Tackling Pilot Contamination in Cell-Free Massive MIMO by Joint Channel Estimation and Linear Multi-User Detection

Roya Gholami
EURECOM
Communication Systems Department
Sophia Antipolis, France
Email: roya.gholami@eurecom.fr

Laura Cottatellucci
Friedrich-Alexander University
Institute for Digital Communications
Erlangen, Germany
Email: laura.cottatellucci@fau.de

Dirk Slock
EURECOM
Communication Systems Department
Sophia Antipolis, France
Email: dirk.slock@eurecom.fr

Abstract—In this paper we consider cell-free (CF) massive MIMO (MaMIMO) systems, which comprise a very large number of geographically distributed access points (APs) serving a much smaller number of users. We exploit channel sparsity to tackle pilot contamination, which originates from the reuse of pilot sequences. Specifically, we consider semi-blind methods for joint channel estimation and data detection. Under the challenging assumption of deterministic parameters, we determine sufficient conditions and necessary conditions for semi-blind identifiability, which guarantee the non-singularity of the Fisher Information Matrix (FIM) and the existence of the Cramer-Rao bound (CRB). We propose a message passing (MP) algorithm which determines the exact channel coefficients in the case of semi-blind identifiability. We show that the system is identifiable if the Karp-Sipser algorithm yields an empty core. Additionally, we propose a Bayesian semi-blind approach which results in an effective algorithm for joint channel estimation and multi-user detection. This algorithm alternates between channel estimation and linear multi-user detection. Numerical simulations verify the analytical derivations.

I. INTRODUCTION

Recently, cell-free (CF) massive MIMO (MaMIMO) systems are attracting extensive research interests as an effective and promising approach for next generation wireless systems thanks to their potential to reap the benefit of both MaMIMO and distributed antenna systems (DAS). CF MaMIMO systems consist of a massive number of access points (APs) which serve a much smaller number of single-antenna users and are geographically distributed over a large coverage area. All the APs are connected through a back-haul network to a central processing unit (CPU). The massive number of antennas improves spectral efficiency [1] whereas energy efficiency [2], [3] and macro-diversity gain result from the distributed topology and ultra-densification. Additionally, since each user is surrounded by a large number of serving APs, with high probability all the users enjoy good channel conditions [4]. Therefore, CF MaMIMO systems are expected to provide significant improvements in terms of spectral/energy efficiency and coverage probability. In [1], [5], the performance of CF MaMIMO and small-cell systems were compared under the assumption of employing maximum ratio (MR) processing. In [6]–[10], the authors advocated the use of more effective

processing than MR processing in CF MaMIMO to guarantee superior performance of CF MaMIMO systems compared to small-cell systems. The performance of CF MaMIMO systems is critically affected by the so-called *pilot contamination*. This impairment originates from the reuse of training sequences or pilots utilized in channel estimation, which prevents the possibility of obtaining an adequate estimate of the channel state information (CSI). The detrimental effects of pilot contamination were highlighted in [11] for centralized MaMIMO systems. Specific features of centralized MaMIMO channels such as channel hardening and favorable propagation or limited angular spread could be exploited to "separate" user channels in power domain [12], angular domain [13], [14], or jointly in power and angular domain [15] and thus, mitigate or annihilate pilot contamination. However, these appealing properties of channels in centralized MaMIMO systems are destroyed in a distributed setting and pilot contamination is still an open and challenging problem in CF MaMIMO systems. Several pilot assignment (PA) methods for mitigating pilot contamination in CF MaMIMO systems were proposed recently in [1], [16]–[18]. In [1], a greedy pilot assignment (GPA) based on knowledge of large-scale fading channel coefficients was proposed. In [16], a location-based greedy (LBG) pilot assignment scheme utilized the location information in a GPA algorithm. The structured PA approach proposed in [17] maximized the minimum geographical distance between users sharing the same pilot sequences. An additional PA method based on graph coloring was proposed in [18]. All these techniques address the pilot contamination problem via a careful assignment of pilots and do not exploit the inherent structure of channels and data in CF MaMIMO systems in contrast to blind or semi-blind estimation and detection techniques. A blind pilot decontamination approach was proposed first in [12] for centralized MaMIMO systems and utilized asymptotic orthogonality of user channels to remove undesired interference including pilot contamination from the received signal. The same property was also exploited for semi-blind channel estimation, e.g., [15], in centralized MaMIMO but it does not hold in CF MaMIMO systems [8], [9]. Blind and semi-blind channel estimation have been thoroughly investi-

gated in general settings, see, e.g., [19]–[22] and references therein. In this context, the concept of identifiability was very relevant since it guarantees the non-singularity of the Fisher information matrix (FIM) and thus, the existence of the Cramer-Rao bound (CRB). The corresponding conditions provide fundamental insights into the feasibility of reliable communications in the analyzed system. Conditions under which channel and data signals are blindly and semi-blindly identifiable have been thoroughly studied in various settings for centralized systems, see, e.g., [23], [24].

CF MaMIMO channels are inherently sparse due to the distribution of APs over a large area and the natural path loss of wireless channels. In this paper, we study semi-blind joint channel estimation and data detection for exploiting the sparsity of the channel support in CF MaMIMO systems to combat pilot contamination. The potential of this approach is analyzed via system identifiability and sets of sufficient conditions and necessary conditions under which channels and data are identifiable are provided. We define a graph that has APs and users as factor and variable nodes and propose a message passing (MP) algorithm over this graph which computes the channel coefficients if the identifiability conditions are satisfied. As by product, we also show that the conditions for semi-blind identifiability are satisfied if the Karp-Sipser algorithm [25]–[27] yields an empty core. Additionally, we derive the FIM and CRB for joint channel estimation and data detection and propose a Bayesian semi-blind method that alternates between channel estimation and linear multi-user detection.

The remainder of this paper is organized as follows. We describe the system and channel model in Section II. The CRB and the identifiability conditions under the assumption of deterministic parameters are presented in Section III and IV, respectively. In Section V, we propose a Bayesian semi-blind iterative algorithm that alternates between channel estimation and linear multi-user detection. Numerical results are illustrated in Section VI. Finally, concluding remarks are drawn in Section VII.

Notation: In the following, superscripts ^{T}, *, and ^{H} stand for transpose, conjugate, and conjugate transpose, respectively. Uppercase and lowercase bold symbols denote matrices and vectors, respectively. The expectation operator is indicated by $\mathbb{E}\{.\}$ and $\mathbf{I}_P$ is the $P \times P$ identity matrix. Here, $\|\cdot\|$ and $\mathrm{diag}(.)$ denote the Euclidean norm operator and the squared diagonal matrix consisting of the diagonal elements of matrix argument, respectively. $\mathrm{vec}(.)$ denotes $\mathrm{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_{:,1}^T & \mathbf{A}_{:,2}^T \cdots \mathbf{A}_{:,n}^T \end{bmatrix}^T$, where $\mathbf{A}_{:,j}$ is the j-th column of matrix $\mathbf{A}$ and $\mathrm{tr}\{.\}$ is the trace operator. The Kronecker operator is denoted by $\otimes$. Finally, $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{CN}(\mu, \sigma^2)$ denote a real and a complex Gaussian distribution with mean μ and variance σ^2, respectively.

II. SYSTEM MODEL

We consider the uplink of a CF MaMIMO system consisting of K users and M APs equipped with a single antenna and randomly distributed over a $D \times D$ square area. We assume that $M \geq K$. The M APs are connected to a central processing unit (CPU) via a back-haul network. The channel matrix between the APs and users is given by $\mathbf{H} \in \mathbb{C}^{M \times K}$, whose (m, k)-element h_{mk} is the channel coefficient between AP m and user k and is modeled as follows

$$h_{mk} = \sqrt{\beta_{mk}}\, g_{mk}, \qquad (1)$$

where β_{mk} represents the large-scale fading coefficient which accounts for path loss and shadowing effects and g_{mk} represents the small-scale fading. We assume that g_{mk}, $m = 1, \cdots M$, $k = 1, \cdots K$, are independent and identically distributed (i.i.d.) complex normal random variables, i.e., $g_{mk} \sim \mathcal{CN}(0, 1)$. Additionally, we assume perfect knowledge of the large-scale fading coefficients β_{mk}, $m = 1, \cdots M$, $k = 1, \cdots K$ at the CPU. Due to the path loss, the channel coefficients are assumed to be negligible at distances higher than a given threshold γ. Then, for each AP m, the CPU is required to estimate only the channels of the users in a disc centered around AP m with radius γ while the signals transmitted from users external to the disc are treated as additive noise. We denote by $\mathcal{K}_I(m)$ and $\mathcal{K}_0(m)$ the sets of users inside the disc centered around AP m and remaining users, respectively. At a global level, this determines a partition of the channel coefficients into two groups, the channel coefficients that have to be detected $\mathcal{K}_I \equiv \{h_{mk} | m = 1, \ldots M, k \in \mathcal{K}_I(m)\}$ and the complement set $\mathcal{K}_0 \equiv \{h_{mk} | m = 1, \ldots M, k \in \mathcal{K}_0(m)\}$. Consistently with this partition, we decompose the channel matrix $\mathbf{H}$ into two matrices $\mathbf{H}_I$ and $\mathbf{H}_0$ such that $\mathbf{H} = \mathbf{H}_I + \mathbf{H}_0$. Then, $\mathbf{H}_I$ and $\mathbf{H}_0$ of size $M \times K$ denote the matrices of the relevant and negligible channel coefficients, respectively. Throughout this paper, we assume that $\gamma \ll D$ and the APs are distributed over the whole region such that matrix $\mathbf{H}_I$ has a large number of zero elements.

In the uplink transmission, each user sends one of P pilot sequences known by the CPU followed by $L - P$ unknown data symbols. The pilot sequences are assumed to be ortho-normal, i.e., orthogonal with unit norm. The L received symbols at the M APs are given by

$$\mathbf{Y} = \sqrt{\rho}\, \mathbf{H}_I\, \mathbf{X} + \sqrt{\rho}\, \mathbf{H}_0\, \mathbf{X} + \mathbf{W}, \qquad (2)$$

where ρ denotes the transmit power at each user terminal normalized by the noise variance. $\mathbf{Y} \in \mathbb{C}^{M \times L}$ is a matrix of the L received symbols at the M APs and $\mathbf{X} \in \mathbb{C}^{K \times L}$ is a matrix of the transmitted symbols. Note that the k-th row corresponds to the signals transmitted by user k. The matrix $\mathbf{W} \in \mathbb{C}^{M \times L}$ is the additive white Gaussian noise (AWGN) with i.i.d. components having zero mean and unit variance.

Let $\mathbf{X}_p \in \mathbb{C}^{K \times P}$ and $\mathbf{X}_d \in \mathbb{C}^{K \times (L-P)}$ denote the pilot sequences and data symbols, respectively. Then, $\mathbf{X} = [\mathbf{X}_p \; \mathbf{X}_d]$. Similarly, $\mathbf{Y} = [\mathbf{Y}_p \; \mathbf{Y}_d]$ where $\mathbf{Y}_p \in \mathbb{C}^{M \times P}$ and $\mathbf{Y}_d \in \mathbb{C}^{M \times (L-P)}$ represent the matrices of received training and data signals, respectively.

III. CRB FOR SEMI-BLIND JOINT CHANNEL ESTIMATION AND DATA DETECTION

To analyze the performance of the semi-blind channel estimation, we derive the CRB in a deterministic framework. In

the deterministic framework, both data signal $\mathbf{X}_d$ and relevant channel coefficients $\mathbf{H}_I$ are modeled as unknown deterministic quantities. Thus, we have

$$\mathbf{y} \sim \mathcal{CN}\big(m_{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{C}_{\mathbf{yy}}\big) \qquad (3)$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$ and $\boldsymbol{\theta} = [\mathbf{h}_I^H \ \text{vec}^H(\mathbf{X}_d)]^H$ is the complex unknown parameter vector to be estimated. Here, $\mathbf{h}_I$ is a vector deduced from the non-zero elements of the matrix $\mathbf{H}_I$, whose support is known. Mean and covariance of received signal $\mathbf{y}$ are given by $m_{\mathbf{y}}(\boldsymbol{\theta}) = \sqrt{\rho}\,\text{vec}(\mathbf{H}_I\mathbf{X})$ and $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}_L \otimes \mathbf{C}_{\mathbf{YY}}$, respectively, with $\mathbf{C}_{\mathbf{YY}} = \mathbf{I}_M + \rho\,\mathbf{C}_0$ and covariance matrix $\mathbf{C}_0$ specified in the following:

$$\mathbf{C}_0 = \mathbb{E}\big\{\mathbf{H}_0 \mathbf{H}_0^H\big\} = \text{diag}\Big(\sum_{k \in \mathcal{K}_0} \beta_{1k}, \cdots, \sum_{k \in \mathcal{K}_0} \beta_{Mk}\Big).$$

The probability density function[1] (pdf) of the observations $\mathbf{Y}$ in the parameter $\boldsymbol{\theta}$ is given by

$$f(\mathbf{Y}|\boldsymbol{\theta}) = \frac{1}{\pi^{ML}(\det(\mathbf{C}_{\mathbf{YY}}))^L}\,\exp\Big(-\|\mathbf{C}_{\mathbf{yy}}^{-1/2}(\mathbf{y}-m_{\mathbf{y}})\|^2\Big).$$

Computing the Jacobian of $m_{\mathbf{y}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, the deterministic complex Fisher information matrix (FIM) denoted as $\mathcal{J}_{\boldsymbol{\theta},\boldsymbol{\theta}}^d$ on the basis of the data $\mathbf{Y}$ is given by

$$\mathcal{J}_{\boldsymbol{\theta},\boldsymbol{\theta}}^d = \Big(\frac{\partial m_{\mathbf{y}}^H}{\partial \boldsymbol{\theta}^*}\Big)\mathbf{C}_{\mathbf{yy}}^{-1}\Big(\frac{\partial m_{\mathbf{y}}^H}{\partial \boldsymbol{\theta}^*}\Big)^H = \rho\,\big[\mathbf{Q}'\ \mathbf{R}'\big]^H\big[\mathbf{Q}'\ \mathbf{R}'\big] \quad (4)$$

where $\mathbf{Q}' = \mathbf{C}_{\mathbf{yy}}^{-1/2}\mathbf{Q}$, $\mathbf{Q} = \frac{1}{\sqrt{\rho}}\frac{\partial m_{\mathbf{y}}}{\partial \mathbf{h}_I^T}$ and $\mathbf{R}' = \mathbf{C}_{\mathbf{yy}}^{-1/2}\mathbf{R}$, $\mathbf{R} = \frac{1}{\sqrt{\rho}}\frac{\partial m_{\mathbf{y}}}{\partial \text{vec}^T(\mathbf{X}_d)}$. Note that $\frac{1}{\sqrt{\rho}}m_{\mathbf{y}} = \text{vec}(\mathbf{H}_I\mathbf{X}) = \mathbf{Q}\,\mathbf{h}_I = \text{vec}(\mathbf{H}_I[\mathbf{X}_p\ \mathbf{0}]) + \mathbf{R}\,\text{vec}(\mathbf{X}_d)$. The FIM $\mathcal{J}_{\boldsymbol{\theta},\boldsymbol{\theta}}^d$ is a 2×2 block matrix. The deterministic CRB^d is obtained as the inverse of the Fisher information matrix $\mathcal{J}_{\boldsymbol{\theta},\boldsymbol{\theta}}^d$

$$\text{CRB}^d = \big(\mathcal{J}_{\boldsymbol{\theta},\boldsymbol{\theta}}^d\big)^{-1}. \qquad (5)$$

The blocks $(1,1)$ and $(2,2)$ of the CRB^d in (5) relative to the estimation of the channel coefficients $\mathbf{h}_I$ and data symbols $\text{vec}(\mathbf{X}_d)$, respectively are given as follows

$$\text{CRB}_{\mathbf{h}_I}^d = \frac{1}{\rho}\big(\mathbf{Q}'^H P_{\mathbf{R}'}^\perp \mathbf{Q}'\big)^{-1} \qquad (6)$$

$$\text{CRB}_{\text{vec}(\mathbf{X}_d)}^d = \frac{1}{\rho}\big(\mathbf{R}'^H P_{\mathbf{Q}'}^\perp \mathbf{R}'\big)^{-1} \qquad (7)$$

where $P_{\mathbf{A}} = \mathbf{A}\big(\mathbf{A}^H\mathbf{A}\big)^{-1}\mathbf{A}^H$ and $P_{\mathbf{A}}^\perp = \mathbf{I} - P_{\mathbf{A}}$ denote the projection matrices on the column space of matrix $\mathbf{A}$ and its orthogonal complement, respectively. In the deterministic identifiability analysis that follows, we shall ignore $\mathbf{C}_0$ ($\mathbf{C}_0 = \mathbf{0}$) and hence $\mathbf{C}_{\mathbf{YY}} = \mathbf{I}_M$, $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}_{ML}$.

IV. Identifiability

In this section, we derive sets of both sufficient and necessary conditions for the identifiability of vector parameter $\boldsymbol{\theta}$ under the assumption that $\boldsymbol{\theta}$ is a deterministic unknown parameter. Then, we propose an MP algorithm over a graph that determines the exact channel coefficients if the sufficient

---

[1]For the sake of compactness, we adopt an identical notation $f(\mathbf{Y}|\boldsymbol{\theta})$ to indicate the pdf of random variable (r.v.) $\mathbf{Y}$ in vector parameter $\boldsymbol{\theta}$ or conditioned to r.v. $\boldsymbol{\theta}$ when $\boldsymbol{\theta}$ is assumed to be a deterministic unknown vector parameter or a r.v., respectively.

identifiability conditions are satisfied. Finally, we show that the system is identifiable via semi-blind algorithms if the Karp-Sipser algorithm applied to the same graph yields an empty core paving the way to an analysis of asymptotically large networks based on core percolation properties.

In the framework of deterministic identifiability, we assume that vector parameter $\boldsymbol{\theta}$ is deterministic and consider channel $\mathbf{H}_0$ negligible. Then, the observation $\mathbf{y}$ is Gaussian distributed, i.e., $\mathbf{y} \sim \mathcal{CN}(m_{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{I}_{ML})$ with covariance matrix independent of $\boldsymbol{\theta}$. The identifiability of $\boldsymbol{\theta}$ relies only on the known mean $m_{\mathbf{y}}(\boldsymbol{\theta})$ and, for semi-blind methods, $\mathbf{X}_d$ and $\mathbf{h}_I$ are said to be *identifiable* [23] if

$$\mathbf{H}_I\mathbf{X} = \mathbf{H}_I'\mathbf{X}' \ \Rightarrow\ \mathbf{h}_I = \mathbf{h}_I' \ \text{and} \ \mathbf{X}_d = \mathbf{X}_d' \qquad (8)$$

Let $m_{\mathbf{Y}\sim^0}$ be the expectation of $\mathbf{Y}$ in (2) obtained assuming $\mathbf{H}_0$ negligible. The identifiability problem reduces to analyze the following bi-linear system of equations in the unknowns $\mathbf{h}_I$ and $\mathbf{X}_d$

$$m_{\mathbf{Y}\sim^0} = \sqrt{\rho}\,\mathbf{H}_I\mathbf{X}$$

and determine under which conditions this system admits a unique solution, which is assumed to exist. These identifiability conditions are summarized in the following proposition.

PROPOSITION 1 **Sufficient Identifiability Conditions** – *Let $\mathcal{S}_k$ denote the support of the channel of user k, i.e., the set of all the indices m such that $\mathbf{H}_{I,m,k} \neq 0$, and let $|\mathcal{S}_k|$ be its cardinality. In a semi-blind joint data detection and channel estimation method, the unknown parameters $\mathbf{h}_I$ and $\mathbf{X}_d$ are identifiable if (i) the $K \times L$ matrix $\mathbf{X}$, with $L \geq K$ has full row rank K, (ii) the channel of each user is sparse and $|\mathcal{S}_k| \leq M - K + 1$, and (iii) for each group of users $\mathcal{G}_p$ utilizing the same ortho-normal pilot sequence $\mathbf{x}_p^{(p)}$, it is possible to identify a sequence $\{\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \ldots \mathcal{G}_{p,s}\}$ satisfying the following properties:*

1) $\bigcup_{j=1}^{s} \mathcal{G}_{p,j} \equiv \mathcal{G}_p$, *i.e., the sequence of subsets is a partition of $\mathcal{G}_p$*
2) *In the support of the channel of each user $k \in \mathcal{G}_{p,i}$ there exists at least an index $j \in \mathcal{S}_k$ that is not contained in any of the channel supports of other users in the same group $\mathcal{G}_{p,i}$ or in the following groups of the sequence $\mathcal{G}_{p,i+1}, \ldots \mathcal{G}_{p,s}$.*

REMARK 1 *Condition iii-2 implies that the signal transmitted by each user k in $\mathcal{G}_{p,i}$ impinges an AP in the disc $\mathcal{M}_k$ centered around user k with radius γ and no other signal transmitted by other users in $\mathcal{G}_{p,i}$ or subsequent subsets $\mathcal{G}_{p,i+1}, \mathcal{G}_{p,i+2}, \ldots \mathcal{G}_{p,s}$ impinges the same AP.*

REMARK 2 *The assumption that $\mathbf{X}$ has full row rank K implies that $\mathbf{X}_d$ has at least rank $K - P$.*

Proof: Observe that since in CF MaMIMO systems $M \gg K$, and the channel matrix $\mathbf{H}_I$ consists of independent channels, we can assume that it has full row rank equal to K with probability 1. Thanks to the assumptions of Proposition 1, also matrix $\mathbf{X}$ has full row rank equal to K as well as matrix $m_{\mathbf{Y}\sim^0}$. Then, the singular value decomposition (SVD) of the noise-free system is given by

$$\frac{1}{\sqrt{\rho}} m_{\mathbf{Y}^{\sim 0}} = \mathbf{H}_I \mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \qquad (9)$$

where $\mathbf{U} \in \mathbb{C}^{M \times K}$ and $\mathbf{V} \in \mathbb{C}^{L \times K}$ are the matrices of the left and right singular-vectors and $\boldsymbol{\Sigma}$ is the $K \times K$ diagonal matrix of singular values. Additionally, the left and right singular value matrices $\mathbf{U}$ and $\mathbf{V}$ span the channel subspace $\mathbf{H}_I$ and the signal space $\mathbf{X}$, respectively. Then, the problem of identifiability reduces to determine a $K \times K$ non-singular matrix $\mathbf{T}$ such that $\mathbf{H}_I = \mathbf{U}\mathbf{T}$ and then, also matrix $\mathbf{X}$ is unequivocally given by $\mathbf{X} = \mathbf{T}^{-1}\boldsymbol{\Sigma}\mathbf{V}^H$. In order to determine matrix $\mathbf{T}$, we utilize the following properties and information:

- The support of each user channel is known and sparse and at least $K-1$ channel coefficients are zero.
- The *contaminated channel.* More specifically, let us consider the linear system of equations corresponding to the transmission of the pilot sequences, i.e., $\frac{m_{\mathbf{Y}_{\tilde p}^{\sim 0}}}{\sqrt{\rho}} = \mathbf{H}_I \mathbf{X}_p$, where $m_{\mathbf{Y}_{\tilde p}^{\sim 0}}$ denotes the expectation of $\mathbf{Y}_p = \sqrt{\rho}\mathbf{H}_I\mathbf{X}_p$. By post-multiplying both sides of the system by the pilot sequence $\mathbf{x}_p^{(p)}$ and exploiting the ortho-normality of the training sequences, $\mathbf{X}_p \mathbf{x}_p^{(p)} = \mathbf{1}_{\mathcal{G}_p}$ where $\mathbf{1}_{\mathcal{G}_p}$ is the K-dimensional vector with elements with indices in $\mathcal{G}_p$, i.e., indices corresponding to users transmitting pilot $\mathbf{x}_p^{(p)}$, equal to one and and zero elsewhere. Then, it is apparent that this system of equations enables to determine exactly at each AP the sum of all the non-zero channel coefficients of the users in each group $\mathcal{G}_p$, $p = 1, \dots P$, i.e., $\frac{1}{\sqrt{\rho}} m_{\mathbf{Y}_{\tilde p}^{\sim 0}} \mathbf{x}_p^{(p)} = \mathbf{H}_I \mathbf{1}_{\mathcal{G}_p}$.

Then, let us focus on a user k in $\mathcal{G}_{p,1}$. Thanks to the assumptions on the partition of $\mathcal{G}_p$, there exists at least an AP m such that $\mathbf{H}_{I,m,:} \mathbf{1}_{\mathcal{G}_p} = h_{m,k} = \frac{1}{\sqrt{\rho}} m_{\mathbf{Y}_{\tilde p}^{\sim 0}} \mathbf{x}_p^{(p)} \neq 0$, where $\mathbf{H}_{I,m,:}$ denotes the m-th row of the matrix $\mathbf{H}_I$. Furthermore, thanks to the assumption on the sparsity of the channels, we can obtain from the system of equations $\mathbf{H}_{I,:,k} = \mathbf{U}\mathbf{T}_{:,k}$ $K-1$ equations where the channel of user k is zero. Then, we can construct a non-homogeneous system of equations in the unknown $\mathbf{T}_{:,k}$ and the vector of constant terms consisting of zeros and at least the non-zero element h_{mk}. This system can be unequivocally solved to determine $\mathbf{T}_{:,k}$. Thanks to the properties of the sequence $\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \dots \mathcal{G}_{p,s}$, it is possible to determine sequentially, the columns of matrix $\mathbf{T}$ corresponding to a certain group, compute exactly the corresponding channels of the users in the group and cancel them from the contaminated channel for group $\mathcal{G}_p$ until the complete computation of all the columns of matrix $\mathbf{T}$ corresponding to all the users in $\mathcal{G}_p$ and the corresponding channels. This approach can be repeated for all the groups up to the complete computation of matrix $\mathbf{T}$ and channel $\mathbf{H}_I$. Then, we observe that $\mathbf{T}$ has full rank K since $\mathbf{H}_I$ has full row rank. The inverse of $\mathbf{T}$ exists and enables the computation of $\mathbf{X}_d$. This concludes the proof. ∎

In the following, let $(\mathbf{H})_{\mathcal{G}_p}$ denote a reduced version of the matrix $\mathbf{H}$ containing only the columns corresponding to the users in $\mathcal{G}_p$.

PROPOSITION 2 *Necessary Identifiability Conditions – Iden-*

tification of $\mathbf{h}_I$ and $\mathbf{X}_d$ from the product $\mathbf{H}_I \mathbf{X}$ leads to the *global necessary identifiability condition*

$$\frac{1}{K} \sum_{k=1}^{K} |\mathcal{S}_k| \leq M - K + P \qquad (10)$$

or the per pilot necessary identifiability condition

$$\frac{1}{|\mathcal{G}_p|} \sum_{k \in \mathcal{G}_p} |\mathcal{S}_k| \leq M - K + \frac{K}{|\mathcal{G}_p|} \quad p = 1, \dots P. \qquad (11)$$

Proof: Consider again the SVD in (9), $\mathbf{H}_I \mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, with $\mathbf{V}^H$ partitioned into P plus $L - P$ columns similar to $\mathbf{X}$, $\mathbf{V}^H = [\mathbf{V}_p^H \; \mathbf{V}_d^H]$. Introducing again the unknown $K \times K$ mixture $\mathbf{T}$, this leads to the equations

$$\mathbf{H}_I = \mathbf{U}\mathbf{T} \;, \quad \mathbf{T}\mathbf{X}_p = \boldsymbol{\Sigma}\mathbf{V}_p^H \qquad (12)$$

which together represent $K(M + P)$ equations in the $\sum_{k=1}^{K} |\mathcal{S}_k|$ unknowns $\mathbf{h}_I$ and the K^2 unknowns $\mathbf{T}$. The proper conditions for solvability of the equations (12), that the number of equations needs to be at least equal to the number of unknowns, then leads to (10). If now we consider the equations for group of users $\mathcal{G}_p$, multiplying $\mathbf{T}\mathbf{X}_p = \boldsymbol{\Sigma}\mathbf{V}_p^H$ by $\mathbf{x}_p^{(p)}$ and exploiting $\mathbf{X}_p \mathbf{x}_p^{(p)} = \mathbf{1}_{\mathcal{G}_p}$ then we get

$$(\mathbf{H}_I)_{\mathcal{G}_p} = \mathbf{U}(\mathbf{T})_{\mathcal{G}_p} \;, \quad \mathbf{T}\mathbf{1}_{\mathcal{G}_p} = (\mathbf{T})_{\mathcal{G}_p} \mathbf{1} = \boldsymbol{\Sigma}\mathbf{V}_p^H \mathbf{x}_p^{(p)} \qquad (13)$$

which represents $M|\mathcal{G}_p| + K$ equations in the $\sum_{k \in \mathcal{G}_p} |\mathcal{S}_k| + K|\mathcal{G}_p|$ unknowns in $(\mathbf{H}_I)_{\mathcal{G}_p}$ and $(\mathbf{T})_{\mathcal{G}_p}$, hence leading to (11). ∎

It is worth noting that the proof of Proposition 1 along with the sufficient conditions for identifiability of the deterministic parameters, provides also a constructive method to determine the unknown parameters $\mathbf{H}_I$ and $\mathbf{X}_d$ if for each $p = 1, \dots P$, the sequence $\{\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \dots \mathcal{G}_{p,s}\}$ partitioning set $\mathcal{G}_p$ were known. In the following, we address this problem and provide an MP algorithm that enables to identify at iteration i the set $\mathcal{G}_{p,i}$ and determine the channel coefficients of all users in the set. Let us focus on the set $\mathcal{G}_p$ and associate to each user k and AP m variable node k and factor node m, respectively. We construct a bipartite graph by connecting a variable node with a factor node if the distance between the corresponding user and AP is lower than γ. We further assume that the factor nodes are initialized with the values of the vector $\mathbf{h}_{I,p}^c = \mathbf{H}_I \mathbf{1}_{\mathcal{G}_p}$, i.e., the sum of all the channel coefficients of users in the corresponding γ-neighborhood. Each variable node knows the matrix $\mathbf{U}$ that spans the channel subspace. The initial step of the MP algorithm starts at the factor nodes. Each factor node m that is a leaf transmits its initialization value $h_{I,p,m}^c$ to its neighbor. It transmits an erasure Δ if it is not a leaf. At iteration i, each variable node k that has received at least a message that is not an erasure solves the system of equations $\mathbf{U}\mathbf{T}_{:,k} = \mathbf{H}_{I,:,k}$ utilizing that value. The construction of a system of K equations to determine $\mathbf{T}_{:,k}$ is detailed in the proof of Proposition 1 and exploits the channel sparsity. Once $\mathbf{T}_{:,k}$ is known, it is possible to determine all the non-zero channel coefficients $\mathbf{H}_{I,:,k}$. Then, variable node

k transmits to all its neighbors the corresponding channel coefficients. Variable node k transmits the same messages in all the following iterations. If variable node k receives all erasures it transmits erasures to all its neighbors. The second step of iteration i determines the messages at the factor nodes. A factor node m computes a message for the output edge $< m, k >$ as the difference between its initialization value $h_{I,p,m}^c$ and all the incoming messages. The resulting message is not an erasure if all the incoming messages are not erasures otherwise the factor node transmits an erasure. The MP algorithm ends when all the channel coefficients have been determined and in this case the identifiability conditions are satisfied or when no additional erasure can be determined and thus the system is not identifiable. Set $\mathcal{G}_{p,i}$ includes all the users/variable nodes that compute their channel coefficients at iteration i. Interestingly, this algorithm is closely related to the MP algorithm for decoding of low density parity check (LDPC) codes in transmissions through binary erasure channels in [28]. It is worth noting that also for random generated CF MaMIMO systems with nodes independently generated, the corresponding graphs have edges intrinsically correlated due to the underlying geometric constraints and the corresponding sparse graphs do not have tree-like neighborhoods in asymptotic conditions. Then, the performance analysis of LDPC codes based on density evolution, see [28], is not directly applicable although the graph is sparse and the message passing yields exact results thanks to the noiseless nature of the considered system and thus the absence of error propagation.

Additionally, let us consider the Karp-Sipser or greedy leaf removal procedure [25]–[27] which consists in removing from a graph sequentially all the leaves and observe that sequential or simultaneous removal of leaves is equivalent in asymptotic conditions. Then, the sufficient identifiability conditions in Proposition 1 are satisfied if the greedy leaf removal procedure yields an empty core.

V. BAYESIAN SEMI-BLIND

Whereas deterministic parameter identifiability allows for consistency in SNR in the approximated model which ignores $\mathbf{C}_0$, in practice performance can be improved by furthermore exploiting prior information. Hence, exploiting the Rayleigh fading channel prior and capturing the uncorrelatedness and constant variance of the data symbols with an i.i.d. Gaussian prior, we get the overall log-likelihood

$$
\begin{aligned}
&\ln f(\mathbf{Y}|\boldsymbol{\theta}) + \ln f(\mathbf{h}_I) + \ln f(\mathbf{X}_d) \\
&= -\text{tr}\{(\mathbf{Y} - \sqrt{\rho}\,\mathbf{H}_I\mathbf{X})^H \mathbf{C}_{\mathbf{YY}}^{-1}(\mathbf{Y} - \sqrt{\rho}\,\mathbf{H}_I\mathbf{X})\} \\
&\quad -\mathbf{h}_I^H \mathbf{C}_{\mathbf{h}_I\mathbf{h}_I}^{-1} \mathbf{h}_I - \text{tr}\{\mathbf{X}_d^H\mathbf{X}_d\} + c^t .
\end{aligned} \tag{14}
$$

where c^t denotes a scalar constant. Alternating optimization with respect to $\mathbf{h}_I$ and $\mathbf{X}_d$ leads to

$$
\begin{aligned}
\widehat{\mathbf{h}}_I &= \sqrt{\rho}\left(\rho\,\widehat{\mathbf{Q}}^H \mathbf{C}_{\mathbf{yy}}^{-1}\widehat{\mathbf{Q}} + \mathbf{C}_{\mathbf{h}_I\mathbf{h}_I}^{-1}\right)^{-1}\widehat{\mathbf{Q}}^H \mathbf{C}_{\mathbf{yy}}^{-1/2}\mathbf{y} \\
\widehat{\mathbf{X}}_d &= \sqrt{\rho}\left(\rho\,\widehat{\mathbf{H}}_I^H \mathbf{C}_{\mathbf{YY}}^{-1}\widehat{\mathbf{H}}_I + \mathbf{I}_K\right)^{-1}\widehat{\mathbf{H}}_I^H \mathbf{C}_{\mathbf{YY}}^{-1}\mathbf{Y}_d
\end{aligned} \tag{15}
$$

where $\widehat{\mathbf{Q}} = \mathbf{Q}(\widehat{\mathbf{X}}_d)$ and $\widehat{\mathbf{H}}_I$ denotes the estimate of the matrix $\mathbf{H}_I$. The relation between $\widehat{\mathbf{h}}_I$ and $\widehat{\mathbf{H}}_I$ is the same as the
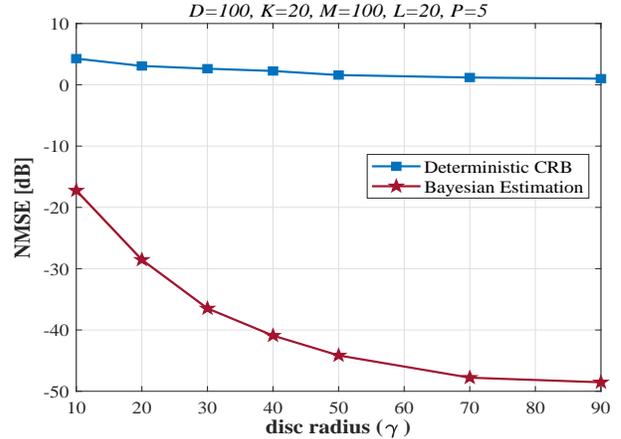
Fig. 1. NMSE [dB] versus disc radius (γ) for Bayesian estimation and deterministic CRB.

relation described for $\mathbf{h}_I$ and $\mathbf{H}_I$. This alternating procedure can be initialized with $\widehat{\mathbf{X}}_d = \mathbf{0}$.

VI. NUMERICAL RESULTS

First, we describe the path loss and shadow fading models used in numerical simulations for performance evaluation. The large-scale fading coefficient β_{mk} in (1) models path loss and shadow fading as follows

$$
\beta_{mk} = 10^{\frac{\text{PL}_{mk}}{10}}\, 10^{\frac{\sigma_{sh} z_{mk}}{10}} \tag{16}
$$

where PL_{mk} represents the path loss (expressed in dB), and $10^{\frac{\sigma_{sh} z_{mk}}{10}}$ represents the shadow fading with standard deviation σ_{sh} and $z_{mk} \sim \mathcal{N}(0,1)$. The three-slope model in [29] is adopted for the path loss. The uplink transmit power is $p = 100\,\text{mW}$, for all users. The performance of the Bayesian estimation is assessed by the normalized mean squared error (NMSE), $\frac{avg\|\mathbf{h}_I - \widehat{\mathbf{h}}_I\|^2}{avg\|\mathbf{h}_I\|^2}$ where avg stands for average. Fig. 1 shows the NMSE versus disc radius γ and compares the NMSE of Bayesian estimation and deterministic CRB, $\frac{avg\,\text{tr}\{\text{CRB}_{\mathbf{h}_I}^d\}}{avg\|\mathbf{h}_I\|^2}$, with $M = 100$ and $K = 20$. The Bayesian estimation outperforms the deterministic CRB.

VII. CONCLUSION

In this paper, we tackled the problem pilot contamination in CF MaMIMO systems leveraging only the channel sparsity. We considered semi-blind methods for joint channel estimation and data detection and derived the FIM and the CRB. Additionally, we determined sufficient conditions and necessary conditions for semi-blind identifiability under the assumption of deterministic parameters. An MP algorithm to verify identifiability and compute the channel coefficients was proposed and the relation with the Karp-Sipser procedure was highlighted. Finally, we proposed a Bayesian semi-blind approach resulting in an algorithm which alternates between channel estimation and linear multi-user detection. We verified the analytical derivations via numerical simulations.

REFERENCES

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

[2] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2112–2127, 2013.

[3] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 954–965, 2014.

[4] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.

[5] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445–4459, 2017.

[6] Z. Chen and E. Björnson, "Can we rely on channel hardening in cell-free massive MIMO?" in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.

[7] ——, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5205–5219, 2018.

[8] R. Gholami, L. Cottatellucci, and D. Slock, "Favorable propagation and linear multiuser detection for distributed antenna systems," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5190–5194.

[9] ——, "Channel models, favorable propagation and multistage linear detection in cell-free massive MIMO," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2942–2947.

[10] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2019.

[11] 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.

[12] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, 2014.

[13] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on selected areas in communications*, vol. 31, no. 2, pp. 264–273, 2013.

[14] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE Journal of selected topics in signal processing*, vol. 8, no. 5, pp. 942–953, 2014.

[15] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2990–3003, 2016.

[16] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive MIMO systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE, 2018, pp. 392–396.

[17] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.

[18] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9180–9184, 2020.

[19] E. De Carvalho and D. T. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *First IEEE signal processing workshop on signal processing advances in wireless communications*. IEEE, 1997, pp. 129–132.

[20] E. de Carvalho and D. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*, vol. 2. IEEE, 1997, pp. 1624–1628.

[21] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind MIMO channel estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 861–869, 2006.

[22] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Processing Letters*, vol. 15, pp. 202–205, 2008.

[23] E. de Carvalho and D. T. Slock, "Blind and semi-blind FIR multichannel estimation:(global) identifiability conditions," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1053–1064, 2004.

[24] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3280–3292, 2003.

[25] R. M. Karp and M. Sipser, "Maximum matching in sparse random graphs," in *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*. IEEE, 1981, pp. 364–375.

[26] J. Aronson, A. Frieze, and B. G. Pittel, "Maximum matchings in sparse random graphs: Karp–sipser revisited," *Random Structures & Algorithms*, vol. 12, no. 2, pp. 111–177, 1998.

[27] J.-H. Zhao and H.-J. Zhou, "Two faces of greedy leaf removal procedure on graphs," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 8, p. 083401, 2019.

[28] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[29] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No. 01CH37202)*, vol. 1. IEEE, 2001, pp. 333–336.