Minimizing Pilot Overhead in Cell-Free Massive MIMO Systems via Joint Estimation and Detection

Author(s):
Song, Haochuan; You, Xiaohu; Zhang, Chuan; Tirkkonen, Olav; Studer, Christoph

Minimizing Pilot Overhead in Cell-Free Massive MIMO Systems via Joint Estimation and Detection

Haochuan Song[1,2,4], Xiaohu You[2], Chuan Zhang[1,2], Olav Tirkkonen[3], and Christoph Studer[4]

[1]Lab of Efficient Architectures for Digital-communication and Signal-processing (LEADS)
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
[3]Aalto University, Aalto, Finland   [4]Cornell Tech, New York, NY, USA
e-mail: [2]{hcsong, xhyu, chzhang}@seu.edu.cn; [3]olav.tirkkonen@aalto.fi; [4]studer@cornell.edu

Abstract—We propose a joint channel estimation and data detection (JED) algorithm for cell-free massive multi-user (MU) multiple-input multiple-output (MIMO) systems. Our algorithm yields improved reliability and reduced latency while minimizing the pilot overhead of coherent uplink transmission. The proposed JED method builds upon a novel non-convex optimization problem that we solve approximately and efficiently using forward-backward splitting. We use simulation results to demonstrate that our algorithm achieves robust data transmission with more than $3\times$ reduced pilot overhead compared to orthogonal training in a 128 antenna cell-free massive MU-MIMO system in which 128 users transmit data over 128 time slots.

I. INTRODUCTION

Cell-free massive multi-user (MU) multiple-input multiple-output (MIMO) wireless systems promise significant improvements in spectral efficiency compared to traditional cellular systems [1]–[3]. While the majority of results on cell-free systems analyze achievable rates [4]–[6], one of the remaining design challenges is the increased training overhead caused by the fact that more user equipments (UE) are communicating simultaneously in a potentially very large area. In fact, naïve channel estimation methods in the uplink (UEs communicate to the distributed antennas) that use orthogonal pilot sequences would require at least one time slot per UE, which inevitably reduces spectral efficiency for cell-free systems with hundreds or even thousands of UEs communicating at the same time. While non-orthogonal training sequences can reduce the training overhead, they generally reduce performance considerably—especially when used together with conventional data detectors [7].

A. Contributions

To reduce the overhead of pilot-based channel estimation for cell-free systems, we propose a novel joint channel estimation and data detection (JED) algorithm that exploits (i) channel sparsity and (ii) the fact that the UE transmit signals are taken from a bounded constellation set (e.g., QPSK). Our JED algorithm uses forward-backward splitting (FBS) [8] to efficiently solve a biconvex optimization problem,

which results in excellent performance in terms of error-vector magnitude (EVM), bit error-rate (BER), and channel-estimation mean-square error (MSE). We perform simulations for cell-free massive MU-MIMO systems, which demonstrate that JED enables near-optimal performance with more than $3\times$ reduced training overhead compared to orthogonal training. We furthermore show that our JED algorithm supports cell-free systems in which the number of UEs is comparable to the number of access-point antennas, which is a relevant scenario in densely-populated (e.g., urban) areas.

B. Relevant Prior Art

Joint channel estimation and data detection has been studied in [9]–[16] for small-scale MIMO systems. For massive single-input multiple-output (SIMO) wireless systems, an efficient JED algorithm and VLSI design has been proposed in [17]. For massive MU-MIMO systems, suitable algorithms have been proposed only recently [18]–[21]. These methods have been designed mostly for Rayleigh fading channels and/or for systems in which the number of basestation (BS) antennas are significantly larger than the number of UEs. In contrast, our proposed JED algorithm specifically exploits sparsity in cell-free massive MU-MIMO channels and the boundedness of QAM constellations, which enables reliable data transmission, even in systems where the number of access-point antennas is equal to the number of UEs.

Most results for cell-free systems [7], [22]–[24] assume limited coherence time or a large number of UEs communicating at the same time, which necessitates non-orthogonal pilot sequences. However, practical, low-complexity data detectors that work satisfactory with non-orthogonal pilot sequences are missing—our algorithm fills in exactly this void by using the data signals together with nonorthogonal training to minimize pilot overhead while enabling reliable data transmission.

C. Notation

Lower case and upper case boldface letters denote matrices and vectors, respectively. We use A_{ij}, $\mathbf{a}_j$, and b_k to represent the entry in the ith row and jth column of the matrix $\mathbf{A}$, the jth column of matrix $\mathbf{A}$, and the kth element in the vector $\mathbf{b}$, respectively. We use the superscripts *, ^{T} and ^{H} to denote the complex conjugate, transpose, and Hermitian conjugate of a matrix or vector. The Frobenius norm of a matrix $\mathbf{A}$ is

$\|\mathbf{A}\|_F$. The quantities $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_{\widetilde{\infty}}$ stand for the ℓ_1 and $\ell_{\widetilde{\infty}}$-norm defined in [25, Sec. V] of the vectorized matrix $\mathbf{A}$.

II. Cell-Free Massive MU-MIMO System Model

We now introduce the cell-free massive MU-MIMO system and summarize the channel model.

A. System Model

We focus on the uplink in a cell-free massive MU-MIMO system with B distributed receive antennas and U single-antenna UEs. We assume a block-fading scenario with a coherence time of $K = T + D$ time slots, where T time slots are reserved for pilot-based training and D time slots are reserved for payload data. The input-output relation of the considered frequency-flat[1] system is given by [26]

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{C}^{B \times K}$ is the receive-signal matrix, $\mathbf{H} \in \mathbb{C}^{B \times U}$ is the MIMO channel matrix, $\mathbf{S} \in \mathbb{C}^{U \times K}$ is the transmit-signal matrix with entries chosen from the constellation set $\mathcal{Q}$ and $\mathbf{N} \in \mathbb{C}^{B \times K}$ is the additive noise matrix, whose entries are assumed to be i.i.d. circularly-symmetric complex Gaussian with variance N_0 per complex entry. To simplify notation, we separate training from payload by rewriting (1) as follows:

$$[\mathbf{Y}_T, \mathbf{Y}_D] = \mathbf{H}[\mathbf{S}_T, \mathbf{S}_D] + \mathbf{N}. \tag{2}$$

Here, the matrices $\mathbf{S}_T \in \mathbb{C}^{U \times T}$ and $\mathbf{S}_D \in \mathbb{C}^{U \times D}$ contain training pilots and data symbols, respectively; the matrices $\mathbf{Y}_T \in \mathbb{C}^{B \times T}$ and $\mathbf{Y}_D \in \mathbb{C}^{B \times T}$ contain the received pilot and data symbols, respectively. Our goal is to jointly estimate the channel matrix $\mathbf{H}$ and detect the entries in $\mathbf{S}_D$ from the received signals in $\mathbf{Y}$ and the known training-pilot matrix $\mathbf{S}_T$.

B. Cell-free Massive MU-MIMO System

To characterize cell-free massive MU-MIMO communication, we use the channel model put forward in [1]. For this model, the MIMO system in (1) can be written as

$$\mathbf{Y} = \sqrt{\rho_u}\mathbf{G}\mathbf{\Lambda}\mathbf{S} + \mathbf{N}, \tag{3}$$

where ρ_u denotes the normalized (uplink) signal-to-noise ratio (SNR), $\mathbf{G} \in \mathbb{C}^{B \times U}$ is the cell-free channel matrix, $\mathbf{\Lambda} \in^{U \times U}$ is a diagonal power control matrix, and the entries of $\mathbf{N}$ are normalized so that $N_0 = 1$. Following the model in [1], the entries of $\mathbf{G}$ are modeled as $G_{ij} = \sqrt{\beta_{ij}}\theta_{ij}$ where β_{ij} and θ_{ij} characterize large-scale and small-scale fading, respectively. Since in a cell-free system, the UEs are close to only a few receive antennas (assuming random placement of UEs and distributed receive antennas in a given area), the matrix $\mathbf{G}$ is sparse, i.e., most of its entries are close to zero—a property which we will exploit. We assume UE-side power control and define the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_U)$ as

$$\lambda_i^2 = \min\left\{\|\mathbf{g}_i\|_2^2, 10^{\frac{P}{10}} \min_{j=1,\ldots,U} \|\mathbf{g}_j\|_2^2\right\} / \|\mathbf{g}_i\|_2^2, \tag{4}$$

where P defines the maximum dynamic range between the weakest and strongest UE receive power in decibels. In summary, the channel matrix in (1) is $\mathbf{H} = \sqrt{\rho_u}\mathbf{G}\mathbf{\Lambda}$.

[1] For frequency-selective channels, we can use orthogonal frequency-division multiplexing (OFDM) to obtain an equivalent system model per subcarrier.

III. Joint Channel Estimation and Data Detection

We now formulate the JED problem and then relax it to a biconvex problem that we solve approximately using FBS [8].

A. The MAP-JED Problem

We first formulate the maximum a-posteriori (MAP) JED problem. To this end, we assume (i) that the channel coefficients in $\mathbf{H}$ follow a sparsity-inducing Laplace prior and (ii) the entries in $\mathbf{N}$ are complex standard normal. For these assumptions, we have the following MAP-JED problem

$$\{\hat{\mathbf{H}}, \hat{\mathbf{S}}_D\} = \underset{\substack{\mathbf{H} \in \mathbb{C}^{B \times U} \\ \mathbf{S}_D \in \mathcal{Q}^{U \times D}}}{\arg\min} \frac{1}{2}\|\mathbf{Y} - \mathbf{H}[\mathbf{S}_T, \mathbf{S}_D]\|_F^2 + \mu\|\mathbf{H}\|_1. \tag{5}$$

Here, the parameter $\mu \in \mathbb{R}_+$ controls the channel's sparsity, where large values promote sparser channel matrices. Note that solving (5) optimally is challenging for two reasons: (i) The entries of the data matrix $\mathbf{S}_D$ are chosen from a discrete set $\mathcal{Q}$; (ii) the objective function contains the product of the channel matrix $\mathbf{H}$ and data matrix $\mathbf{S}_D$—for a fixed data matrix $\mathbf{S}_D$, however, the MAP-JED problem is convex in $\mathbf{H}$.

B. Biconvex Relaxation

We now show how the MAP-JED problem can be relaxed to a biconvex optimization problem [27], i.e., a problem that is convex in $\mathbf{H}$ for a fixed $\mathbf{S}_D$ and vice versa. We will show in Section III-C how to solve this problem approximately.

We start by relaxing the discrete constellation $\mathcal{Q}$ to its convex hull, which is defined as [17]

$$\mathcal{C} = \left\{ \sum_{i=1}^{|\mathcal{Q}|} \alpha_i s_i \mid (\alpha_i \in \mathbb{R}_+, \forall i) \wedge \sum_{i=1}^{|\mathcal{Q}|} \alpha_i = 1) \right\}, \tag{6}$$

where s_i is the ith symbol in the constellation $\mathcal{Q}$. Note that for QPSK with $\{\pm 1 \pm j\}$ the convex hull $\mathcal{C}$ is a tight square region (box) around the four constellation points. This relaxation results in the following biconvex optimization problem

$$\{\hat{\mathbf{H}}, \hat{\mathbf{S}}_D\} = \underset{\substack{\mathbf{H} \in \mathbb{C}^{B \times U} \\ \mathbf{S}_D \in \mathcal{C}^{U \times D}}}{\arg\min} \frac{1}{2}\|\mathbf{Y} - \mathbf{H}[\mathbf{S}_T, \mathbf{S}_D]\|_F^2 + \mu\|\mathbf{H}\|_1. \tag{7}$$

We note that a similar convexification strategy has been used recently for conventional massive MIMO data detectors that separate data detection from channel estimation [28]–[31].

To improve the performance of this convex relaxation, we additionally use a technique from [27], which gently pushes the relaxed constellation points towards the corners of the convex hull. For QPSK, we are favoring solutions at the four corner points $\{\pm 1 \pm j\}$. Concretely, we include a concave regularizer $-\frac{\gamma}{2}\|\mathbf{S}_D\|_F^2$ with parameter $\gamma \in \mathbb{R}_+$ to the objective function of (7), which is small for larger values in the data matrix $\mathbf{S}_D$. The resulting optimization problem is given by

$$\{\hat{\mathbf{H}}, \hat{\mathbf{S}}_D\} = \underset{\substack{\mathbf{H} \in \mathbb{C}^{B \times U} \\ \mathbf{S}_D \in \mathcal{C}^{U \times D}}}{\arg\min} \frac{1}{2}\|\mathbf{Y} - \mathbf{H}[\mathbf{S}_T, \mathbf{S}_D]\|_F^2 + \mu\|\mathbf{H}\|_1 - \frac{\gamma}{2}\|\mathbf{S}_D\|_F^2. \tag{8}$$

While the problem (8) remains nonconvex, it is biconvex in $\mathbf{H}$ and $\mathbf{S}_D$ as long as the parameter γ is sufficiently small (see, e.g., [17] for a precise justification of this property). We next show an FBS-based approach to solve the problem (8) approximately and at low complexity.

C. JED via Forward-Backward Splitting

FBS is an efficient numerical method to iteratively solve convex optimization problems of the following form:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}). \qquad (9)$$

Here, the function f is differentiable and convex, and g is a more general (not necessarily smooth or bounded) convex function. After initializing $\mathbf{x}^{(1)}$, FBS solves the problem in (9) for the iterations $t = 1, 2, \ldots$ until convergence, by computing

$$\mathbf{x}^{(t+1)} = \text{prox}_g(\mathbf{x}^{(t)} - \tau^{(t)}\nabla f(\mathbf{x}^{(t)}); \tau^{(t)}), \qquad (10)$$

where $\nabla f(\mathbf{x})$ is the gradient of $f(\mathbf{x})$ and $\tau^{(t)}$ is a per-iteration step size; we use the adaptive step-size strategy from FASTA [8]. The proximal operator for $g(\mathbf{x})$ is defined as

$$\text{prox}_g(\mathbf{z}; \tau) = \arg\min_{\mathbf{x}} \tau g(\mathbf{x}) + \tfrac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2. \qquad (11)$$

While FBS exactly solves convex optimization problems, it can be used to approximately solve nonconvex problems [8]. We use FBS to jointly solve for $\mathbf{H}$ and $\mathbf{S}_D$ in (8). To this end, we define the matrix $\mathbf{Z} = [\mathbf{H}^H \ \mathbf{S}]^H$, where $\mathbf{S} = [\mathbf{S}_T, \mathbf{S}_D]$ and $\mathbf{S}_T$ is known and kept constant during the iterations; the matrices $\mathbf{H}$ and $\mathbf{S}_D$ contain the optimization variables. We start by defining the functions f and g in (9) as

$$f(\mathbf{Z}) = f(\mathbf{H}, \mathbf{S}_D) = \tfrac{1}{2}\|\mathbf{Y} - \mathbf{H}\mathbf{S}\|_F^2, \qquad (12)$$

$$g(\mathbf{Z}) = g(\mathbf{H}, \mathbf{S}_D) = \mu\|\mathbf{H}\|_1 - \tfrac{\gamma}{2}\|\mathbf{S}_D\|_F^2 + \chi_{\mathcal{C}}(\mathbf{S}_D), \quad (13)$$

where the indicator function

$$\chi_{\mathcal{C}}(\mathbf{S}_D) = \begin{cases} 0 & \mathbf{S}_D \in \mathcal{C} \\ \infty & \mathbf{S}_D \notin \mathcal{C}, \end{cases} \qquad (14)$$

replaces the constraint $\mathbf{S}_D \in \mathcal{C}^{U \times D}$ in (8). For these definitions, the joint gradient of f in $\mathbf{Z}$ is given by

$$\nabla f(\mathbf{Z}) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{H}} \\ \frac{\partial f}{\partial \mathbf{S}^H} \end{bmatrix} = \begin{bmatrix} (\mathbf{H}\mathbf{S} - \mathbf{Y})\mathbf{S}^H \\ (\mathbf{H}\mathbf{S} - \mathbf{Y})^H \mathbf{H} \end{bmatrix}. \qquad (15)$$

The proximal operator for $\mathbf{H}$ is element-wise shrinkage [8]:

$$\eta(H_{ij}; \mu\tau^{(t)}) = \frac{H_{ij}}{|H_{ij}|}\max\{|H_{ij}| - \mu\tau^{(t)}, 0\}, \qquad (16)$$

where μ is the sparsity parameter, $\tau^{(t)}$ is the step size, and we define $x/|x| = 0$ for $x = 0$.

The proximal operator for $\mathbf{S}_D$ is a bit more involved but can be derived from (11) as follows:

$$\text{prox}_g(\mathbf{S}_D; \tau^{(t)}) \qquad (17)$$

$$= \arg\min_{\mathbf{X} \in \mathcal{C}^{U \times D}} -\tau^{(t)}\tfrac{\gamma}{2}\|\mathbf{X}\|_F^2 + \tfrac{1}{2}\|\mathbf{X} - \mathbf{S}_D\|_F^2, \qquad (18)$$

where we moved the indicator function in $g(\mathbf{Z})$ to the constraint. By completing the square, (18) is given by

$$\text{prox}_g(\mathbf{S}_D; \tau^{(t)}) = \arg\min_{\mathbf{X} \in \mathcal{C}^{U \times D}} \left\|\mathbf{X} - \tfrac{1}{1-\rho}\mathbf{S}_D\right\|_F^2, \qquad (19)$$

where $\rho = \tau^{(t)}\lambda \in [0, 1)$. As a consequence, we have

$$\text{prox}_g(\mathbf{S}_D; \tau^{(t)}) = \text{proj}_{\mathcal{C}}\left(\tfrac{1}{1-\rho}\mathbf{S}_D; \tau^{(t)}\right). \qquad (20)$$

The right-hand side proximal operator for the convex hull $\mathcal{C}$ is applied independently to real and imaginary parts as

$$\text{proj}_{\mathcal{C}}(\Re\{S_{ij}\}) = \min\{\max\{|\Re\{S_{ij}\}|, -\alpha\}, \alpha\} \qquad (21)$$

$$\text{proj}_{\mathcal{C}}(\Im\{S_{ij}\}) = \min\{\max\{|\Im\{S_{ij}\}|, -\alpha\}, \alpha\}. \qquad (22)$$

Here, α defines the box $\mathcal{C}$ and depends on the transmit symbols. For QPSK with $\mathcal{Q} = \{\pm 1 \pm j\}$, we have $\alpha = 1$. In summary, the joint proximal operator is defined as follows:

$$\text{prox}_g(\mathbf{Z}; \tau^{(t)}) = \begin{bmatrix} \eta(\mathbf{H}, \mu\tau^{(t)}) \\ \text{proj}_{\mathcal{C}}\left(\tfrac{1}{1-\rho}\mathbf{S}^H\right) \end{bmatrix}. \qquad (23)$$

Since the pilot sequences in the matrix $\mathbf{S}_T$ are fixed and known, we enforce $\mathbf{S}_T$ at the end of every FBS iteration.

IV. RESULTS

We now demonstrate the efficacy of our JED algorithm. We first detail the simulation setup and then show results.

A. Simulation Setup and Performance Metrics

We evaluate our algorithm with the cell-free channel model detailed Section II-B. As in [1], we consider a square area of $1\,\text{km}^2$ with $B = 128$ randomly positioned receive antennas and $U = 128$ randomly positioned UEs. The maximum UE transmission power is $100\,\text{mW}$ and we use per-UE power control with $P = 12\,\text{dB}$ as in (4). The carrier frequency is $1.9\,\text{GHz}$ and the bandwidth $20\,\text{MHz}$. Equiangular tight frames were used for the pilot sequences in $\mathbf{S}_T$ [32]. The receive and UE antennas are at a height of $15\,\text{m}$ and $1.65\,\text{m}$, respectively. We use the three-slope path-loss model defined in [33]. The small-scale fading and large-scale fading parameters between the ith antenna and the jth UE are $\theta_{ij} \sim \mathcal{CN}(0, 1)$ and $\beta_{ij} = PL_{ij}10^{\frac{\sigma_{\text{sh}}z_{ij}}{10}}$ where PL_{ij} is the path loss, σ_{sh} is $8\,\text{dB}$, and $z_{ij} \sim \mathcal{N}(0, 1)$ is shadow fading with variance σ_{sh}^2.

B. Performance Metrics and Baseline Algorithms

Since we evaluate a nonconvex JED algorithm, a spectral efficiency analysis is non-trivial. Furthermore, bit-error rate (BER) vs. SNR plots cannot be generated since the UE's experience different SNR. Hence, we characterize the per-UE cumulative density function (CDF) for the EVM, BER, and channel estimation MSE. For the jth UE, we define the BER as $BER_j = \frac{\varepsilon_j}{n_q \cdot D}$, where ε_j is the total number of bit errors for UE j over D time slots and q the number of bits per symbol. We define the EVM over the D payload data slots as

$$EVM_j = \sqrt{\frac{\sum_{k=1}^D |[\hat{\mathbf{S}}_D]_{jk} - [\mathbf{S}_D]_{jk}|^2}{\sum_{k=1}^D |[\hat{\mathbf{S}}_D]_{jk}|^2}}, \qquad (24)$$

where $[\hat{\mathbf{S}}_D]_{jk}$ and $[\mathbf{S}_D]_{jk}$ denote the estimated and transmitted data symbols of the jth UE at time slot k, respectively. We define the channel estimation MSE of the jth UE as $MSE_j = \frac{1}{B}\mathbb{E}\left[\|\hat{\mathbf{h}}_j - \mathbf{h}_j\|_2^2\right]$, where $\hat{\mathbf{h}}_j$ and $\mathbf{h}_j$ are the estimated and true channel vectors, respectively. By treating all these quantities as random variables, we use Monte-Carlo simulations to characterize their CDF over multiple UE and antenna placements, noise realizations, and data transmissions. The fraction of Monte-Carlo trials for which the per-UE EVM
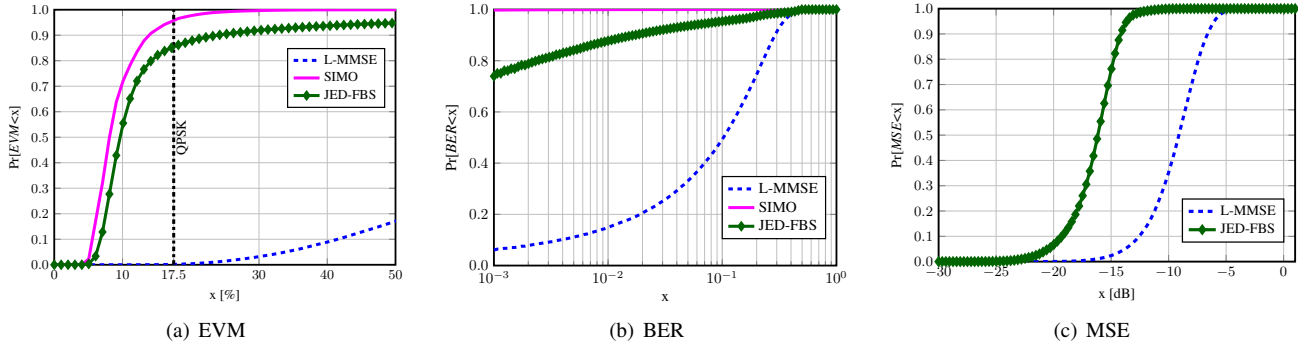
Fig. 1. EVM (a), BER (b), and MSE (c) performance for a cell-free massive MU-MIMO system with $B = 128$ receive antennas, $U = 128$ UEs transmitting QPSK, $K = 128$ time slots, and only $T = 32$ (25%) nonorthogonal training symbols. The proposed JED algorithm supports over 87% of the UEs with an EVM of 17.5%; ℓ_1-norm training-based L-MMSE data detection is unable to achieve satisfactory performance.
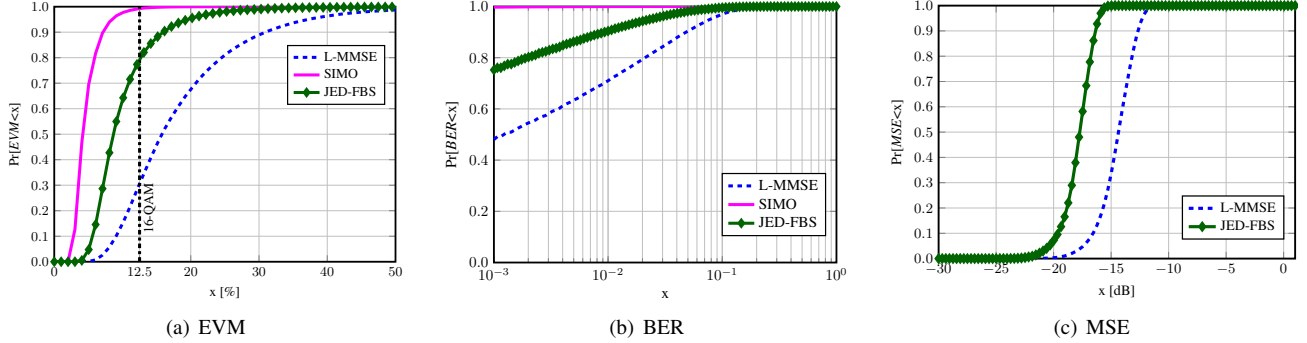

Fig. 2. EVM (a), BER (b), and MSE (c) performance for a cell-free massive MU-MIMO system with $B = 256$ receive antennas, $U = 128$ UEs transmitting 16-QAM, $K = 128$ time slots, and $T = 64$ (50%) nonorthogonal training symbols. The proposed JED algorithm supports over 90% of the UEs with an EVM of 12.5%; ℓ_1-norm training-based L-MMSE data detection does not perform as well as JED.

was below x is defined as $\Pr[EVM < x]$; the quantities $\Pr[BER < x]$ and $\Pr[MSE < x]$ are defined analogously.

To characterize the performance of our JED algorithm, we use the single-input multiple-output (SIMO) lower bound (also known as matched filter bound), which perfectly cancels MU interference in a genie-aided fashion [34]. We also compare our algorithm to that of a conventional training-based linear minimum mean-square error (L-MMSE) data detector. For this algorithm, we use a regularized ℓ_1-norm channel estimator that exploits sparsity in cell-free massive MIMO systems:

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{H} \in \mathbb{C}^{B \times U}} \frac{1}{2}\|\mathbf{Y}_T - \mathbf{H}\mathbf{S}_T\|_F^2 + \mu_1\|\mathbf{H}\|_1, \quad (25)$$

where we tune the sparsity parameter μ_1 for each scenario.

C. EVM, BER, and MSE Performance Results

Figure 1 shows simulation results for a $B = 128$ antenna system with $U = 128$ UEs transmitting pilots and QPSK payload data over $K = 128$ time slots, where only $T = 32$ (25%) pilots are used; the pilots form an equiangular tight frame. Figure 2 shows simulation results for a $B = 256$ antenna system with $U = 128$ UEs transmitting pilots and 16-QAM payload data over $K = 128$ time slots, where $T = 64$ (50%) pilots are used. In Fig. 1(a) and Fig. 2(a), we see that JED enables more than 85% and 80% of all UEs to meet the EVM of $R = 17.5\%$ for QPSK and of $R = 12.5\%$ for 16-QAM, respectively, which is similar to the requirements of BS purity in the 3GPP 5G NR technical specification [35,

Tbl. 6.5.2.2-1]. In Fig. 1(b) and Fig. 2(b), we see that JED enables an 1% uncoded BER for 87% and 90% of the UEs for QPSK and 16-QAM, respectively. In Fig. 1(c), we see that JED provides more than $5\,\mathrm{dB}$ lower channel estimation MSE than the ℓ_1-norm based approach in (25); in Fig. 2(c), JED provides more than $3\,\mathrm{dB}$ SNR lower MSE.

D. Pilot Overhead vs. EVM Trade-off

Figure 3 shows the trade-off between between the percentage of UEs that achieve the target EVM R and the amount of used pilots (as a fraction of orthogonal training). We show results for the $B = 128$ system with QPSK and a target EVM of $R = 17.5\%$ and for the $B = 256$ system with 16-QAM and a target EVM of $R = 12.5\%$; both systems use $K = 128$ time slots. For 50% pilot overhead ($T = 64$), both scenarios are able to meet the minimum EVM requirements for more than 80% of the UEs. Reducing the pilot overhead to 36% ($T = 46$), JED is still able to meet the minimum EVM requirements for more than 70% of the UEs. In stark contrast, conventional L-MMSE data detection with ℓ_1-norm-based channel estimation is unable to meet the EVM requirements for most of the UEs.

V. CONCLUSIONS

We have proposed a novel joint channel estimation and data detection (JED) algorithm for cell-free massive MU-MIMO systems. Our method formulates JED as a biconvex optimization problem which we solve approximately using forward-backward splitting. By exploiting channel sparsity and
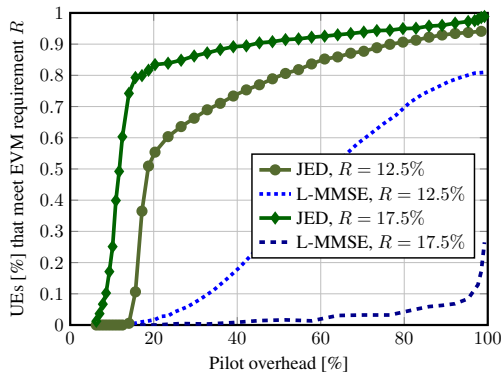
Fig. 3. Trade-off between pilot overhead and fraction of UEs that meet the minimum EVM requirements of 3GPP 5G NR [35] for two $U = 128$ UE cell-free systems with $K = 128$ time slots: (solid lines) $B = 128$ antennas with QPSK EVM requirement $R = 17.5\%$; (dashed lines) $B = 256$ antennas with 16-QAM EVM requirement $R = 12.5\%$ JED provides over 50% of the UEs with the minimal EVM requirements for only 20% pilots (compared to orthogonal training); L-MMSE detection requires orthogonal training.

the boundedness of QAM, our algorithm enables more than 80% of the UEs to achieve the minimum EVM requirements specified by 3GPP 5G NR while using $2\times$ fewer pilots than orthogonal training. Since baseband processing for cell-free systems is likely to be performed in a centralized processor with extensive computing resources, JED has potential to improve spectral efficiency while minimizing the pilot overhead.

REFERENCES

[1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[2] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[3] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, Oct. 2018.

[4] Ö. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive MIMO with Rician fading and phase shifts," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5299–5315, Nov. 2019.

[5] J. Francis, P. Baracca, S. Wesemann, and G. Fettweis, "Downlink power control in cell-free massive MIMO with partially distributed access points," in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, Sept. 2019, pp. 1–7.

[6] J. M. Palacios, O. Raeesi, A. Gokceoglu, and M. Valkama, "Impact of channel non-reciprocity in cell-free massive MIMO," *IEEE Wireless Commun. Lett.*, Nov. 2019, Early Access.

[7] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[8] T. Goldstein, C. Studer, and R. G. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," Nov. 2014. [Online]. Available: https://arxiv.org/abs/1411.3406

[9] P. Stoica and G. Ganesan, "Space-time block codes: trained, blind and semi-blind detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, May 2002, p. 1609.

[10] H. Vikalo, B. Hassibi, and P. Stoica, "Efficient joint maximum-likelihood channel estimation and signal detection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1838–1845, July 2006.

[11] H. A. J. Alshamary, T. Al-Naffouri, A. Zaib, and W. Xu, "Optimal non-coherent data detection for massive SIMO wireless systems: A polynomial complexity solution," in *Proc. IEEE Signal Process. Signal Process. Edu. Workshop*, Aug. 2015, pp. 172–177.

[12] W. Xu, M. Stojnic, and B. Hassibi, "On exact maximum-likelihood detection for non-coherent MIMO wireless systems: a branch-estimate-bound optimization framework," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2008, pp. 2017–2021.

[13] T.-H. Pham, Y.-C. Liang, and A. Nallanathan, "A joint channel estimation and data detection receiver for multiuser MIMO IFDMA systems," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1857–1865, June 2009.

[14] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5369–5382, Oct. 2015.

[15] E. Kofidis, C. Chatzichristos, and A. L. de Almeida, "Joint channel estimation/data detection in MIMO-FBMC/OQAM systems—a tensor-based approach," in *Proc. IEEE European Signal Process. Conf. (EUSIPCO)*, Sept. 2017, pp. 420–424.

[16] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, Jul. 2015.

[17] O. Castañeda, T. Goldstein, and C. Studer, "VLSI designs for joint channel estimation and data detection in large SIMO wireless systems," *IEEE Trans. Circuits Syst. I*, vol. 65, no. 3, pp. 1120–1132, Mar. 2017.

[18] B. Yilmaz and A. Erdogan, "Channel estimation for massive MIMO: A semiblind algorithm exploiting QAM structure," presented at the Asilomar Conf. Signals, Syst., Comput., Nov. 2019.

[19] A. Schenk and R. F. Fischer, "Noncoherent detection in massive MIMO systems," in *Int. ITG Workshop Smart Antennas*, Mar. 2013, pp. 1–8.

[20] G. Yammine and R. F. Fischer, "Soft-decision decoding in noncoherent massive MIMO systems," in *Int. ITG Workshop Smart Antennas*, Mar. 2016, pp. 1–7.

[21] J. Feng, H. Gao, T. Wang, T. Lv, and W. Guo, "A noncoherent differential transmission scheme for multiuser massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[22] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, July 2015, pp. 201–205.

[23] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 264–11 268, Nov. 2018.

[24] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green. Comm. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2017.

[25] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.

[26] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space–time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[27] S. Shah, A. K. Yadav, C. D. Castillo, D. W. Jacobs, C. Studer, and T. Goldstein, "Biconvex relaxation for semidefinite programming in computer vision," in *Eur. Conf. Comput. Vision*, Sep. 2016, pp. 717–735.

[28] C. Jeon, A. Maleki, and C. Studer, "On the performance of mismatched data detection in large MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, May 2016.

[29] E. Abbasi, F. Salehi, and B. Hassibi, "Performance analysis of convex data detection in MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 4554–4558.

[30] C. Thrampoulidis, E. Abbasi, W. Xu, and B. Hassibi, "BER analysis of the box relaxation for BPSK signal recovery," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 3776–3780.

[31] S. Shahabuddin, M. Juntti, and C. Studer, "ADMM-based infinity norm detection for large MU-MIMO: Algorithm and vlsi architecture," in *Proc. IEEE Int. Symp. Circuits and Syst. (ISCAS)*, May 2017, pp. 1–4.

[32] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, 2005.

[33] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, vol. 1, May 2001, pp. 333–336.

[34] J. Zhang, "Non-asymptotic capacity lower bound for non-coherent SIMO channels with memory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2006, pp. 1272–1276.

[35] 3GPP, "5G; NR; base station (BS) radio transmission and reception," May 2019, TS 38.104 version 15.5.0 Rel. 15.