]. Letting the first-order derivative of $\mathcal{L}$ w.r.t. q_l be zero yields

$$\mathcal{L}'\left(q_l\right)=\left(1-\vartheta\right)\overline{R}'\left(q_l\right)-\eta/T_{\mathrm{c}}-\mu+\nu\overline{R}'\left(q_l\right)=0,\ \forall l \tag{42}$$

From (42), we have

$$\overline{R}'\left(q_l\right)=\overline{R}'\left(q_j\right),\ \forall l\neq j \tag{43}$$

where $\overline{R}'\left(q_l\right)$ is given by

$$\overline{R}'\left(q_l\right)=\frac{\delta\mathbf{Z}\left(l,l\right)}{\left(q_l+\mathbf{S}\left(l,l\right)\right)^{2}\ln 2}\cdot\left(1+\delta N-\delta\sum_{i=1}^{L}\frac{\mathbf{Z}\left(i,i\right)}{q_i+\mathbf{S}(i,i)}-\delta\sum_{i=L+1}^{K}\lambda_i\right)^{-1},\ \forall l \tag{44}$$

Combining (43) and (44) yields

$$q_j=\sqrt{\frac{\mathbf{Z}\left(j,j\right)}{\mathbf{Z}\left(l,l\right)}}\left(q_l+\mathbf{S}\left(l,l\right)\right)-\mathbf{S}\left(j,j\right) \tag{45}$$

By substituting (44) and (45) into (42) and simplifying it, one can find that the optimal $q_l^{\star}$ is equivalent to solve the root of a quadratic polynomial equation (46), as shown at the top of this page.

By using the standard root of quadratic equation, we have

$$\frac{\sqrt{\mathbf{Z}\left(l,l\right)}}{q_l^{\star}+\mathbf{S}\left(l,l\right)}=\frac{\sqrt{\left(\left(\frac{\eta}{T_{\mathrm{c}}}+\mu\right)b\right)^{2}+4ac\left(\frac{\eta}{T_{\mathrm{c}}}+\mu\right)}-\left(\frac{\eta}{T_{\mathrm{c}}}+\mu\right)b}{2a}$$

where a, b, and c are given in (36). Therefore, the closed form of $q_l^{\star}$ is derived in (35) after simple transformation.

This completes the proof. ■

APPENDIX D
PROOF OF PROPOSITION 1

For simplicity of analysis, ϖ in (36) can be simplified as

$$\varpi=\frac{2c}{b+\sqrt{b^{2}+\frac{4ac}{\eta/T_{\mathrm{c}}+\mu}}} \tag{47}$$

Then, the following two statements hold

(a) ϖ is monotonically increasing in μ;

(b) $\varpi^{\infty}=\lim_{\mu\to\infty}\varpi=\frac{c}{b}=\frac{\delta^{-1}+N-\sum_{i=L+1}^{K}\lambda_i}{\sum_{i=1}^{L}\sqrt{\mathbf{Z}(i,i)}}$.

Therefore, one can find that ϖ^{∞} is irrelevant of η and $\varpi\leq\varpi^{\infty}$. Accordingly, the optimal power in (35) has a lower bound

$$q_l^{\star}\geq\left[\frac{\sqrt{\mathbf{Z}\left(l,l\right)}}{\varpi^{\infty}}-\mathbf{S}\left(l,l\right)\right]^{+} \tag{48}$$

Eq. (48) contains two-fold meanings:

- When the equality is achieved, the required minimum total power for maximizing EE with training length L is given by

$$P_{\min}=\sum_{l=1}^{L}q_l^{\star}=\sum_{l=1}^{L}\left[\frac{\sqrt{\mathbf{Z}\left(l,l\right)}}{\varpi^{\infty}}-\mathbf{S}\left(l,l\right)\right]^{+}; \tag{49}$$

- When the total power constraint, P, is less than $P_{\min}$, there is no solution for problem $\mathcal{P}_1$. That is to say, the optimization problem (30) cannot converge.

This completes the proof. ■

REFERENCES

[1] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] J. Zhu, R. Schober, and V. K. Bhargava, "Secure transmission in multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4766–4781, Sep. 2014.

[5] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, no. 081301, pp. 1–18, Aug. 2016.

[6] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2181–2194, Apr. 2017.

[7] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Performance scaling law for multicell multiuser massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9890–9903, Nov. 2017.

[8] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[9] 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.

[10] 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.

[11] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[12] J. Pan and W.-K. Ma, "Constant envelope precoding for single-user large-scale MISO channels: Efficient precoding and optimal designs," *IEEE J. Sel. Areas Signal Process.*, vol. 8, no. 5, pp. 982–995, Oct. 2014.

[13] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4879–4893, Dec. 2015.

[14] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.

[15] A. Minasian, R. S. Adve, and S. ShahbazPanahi, "The impact of hardware calibration errors on the performance of massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[16] S. Zarei, W. H. Gerstacker, J. Aulin, and R. Schober, "Multi-cell massive MIMO systems with hardware impairments: Uplink-downlink duality and downlink precoding," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5115–5130, Aug. 2017.

[17] Wikipedia. (2014). *List of LTE Networks*. Accessed: Sep. 3, 2014. [Online]. Available: http://en.wikipedia.org/wiki/List_of_LTE_networks

[18] P. W. C. Chan *et al.*, "The evolution path of 4G networks: FDD or TDD?" *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42–50, Dec. 2006.

[19] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[20] X. Zhang, J. Tadrous, E. Everett, F. Xue, and A. Sabharwal, "Angle-of-arrival based beamforming for FDD massive MIMO," in *Proc. Allerton Conf. Commun. Control Comput.*, Nov. 2015, pp. 704–708.

[21] A. J. Duly, T. Kim, D. J. Love, and J. V. Krogmeier, "Closed-loop beam alignment for massive MIMO channel estimation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1439–1442, Aug. 2014.

[22] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[23] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[24] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.

[25] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[26] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[27] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[28] L. Dai, Z. Gao, and Z. Wang, "Joint channel estimation and feedback with low overhead for FDD massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–6.

[29] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[30] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 33–56, 1st Quat., 2017.

[31] Y. Wang, C. Li, Y. Huang, D. Wang, T. Ban, and L. Yang, "Energy-efficient optimization for downlink massive MIMO FDD systems with transmit-side channel correlation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7228–7243, Sep. 2016.

[32] Y. Kim, G. Miao, and T. Hwang, "Energy efficient pilot and link adaptation for mobile users in TDD multi-user MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 382–393, Jan. 2014.

[33] Z. Xu, G. Y. Li, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient power allocation for pilots in training-based downlink OFDMA systems," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3047–3058, Oct. 2012.

[34] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[35] 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.

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

[37] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[38] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[39] C. Li, L. Yang, and Y. Shi, "An asymptotically optimal cooperative relay scheme for two-way relaying protocol," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 145–148, Feb. 2010.

[40] Q. Wu, W. Chen, M. Tao, J. Li, H. Tang, and J. Wu, "Resource allocation for joint transmitter and receiver energy efficiency maximization in downlink OFDMA systems," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 416–430, Feb. 2015.

[41] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[42] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.

[43] Y. Hu, B. Ji, Y. Huang, F. Yu, and L. Yang, "Energy-efficient resource allocation in uplink multiuser massive MIMO systems," *Int. J. Antennas Propag.*, vol. 2015, no. 4, pp. 1–9, Aug. 2015.

[44] H. Q. Ngo *et al.*, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 605–608, Dec. 2014.

[45] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[46] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.

[47] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Boston, MA, USA: Academic, 1979.
[48] E. Björnson 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.
[49] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K: Cambridge Univ. Press, 2004.
[50] D. Wang, C. Ji, X. Gao, S. Sun, and X. You, "Uplink sum-rate analysis of multi-cell multi-user massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 5404–5408.
[51] D. Wang, C. Ji, S. Sun, and X. You, "Spectral efficiency of multi-cell multi-user DAS with pilot contamination," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 3208–3212.

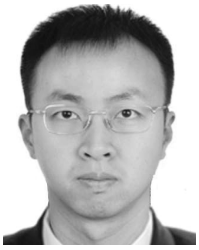
YI WANG received the B.S. degree from PLA Information Engineering University, Zhengzhou, China, in 2006, and the M.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, China, in 2009 and 2016, respectively. Since 2016, he has been with the School of Electronics and Communication Engineering, Zhengzhou University of Aeronautics, China. His current research interests include signal processing, energy-efficient communication, relay-aided system, and massive MIMO. He received the Best Paper Awards of the IEEE WCSP in 2015.

PENGGE MA received the B.S. degree from Beihang University, Beijing, China, in 1996, the M.S. degree from the School of Physical Science and Engineering, Zhengzhou University, in 2004, and the Ph.D. degree from the School of Information Engineering, Zhengzhou University, China, in 2012. Since 1996, he has been with the Zhengzhou University of Aeronautics, China. His current research interests include signal processing, photoelectric detection, target detection, and recognition. He received the first prize of Henan Natural Science Thesis and the second prize of Henan Natural Science Thesis in 2013. He presided over and received the second prize of scientific and technological progress in Henan Province and the first prize of scientific and technological achievements of Henan Provincial Department of Education in 2014. He was recognized as the Academic and Technical Leader of Henan Provincial Education Department in 2015. He presided over and received the third prize of scientific and technological progress in Henan Province in 2017.

RUI ZHAO (M'12) received double bachelor's degrees from the Harbin Institute of Technology in 2003, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, China, in 2006 and 2010, respectively. After graduation, he joined the School of Information Science and Engineering, Huaqiao University, China, where he is currently an Associate Professor. From 2014 to 2015, he visited the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, where he was a Visiting Research Scholar. He has published many papers in international journals, such as the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON COMMUNICATIONS and papers in conferences, such as the IEEE Globecom and IEEE ICC. His current research interests include cooperative communications, physical-layer security, and MIMO communication systems.

YONGMING HUANG (SM'17) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2007. Since 2007, he has been a Faculty Member with the School of Information Science and Engineering, Southeast University. In 2008 and 2009, he visited the Signal Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include spaceCtime wireless communications, cooperative wireless communications, energy-efficient wireless communications, and optimization theory. He serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal on Advances in Signal Processing*, and the *EURASIP Journal on Wireless Communications and Networking*.

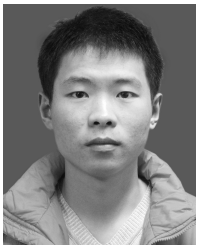
CHUNGUO LI (SM'16) received the bachelor's degree in wireless communications from Shandong University in 2005 and the Ph.D. degree in wireless communications from Southeast University in 2010. In 2010, he joined the Faculty of Southeast University in Nanjing, where he has been an Associate Professor since 2012. From 2012 to 2013, he did the postdoctoral research with Concordia University, Montreal, Canada. Since 2013, he has been with the DSL laboratory supervised by Prof. J. M. Cioffi. His research interests are in the MIMO relay communications, green communications, and next generation of Wi-Fi. He is a Senior Member of the Chinese Institute of Electronics. He was a recipient of the Southeast University Excellent Young Professor Award in 2015, the Science and Technology Progress Award of the National Education Ministry of China in 2014, the Excellent Visiting Associate Professor Award at Stanford in 2014, the Excellent Foreign Postdoc Award of Canada in 2012, the Best Ph.D. Thesis Award of Southeast University in 2010, and several conference best paper awards. He has served for many IEEE conferences, including the IEEE 16th International Symposium on Communications and Information Technologies as the Track Chair of the Wireless Communications, the International Conference on Communications, the International Conference on Acoustics, Speech, and Signal Processing as the TPC member. He is currently an Area Editor of the *AEU-International Journal of Electronics and Communications* (Elsevier), an Associate Editor of the *Circuits, Systems, and Signal Processing*, an Editor of the *KSII Transactions on Internet and Information Systems*. He is currently a reviewer of many IEEE journals.



JUN ZHU (S'10) was born in Nanjing, China. He received the B.Sc. degree (Hons.) in information engineering from Southeast University, Nanjing, China, in 2008, the M.A.Sc. degree (Hons.) in electrical engineering from the University of Victoria, Victoria, Canada, in 2011, and the Ph.D. degree (Hons.) in electrical engineering from The University of British Columbia, Vancouver, Canada, in 2016. He was a Visiting Researcher with the Institute for Digital Communications, Friedrich Alexander University Erlangen-Nürnberg, Erlangen, Germany, in 2014 and 2015. He was a Post-Doctoral Research Fellow with The University of British Columbia in 2016. He is currently a Senior System Engineer at Qualcomm, San Diego, USA. His main focus is on 5G wireless system design. His research interests include MIMO-OFDM wireless systems, massive MIMO, energy-efficient (green) communications, and physical layer security. He has served on the technical program committees of many international conferences, including the IEEE Globecom, the IEEE PacRim, and the IEEE WCSP.

Dr. Zhu was a finalist of Lieutenant Governor's Silver Medal for Outstanding Master Thesis at University of Victoria and a recipient of the Dr. Esme Foord Scholarship in 2011, and has been a recipient of Four-Year-Fellowship at The University of British Columbia since 2011. He was also the holder of Pei-Huang Tung and Tan-Wen Tung Fellowship in 2012, the Graduate Support Initiative Award in 2013, the Chinese Government Award for Outstanding Self-Financed Students Abroad in 2014, German Academic Exchange Service (DAAD) Grant, and the UBC International Research Award in 2015. He received the QualStar Award at Qualcomm three times in 2017.

KAIZHI HUANG received the Ph.D. degree in communication and information system from Tsinghua University, Beijing, China. She is currently a Professor and a Doctoral Supervisor of the National Digital Switching System Engineering and Technological Research Center, Zhengzhou, China. Her research interests include wireless mobile communication network and physical layer security.



BING WANG received the B.S. degree in communication engineering from the Wuhan University of Technology, Wuhan, China, in 2015. He is currently pursuing the B.S. degree with the National Digital Switching System Engineering and Technological Research Center, Zhengzhou. His research interests include wireless communication, physical layer security, and stochastic geometry.