On Orthogonal and Superimposed Pilot Schemes in Massive MIMO NOMA Systems

Junjie Ma, Chulong Liang, Chongbin Xu, and Li Ping, *Fellow, IEEE*

***Abstract*— This paper is concerned with pilot transmission schemes in a large antenna system with non-orthogonal multiple-access (NOMA). We investigate two pilot structures—orthogonal pilot (OP) and superimposed pilot (SP). In OP, pilots occupy dedicated time (or frequency) slots, while in SP, pilots are superimposed with data. We study an iterative data-aided channel estimation (IDACE) receiver, where partially decoded data are used to refine channel estimation. We analyze the achievable rates for systems with IDACE receivers for both OP and SP. We show that the optimal portion of pilot power tends to zero for SP with Gaussian signaling. This result is consistent with existing findings obtained via the replica method in statistical physics. The latter involves multiple codes, which is convenient for theoretical analysis but difficult to implement. As a comparison, IDACE is potentially implementable in practice. We demonstrate that, with code optimization, SP can outperform OP in a high mobility environment with a large number of users. We provide numerical examples to verify our analysis.**

***Index Terms*— Achievable rate, iterative data-aided channel estimation, massive multiple-input multiple-output (MIMO) system, non-orthogonal multiple-access (NOMA), orthogonal pilot, superimposed pilot.**

I. Introduction

In A large multiple-input multiple-output (MIMO) system [1], increasing the number of simultaneously active users, denoted by K below, can efficiently exploit the rich spatial diversity offered by large MIMO and significantly enhance system sum-rate. This is illustrated in Fig. 1 which plots the sum-rate capacity of a quasi-static multi-user uplink system with M antennas at the base station (BS) and one antenna at each mobile terminal (MT). We can see that the sum-rate capacity increases rapidly when K is relatively small compared with M (say, $K \leq M/4$). When K increases with M with a fixed ratio, as shown in Fig. 1 using a dashed line, the sum-rate capacity grows almost linearly with M. Perfect channel state information at receiver (CSIR) is assumed in Fig. 1. Under this assumption, the performance in Fig. 1 can be approximately achieved using, e.g., zero forcing (ZF) for a massive MIMO system. With ZF, users are orthogonal in space, which avoids the interference problem in a multi-user system.

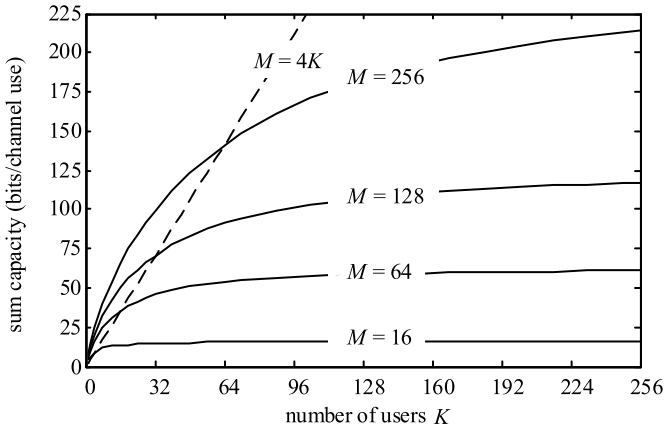


Fig. 1. Capacity of a single-cell quasi-static multi-user up-link system with perfect CSIR under equal-receive-power constraint at signal-to-noise-ratio (SNR) = 0 dB. Large scale fading is not included.

However, accurate channel state information (CSI) is a demanding requirement when both K and M are large. To see the problem, consider a conventional method where a dedicated pilot slot is assigned to each MT. The pilot slots for different users are orthogonal in time [1]. This is referred to as an orthogonal pilot (OP) method in this paper. OP may result in rate loss due to the use of dedicated pilot slots. Such loss can be severe when K is large or in a high mobility environment with short channel coherent time, denoted as T_c below. Also, the power used for pilot constitutes an extra overhead.

The rate loss can be potentially avoided by superimposing pilots with data [2]–[11], so that all the available time resource is used for data. This is referred to as the superimposed pilot (SP) method. SP introduces interference among pilot and data, which may seriously affect estimation accuracy.

An important issue for SP is to optimize the portion of power used for pilot. This problem has been studied in [3], [6], [9], and [11] under various criteria. Using the replica method, it is shown in [12] that the optimal pilot power overhead tends to zero when K, M and T_c go to infinity simultaneously with fixed ratios. This scheme employs T_c forward error-correction (FEC) codes (each with a different code rate) and superimposed pilots. An extra multiplexed pilot signal is used to enhance initial CSI. The detection process starts from the code with the lowest rate (that is easiest to decode)

Manuscript received January 26, 2017; revised May 14, 2017; accepted May 23, 2017. Date of publication July 11, 2017; date of current version December 22, 2017. This work was supported by the University Grants Committee of the Hong Kong Special Administrative Region, China, under Project AoE/E-02/08, Project CityU 11208114, Project CityU 11217515, and Project CityU 11280216. *(Corresponding author: Chulong Liang.)*

J. Ma was with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. He is now with the Department of Statistics, Columbia University, New York City, NY 10027-6902 USA (e-mail: junjiema2-c@my.cityu.edu.hk).

C. Liang and L. Ping are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: chuliang@cityu.edu.hk; eeliping@cityu.edu.hk).

C. Xu is with the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China (e-mail: chbinxu@fudan.edu.cn).

Digital Object Identifier 10.1109/JSAC.2017.2726019

and progress towards higher rate ones successively. The successfully decoded codewords are used together with the pilots to refine channel estimation. The scheme in [12] is mainly devised for theoretical analysis. It is difficult to implement in practice due to the use of multiple FEC codes of different rates. A similar result is observed in [13, Section V] based on Bayes-optimal estimation. However, [13] considers a system without FEC codes (which are essential for reliable communications).

In this paper, we study the use of OP and SP in non-orthogonal multiple-access (NOMA) [14]–[18] systems. Interleave division multiple access (IDMA) [19], [20] is adopted to separate multiple concurrent transmitting MTs. There is no effort to establish spatial orthogonality. Interference is allowed among data from different MTs, and hence the scheme is a special case of NOMA. Our focus is to compare the performances of OP and SP involving iterative data-aided channel estimation (IDACE) [2], [4], [21]–[24]. As shown in [24], IDACE can provide significantly improved channel estimation and alleviate the pilot contamination problem in massive MIMO.

IDACE works as follows. At the beginning, limited CSIR is acquired using pilots, based on which data are estimated and decoded. Partially decoded data are then used to refine CSIR, which further improves the performances of data estimation and decoding. This process proceeds iteratively until convergence.

We will derive the achievable rate for the IDACE scheme using the relationship between the mutual information-minimum mean squared error (MMSE) (I-MMSE relationship) [25]. A main contribution of this paper is a proof that, under some standard assumptions on message passing decoding, the optimal portion of pilot power tends to zero in SP with Gaussian signaling. This is consistent with the findings in [12] and [13].

Compared with the successive decoding scheme in [12], the IDACE scheme in this paper has much lower complexity and is potentially implementable in practice. We will also show by simulation that, with code optimization, SP can outperform OP when channel coherent time T_c is small and mobility is high, and the difference is significant when K is large. Theoretically, SP has the advantage that its optimal pilot power ratio for Gaussian signaling does not vary with system settings. Thus, the design problem for SP can be much simpler than that for OP.

Notations: Boldface symbols denote matrices or vectors; $x(j)$ denotes the j^{th} component of a vector $\boldsymbol{x}$; $\mathbf{0}$ denotes an all zero matrix; $\text{CN}(\boldsymbol{\mu}, \boldsymbol{C})$ represents circular symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$; $(\cdot)^*$, $(\cdot)^{\text{T}}$ and $(\cdot)^{\text{H}}$ are for complex conjugate, transpose and conjugate transpose, respectively; $\|\cdot\|$ represents the 2-norm of a vector; $\text{diag}\{x_1, x_2, \ldots, x_J\}$ denotes a diagonal matrix with diagonal entries given by $x_1, x_2, \ldots, x_J$.

II. System Model

In this section, we first present the system model. We then introduce two pilot transmission structures: orthogonal pilot (OP) and superimposed pilot (SP).

A. Channel Model

We focus on the uplink of a particular cell with K MTs. Each MT is equipped with a single antenna. The base station (BS) is equipped with M antennas. A received signal at time j on M BS antennas is a length-M vector $\boldsymbol{y}(j)$:

$$\boldsymbol{y}(j) = \sum_{k=1}^{K} \boldsymbol{h}_k x_k(j) + \boldsymbol{\psi}(j), \quad j = 1, \cdots, J, \tag{1a}$$

where $\boldsymbol{h}_k$ is a length-M vector of the channel coefficients from MT k to the BS antennas, $x_k(j)$ a symbol transmitted by MT k at time j and $\boldsymbol{\psi}(j)$ a length-M vector of combined out-of-cell interference and additive white Gaussian noise (AWGN) samples. Define $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]$ and $\boldsymbol{x}(j) = [x_1(j), \cdots, x_K(j)]^{\text{T}}$. We rewrite (1a) as

$$\boldsymbol{y}(j) = \boldsymbol{H}\boldsymbol{x}(j) + \boldsymbol{\psi}(j), \quad j = 1, \cdots, J. \tag{1b}$$

We can further rewrite (1b) in an augmented matrix form

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{\Psi}, \tag{1c}$$

where the j^{th} columns of $\boldsymbol{Y} \in \mathbb{C}^{M\times J}$, $\boldsymbol{X} \in \mathbb{C}^{K\times J}$ and $\boldsymbol{\Psi} \in \mathbb{C}^{M\times J}$ are, respectively, $\boldsymbol{y}(j)$, $\boldsymbol{x}(j)$ and $\boldsymbol{\psi}(j)$. We assume that $\boldsymbol{\Psi}$ consists of uncorrelated entries with zero mean and variance v_Ψ. The k^{th} row of $\boldsymbol{X}$ forms a transmitted frame from MT k. We assume that $\boldsymbol{H}$ is quasi-static; it remains constant within T_c symbols and changes independently in different frames.

For simplicity, we assume throughout this paper that the transmission block length J is equal to the channel coherence time T_c.

B. Power Control

Denote the entries of $\boldsymbol{H}$ by $\{H_{m,k}, \forall m, k\}$. Throughout this paper, we will only consider Rayleigh fading with $H_{m,k} \sim \text{CN}(0, 1)$ and $H_{m,k}$ being independent and identically distributed (IID) for different m and k. Our results in this paper may shed light on other types of power control, but the detailed discussions are beyond the scope of this paper.

C. Transmitter Structure

At the transmitter of MT k, the information sequence $\boldsymbol{b}_k$ is processed by encoder k, producing a symbol sequence $\boldsymbol{d}_k = \{d_k(j), \forall j\}$. We assume that each encoder includes conventional binary FEC coding, user-specific random interleaving (based on interleave division multiple access (IDMA) [19]) and signal mapping (based on a proper constellation). A pilot sequence $\boldsymbol{p}_k = \{p_k(j), \forall j\}$ is then inserted to form the transmit sequence $\boldsymbol{x}_k$. Here $\boldsymbol{x}_k$ is the k^{th} row of $\boldsymbol{X}$ defined in (1c). We also defined a data matrix $\boldsymbol{D}$ and a pilot matrix $\boldsymbol{P}$. Their k^{th} rows are, respectively, $\boldsymbol{d}_k$ and $\boldsymbol{p}_k$. The $(k, j)^{\text{th}}$ entries of $\boldsymbol{P}$ and $\boldsymbol{D}$ denote respectively the pilot and data symbols transmitted at time j by MT k. We assume that the entries of $\boldsymbol{D}$ have zero mean.

D. Orthogonal and Superimposed Pilot Schemes

Let J_d and J_p be the lengths of $\boldsymbol{d}_k$ and $\boldsymbol{p}_k$, respectively. We consider two pilot schemes as illustrated in Fig. 2.

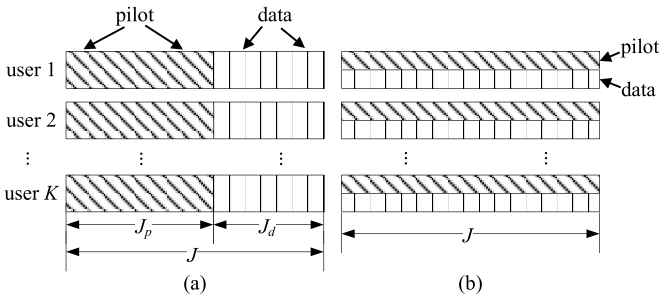


Fig. 2. Frame structures. (a) Orthogonal pilot (OP) scheme. (b) Superimposed pilot (SP) scheme.

1) Orthogonal Pilot (OP) Scheme: With OP, we set $J_p = K$ and $J_d = J - K$. The transmitted signal matrix is given by

$$\boldsymbol{X} = \left[\sqrt{\alpha_P}\boldsymbol{P}, \sqrt{\alpha_D}\boldsymbol{D}\right], \tag{2}$$

where $\boldsymbol{P} \in \mathbb{C}^{K\times J_p}$ and $\boldsymbol{D} \in \mathbb{C}^{K\times J_d}$ are the pilot matrix and data matrix, respectively, and α_P and α_D the corresponding power control factors. Clearly, in OP, pilot and data are orthogonal in time. For OP, we further assume that $\boldsymbol{P}$ contains orthogonal rows, namely, orthogonal pilot sequences are assigned to different users.

2) Superimposed Pilot (SP) Scheme: For SP, we set $J_p = J_d = J$. The transmitted signal matrix is given by

$$\boldsymbol{X} = \sqrt{\alpha_P}\boldsymbol{P} + \sqrt{\alpha_D}\boldsymbol{D}. \tag{3}$$

For SP, we assume that $\boldsymbol{P}$ contains IID entries generated from the same signal constellation as $\boldsymbol{D}$. Note that we can also employ orthogonal pilot sequences for SP, but the IID assumption simplifies our analysis in Section IV.

For ease of analysis, we assume

$$\mathrm{E}\left[|p_k(j)|^2\right] = \mathrm{E}\left[|d_k(j)|^2\right] = 1, \quad \forall k, j. \tag{4a}$$

We also assume that

$$\mathrm{E}\left[|x_k(j)|^2\right] = 1, \quad \forall k, j. \tag{4b}$$

The selections of

$$\alpha_P = \frac{J\cdot t}{J_p} \text{ and } \alpha_D = \frac{J\cdot(1-t)}{J_d} \tag{5a}$$

for OP and

$$\alpha_P = t \text{ and } \alpha_D = 1 - t \tag{5b}$$

for SP ensure $\mathrm{E}\left[|x_k(j)|^2\right] = 1$ for both SP and OP. In (5), $t \in [0, 1]$ represents the ratio of pilot power to total power

$$t \triangleq \frac{\text{pilot power}}{\text{pilot power} + \text{data power}}. \tag{6}$$

III. Iterative Data-Aided Channel Estimation

The receiver structure in Fig. 3 involves the following three modules:

- *Channel Estimator (CE):* Estimate $\boldsymbol{H}$ based on the outputs of DEC.
- *Signal Estimator (SE):* Estimate $\boldsymbol{D}$ based on the outputs of CE and DEC. The FEC coding constraint is ignored in this step.

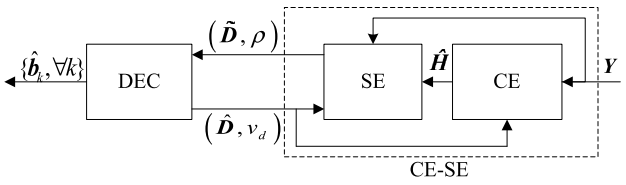


Fig. 3. Receiver structure for IDACE. CE-SE denotes the channel estimation (CE) and the signal estimation (SE) module. DEC consists of a bank of K single-user decoders.

- *Decoder (DEC):* Estimate for $\boldsymbol{D}$ again based on the FEC coding constraint.

The three modules are executed iteratively, forming an iterative data-aided channel estimation (IDACE) process. More details on the function blocks and notations (i.e., $\hat{\boldsymbol{H}}$, $\tilde{\boldsymbol{D}}$, ρ, $\hat{\boldsymbol{D}}$, v_d and $\hat{\boldsymbol{b}}_k$) in Fig. 3 are explained in the subsequent subsections.

A. Modeling of DEC Outputs

Channel estimation (CE) is performed based on DEC feedback $\hat{\boldsymbol{D}}$ and v_d. (The details in generating $\hat{\boldsymbol{D}}$ and v_d will be discussed in Section III-D.) Denote by $\hat{d}_k(j)$ the $(k, j)^{\text{th}}$ entry of $\hat{\boldsymbol{D}}$. Similar to [24], we make the following assumptions on $\{\hat{d}_k(j), \forall k, j\}$.

Assumption 1: (i) Each $\hat{d}_k(j)$ is generated from an observation $s_k(j)$ as:

$$\hat{d}_k(j) = \mathrm{E}\left[d_k(j)\,|s_k(j)\right], \tag{7a}$$

where $s_k(j)$ is modeled as

$$s_k(j) = d_k(j) + \xi_k(j), \quad \forall k, j, \tag{7b}$$

and $\xi_k(j) \sim \mathrm{CN}(0, v_\xi)$ is independent of $d_k(j)$.

(ii) Both $\{d_k(j), \forall k, j\}$ and $\{\xi_k(j), \forall k, j\}$ contain IID entries. Denote by v_d the MSE of $\hat{d}_k(j)$, i.e., $v_d = \mathrm{E}[|d_k(j) - \hat{d}_k(j)|^2]$.

When $d_k(j)$ is BPSK modulated, $s_k(j)$ can be understood as a scaled version of the extrinsic log-likelihood ratio (LLR) of the decoder. Gaussian assumption on the extrinsic LLR is widely adopted in the literature of iterative decoding [26]–[28]. Assumption 1-(i) above can be seen as a generalization of the BPSK case. Assumption 1-(ii) can be approximately ensured by coding over multiple coherence blocks and random interleaving. The identical-variance assumption across k is due to the equal-receive power policy stated earlier (see (4)). This assumption is generally not applicable if the arrival powers are different among MTs, since then the decoding reliabilities vary with k.

The following property is a direct consequence of Assumption 1.

Property 1: When $\{d_k(j), \forall k, j\}$ are IID Gaussian, $\{\hat{d}_k(j), \forall k, j\}$ are also Gaussian since from (7)

$$\hat{d}_k(j) = \mathrm{E}\left[d_k(j)|s_k(j)\right] = \frac{1}{1+v_\xi}s_k(j). \tag{8}$$

B. CE Module

For the CE module, the DEC feedback is treated as the *a priori* mean of the data. At the beginning of the iterative

receiver, $\boldsymbol{D}$ is completely unknown, so $\hat{\boldsymbol{D}} = \boldsymbol{0}$ is used. During iterative process, DEC provides information on $\boldsymbol{D}$ to update $\hat{\boldsymbol{D}}$. Since $\boldsymbol{P}$ is known, we update $\hat{\boldsymbol{X}}$ (*a priori* mean of $\boldsymbol{X}$) as (based on (2) and (3))

$$\hat{\boldsymbol{X}} \equiv \begin{cases} \left[\sqrt{\alpha_P}\boldsymbol{P}, \sqrt{\alpha_D}\hat{\boldsymbol{D}}\right], & \text{for OP,} \\ \sqrt{\alpha_P}\boldsymbol{P} + \sqrt{\alpha_D}\hat{\boldsymbol{D}}, & \text{for SP.} \end{cases} \tag{9}$$

Notice that the selections of α_P and α_D are different from OP and SP, see (5).

CE estimates $\boldsymbol{H}$ by treating $\hat{\boldsymbol{X}}$ as equivalent pilots. For this purpose, we define the unknown part of $\boldsymbol{X}$ as $\Delta\boldsymbol{X} = \boldsymbol{X} - \hat{\boldsymbol{X}}$ and rewrite (1c) as

$$\boldsymbol{Y} = \boldsymbol{H}\hat{\boldsymbol{X}} + \tilde{\boldsymbol{\Psi}}, \tag{10a}$$

where

$$\tilde{\boldsymbol{\Psi}} = \boldsymbol{H}\Delta\boldsymbol{X} + \boldsymbol{\Psi}. \tag{10b}$$

The linear minimum mean squared error (LMMSE) estimator of $\boldsymbol{H}$ based on (10) is given by [29]

$$\hat{\boldsymbol{H}} = \boldsymbol{Y}\boldsymbol{V}_{\tilde{\Psi}}^{-1}\hat{\boldsymbol{X}}^{\mathrm{H}}\left(\boldsymbol{I} + \hat{\boldsymbol{X}}\boldsymbol{V}_{\tilde{\Psi}}^{-1}\hat{\boldsymbol{X}}^{\mathrm{H}}\right)^{-1}, \tag{11}$$

where $\boldsymbol{V}_{\tilde{\Psi}}$ is the covariance matrix of the rows of $\tilde{\boldsymbol{\Psi}}$ and $\boldsymbol{I}$ an identity matrix with an appropriate size. We next provide detailed derivations of $\boldsymbol{V}_{\tilde{\Psi}}$.

From (10b), the m^{th} row of $\tilde{\boldsymbol{\Psi}}$ (denoted as $\tilde{\boldsymbol{\psi}}_m$) is given by

$$\tilde{\boldsymbol{\psi}}_m = \boldsymbol{h}_m\Delta\boldsymbol{X} + \boldsymbol{\psi}_m, \tag{12}$$

where $\boldsymbol{h}_m$ and $\boldsymbol{\psi}_m$ represent the m^{th} rows of $\boldsymbol{H}$ and $\boldsymbol{\Psi}$, respectively. Noting that $\boldsymbol{h}_m \sim \mathrm{CN}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{\psi}_m$ consists of uncorrelated entries with zero-mean and variance υ_ψ, the covariance matrix of $\tilde{\boldsymbol{\psi}}_m$ in (12) is given by

$$\boldsymbol{V}_{\tilde{\psi}} \triangleq \mathrm{E}\left[\tilde{\boldsymbol{\psi}}_m^{\mathrm{T}}\tilde{\boldsymbol{\psi}}_m^*\right] = \mathrm{E}\left[\Delta\boldsymbol{X}^{\mathrm{T}}\boldsymbol{h}_m^{\mathrm{T}}\boldsymbol{h}_m^*\Delta\boldsymbol{X}^*\right] + \upsilon_\psi\boldsymbol{I} \tag{13a}$$

$$= \mathrm{E}\left[\Delta\boldsymbol{X}^{\mathrm{T}}\Delta\boldsymbol{X}^*\right] + \upsilon_\psi\boldsymbol{I} \tag{13b}$$

$$= K \cdot \mathrm{diag}\left[\upsilon_x(1), \cdots, \upsilon_x(J)\right] + \upsilon_\psi\boldsymbol{I}, \tag{13c}$$

where $\upsilon_x(j) \triangleq \mathrm{E}\left[|\Delta x_k(j)|^2\right]$ and (13c) follows from Assumption 1-(ii).

For OP, each transmitted sequence is divided into a pilot part and a data part. The pilot part is known at the receiver so its variance is zero, namely,

$$\text{OP:} \quad \upsilon_x(j) = \begin{cases} 0, & \text{for } 1 \le j \le J_p, \\ \alpha_D \cdot \upsilon_d, & \text{for } J_p + 1 \le j \le J. \end{cases} \tag{14}$$

For SP, each transmitted symbols is a combination of pilot and data, and so

$$\text{SP:} \quad \upsilon_x(j) = \alpha_D \cdot \upsilon_d, \quad \forall j, \tag{15}$$

where υ_d will be discussed in Section III-D.

Note: It is well known that the use of extrinsic information can improve the performance of a turbo-type iterative receiver [30]. Based on this principle, an extrinsic message technique has been devised in [24] for iterative channel estimation and data detection. The technique can be extended to the multiple-user scenario considered in this paper. We omit the details to save space. All numerical results in this paper are based on this extrinsic message technique.

C. SE Module

The FEC constraint is ignored in the SE module to reduce complexity. We consider an SE module performing soft-interference cancelation followed by maximum ratio combining (MRC) [31]:

$$\tilde{\boldsymbol{D}} = \hat{\boldsymbol{D}} + \frac{1}{\sqrt{\alpha_D}} \cdot \left[\left(\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{H}}\right)_{\mathrm{diag}}\right]^{-1}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{Y}}, \tag{16}$$

where $(\cdot)_{\mathrm{diag}}$ denotes an operation that sets all the off-diagonal entries of a matrix to zero. The matrix $\hat{\boldsymbol{Y}}$ is given by

$$\hat{\boldsymbol{Y}} = \begin{cases} \boldsymbol{Y}^D - \hat{\boldsymbol{H}}\hat{\boldsymbol{D}}, & \text{for OP,} \\ \boldsymbol{Y} - \hat{\boldsymbol{H}}\hat{\boldsymbol{X}}, & \text{for SP,} \end{cases} \tag{17}$$

where $\boldsymbol{Y}^D \triangleq \left[\boldsymbol{y}(J_p + 1), \cdots, \boldsymbol{y}(J)\right]$. Notice that $\tilde{\boldsymbol{D}}$, $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{Y}}$ have different sizes for OP and SP.

In (16) and (17), $\hat{\boldsymbol{H}}$ and $\hat{\boldsymbol{X}}$ (resp. $\hat{\boldsymbol{D}}$) are the estimates of $\boldsymbol{H}$ and $\boldsymbol{X}$ (resp. $\boldsymbol{D}$) produced by CE and DEC respectively. We treat $\hat{\boldsymbol{H}}\hat{\boldsymbol{X}}$ (or $\hat{\boldsymbol{H}}\hat{\boldsymbol{D}}$) as an estimate of the received signal. The second term in (16) is a standard MRC operation [31], performing a coherent spatial combining of the information about $\boldsymbol{D}$. The first term in (16) adds back the known part of the useful signal.

In Appendix A-A, we express $\tilde{\boldsymbol{D}}$ into the following form:

$$\tilde{\boldsymbol{D}} = \boldsymbol{D} + \boldsymbol{W}. \tag{18}$$

Assumption 2: $\boldsymbol{W}$ is independent of $\boldsymbol{D}$. Its entries are IID with distribution $w_k(j) \sim \mathrm{CN}\left(0, \rho^{-1}\right)$.

Assumption 2 is commonly adopted in the literature of iterative decoding [28]. More discussions on Assumption 2 can be found in Appendix A-B.

From Assumption 2, the SNR ρ in $\tilde{d}_k(j)$ (the $(k, j)^{\text{th}}$ entry of $\tilde{\boldsymbol{D}}$) is given by

$$\rho = \frac{\mathrm{E}\left[|d_k(j)|^2\right]}{\mathrm{E}\left[|w_k(j)|^2\right]} \equiv \phi(\upsilon_d, t), \tag{19}$$

where $d_k(j)$ and $w_k(j)$ represent the $(k, j)^{\text{th}}$ entry of $\boldsymbol{D}$ and $\boldsymbol{W}$ in (18), respectively. In (19), we write the SNR as $\phi(\upsilon_d, t)$ to emphasize that it is a function of both υ_d and t.

Assumption 3: Fixing t, $\rho = \phi(\upsilon_d, t)$ in (19) is a monotonically decreasing function of υ_d.

Assumption 3 means that the SNR at the output of the CE-SE module increases when the feedback accuracy of DEC improves. This is intuitively reasonable although a rigorous justification might be complicated.

The outputs of SE are $\tilde{\boldsymbol{D}}$ and ρ. They are fed to DEC for further processing. In practice, ρ should be estimated. The simple estimator in [22, eq. (6)] is used for our simulation results in Section V.

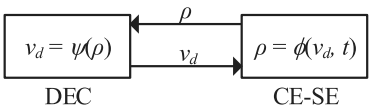


Fig. 4. Transfer function evolution between the CE-SE module and the DEC module.

D. DEC Module

DEC refines the estimate of $\boldsymbol{D}$ using the FEC coding constraint. Recall that the k^{th} row of $\boldsymbol{D}$, denoted as $\boldsymbol{d}_k$ in Section II-C, is the transmitted symbol sequence from MT k. DEC consists of a bank of K constituent decoders for K MTs. Decoding is carried out in a user-by-user way. For convenience, we also include the de-mapping and re-mapping operations in DEC if the entries in $\boldsymbol{D}$ are modulated using a multi-ary constellation. Such decoders are widely discussed for bit-interleaved coded modulation (BICM) [32] and superposition coded modulation (SCM) [33]. We will therefore omit the details.

After decoding, using the extrinsic probability values for the constellation points for each transmitted symbol, we generate an estimate of $\boldsymbol{D}$, denoted by $\hat{\boldsymbol{D}}$. The accuracy of $\hat{\boldsymbol{D}}$ is measured by:

$$\upsilon_d \triangleq \frac{1}{KJ_d}\sum_{k=1}^{K}\sum_{j=1}^{J_d}\mathrm{E}\left[\left|\hat{d}_k(j)-d_k(j)\right|^2\right] \triangleq \psi(\rho). \quad (20)$$

In practice, we need to generate an estimate for υ_d in (20). This can be obtained by using the variances corresponding to the extrinsic probabilities. The details on the calculations for $\hat{\boldsymbol{D}}$ and υ_d can be found in [33]. In the last iteration, $\left\{\hat{\boldsymbol{b}}_k, \forall k\right\}$ (see Fig. 3) are generated using hard decision.

IV. Achievable Rate Analysis

From the above discussions, we can characterize the CE-SE module and the DEC module by the input-output relationship between ρ and υ_d. We use the following transfer functions (cf. (19) and (20))

$$\rho = \phi(\upsilon_d, t) \text{ and } \upsilon_d = \psi(\rho) \quad (21)$$

for the CE-SE module and the DEC module, respectively. See Fig. 4 for an illustration. Note that t is fixed during the process, so (21) defines a recursion between υ_d and ρ. In this section, we will show that the optimal value of t for SP with Gaussian signaling is $t = 0$ based on the matching principle of ϕ and ψ. In general, both ϕ and ψ can be computed using Monte Carlo methods.

A. MSE for DEC

Our analysis is based on the I-MMSE relationship developed in [25]. We first discuss the MSE for the DEC module.

We will assume that DEC provides optimal decoding of $\boldsymbol{d}_k$ based on $\tilde{\boldsymbol{d}}_k$(the k^{th} row of $\tilde{\boldsymbol{D}}$). Note that $\hat{\boldsymbol{d}}_k$ (the k^{th} row of $\hat{\boldsymbol{D}}$) is obtained from the extrinsic probabilities generated by the DEC. From the IID assumptions in Assumption 1 and Assumption 2, we only need to consider a single entry. For notational brevity, we will omit the user and time indices in this subsection.

From Assumption 1, $\hat{d}$ is modeled as an MMSE estimate of d from an effective AWGN observation. Also, from Assumption 2, $\tilde{d}$ is also an AWGN observation for d. Since $\tilde{d}$ is produced by extrinsic probabilities, the noise terms in $\tilde{d}$ is independent of that in $\hat{d}$ [27]. Hence, the information in $\tilde{d}$ and $\hat{d}$ can be combined [27], [28] to produce the *a posteriori* estimate. The MMSE for this *a posteriori* estimate is given by:

$$\mathrm{mmse}(\rho) = \gamma\left[\gamma^{-1}(\psi(\rho)) + \rho\right], \quad (22a)$$

where γ is the scalar MMSE function [25]:

$$\gamma(\rho) = \mathrm{E}\left[|d - \mathrm{E}[d \mid d + w]|^2\right], \quad (22b)$$

and d is a scalar data symbol independent of $w \sim \mathrm{CN}\left(0, \rho^{-1}\right)$.

Here is the rationale behind (22a). From Assumption 1-(i), $\hat{d}$ is obtained as $\hat{d} = \mathrm{E}[d \mid s = d + \xi]$, where $\xi \sim \mathrm{CN}\left(0, \upsilon_\xi\right)$. Also, the MSE of $\hat{d}$ is $\psi(\rho)$ (see (20)). Further, from the IID assumption in Assumption 1, the MSE in (20) reduces to a scalar MMSE:

$$\psi(\rho) = \mathrm{E}\left\{|d - \mathrm{E}[d \mid d + \xi]|^2\right\} = \gamma\left(\upsilon_\xi^{-1}\right), \quad (23)$$

where the second equality follows from the definition of $\gamma(\cdot)$ in (22b). From (23), $\upsilon_\xi^{-1} = \gamma^{-1}(\psi(\rho))$. Combining $\tilde{d}$ and $\hat{d}$ results in an AWGN observation for d with effective SNR $\upsilon_\xi^{-1} + \rho = \gamma^{-1}(\psi(\rho)) + \rho$. The final MSE is therefore given by $\gamma\left(\upsilon_\xi^{-1} + \rho\right) = \gamma\left[\gamma^{-1}(\psi(\rho)) + \rho\right]$, which is the right hand side of (22a).

B. Curve Matching Principle

It is well known that $\psi(\rho)$ should be matched to $\phi(\upsilon_d, t)$ to maximize code rate [27], [28]:

$$\psi(\rho) = \phi^{-1}(\rho, t), \text{ for } \rho \in \left[\rho_{\text{low}}, \rho_{\text{high}}\right], \quad (24)$$

where $\phi^{-1}(\rho, t)$ is the inverse function of $\phi(\cdot, t)$ (with t fixed). The existence of the inverse function is guaranteed by Assumption 3. Note that $\phi(\upsilon_d, t)$ is defined for $\upsilon_d \in [0, 1]$, and the critical values in (24) are given by

$$\rho_{\text{low}} = \phi(1, t) \text{ and } \rho_{\text{high}} = \phi(0, t). \quad (25)$$

For ρ outside the range $\left[\rho_{\text{low}}, \rho_{\text{high}}\right]$, $\psi(\rho)$ is given by [28]

$$\psi(\rho) = \begin{cases} 1, & \text{if } \rho < \rho_{\text{low}}, \\ 0, & \text{if } \rho > \rho_{\text{high}}. \end{cases} \quad (26)$$

Overall,

$$\psi(\rho) = \begin{cases} 1, & \text{if } \rho < \rho_{\text{low}}, \\ \phi^{-1}(\rho, t), & \text{if } \rho_{\text{low}} \le \rho \le \rho_{\text{high}}, \\ 0, & \text{if } \rho > \rho_{\text{high}}. \end{cases} \quad (27)$$

The corresponding MMSE in (22a) becomes

$$\mathrm{mmse}(\rho) = \begin{cases} \gamma(\rho), & \text{if } \rho < \rho_{\text{low}}, \\ \gamma\left[\gamma^{-1}\left(\phi^{-1}(\rho, t)\right) + \rho\right], & \text{if } \rho_{\text{low}} \le \rho \le \rho_{\text{high}}, \\ 0, & \text{if } \rho > \rho_{\text{high}}. \end{cases} \quad (28)$$

C. Achievable Rate

Following the I-MMSE relationship developed in [25], the rate for an FEC code in an AWGN channel can be expressed as [27, Corollary 1], [28][1]

$$R(t) = \int_0^{\infty} \text{mmse}\,(\rho)\,\mathrm{d}\rho \tag{29a}$$

$$= \int_0^{\rho_{\text{low}}} \gamma\ (\rho)\mathrm{d}\rho + \int_{\rho_{\text{low}}}^{\rho_{\text{high}}} \gamma\left[\gamma^{-1}\left(\phi^{-1}(\rho,t)\right)+\rho\right]\mathrm{d}\rho, \tag{29b}$$

where (29b) follows from (28).

Here, $R(t)$ is the rate per user per channel use. Taking the pilot overhead into account, the effective data rate for all K MTs is

$$R_{\text{eff}}\,(t) = \begin{cases} \dfrac{J-K}{J} R\,(t)\cdot K, & \text{for OP,} \\ R\,(t)\cdot K, & \text{for SP.} \end{cases} \tag{30}$$

D. Pilot Power Optimization for SP With Gaussian Signaling

We now focus on SP with Gaussian signaling, i.e., when both $\boldsymbol{P}$ and $\boldsymbol{D}$ contain IID standard Gaussian entries.[2] In this case, the achievable rate in (29) reduces to

$$R\,(t) = \int_0^{\rho_{\text{low}}} \frac{1}{1+\rho}\mathrm{d}\rho + \int_{\rho_{\text{low}}}^{\rho_{\text{high}}} \frac{\phi^{-1}\,(\rho,t)}{1+\phi^{-1}\,(\rho,t)\cdot\rho}\mathrm{d}\rho \tag{31}$$

by noting that $\gamma\,(\rho) = (1+\rho)^{-1}$ (see (8)) for Gaussian signaling [28].

Consider the following rate maximization problem:

$$\max_{t}\ R_{\text{eff}}\,(t)$$
$$s.t.\ 0 \le t \le 1. \tag{32}$$

Since (32) only involves one optimization variable, we could solve (32) by simple grid search. It is an interesting future work to develop faster techniques to optimize t. In general, there is no closed form solution to (32). However, there exists a nice result for SP with Gaussian signaling.

Theorem 1: For SP with Gaussian signaling, $R_{\text{eff}}\,(t)$ is maximized at $t = 0$.

Proof: See Appendix B. ∎

Intuitively, the main difference between data and pilot is that the former is unknown at the receiver while the latter is known. In IDACE, data is gradually turned from "unknown" to "known". A proper portion of transmission power should be used for pilot to trigger the iterative process. Theorem 1 indicates that this portion, represented by t, approaches to zero under Assumptions 1-3. However, perfect curve-matching, which is a condition of Theorem 1, implies an infinite number of iterations. Therefore, in practice, t cannot really vanish given complexity constraint.

[1]For notational brevity, we use nats as the rate unit here. We will change the unit to bits for the numerical results in Section IV-E.

[2]Strictly speaking, entries in the same row of $\boldsymbol{D}$ are correlated due to FEC coding. However, based on the same random interleaving argument as in Appendix A-B, we may assume that the coded symbols are locally independent within a data block.

It is interesting to note that Theorem 1 is consistent with the result in [12], where a hybrid multiplexed-superimposed pilot scheme is considered. (The "superimposed pilot" part is realized through biased signaling.). In [12], it is proved that, when K, M and $T_c \to \infty$ with fixed ratios, the optimal portion of pilots (both the multiplexed part and the superimposed part) goes to zero. Notice that this result is trivial if K is fixed and $T_c \to \infty$, since then the amount of channel coefficients is limited. The required pilot overhead to accurately estimate these coefficients, however large it is, becomes negligible when $T_c \to \infty$. The problem with K, M and $T_c \to \infty$ together is not trivial since then the amount of channel coefficients to be estimated also become infinite.

Compared with the scheme in [12], the system considered in this paper is more practical. The reason is that the scheme in [12] involves a large number of FEC codes with different rates. This makes it difficult to implement in practice. In contrast, our scheme is based on a single FEC code and is therefore easier to realize.

There is another subtle point. Compared with the result in [12], Theorem 1 above does not explicitly require K, M and T_c go to infinity simultaneously. However, Theorem 1 is established based on Assumptions 1-3. We conjecture that Assumptions 1-3 approximately hold when K, M and T_c are large. However, a rigorous analysis is a difficult issue. (Note that correlation analysis is an open problem for the turbo decoding technique as well as other related iterative algorithms.) Nevertheless, numerical results show that quite small t values can indeed ensure good performance in systems with reasonably large K, M and T_c, as discussed later.

E. Numerical Results

Recall that $\boldsymbol{\Psi}$ in (1c) contains both out-of-cell interference and AWGN samples. Let $\boldsymbol{\Psi} = \boldsymbol{\Psi}_I + \boldsymbol{\Psi}_N$ with $\boldsymbol{\Psi}_I$ being out-of-cell interference and $\boldsymbol{\Psi}_N$ being AWGN samples. Further, assume that $\boldsymbol{\Psi}_N$ consists of IID zero-mean Gaussian samples with variance N_0. In the following discussions, the channel SNR is defined as $\text{SNR} = K/N_0$. This (channel) SNR should not be confused with the "effective" SNR ρ at the output of the SE module.

Our modeling of the multi-cell system is similar to that in [24]. The only difference is that [24] considers a single-user scenario while this paper focuses on the multi-user scenario. For OP, we assume that orthogonal pilots are used for the same-cell users, and the set of orthogonal pilots is reused for all cells. In all simulations, we consider a 7-cell cellular system. The large scaling fading parameter for out-of-cell users is 0.1. The total out-of-cell interference power is proportional to $0.6K$. This setting roughly corresponds to a common situation in a cellular system [34].

In the following, we give examples to compare the achievable rates between OP and SP. We use Monte Carlo simulation to obtain the transfer function $\rho = \phi\,(\upsilon_d, t)$. Given t and υ_d, we generate IID extrinsic information $\{s_k(j), \forall k, j\}$ according to (7b) with $\upsilon_\xi = 1/\gamma^{-1}\,(\upsilon_d)$. We then compute $\{\hat{d}_k(j), \forall k, j\}$ according to (7a). We perform channel estimation using the extrinsic version of (11) (see discussions

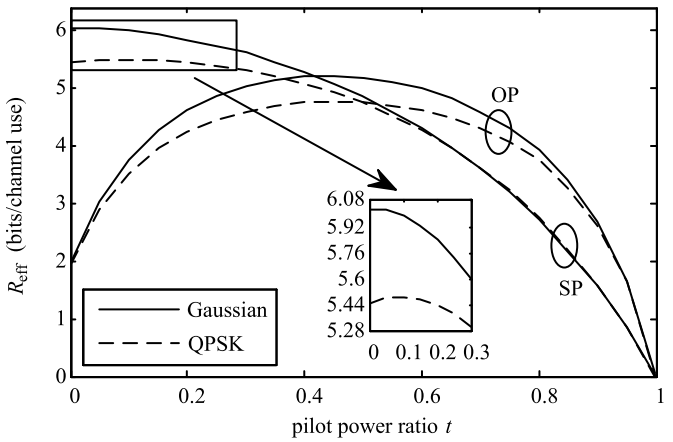


Fig. 5. Achievable rate R_{eff} versus pilot power ratio t at SNR $= 0$ dB for OP and SP under Gaussian and QPSK signaling. $M = 64$, $J = T_c = 16$, $K = 8$.

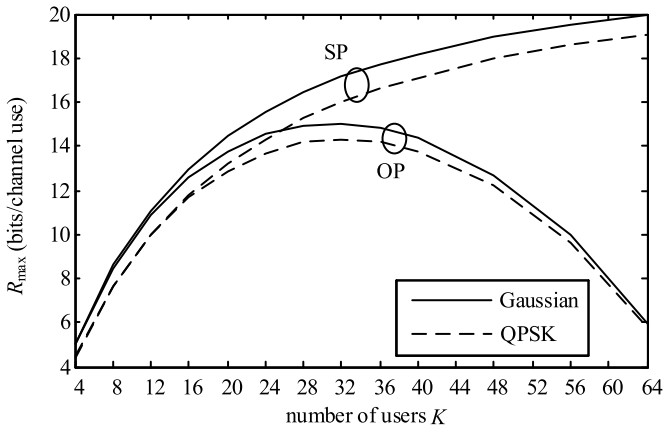


Fig. 6. Maximum achievable rate $R_{\max}$ versus K for both OP and SP under Gaussian and QPSK signaling. SNR $= 0$ dB. $M = 4K$. $J = T_c = 72$.

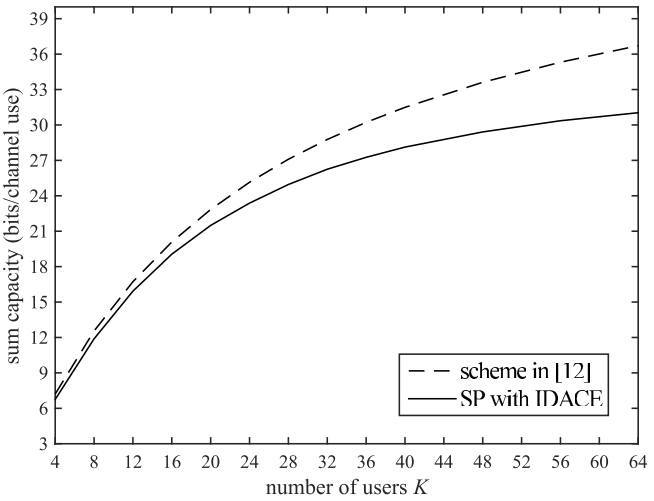


Fig. 7. Comparison of the achievable of [12] and SP with IDACE. Gaussian signaling is employed and $t = 0$. SNR $= 0$ dB. The parameters are the same as those in Fig. 6 except that a single-cell scenario is considered.

at the end of Section III-B) and perform signal estimation using (16). Finally, we measure the SNR ρ according to (19). The achievable rate $R(t)$ is obtained by numerically computing (29b) with the obtained transfer function $\rho = \phi(v_d, t)$. The effective date rate $R_{\text{eff}}(t)$ is then calculated according to (30).

Fig. 5 shows R_{eff} against the pilot power ratio t at SNR $= 0$ dB for both Gaussian and quadrature phase-shift keying (QPSK) signaling. We set $M = 64$, $J = 16$, $K = 8$. From Fig. 5, we have the following observations.

- SP achieves the maximum rate at $t = 0$ when Gaussian signaling is used. This verifies Theorem 1. However, this is not the case for SP with general constrained signaling (QPSK in the figure) and OP.
- For a fixed pilot power ratio t, SP does not always achieve a larger rate than OP.
- For SP with QPSK signaling, $t = 0$ is close to be optimal. Note that this is not the case in the high SNR region; see examples in [35, Ch. 4].

In the following, we compare the achievable rates of OP and SP under their respective optimal pilot power ratios. We denote this maximum rate by $R_{\max}$. We solve the optimization problem in (32) using grid search with grid step $\Delta t = 0.05$.

Fig. 6 shows the maximum achievable rate against K with J fixed. Here we fix $M = 4K$ (see also Fig. 1). Some observations from Fig. 6 are as follows.

- As K increases, the effective data rate of SP monotonically increases. However, this is not true for OP.
- When K is small, the gap between OP and SP is negligible.

The reason for the above observations is that, for OP, the rate loss incurred by pilots is negligible when K is small but becomes serious when K is large. This result shows that SP is more advantageous for systems with a large number of users. As discussed in Introduction (Fig. 1), multi-user concurrent transmission is key to exploit the full potential of a massive MIMO system. Hence, SP is more promising for massive MIMO NOMA systems.

Finally, Fig. 7 compares the achievable rates of the scheme in [12] and SP with an IDACE receiver. Both schemes employ Gaussian signaling with $t \to 0$. We use a single-cell scenario as in [12] (other parameters remain the same as in Fig. 6). We can see that the achievable rate of IDACE is slightly lower than that of the bound derived in [12]. However, IDACE only involves a single FEC code, while the scheme in [12] employs multiple codes with different rates. To reach the bound derived in [12], the number of codes used needs to go to infinity. This is convenient for theoretical analysis but not for practical use. Therefore IDACE indeed provides a promising direction towards a practical solution.

V. System Design via Irregular LDPC Optimization

In previous section, we have compared the achievable rate of two pilot transmission schemes numerically. In this section, we design practical FEC codes to realize the promised rate for general constellations.

A. Code Optimization

When the pilot power ratio t and other system parameters are fixed, the transfer function $\phi(\cdot, t)$ for the CE-SE module

is fixed. System optimization then becomes designing an FEC code for which the transfer function $\psi(\cdot)$ is matched to $\phi^{-1}(\cdot, t)$ (cf. (27)). A standard approach for this purpose is to adopt an irregular binary low-density parity-check (LDPC) code with properly designed degree distributions [36]–[38]. Generally, we have to optimize degree distributions for both variable nodes and check nodes. However, this is very difficult for optimization. Following the method in Appendix 5G of [38], we set the check node distribution manually and optimize the variable degree distribution using standard linear programming. Below, we denote the variable degree distribution and the check node distribution by $\lambda(x)$ and $\eta(x)$, respectively.

Given t and a target effective transmission rate (denoted by R_{target}), the procedure for designing irregular LDPC codes are summarized as follows.

(i) Notice that $\phi(v_d, t)$ depends on N_0 implicitly. Here, we temporarily introduce a notation $\phi(v_d, t, N_0)$. Fixing other parameters, there exist a one-to-one correspondence between N_0 and $R_{\text{eff}}(t)$ in (30), and so the operating value of N_0 (denoted by N_0^*) can be determined from the target rate R_{target}. The function $\phi(v_d, t) = \phi\left(v_d, t, N_0^*\right)$ is the target transfer function to be matched.

(ii) Design an LDPC code whose $\psi(\rho)$ function is matched to $\phi(v_d, t)$ (i.e., (27) is satisfied). This is achieved by properly choosing $\eta(x)$ and $\lambda(x)$ using the method described in Appendix 5G of [38].

In general, $\text{SNR}_{\text{target}} = K/N_0^*$ is different for different t. The optimal t corresponds to the one that minimizes $\text{SNR}_{\text{target}}$.

B. Simulation Examples

We now provide simulation results for both OP and SP with optimized LDPC codes. All simulation results here are based on QPSK signaling. Gaussian signaling can be approached by superposition coded modulation (SCM), see examples in [28].

For the simulation results in this section, we fix the transmission rate to be 6.4 bits/channel use (i.e., 0.8 bits/user/channel use). The minimum SNRs to achieve this rate are found to be 5.7 dB for OP and 2.0 dB for SP. The corresponding optimal pilot power ratios are 0.5 and 0.15, respectively. Note that the optimal pilot power for SP is not zero (but still smaller than that for OP). The reason is that we use QPSK modulation here, not Gaussian.

For OP, the optimized check node distribution and variable node distribution are given by $\eta(x) = x^{19}$ and $\lambda(x) = 0.3979x + 0.0542x^2 + 0.2380x^{13} + 0.0362x^{14} + 0.2737x^{49}$, respectively. For SP, the check and variable nodes distributions are $\eta(x) = 0.15x^2 + 0.35x^5 + 0.5x^{14}$ and $\lambda(x) = 0.3820x + 0.0768x^2 + 0.1386x^{11} + 0.0846x^{12} + 0.3180x^{49}$, respectively.

Fig. 8 shows the simulation results for OP and SP with IDACE receivers. The sum-product algorithm (SPA) [39] is used for LDPC decoding. To reduce the computational complexity, we merge the inner SPA iterations with the outer IDACE iterations, i.e., one SPA iteration per IDACE iteration. From Fig. 8, we see that SP outperforms OP by around 2 dB at BER $= 10^{-5}$. Fig. 8 also plots the BER performances

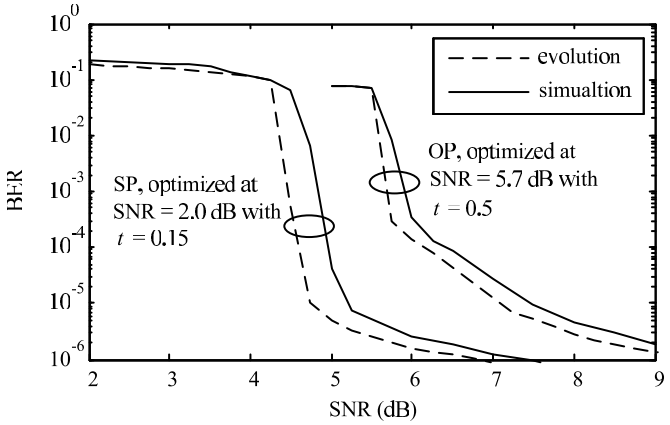


Fig. 8. Simulation results (solid lines) and evolution analysis (dashed lines) for optimized LDPC codes for OP and SP. The transmission rate is 0.8 bits per user. The number of iterations is 50. $M = 64$, $J = T_c = 16$, $K = 8$. Coding rates are 0.7933 and 0.4122 for OP and SP respectively. After QPSK modulation and pilot insertion, the rates per user are approximately 0.8 for both SP and OP. The codeword length is fixed at 2^{17}. (Note that, due to rate difference, a codeword contains less information bits in SP than that in OP.)

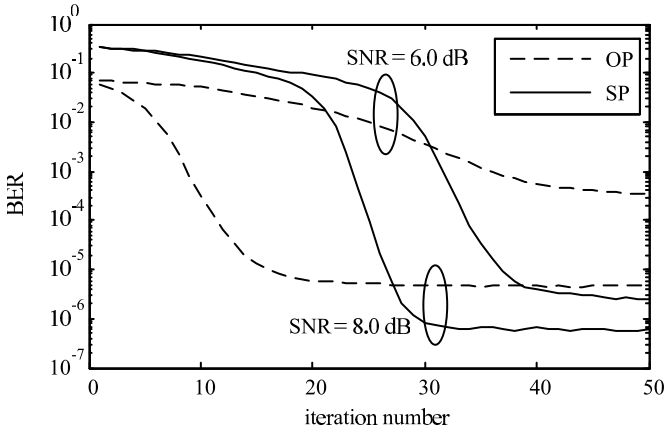


Fig. 9. Convergence behaviors of OP and SP. The parameters are the same as those in Fig. 8. The inner SPA iterations for LDPC decoding are merged into the outer IDACE loop.

predicted using evolution [24]. We see that the prediction is reasonably accurate.

Fig. 9 shows the convergence behaviors for OP and SP at different SNRs. At low SNR (= 6.0 dB), OP and SP has a similar convergence speed. At high SNR (= 8.0 dB), SP converges slower than OP. However, SP outperforms OP in terms of convergent BERs for both cases. We have observed that (results not shown here) the number of iterations can be significantly reduced for a relatively large t.

The above examples show that SP can outperform OP both in theory and in practice when channel coherent time T_c is small in a high mobility environment. Despite this advantage, we found that, when t is very small, it is difficult to match $\psi(\rho)$ with $\phi(v_d, t)$ for SP using LDPC code design. Other channel coding schemes [40], [41] may help on this issue. Also, the spatial coupling technique [42] is promising in relaxing the requirement of tight curve matching, which is an interesting future work. Overall, we observed that, without code optimization, OP with IDACE can perform well if T_c

is sufficiently large relative to K, since then the loss due to dedicated pilot slots is relatively low and IDACE can suppress a major part of interference.

VI. Conclusions

In this paper, we presented an iterative data-aided channel estimation (IDACE) scheme. Two pilot structures, OP and SP, are discussed. Based on several assumptions, we proved that the optimal portion of pilot power tends to zero for SP with Gaussian signaling. This implies that SP can potentially avoid the rate loss and power overhead related to the use of pilots. We have provided numerical results to verify the above statement. We showed that SP can outperform OP in a high mobility environment with a large K. However, as seen from Figs. 5 and 8, the advantages of SP over OP rely on the optimizations on t as well as code structure.

The work in this paper is preliminary. We observed several issues in system design. First, it is difficult to match the transfer function of an FEC code with that of the CE-SE module. We are investigating alternative approaches, such as spatial coupling, to the problem. Second, Theorem 1 in our paper requires Gaussian signaling. We are considering using superposition coded modulation (SCM) [33], [43] to approach Gaussian signaling. Third, we are seeking the extension of the results in this paper to unequal receive-power scenarios.

Appendix A
Estimation Distortion of SE Module

For simplicity, we only discuss SP. Justifications of Assumption 1 are similar for OP.

A. SE Module for SP

Recall from (16), for SP, we have

$$\tilde{\boldsymbol{D}} = \hat{\boldsymbol{D}} + \frac{1}{\sqrt{\alpha_D}} \cdot \left[\left(\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{H}} \right)_{\mathrm{diag}} \right]^{-1} \hat{\boldsymbol{H}}^{\mathrm{H}} \left(\boldsymbol{Y} - \hat{\boldsymbol{H}} \hat{\boldsymbol{X}} \right). \quad (33)$$

Consider the $(k, j)^{\mathrm{th}}$ entry of (33)

$$\tilde{d}_k(j) = \hat{d}_k(j) + \frac{1}{\sqrt{\alpha_D} \left\| \hat{\boldsymbol{h}}_k \right\|^2} \hat{\boldsymbol{h}}_k^{\mathrm{H}} \left(\boldsymbol{y}(j) - \sum_{k'=1}^{K} \hat{\boldsymbol{h}}_{k'} \hat{x}_{k'}(j) \right), \quad (34)$$

where $\hat{x}_{k'}(j)$ is the $(k', j)^{\mathrm{th}}$ entry of $\hat{\boldsymbol{X}}$. For brevity, we omit the time index j in the rest of this Appendix. Using the definitions of $\boldsymbol{y}(j)$ in (1a), we can rewrite (34) as

$$\tilde{d}_k = \hat{d}_k + \frac{\hat{\boldsymbol{h}}_k^{\mathrm{H}}}{\sqrt{\alpha_D} \left\| \hat{\boldsymbol{h}}_k \right\|^2} \left(\sum_{k'=1}^{K} \boldsymbol{h}_{k'} x_{k'} + \boldsymbol{\psi} - \sum_{k'=1}^{K} \hat{\boldsymbol{h}}_{k'} \hat{x}_{k'} \right) \quad (35a)$$

$$= \hat{d}_k + \frac{\Delta x_k}{\sqrt{\alpha_D}} + \frac{\hat{\boldsymbol{h}}_k^{\mathrm{H}} \Delta \boldsymbol{h}_k x_k}{\sqrt{\alpha_D} \left\| \hat{\boldsymbol{h}}_k \right\|^2} + \frac{\sum_{k' \neq k}^{K} \hat{\boldsymbol{h}}_k^{\mathrm{H}} \left(\boldsymbol{h}_{k'} x_{k'} - \hat{\boldsymbol{h}}_{k'} \hat{x}_{k'} \right) + \hat{\boldsymbol{h}}_k^{\mathrm{H}} \boldsymbol{\psi}}{\sqrt{\alpha_D} \left\| \hat{\boldsymbol{h}}_k \right\|^2} \quad (35b)$$

where $\Delta \boldsymbol{h}_k \equiv \boldsymbol{h}_k - \hat{\boldsymbol{h}}_k$ and $\Delta x_k \equiv x_k - \hat{x}_k$. From (3) and (9), we have

$$\Delta x_k = x_k - \hat{x}_k \quad (36a)$$

$$= \left(\sqrt{\alpha_P} p_k + \sqrt{\alpha_D} d_k \right) - \left(\sqrt{\alpha_P} p_k + \sqrt{\alpha_D} \hat{d}_k \right) \quad (36b)$$

$$= \sqrt{\alpha_D} \left(d_k - \hat{d}_k \right). \quad (36c)$$

Substituting (36) into (35) yields

$$\tilde{d}_k = d_k + w_k = d_k + \frac{1}{\sqrt{\alpha_D}} \tilde{w}_k, \quad (37a)$$

where

$$\tilde{w}_k \triangleq \frac{\hat{\boldsymbol{h}}_k^{\mathrm{H}} \Delta \boldsymbol{h}_k}{\left\| \hat{\boldsymbol{h}}_k \right\|^2} x_k + \frac{\hat{\boldsymbol{h}}_k^{\mathrm{H}} \sum_{k' \neq k}^{K} \left(\boldsymbol{h}_{k'} x_{k'} - \hat{\boldsymbol{h}}_{k'} \hat{x}_{k'} \right)}{\left\| \hat{\boldsymbol{h}}_k \right\|^2} + \frac{\hat{\boldsymbol{h}}_k^{\mathrm{H}} \boldsymbol{\psi}}{\left\| \hat{\boldsymbol{h}}_k \right\|^2}. \quad (37b)$$

B. Justification of Property 1

The first term in (37b) is a self-interference term resulting from inaccurate channel estimate of $\boldsymbol{h}_k$. The second term and third term in (37b) represent intra-cell interferences and out-of-cell interferences (and noise), respectively, to user k. When K is large, based on the central limit theorem, we can approximate w_k by a Gaussian random variable.

The independency assumption (between d_k and w_k, and also among $\{w_k, \forall k\}$) is a standard approximation to simplify analysis for iterative systems [28], [33].

The independency assumption on $\{w_k(j), \forall k, j\}$ for different j can be ensured using random interleaving after FEC coding, and transmitting a codeword over many coherence blocks. Similar treatments have been widely used in the analysis for turbo-type iterative decoding algorithm [28], [33].

C. A Property of $\phi(v_d, t)$ for SP with Gaussian Signaling

In this section, we discuss a property of $\phi(v_d, t)$ defined in (19) for SP with Gaussian signaling. This property is crucial for the proof of Theorem 1 in Appendix B.

Before the discussions, we first show that $\hat{\boldsymbol{X}}$ is Gaussian when both $\boldsymbol{P}$ and $\boldsymbol{D}$ are Gaussian. Recall from (9) and (5b) that

$$\hat{x}_k = \sqrt{t} \cdot p_k + \sqrt{1-t} \cdot \hat{d}_k. \quad (38)$$

From Property 1, $\hat{d}_k$ is Gaussian when d_k is Gaussian. When p_k is also Gaussian, $\hat{x}_k$ in (38) is Gaussian. Then, the distribution of $\hat{x}_k$ is completely determined by its variance[3]:

$$\mathrm{E}\left[|\hat{x}_k|^2 \right] = t \cdot \mathrm{E}\left[|p_k|^2 \right] + (1-t) \cdot \mathrm{E}\left[|\hat{d}_k|^2 \right] \quad (39a)$$

$$= t + (1-t)(1 - v_d) \quad (39b)$$

$$= 1 - v_x \quad (39c)$$

where $v_x = (1-t)v_d$ and (39b) follows from the orthogonality property for MMSE estimation [29]: $\mathrm{E}\left[|\hat{d}_k|^2 \right] = \mathrm{E}\left[|d_k|^2 \right] - \mathrm{E}\left[|\hat{d}_k - d_k|^2 \right] = 1 - v_d$.

[3]All Gaussian variables discussed in this section have zero-means.

Property 2: When Gaussian signaling is employed, $\phi(v_d, t)$ for SP can written as

$$\phi(v_d, t) \equiv (1-t) \cdot \theta[(1-t) \cdot v_d], \tag{40}$$

where $\theta(v_x)$ is not a function of t if v_x is fixed.

The justifications of Property 2 are as follows. Recall that $\mathrm{E}\left[|d_k|^2\right] = 1$ in (4a) and using $\alpha_D = 1 - t$ (cf. (5b)), we can compute the average SNR for the modeling in (37) as

$$\phi(v_d, t) = \frac{\mathrm{E}\left[|d_k|^2\right]}{\mathrm{E}\left[|\tilde{d}_k - d_k|^2\right]} = (1-t) \cdot \underbrace{\frac{1}{\mathrm{E}\left[|\tilde{w}_k|^2\right]}}_{\theta(v_d, t)}. \tag{41}$$

Here, $\phi(v_d, t)$ is not a function of k since, due to the symmetry of the problem, all $\{\tilde{w}_k, \forall k\}$ have identical distributions. In (41), we have written the SNR into the following form

$$\phi(v_d, t) = (1-t) \cdot \theta(v_d, t). \tag{42}$$

From (10) and (11), $\hat{\boldsymbol{H}}$ is completely determined by $\{\boldsymbol{H}, \boldsymbol{X}, \hat{\boldsymbol{X}}, \boldsymbol{\Psi}\}$. Therefore, from (37b), $\tilde{w}_k$ is also determined by $\{\boldsymbol{H}, \boldsymbol{X}, \hat{\boldsymbol{X}}, \boldsymbol{\Psi}\}$. Since the entries of $\hat{\boldsymbol{X}}$ are IID zero-mean Gaussian, their distribution is completely specified by the variance $1 - v_x$ (see (39)). Although $v_x = (1-t)v_d$ is a function of t, once v_x is fixed, the distribution of $\hat{\boldsymbol{X}}$ does not depend on the individual values of v_d and t.

Remarks: (i) Property 2 does not hold for general non-Gaussian signaling. This is because $\hat{x}_k$ in (38) is a combination of two random variables that have different distributions, and the distribution after combining clearly depends on the weighting coefficient. (ii) Property 2 does not hold for OP even with Gaussian signaling. In this case, $\hat{\boldsymbol{x}}_k = \left[\sqrt{\alpha_P}\boldsymbol{p}_k, \sqrt{\alpha_D}\hat{\boldsymbol{d}}_k\right]$ is still Gaussian distributed. However, the entries of $\hat{\boldsymbol{x}}$, consisting of both pilots and data estimates, have different variances. Hence, the distribution of the sequence $\hat{\boldsymbol{x}}_k$ depends on both v_d and t, and cannot be determined by a single variable $v_x = (1-t)v_d$.

APPENDIX B
PROOF OF THEOREM 1

For SP, maximizing $R_{\mathrm{eff}}(t)$ is equivalent to maximizing $R(t)$. From Property 2 in Appendix A, with Gaussian signaling, $\phi(v_d, t)$ for SP can expressed as

$$\rho = \phi(v_d, t) \equiv (1-t) \cdot \theta[(1-t)v_d], \tag{43}$$

where $\theta(v_x)$ is not a function of t if v_x is fixed (namely, the mapping does not depend on t). From (43), we have

$$v_d = \phi^{-1}(\rho, t) = \frac{1}{1-t}\theta^{-1}\left(\frac{\rho}{1-t}\right), \tag{44}$$

where $\phi^{-1}(\cdot,\cdot)$ is the inverse function of $\phi(\cdot,\cdot)$ for the first argument. Substituting (44) into (31), the achievable rate for Gaussian signaling can be written as

$$R(t) = \int_0^{\rho_{\mathrm{low}}} \frac{1}{1+\rho}\mathrm{d}\rho + \int_{\rho_{\mathrm{low}}}^{\rho_{\mathrm{high}}} \frac{\theta^{-1}\left(\frac{\rho}{1-t}\right)}{1+\theta^{-1}\left(\frac{\rho}{1-t}\right)\frac{\rho}{1-t}} \frac{1}{1-t}\mathrm{d}\rho \tag{45a}$$

$$= \int_0^{\rho_{\mathrm{low}}} \frac{1}{1+\rho}\mathrm{d}\rho + \int_{\frac{\rho_{\mathrm{low}}}{1-t}}^{\frac{\rho_{\mathrm{high}}}{1-t}} \frac{\theta^{-1}(\rho')}{1+\theta^{-1}(\rho')\cdot\rho'}\mathrm{d}\rho', \tag{45b}$$

where we made a variable change $\rho' = \frac{\rho}{1-t}$ in (45b). From (25) and (43), the upper limit of the second integral in (45b) becomes:

$$\frac{\rho_{\mathrm{high}}}{1-t} = \frac{\phi(0,t)}{1-t} = \theta[(1-t)\cdot 0] = \theta(0), \tag{46}$$

which is not a function of t since the function $\theta(\cdot)$ does not depend on t. Therefore,

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \frac{\mathrm{d}\rho_{\mathrm{low}}}{\mathrm{d}t}\frac{1}{1+\rho_{\mathrm{low}}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\rho_{\mathrm{low}}}{1-t}\left.\frac{\theta^{-1}(\rho')}{1+\theta^{-1}(\rho')\rho'}\right|_{\rho'=\frac{\rho_{\mathrm{low}}}{1-t}} \tag{47a}$$

$$= \frac{\mathrm{d}\rho_{\mathrm{low}}}{\mathrm{d}t}\frac{1}{1+\rho_{\mathrm{low}}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\rho_{\mathrm{low}}}{1-t}\frac{1-t}{1+\rho_{\mathrm{low}}} \tag{47b}$$

where (47b) is from (see (25) and (43))

$$\theta^{-1}\left(\frac{\rho_{\mathrm{low}}}{1-t}\right) = 1-t. \tag{48}$$

Further manipulation of (47) shows that

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \frac{\mathrm{d}\rho_{\mathrm{low}}}{\mathrm{d}t}\frac{1}{1+\rho_{\mathrm{low}}} + \left(\frac{-1}{1-t}\frac{\mathrm{d}\rho_{\mathrm{low}}}{\mathrm{d}t} - \frac{\rho_{\mathrm{low}}}{(1-t)^2}\right)\frac{1-t}{1+\rho_{\mathrm{low}}}$$

$$= \frac{-\rho_{\mathrm{low}}}{(1+\rho_{\mathrm{low}})(1-t)} < 0. \tag{49}$$

The last inequality holds since $\rho_{\mathrm{low}} \geq 0$ and $1 - t \geq 0$. This concludes our proof.

REFERENCES

[1] 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.

[2] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Rio de Janeiro, Brazil, Dec. 1999, pp. 2162–2166.

[3] J. K. Tugnait and X. Meng, "On superimposed training for channel estimation: Performance analysis, training power allocation, and frame synchronization," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 752–765, Feb. 2006.

[4] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave division multiple access (IDMA) communication systems," in *Proc. 3rd Int. Symp. Turbo Codes Rel. Topics*, Brest, France, Sep. 2003, pp. 173–180.

[5] H. Schoeneich and P. A. Hoeher, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP J. Adv. Signal Process.*, vol. 2006, pp. 081729-1–081729-15, Aug. 2006.

[6] M. Coldrey and P. Bohlin, "Training-based MIMO systems—Part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[7] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, Jun. 2014.

[8] Y. Chen, T. Wild, and F. Schaich, "Realizing asynchronous massive MIMO with trellis-based channel estimation and superimposed pilots," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1601–1606.

[9] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multicell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[10] B. Mansoor, S. J. Nawaz, and S. M. Gulfam, "Massive-MIMO sparse uplink channel estimation using implicit training and compressed sensing," *Appl. Sci.*, vol. 7, no. 1, p. 63, Jan. 2017.

[11] K. Upadhya, S. A. Vorobyov, and M. Vehkaperä, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[12] K. Takeuchi, R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[13] C.-K. Wen, Y. Wu, K.-K. Wong, R. Schober, and P. Ting, "Performance limits of massive MIMO systems based on Bayes-optimal inference," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1783–1788.

[14] P. Wang, J. Xiao, and L. Ping, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 4–11, Sep. 2006.

[15] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[16] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98-B, no. 3, pp. 403–414, Mar. 2015.

[17] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[18] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, "Wireless-powered communications with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8422–8436, Dec. 2016.

[19] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[20] P. Wang and L. Ping, "On maximum eigenmode beamforming and multi-user gain," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4170–4186, Jul. 2011.

[21] P. D. Alexander and A. J. Grant, "Iterative channel and information sequence estimation in CDMA," in *Proc. IEEE 6th Int. Symp. Spread Spectr. Techn. Appl.*, vol. 2. Parsippany, NJ, USA, Sep. 2000, pp. 593–597.

[22] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1450–1460, Aug. 2001.

[23] M. Zhao, Z. Shi, and M. C. Reed, "Iterative turbo channel estimation for OFDM system over rapid dispersive fading channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3174–3184, Aug. 2008.

[24] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[25] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[26] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[27] K. Bhattad and K. R. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 22–38, Jan. 2007.

[28] X. Yuan, L. Ping, C. Xu, and A. Kavcic, "Achievable rates of MIMO systems with linear precoding and iterative LMMSE detection," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7073–7089, Nov. 2014.

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

[30] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[32] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.

[33] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 995–1004, Aug. 2009.

[34] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley, III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, May 1991.

[35] J. Ma, "Iterative channel estimation methods in large antenna systems," Ph.D. dissertation, Dept. Electron. Eng., City Univ. Hong Kong, Hong Kong, 2015.

[36] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.

[37] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[38] X. Yuan, "Low-complexity iterative detection in coded linear systems," Ph.D. dissertation, Dept. Electron. Eng., City Univ. Hong Kong, 2008.

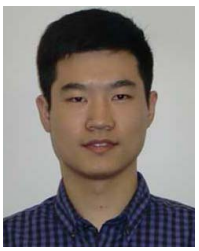
[39] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[40] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd Int. Symp. Turbo Codes Rel. Topics*, Brest, France, Sep. 2000, pp. 1–8.

[41] J. Wang, S. X. Ng, A. Wolfgang, L. L. Yang, S. Chen, and L. Hanzo, "Near-capacity three-stage MMSE turbo equalization using irregular convolutional codes," in *Proc. 4th Int. Symp. Turbo Codes Rel. Topics*, Munich, Germany, Apr. 2006, pp. 1–6.

[42] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "Universal codes for the Gaussian MAC via spatial coupling," in *Proc. 49th Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2011, pp. 1801–1808.

[43] P. A. Hoeher and T. Wo, "Superposition modulation: Myths and facts," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 110–116, Dec. 2011.



Junjie Ma received the B.E. degree from Xidian University, China, in 2010, and the Ph.D. degree from the City University of Hong Kong in 2015. He was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong, from 2015 to 2016. Since 2016, he has been a Post-Doctoral Researcher with the Department of Statistics, Columbia University. His current research interests include statistical signal processing, compressed sensing, and optimization methods.

Chulong Liang received the B.E. degree in communication engineering and the Ph.D. degree in communication and information systems from Sun Yat-sen University, Guangzhou, China, in 2010 and 2015, respectively. He is currently a Post-Doctoral Fellow with the City University of Hong Kong, Hong Kong, China. His current research interests include channel coding theory and its applications to communication systems.

Chongbin Xu received the B.S. degree in information engineering from Xi'an Jiaotong University in 2005 and the Ph.D. degree in information and communication engineering from Tsinghua University in 2012. Since 2014, he has been with the Department of Communication Science and Engineering, Fudan University, China. His research interests are in the areas of signal processing and communication theory, including linear precoding, iterative detection, and random access techniques.

Li Ping (S'87–M'91–SM'06–F'10) received the Ph.D. degree from Glasgow University in 1990. He was a Lecturer with the Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and a member of research staff with Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since 1996, where he is currently a Chair Professor. He received the British Telecom-Royal Society Fellowship in 1986, the IEE J. J. Thomson premium in 1993, the Croucher Foundation Award in 2005, and the British Royal Academy of Engineering Distinguished Visiting Fellowship in 2010. He served as a member of the Board of Governors of the IEEE Information Theory Society from 2010 to 2012.