

Joint Active User Detection, Channel Estimation, and Data Detection for Massive Grant-Free Transmission in Cell-Free Systems

Gangle Sun^{1,2}, Mengyao Cao¹, Wenjin Wang^{1,2}, Wei Xu^{1,2}, and Christoph Studer³

¹*National Mobile Communications Research Laboratory, Southeast University, Nanjing, China*

²*Purple Mountain Laboratories, Nanjing, China*

³*Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland*
email: sungangle@seu.edu.cn, cmengyao64@gmail.com, {wangwj, wxu}@seu.edu.cn, and studer@ethz.ch

Abstract—Cell-free communication has the potential to significantly improve grant-free transmission in massive machine-type communication, wherein multiple access points jointly serve a large number of user equipments to improve coverage and spectral efficiency. In this paper, we propose a novel framework for joint active user detection (AUD), channel estimation (CE), and data detection (DD) for massive grant-free transmission in cell-free systems. We formulate an optimization problem for joint AUD, CE, and DD by considering both the sparsity of the data matrix, which arises from intermittent user activity, and the sparsity of the effective channel matrix, which arises from intermittent user activity and large-scale fading. We approximately solve this optimization problem with a box-constrained forward-backward splitting algorithm, which significantly improves AUD, CE, and DD performance. We demonstrate the effectiveness of the proposed framework through simulation experiments.

I. INTRODUCTION

Massive machine-type communications (mMTC) is a central scenario in fifth-generation (5G) wireless communication systems, in which user equipments (UEs) transmit data intermittently to an infrastructure base station (BS). Massive grant-free transmission techniques are suitable for mMTC scenarios as they reduce excessive signaling overhead, network congestion, and high transmission latency by allowing the active UEs to transmit signals over shared resource elements directly without sophisticated scheduling mechanisms [?].

In order to improve coverage for UEs in mMTC scenarios, cell-free communication techniques have emerged as a powerful solution [?], [?]. Cell-free communication mitigates inter-cell interference and improves spectral efficiency by jointly

processing all of the information acquired at a large number of distributed access points (APs) that are connected to a central processing unit (CPU) [?], [?], [?]. One of the key tasks for massive grant-free transmission in cell-free systems involves detecting the set of active UEs, estimating their channels, and detecting their transmitted data at the CPU side.

A. Contributions

This paper proposes a novel framework for joint active user detection, channel estimation, and data detection (JACD) suitable for massive grant-free transmission in cell-free systems. We formulate the JACD problem as a nonconvex optimization problem, accounting for sparsity in the data matrix arising from UEs' sporadic activity and sparsity in the effective channel matrix resulting from both the UEs' sporadic activity and large-scale fading. We relax the discrete constellation constraints in our problem formulation, which enables the use of computationally-efficient gradient-type solvers. We then develop a forward-backward splitting (FBS) algorithm to approximately solve the JACD problem, leading to a significant improvement in joint estimation accuracy. Finally, we demonstrate the effectiveness of the proposed algorithm in terms of active user detection (AUD), channel estimation (CE), and data detection (DD) through system simulations.

B. Relevant Prior Art

Recent research has focused on AUD, CE, and DD for massive grant-free transmission in cell-free communication systems [?], [?], [?], [?], [?], [?]. Reference [?] proposed two different AUD algorithms based on dominant APs and clustering, respectively, demonstrating that cell-free communication can surpass co-located schemes for AUD in large coverage areas. Reference [?] proposed an expectation-maximization approximate message passing (AMP) algorithm for CE and detected active UEs using posterior support probabilities. Reference [?] developed a Bayesian AMP algorithm based on a single measurement vector for joint AUD and CE, processing received signals at each AP separately. Reference [?] presented an AMP

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62341110, 62022026 and 62211530108; in part by the Jiangsu Province Basic Research Project under Grant BK20192002; in part by the Fundamental Research Funds for the Central Universities under Grants 2242022k30005, 2242022k60002 and 2242023k5003. The work of Gangle Sun was supported in part by China Scholarship Council (CSC) under Grant 202206090074. The work of CS was supported in part by the U.S. National Science Foundation (NSF) under grants CNS-1717559 and ECCS-1824379, and in part by an ETH Research Grant.

The authors would like to thank Victoria Palhares and Haochuan Song for discussing channel modeling of cell-free systems. We also acknowledge Gian Marti and Sueda Taner for their suggestions on deriving the FBS algorithm.

algorithm for joint AUD and CE, accounting for quantization artifacts and exploited the sparsity structure in the channel matrix. Reference [?] introduced a two-stage CE and AUD method using the vector AMP algorithm for AUD followed by linear minimum mean square error-based CE. Reference [?] introduced a bilinear Gaussian belief propagation algorithm for JACD, combining successive interference cancellation and Bayesian message passing methods. Unlike most prior works, we address the JACD problems for mMTC in cell-free systems and exploit sparsity in the data matrix arising from intermittent UE activity to improve the joint estimation performance.

JACD for single-cell massive grant-free transmission has been explored as well in [?], [?], [?], [?]. Reference [?] applied bilinear generalized AMP (BiG-AMP) and belief propagation algorithms for JACD in massive grant-free systems with low-precision data converters. Reference [?] proposed the use of bilinear message-scheduling generalized AMP to enable JACD and data detection by utilizing channel decoder beliefs to improve AUD and detection performance. Reference [?] developed a BiG-AMP algorithm based on the row-sparse channel matrix structure for JACD by leveraging channel correlation across different antennas. Reference [?] utilized AMP to decouple transmissions of different UEs and addressed the nonlinear coupling of each UE's activity, channel coefficient, and data separately. In contrast, our work focuses on JACD in cell-free systems, utilizes an optimization-based problem formulation, deploys gradient-type algorithms, and considers both sources of sparsity in the channel matrix due to sporadic UE activity in mMTC scenarios and large-scale fading in cell-free wireless communication systems.

C. Notation

Uppercase and lowercase boldface letters denote matrices and column vectors, respectively; $\mathbf{A}(m, n)$ and $\mathbf{a}(m)$ correspond to the element in the m th row and n th column of the matrix $\mathbf{A}$ and the m th element of the vector $\mathbf{a}$, respectively; $\mathbf{1}_{M \times N}$ represents the $M \times N$ all-ones matrix. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively; $\|\cdot\|_F$ is the Frobenius norm, $|\mathcal{Q}|$ the number of elements in the set $\mathcal{Q}$, and $\odot$ the Hadamard product. Proportional relationships are denoted by $\propto$. The indicator function $\mathbb{I}\{\cdot\}$ is 1 if the condition is true and 0 otherwise; $\text{Re}\{x\}$ and $\text{Im}\{x\}$ are the real and image parts of $x \in \mathbb{C}$, respectively. $\mathbb{P}\{\cdot\}$ denotes probability.

II. SYSTEM MODEL

We consider a cell-free mMTC system consisting of P distributed APs with M antennas each and N single-antenna UEs, where $N_a \ll N$ active UEs transmit signals simultaneously to the APs over R shared resource elements. In what follows, we assume that all UEs are perfectly synchronized.

We consider frequency-flat and block-fading channels with the following input-output relation [?], [?], [?]:

$$\mathbf{Y} = \sum_{n=1}^N \xi_n \mathbf{h}_n \mathbf{x}_n^T + \mathbf{N}. \quad (1)$$

Here, $\mathbf{Y} \in \mathbb{C}^{MP \times R}$ contains the received signals of all APs for the R resource elements, $\xi_n \in \{0, 1\}$ is n th UE's activity

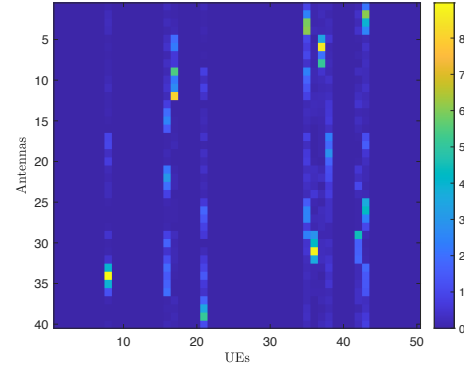


Fig. 1. Amplitude of entries of the efficient channel matrix $\mathbf{H}$ between 10 APs with 4 antennas each and 50 UEs, where only 10 UEs are active.

indicator with $\xi_n = 1$ if the n th UE is active and $\xi_n = 0$ otherwise. The vector $\mathbf{h}_n = [\mathbf{h}_{n,1}^T, \mathbf{h}_{n,2}^T, \dots, \mathbf{h}_{n,P}^T]^T \in \mathbb{C}^{MP \times 1}$ represents the channel vector between the n th UE and all APs while $\mathbf{h}_{n,p} \in \mathbb{C}^{M \times 1}$ is the channel vector between the n th UE and the p th AP, following the model in [?], [?]. The vector $\mathbf{x}_n = [\mathbf{x}_{p,n}^T, \mathbf{x}_{d,n}^T]^T \in \mathbb{C}^{R \times 1}$ contains the n th UE's pilots $\mathbf{x}_{p,n} \in \mathbb{C}^{R_P \times 1}$ and data signals $\mathbf{x}_{d,n} \in \mathbb{Q}^{R_D \times 1}$, where $R = R_P + R_D$ and $\mathbb{Q}$ is a discrete constellation set. The matrix $\mathbf{N} \in \mathbb{C}^{MP \times R}$ models noise with i.i.d. circularly-symmetric complex Gaussian entries of variance $N_0 = 1$.

Since $\xi_n \mathbf{h}_n \mathbf{x}_{d,n}^T = (\xi_n \mathbf{h}_n)(\xi_n \mathbf{x}_{d,n}^T)$, the data vector $\mathbf{x}_{d,n}$ can also be treated as a zero vector if n th UE is inactive. As such, we can rewrite (??) as follows:

$$\mathbf{Y} = \mathbf{H}[\mathbf{X}_P, \mathbf{X}_D] + \mathbf{N} = \mathbf{H}\mathbf{X} + \mathbf{N}. \quad (2)$$

Here, $\mathbf{H} \triangleq [\xi_1 \mathbf{h}_1, \xi_2 \mathbf{h}_2, \dots, \xi_N \mathbf{h}_N] \in \mathbb{C}^{MP \times N}$ is the effective channel matrix, $\mathbf{X}_P = [\mathbf{x}_{p,1}, \mathbf{x}_{p,2}, \dots, \mathbf{x}_{p,N}]^T \in \mathbb{C}^{N \times R_P}$ is the pilot matrix, and $\mathbf{X}_D = [\bar{\mathbf{x}}_{d,1}, \bar{\mathbf{x}}_{d,2}, \dots, \bar{\mathbf{x}}_{d,N}]^T \in \bar{\mathbb{Q}}^{N \times R_D}$ is the sparse data matrix, where $\bar{\mathbf{x}}_{d,n} \triangleq \xi_n \mathbf{x}_{d,n}$ and $\bar{\mathbb{Q}} \triangleq \{\mathbb{Q}, 0\}$. To simplify notation, we define $\mathbf{X} \triangleq [\mathbf{X}_P, \mathbf{X}_D]$ in (??). Note that, in signal matrix $\mathbf{X}$, we only consider sparsity in the data matrix $\mathbf{X}_D$ and not in the pilot matrix $\mathbf{X}_P$ as our optimization problem will leverage all of the available pilot information—nonetheless, inactive UEs will not transmit any pilots. In addition, as illustrated in Fig. ??, there are two sources of sparsity in the effective channel matrix $\mathbf{H}$: (i) column sparsity caused by the UEs' sporadic activity and (ii) inherent channel sparsity among different APs caused by the fact that each UE is only nearby a few APs. In the following, we will explore both of these sources of sparsity.

III. JOINT ACTIVE USER DETECTION, CHANNEL ESTIMATION, AND DATA DETECTION

A. Problem Formulation

According to system model (??), the JACD optimization problem for mMTC in cell-free systems can be formulated as

$$\{\hat{\mathbf{H}}, \hat{\mathbf{X}}_D\} = \arg \max_{\substack{\mathbf{H} \in \mathbb{C}^{MP \times N} \\ \mathbf{X}_D \in \bar{\mathbb{Q}}^{N \times R_D}}} P(\mathbf{Y}|\mathbf{H}, \mathbf{X}_D)P(\mathbf{H})P(\mathbf{X}_D) \quad (3)$$

$$\mathcal{P}_1 : \{\hat{\mathbf{H}}, \hat{\mathbf{X}}_D\} = \arg \min_{\substack{\mathbf{H} \in \mathbb{C}^{MP \times N} \\ \mathbf{X}_D \in \bar{\mathcal{Q}}^{N \times R_D}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{H}[\mathbf{X}_P, \mathbf{X}_D]\|_F^2 + \mu_h \sum_{n=1}^N \sum_{p=1}^P \|\mathbf{h}_{n,p}\|_F + \mu_x \sum_{n=1}^N \|\bar{\mathbf{x}}_{D,n}\|_F. \quad (6)$$

with the channel law

$$P(\mathbf{Y}|\mathbf{H}, \mathbf{X}_D) \propto \exp\left(-\|\mathbf{Y} - \mathbf{H}[\mathbf{X}_P, \mathbf{X}_D]\|_F^2 / N_0\right). \quad (4)$$

To take into account the sparsity in both $\mathbf{H}$ and $\mathbf{X}_D$, we utilize the following models for $\mathbf{H}$ and $\mathbf{X}_D$:

$$P(\mathbf{H}) \propto \prod_{n=1}^N \prod_{p=1}^P \exp(-2\mu_h \|\mathbf{h}_{n,p}\|_F), \quad (5a)$$

$$P(\mathbf{X}_D) \propto \prod_{n=1}^N \exp(-2\mu_x \|\bar{\mathbf{x}}_{D,n}\|_F), \mathbf{X}_D \in \bar{\mathcal{Q}}^{N \times R_D}, \quad (5b)$$

where μ_h and μ_x are parameters that determine the amount of sparsity. The sparsity in $\mathbf{H}$ due to UE activity and large-scale fading is modeled by a complex-valued block-Laplace prior in (??), whereas the sparsity of $\mathbf{X}_D$ due to UE activity is modeled by a complex-valued Laplace prior in (??) [?]. Finally, by plugging (??) and (??) into (??) and taking the logarithm, we obtain the optimization problem $\mathcal{P}_1$ in (??).

B. Problem Relaxation

The discrete set $\bar{\mathcal{Q}}^{N \times R_D}$ of $\mathbf{X}_D$ renders $\mathcal{P}_1$ a discrete-valued optimization problem, and a naïve exhaustive search would be impractical. To arrive at a tractable optimization problem, we relax the set $\bar{\mathcal{Q}}$ to its convex hull $\mathcal{B}$ as in [?]

$$\mathcal{B} = \left\{ \sum_{i=1}^{|\bar{\mathcal{Q}}|} \delta_i q_i : q_i \in \bar{\mathcal{Q}}, \delta_i \geq 0, \forall i; \sum_{i=1}^{|\bar{\mathcal{Q}}|} \delta_i = 1 \right\}. \quad (7)$$

For simplicity of exposition, we consider quadrature phase shift keying (QPSK) with $\mathcal{B} = \{x \in \mathbb{C} : -B \leq \text{Re}\{x\} \leq B, -B \leq \text{Im}\{x\} \leq B\}$ and $B > 0$. Nonetheless, our method can be generalized to other, higher-order quadrature amplitude modulation constellation sets.

By relaxing $\mathcal{P}_1$ into a continuous optimization problem, we still want the estimated elements of $\mathbf{X}_D$ to be in the set $\bar{\mathcal{Q}}$. To this end, we introduce a regularizer $\mathcal{C}(\mathbf{X}_D) = -\|\mathbf{X}_D \odot \mathbf{X}_D^* - B^2 \mathbf{1}_{N \times R_D}\|_F^2$ to $\mathcal{P}_1$, which pushes elements of $\mathbf{X}_D$ into $\bar{\mathcal{Q}}$ [?], where $\mathcal{C}(\mathbf{X}_D)$ will take its minimum only if all elements in $\mathbf{X}_D$ fall in $\bar{\mathcal{Q}}$. As such, we can transform $\mathcal{P}_1$ into

$$\begin{aligned} \mathcal{P}_2 : \{\hat{\mathbf{H}}, \hat{\mathbf{X}}_D\} = \arg \min_{\substack{\mathbf{H} \in \mathbb{C}^{MP \times N} \\ \mathbf{X}_D \in \mathcal{B}^{N \times R_D}}} & \frac{1}{2} \|\mathbf{Y} - \mathbf{H}[\mathbf{X}_P, \mathbf{X}_D]\|_F^2 \\ & + \mu_h \sum_{n=1}^N \sum_{p=1}^P \|\mathbf{h}_{n,p}\|_F \\ & + \mu_x \sum_{n=1}^N \|\bar{\mathbf{x}}_{D,n}\|_F + \lambda \mathcal{C}(\mathbf{X}_D), \end{aligned} \quad (8)$$

which can be solved approximately with FBS [?], [?].

C. Box-Constrained FBS Algorithm

FBS splits the objective function of a convex optimization problem into a smooth function $f(\mathbf{S})$ and an arbitrary (not necessarily smooth) function $g(\mathbf{S})$ [?], [?]:

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

The principle is to alternate between a gradient step for the smooth function $f(\mathbf{S})$ (forward step) and a proximal operation to find a point near the minimizer of the non-smooth function $g(\mathbf{S})$ (backward step). This process is repeated until

convergence (e.g., a stopping criterion is met). We apply this technique to the nonconvex problem $\mathcal{P}_2$.

1) *Problem Splitting*: With the definition $\mathbf{S} \triangleq [\mathbf{H}^H, \mathbf{X}_D]^H \in \mathbb{C}^{(MP+RD) \times N}$, we split the objective function in $\mathcal{P}_2$ into

$$\begin{aligned} f(\mathbf{S}) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{H}[\mathbf{X}_P, \mathbf{X}_D]\|_F^2 + \lambda \mathcal{C}(\mathbf{X}_D), \\ g(\mathbf{S}) &= \mu_h \sum_{n=1}^N \sum_{p=1}^P \|\mathbf{h}_{n,p}\|_F + \mu_x \sum_{n=1}^N \|\bar{\mathbf{x}}_{D,n}\|_F \\ &\quad + \mathcal{X}(\mathbf{X}_D), \end{aligned} \quad (10)$$

where $\mathcal{X}(\mathbf{X}_D)$ enforces the data to be within the convex set $\mathcal{B}$:

$$\mathcal{X}(\mathbf{X}_D) \triangleq \begin{cases} +\infty & , \exists \mathbf{X}_D(n, r) \notin \mathcal{B}, \forall n, r \\ 0 & , \mathbf{X}_D(n, r) \in \mathcal{B}, \forall n, r. \end{cases} \quad (11)$$

2) *Forward Step*: The forward step is given by

$$\hat{\mathbf{S}}^k = \mathbf{S}^k - \tau^k \nabla f(\mathbf{S}^k), \quad (12)$$

where the superscript k indicates the k th iteration, $\hat{\mathbf{S}}^k = [(\hat{\mathbf{H}}^k)^H, \hat{\mathbf{X}}_D^k]^H$, τ^k is the step size of the k th iteration [?], and gradient of $f(\mathbf{S})$ with respect to $\mathbf{S}$ is given by $\nabla f(\mathbf{S}) = [(\frac{\partial f}{\partial \mathbf{H}^*})^T, (\frac{\partial f}{\partial \mathbf{X}_D^T})^T]^T$ with

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{H}^*} &= -(\mathbf{Y} - \mathbf{H}\mathbf{X})\mathbf{X}^H, \\ \frac{\partial f}{\partial \mathbf{X}_D^T} &= -(\mathbf{Y}_D - \mathbf{H}\mathbf{X}_D)^H \mathbf{H} + \lambda \frac{\partial \mathcal{C}(\mathbf{X}_D)}{\partial \mathbf{X}_D^T}, \end{aligned} \quad (13)$$

where $\frac{\partial \mathcal{C}(\mathbf{X}_D)}{\partial \mathbf{X}_D^T} = -4(\mathbf{X}_D^* \odot (\mathbf{X}_D \odot \mathbf{X}_D^* - B^2 \mathbf{1}_{N \times R_D}))^T$.

3) *Backward Step*: The proximal operator for $\mathbf{H}$ is

$$\begin{aligned} \mathbf{H}^{k+1} &= \arg \min_{\mathbf{H} \in \mathbb{C}^{MP \times N}} \frac{1}{2} \|\mathbf{H} - \hat{\mathbf{H}}^k\|_F^2 \\ &\quad + \tau^k \mu_h \sum_{n=1}^N \sum_{p=1}^P \|\mathbf{h}_{n,p}\|_F, \end{aligned} \quad (14)$$

which has the following closed-form solution [?], [?], [?]:

$$\mathbf{h}_{n,p}^{k+1} = \hat{\mathbf{h}}_{n,p}^k \frac{\max\{\|\hat{\mathbf{h}}_{n,p}^k\|_F - \tau^k \mu_h, 0\}}{\|\hat{\mathbf{h}}_{n,p}^k\|_F}. \quad (15)$$

The proximal operator for $\mathbf{X}_D$ is given by

$$\begin{aligned} \mathbf{X}_D^{k+1} &= \arg \min_{\mathbf{X}_D \in \mathcal{B}^{N \times R_D}} \frac{1}{2} \|\mathbf{X}_D - \hat{\mathbf{X}}_D^k\|_F^2 \\ &\quad + \tau^k \mu_x \sum_{n=1}^N \|\bar{\mathbf{x}}_{D,n}\|_F, \end{aligned} \quad (16)$$

which is a convex optimization problem. To obtain the optimal solution of (??), we first decompose the problem (??) as N independent subproblems

$$\begin{aligned} \mathbf{r}_n &= \arg \min_{\mathbf{r}_n \in \mathbb{C}^{2R_D \times 1}} \frac{1}{2} \|\mathbf{r}_n - \hat{\mathbf{r}}_n^k\|_F^2 + \tau^k \mu_x \|\mathbf{r}_n\|_F \\ \text{s.t. } & -B \leq \mathbf{r}_n(d) \leq B, \forall d \in \{1, 2, 3, \dots, 2R_D\}, \end{aligned} \quad (17)$$

where $\mathbf{r}_n \triangleq [\text{Re}\{\bar{\mathbf{x}}_{D,n}\}^T, \text{Im}\{\bar{\mathbf{x}}_{D,n}\}^T]^T \in \mathbb{R}^{2R_D \times 1}$ and $\hat{\mathbf{r}}_n^k \triangleq [\text{Re}\{\hat{\mathbf{x}}_{D,n}^k\}^T, \text{Im}\{\hat{\mathbf{x}}_{D,n}^k\}^T]^T \in \mathbb{R}^{2R_D \times 1}$. Let $L(\mathbf{r}_n, \mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{r}_n - \hat{\mathbf{r}}_n^k\|_F^2 + \tau^k \mu_x \|\mathbf{r}_n\|_F + \sum_d \mathbf{p}(d)(\mathbf{r}_n(d) - B) - \sum_d \mathbf{q}(d)(\mathbf{r}_n(d) + B)$ be the Lagrangian, then the Karush-Kuhn-Tucker (KKT) conditions are as follows:

$$\mathbf{r}_n + \frac{\tau^k \mu_x}{\|\mathbf{r}_n\|_F} \mathbf{r}_n - \hat{\mathbf{r}}_n^k + \mathbf{p} - \mathbf{q} = 0, \quad (18a)$$

$$\mathbf{r}_n(d) - B \leq 0, -\mathbf{r}_n(d) - B \leq 0, \forall d, \quad (18b)$$

$$\mathbf{p}(d) \geq 0, \mathbf{q}(d) \geq 0, \forall d, \quad (18c)$$

$$\mathbf{p}(d)(\mathbf{r}_n(d) - B) = 0, \forall d, \quad (18d)$$

$$\mathbf{q}(d)(\mathbf{r}_n(d) + B) = 0, \forall d. \quad (18e)$$

To obtain the optimal solution that satisfies these KKT conditions, we can first solve (??) without constraints, i.e., $\mathbf{r}_{n,\text{tmp}}^{k+1} = \frac{\max\{\|\hat{\mathbf{r}}_n^k\|_F - \tau^k \mu_x, 0\}}{\|\hat{\mathbf{r}}_n^k\|_F} \hat{\mathbf{r}}_n^k$. It is clear that if $\mathbf{r}_{n,\text{tmp}}^{k+1}$ satisfies the conditions in (??), the problem is solved. However, if $\mathbf{r}_{n,\text{tmp}}^{k+1}$ does not satisfy these conditions, we must consider different cases. We first define two index sets $\mathcal{S}_p \triangleq \{d : \mathbf{r}_{n,\text{tmp}}^{k+1}(d) > B\}$ and $\mathcal{S}_q \triangleq \{d : \mathbf{r}_{n,\text{tmp}}^{k+1}(d) < -B\}$ and consider

- 1) If $d \in \mathcal{S}_p$, then we should set $\mathbf{p}(d) > 0$ and $\mathbf{q}(d) = 0$ to reduce its absolute value and ensure the corresponding value of optimal vector $\mathbf{r}_n^{k+1}(d) = B$.
- 2) If $d \in \mathcal{S}_q$, then we should set $\mathbf{p}(d) = 0$ and $\mathbf{q}(d) > 0$ to reduce its absolute value and ensure the corresponding value of optimal vector $\mathbf{r}_n^{k+1}(d) = -B$.
- 3) If $d \notin \mathcal{S}_p \cup \mathcal{S}_q$, then we should set $\mathbf{p}(d) = 0$ and $\mathbf{q}(d) = 0$ because positive $\mathbf{p}(d)$ and $\mathbf{q}(d)$ in cases 1) and 2) result in a smaller proximal coefficient $\frac{\max\{\|\hat{\mathbf{r}}_n^k - \mathbf{p} + \mathbf{q}\|_F - \tau^k \mu_x, 0\}}{\|\hat{\mathbf{r}}_n^k - \mathbf{p} + \mathbf{q}\|_F}$, causing $|\mathbf{r}_n^{k+1}(d)| < |\mathbf{r}_{n,\text{tmp}}^{k+1}(d)|$. Consequently, $\mathbf{r}_n^{k+1}(d)$ would still satisfy the conditions.

The next step is to find the value of $\mathbf{p}$ and $\mathbf{q}$. Let $\mathbf{m} = \hat{\mathbf{r}}_n^k - \mathbf{p} + \mathbf{q}$, and $a = \frac{\max\{\|\mathbf{m}\|_F - \tau^k \mu_x, 0\}}{\|\mathbf{m}\|_F}$. Then, there are two cases:

- i) If $\|\mathbf{m}\|_F > \tau^k \mu_x$, then $a = \frac{\|\mathbf{m}\|_F - \tau^k \mu_x}{\|\mathbf{m}\|_F} \in (0, 1]$ and we have $\mathbf{r}_n^{k+1} = a \mathbf{m}$, i.e., $\mathbf{m}(d) = \hat{\mathbf{r}}_n^k(d) \mathbb{I}\{d \notin \mathcal{S}_p \cup \mathcal{S}_q\} + \frac{B}{a} \mathbb{I}\{d \in \mathcal{S}_p\} - \frac{B}{a} \mathbb{I}\{d \in \mathcal{S}_q\}$. Accordingly, $\|\mathbf{m}\|_F$ can be rewritten as $\|\mathbf{m}\|_F = \sqrt{\frac{|\mathcal{S}_p \cup \mathcal{S}_q| B^2}{a^2} + \sum_{d \notin \mathcal{S}_p \cup \mathcal{S}_q} \hat{\mathbf{r}}_n^k(d)^2}$. This can be substituted into $a = \frac{\|\mathbf{m}\|_F - \tau^k \mu_x}{\|\mathbf{m}\|_F}$ to obtain a quartic equation with respect to a . Among the four solutions, the desired one lies within the range $(0, 1]$.
- ii) If the aforementioned quartic equation has no solution in the range $(0, 1]$, then we can only consider $\|\mathbf{m}\|_F \leq \tau^k \mu_x$, i.e., $a = 0$ and $\mathbf{r}_n^{k+1} = \mathbf{0}$.

After a maximum of K iterations of the box-constrained FBS algorithm, we obtain $\mathbf{S}^{K+1} = [(\mathbf{H}^{K+1})^H, \mathbf{X}_D^{K+1}]^H \in \mathbb{C}^{(MP+R_D) \times N}$ and the active set of UEs is identified by comparing the UEs' channel energy to a threshold; expressly, we set $\hat{\xi}_n = 1$ if $\|\hat{\mathbf{h}}_n^{K+1}\|_F^2 \geq T_{th}$ and $\hat{\xi}_n = 0$ otherwise.

IV. SIMULATION RESULTS

A. Simulation Setup

We consider a cell-free communication system containing $N = 400$ uniformly distributed UEs at a height of 1.65 m, and P uniformly distributed APs (we vary P from 20 to 100) each with $M = 4$ antennas and at a height 15 m in an area of $500 \text{ m} \times 500 \text{ m}$. We assume that the UE activity $\{\xi_n\}_{\forall n}$ follows an i.i.d. Bernoulli distribution with $\mathbb{P}\{\xi_n = 1\} = 0.2$, $\forall n$. Each active UE transmits $R_p = 50$ pilots generated from a complex equiangular tight frame [?] with $\|\mathbf{x}_{P,n}\|_F^2 = R_P$, and $R_D = 200$ data signals with unit amplitude, i.e., $B = \sqrt{0.5}$, over the channel with a bandwidth of 20 MHz and a carrier frequency of 1.9 GHz. The transmit power of the UEs is 0.1 W. We consider power control with a maximum dynamic power range of 12 dB between the weakest and strongest UE. Furthermore, shadow fading variance, noise figure, and noise temperature are 8 dB,

9 dB, and 290 K, respectively. For the FBS algorithms, we set the maximum number of iterations to $K = 200$ and a stopping tolerance of 10^{-3} . In our simulations, we perform $5 \cdot 10^3$ Monte-Carlo trials.

B. Performance Metrics and Baseline Algorithms

We consider the following performance metrics: user mis-detection rate (UMR), channel estimation normalized mean square error (NMSE), average symbol error rate (ASER), and cumulative symbol error rate (CSER), which are defined as

$$UMR = \frac{1}{N} \sum_{n=1}^N |\xi_n - \hat{\xi}_n|, \quad (19a)$$

$$NMSE = \|\mathbf{H} - \mathbf{H}^{K+1}\|_F^2 / \|\mathbf{H}\|_F^2, \quad (19b)$$

$$ASER = \frac{1}{R_D N_a} \sum_{n,r} \xi_n \mathbb{I}\{\mathbf{X}_D(n, r) \neq \hat{\mathbf{X}}_D^{K+1}(n, r)\}, \quad (19c)$$

$$CSER(x) = \sum_{n_a=1}^x ASER(n_a) \mathbb{P}(N_a = n_a). \quad (19d)$$

Here, $ASER(n_a)$ is the ASER for a specific number of active UEs $N_a = n_a$. CSER measures the impact of the number of active UEs on the symbol error rate.

To confirm the effectiveness of our algorithm, we compare it to different baselines, including “Joint AUD-CE via [?], then DD,” “Joint AUD-CE via [?], then DD,” and “Joint AUD-CE-DD via [?].” Given the estimated channel matrix and active UEs via [?] and [?], data detection is implemented by first performing zero-forcing equalization followed by mapping the result to the nearest QPSK symbol. To improve the convergence of JACD, we take the result of the baseline “Joint AUD-CE via [?], then DD” as the starting point for “Joint AUD-CE-DD via [?]” and our proposed FBS algorithm.

C. Simulation Results

In Fig. ??, we compare the different methods in terms of AUD, CE, and DD as the number of APs varies. Our proposed algorithm outperforms all considered baseline methods in terms of AUD, CE, and DD in most scenarios. Additionally, as the number of APs increases, the performance of all algorithms in AUD, CE, and DD improves and eventually stabilizes. We also observe that our FBS algorithm can accurately detect the set of active UEs and their data when the number of APs is no less than 60.

When comparing the “Joint AUD-CE, then DD” scheme, we note that the “Joint AUD-CE-DD” schemes generally outperform them in terms of DD performance. This can be attributed to the received data signal containing implicit information about UEs' channels and activity, which helps to improve DD performance. Moreover, compared with the “Joint AUD-CE-DD via [?]” baseline, we find that exploiting sparsity in the data matrix can improve JACD performance by providing additional UE sparse activity information to the proposed FBS algorithm.

To illustrate the probability distribution of data detection errors for different numbers of active UEs N_a , we plot the CSER of different methods for $P = 20$ in Fig. ?. Our proposed algorithm exhibits the best CSER performance regardless of the value of x , followed by the “Joint AUD-CE-DD via [?]” baseline, which implies that our method achieves the best DD performance regardless of the number of active UEs.

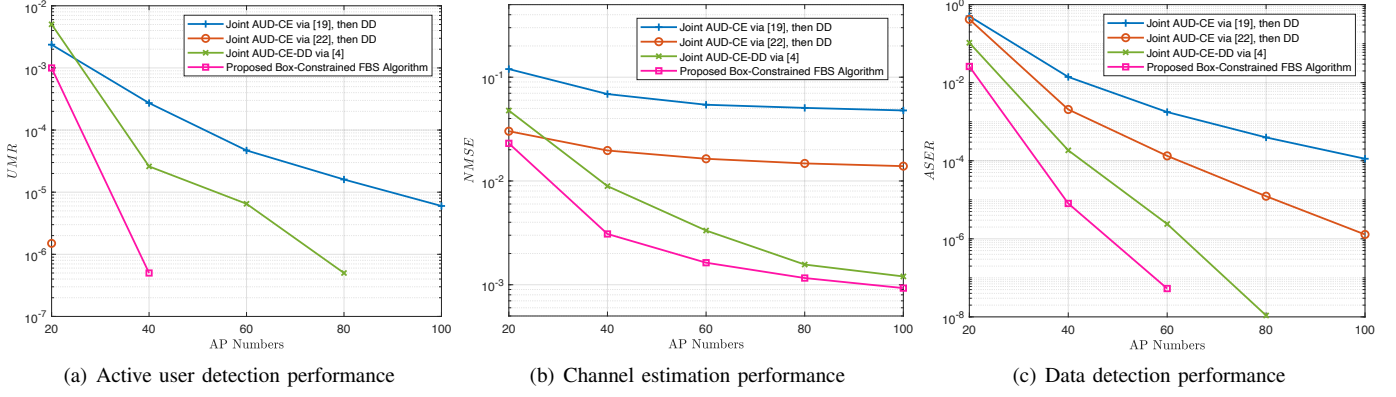


Fig. 2. Active user detection, channel estimation, and data detection performance comparison versus number of APs.

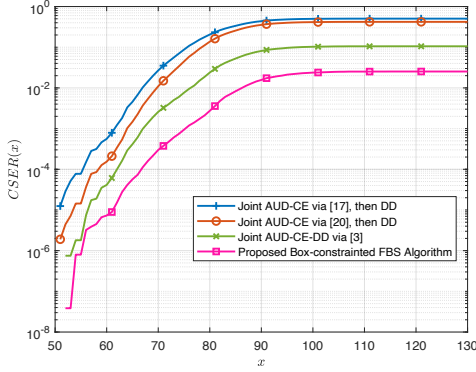


Fig. 3. Cumulative symbol error rate comparison under $P = 20$.

V. CONCLUSIONS

We have proposed a novel framework for joint active user detection, channel estimation, and data detection for massive grant-free transmission in cell-free wireless communication systems. The proposed box-constrained forward-backward splitting algorithm outperforms existing methods in terms of active user detection, channel estimation, and data detection in most scenarios by exploiting the sparsity of both the cell-free channels and users' intermittent activity.

REFERENCES

- [1] G. Sun, Y. Li, X. Yi, W. Wang, X. Gao, L. Wang, F. Wei, and Y. Chen, "Massive grant-free OFDMA with timing and frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3365–3380, May 2022.
- [2] A. Mishra, Y. Mao, L. Sanguinetti, and B. Clerckx, "Rate-splitting assisted massive machine-type communications in cell-free massive MIMO," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1358–1362, Jun. 2022.
- [3] W. Xu, Y. Huang, W. Wang, F. Zhu, and X. Ji, "Toward ubiquitous and intelligent 6G networks: from architecture to technology," *Sci. China Inf. Sci.*, vol. 66, no. 3, p. 130300, Feb. 2023.
- [4] H. Song, T. Goldstein, X. You, C. Zhang, O. Tirkkonen, and C. Studer, "Joint channel estimation and data detection in cell-free massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4068–4084, Jun. 2022.
- [5] M. Ke, Z. Gao, Y. Wu, X. Gao, and K.-K. Wong, "Massive access in cell-free massive MIMO-based Internet of Things: Cloud computing and edge computing paradigms," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 756–772, Mar. 2021.
- [6] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, Jan. 2023.
- [7] U. K. Ganesan, E. Björnson, and E. G. Larsson, "Clustering-based activity detection algorithms for grant-free random access in cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 69, no. 11, pp. 7520–7530, Nov. 2021.
- [8] S. Jiang, J. Dang, Z. Zhang, L. Wu, B. Zhu, and L. Wang, "EM-AMP-based joint active user detection and channel estimation in cell-free system," *IEEE Syst. J.*, pp. 1–12, Early Access, 2023.
- [9] M. Guo and M. C. Gursoy, "Joint activity detection and channel estimation in cell-free massive MIMO networks with massive connectivity," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 317–331, Jan. 2022.
- [10] X. Wang, A. Ashikhmin, Z. Dong, and C. Zhai, "Two-stage channel estimation approach for cell-free IoT with massive random access," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1428–1440, May 2022.
- [11] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, Nov. 2021.
- [12] Q. Zou, H. Zhang, D. Cai, and H. Yang, "A low-complexity joint user activity, channel and data estimation for grant-free massive MIMO systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 1290–1294, Jul. 2020.
- [13] R. B. Di Renna and R. C. de Lamare, "Joint channel estimation, activity detection and data decoding based on dynamic message-scheduling strategies for mMTC," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2464–2479, Apr. 2022.
- [14] S. Zhang, Y. Cui, and W. Chen, "Joint device activity detection, channel estimation and signal detection for massive grant-free access via BiGAMP," *IEEE Trans. Signal Process.*, vol. 71, pp. 1200–1215, Apr. 2023.
- [15] S. Jiang, X. Yuan, X. Wang, C. Xu, and W. Yu, "Joint user identification, channel estimation, and signal detection for grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6960–6976, Oct. 2020.
- [16] G. Sun, Y. Li, X. Yi, W. Wang, X. Gao, and L. Wang, "OFDMA based massive grant-free transmission in the presence of timing offset," in *Proc. 13th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Changsha, China, Oct. 2021, pp. 1–6.
- [17] 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.
- [18] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit massive MU-MIMO precoding in VLSI," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 508–522, Dec. 2017.
- [19] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," *arXiv preprint arXiv:1411.3406*, Nov. 2014.
- [20] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [21] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.
- [22] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.