Message Passing for a Bayesian Semi-Blind Approach to Cell-Free Massive MIMO

Roya Gholami,[1] Laura Cottatellucci,[2] and Dirk Slock[1]
[1] EURECOM, Communication Systems Department, Sophia Antipolis, France
[2] Friedrich-Alexander University, Institute for Digital Communications, Erlangen, Germany
Emails:{roya.gholami, dirk.slock}@eurecom.fr, laura.cottatellucci@fau.de

***Abstract*—In this paper, we consider cell-free (CF) massive MIMO systems employing a massive number of access points (APs) geographically distributed over a wide area to jointly serve a smaller number of users over the same time-frequency resources. We consider semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data to resolve the pilot contamination. This task is further aided by exploiting prior channel information in a Bayesian formulation. We propose a variable level expectation propagation (VL-EP) algorithm for message passing (MP) style semi-blind channel estimation which provides an approximate minimum mean square error (MMSE) channel estimator (which itself can not be found analytically). Numerical simulations verify the analytical derivations and the proposed algorithm.**

***Index Terms*—Cell-free massive MIMO, expectation propagation, channel estimation, semi-blind, pilot contamination**

I. Introduction

In recent years, cell-free (CF) massive MIMO systems have drawn extensive attention as a potential paradigm shift for beyond 5G networks thanks to their potential to reap the benefit of both massive MIMO and distributed antenna systems (DAS) [1]–[3]. In a CF massive MIMO system, a massive number of geographically distributed access points (APs) jointly serve a much smaller number of single-antenna users distributed over a wide area on the same time-frequency resources [1]. All the APs are connected to a central processing unit (CPU) through a back-haul network. CF massive MIMO systems have shown a great potential in improving the network performance in various perspectives compared to the co-located massive MIMO and conventional small-cell systems [1], [4], [5]. As in centralized massive MIMO [6], CF massive MIMO systems exploit phenomena such as the favorable propagation [7], [8] and channel hardening [9], [10] as the number of APs grows large. Additionally, in CF massive MIMO systems each user is surrounded by a large number of serving APs which yields a high degree of macro-diversity [1] and with high probability all the users enjoy good channel conditions [2]. Therefore, CF massive MIMO systems are expected to provide significant improvements in terms of spectral/energy efficiency and coverage probability. Channel state information (CSI) in multiple antenna systems, both cellular and CF, is crucial for accomplishing successful transmission under various channel conditions. Ideally, training sequences or pilots should be selected to be mutually orthogonal in the channel estimation. However, in most practical scenarios the number of users is greater than the number of orthogonal training sequences and a given training sequence can be assigned to more than one user, therefore leading to the so-called *pilot contamination* which prevents obtaining an adequate estimate of CSI. Several pilot assignment (PA) techniques for suppressing pilot contamination in CF massive MIMO systems were proposed recently in [1], [11]–[15]. A semi-blind pilot decontamination approach was proposed in [16] exploiting the inherent structure of channels and data in CF massive MIMO systems. An efficient data detection algorithm with affordable complexity to achieve the optimum performance is of highly interest in large-scale networks, such as CF massive MIMO. An extensive attention has been dedicated to the design of detectors relying on message passing (MP) algorithms in recent years. The expectation propagation (EP) which is a kind of MP algorithm is a generalization of sum-product belief propagation (BP) algorithm. The EP algorithm attempts to find the closest approximation for a computationally intractable target probability distribution from a tractable family of distributions in an iterative refinement procedure by minimizing a Kullback-Leiber (KL) distance [17]–[20]. The EP algorithm was first proposed in [21] and summarized in, e.g., [22] for approximate inference in probabilistic graphical models. The method of EP was firstly applied to MIMO detection in [20], where an EP-based MIMO detector shows near-optimal performance with acceptable complexity under specific conditions. With EP-based MIMO detector, a Gaussian approximation is constructed for the posterior distribution of the transmitted symbols by an iterative procedure based on moment matching. The EP has been proposed as a low complexity algorithm for symbol detection in massive MIMO systems [23]–[26]. In [23], the EP principle was exploited for designing efficient detector of extra-large-scale massive MIMO systems with the subarray-based processing architecture. They represented the a posteriori distribution as a factor graph and developed the iterative algorithm by computing and transferring messages among different nodes on the factor graph. A non-coherent detection scheme for SIMO systems based on the EP algorithm was proposed in [24]. The proposed EP detector iteratively searches for the best approximation of the joint probability density function of the channel coefficients and the transmitted symbols. The output probability density function is used for direct estimation of the channel coefficients, as well as the transmitted symbols. In [26], an efficient MP algorithm for non-coherent multi-user detection scheme based on EP was proposed. In [27], a non-linear detector for CF massive MIMO networks based on the EP principle with a distributed approach [23] [28] was proposed. It was shown that such detector can achieve better performance than other linear receivers for both original and scalable CF massive MIMO networks. In this paper, we propose a new variable level expectation propagation (VL-EP) algorithm to iteratively conduct the semi-blind method for channel estimation in the presence of Gaussian i.i.d. data to mitigate the pilot contamination problem originating from the reuse of training sequences. The proposed VL-EP algorithm provides an approximate minimum mean square error (MMSE) channel estimator which can not be found analytically.

The remainder of this paper is organized as follows. We describe the system and channel model in Section II. A VL-EP algorithm is presented in Section III. In Section IV, we propose channel VL-EP for Gaussian-Gaussian semi-blind channel estimation. Numerical results are illustrated in Section V. Finally, concluding remarks are drawn in Section VI.

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 $\mathrm{vec}(\mathbf{A}) = \left[\mathbf{A}_{:,1}^T \;\; \mathbf{A}_{:,2}^T \cdots \mathbf{A}_{:,n}^T\right]^T$, where $\mathbf{A}_{:,j}$ is the j-th column of matrix $\mathbf{A}$. $\mathrm{tr}\{.\}$ is the trace operator and $\Re\{.\}$ denotes real part operator. 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 a CF massive MIMO system in uplink in which M APs serve K users in the same time-frequency resource. All APs and users equipped with a single antenna are randomly distributed over a $D \times D$ square area. Furthermore, all APs are connected to a CPU via a back-haul network. The channel is assumed to remain constant over L consecutive symbol intervals, i.e., a block. In the uplink transmission, each user k sends 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 received signal $\mathbf{Y} \in \mathbb{C}^{M\times L}$ at the M APs over the block interval is given by

$$\mathbf{Y} = \sqrt{\rho}\,\mathbf{H}\,\mathbf{X}^T + \mathbf{V}, \tag{1}$$

where ρ denotes the transmit power at each user terminal normalized by the noise variance, $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_K] \in \mathbb{C}^{L\times K}$ is a matrix of the transmitted symbols and $\mathbf{x}_k \in \mathbb{C}^{L\times 1}$ is the signal vector sent by user k. The channel vector between user k and M APs is denoted by $\mathbf{h}_k = [h_{1k} \dots h_{Mk}]^T \in \mathbb{C}^{M\times 1}$, then the channel matrix between the APs and users is given by $\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_K] \in \mathbb{C}^{M\times K}$. The matrix $\mathbf{V} \in \mathbb{C}^{M\times L}$ represents the additive white Gaussian noise (AWGN) with i.i.d. components having zero mean and unit variance. The channel coefficient h_{mk} between AP m and user k is modeled as follows

$$h_{mk} = \sqrt{\beta_{mk}}\, g_{mk}, \tag{2}$$

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., $g_{mk} \sim \mathcal{CN}(0,1)$. Furthermore, we assume perfect knowledge of the large-scale fading coefficients β_{mk}, $m = 1,\cdots M$, $k = 1,\cdots K$ at the CPU.

Let the matrices $\mathbf{X}_p \in \mathbb{C}^{P\times K}$ and $\mathbf{X}_d \in \mathbb{C}^{(L-P)\times K}$ denote the pilot sequences and data symbols, respectively. Then, $\mathbf{X} = [\mathbf{X}_p^T \;\; \mathbf{X}_d^T]^T$ and $\mathbf{x}_k = [\mathbf{x}_{p,k}^T \;\; \mathbf{x}_{d,k}^T]^T$. 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. Variable Level expectation propagation algorithm

Let $\boldsymbol{\theta} = [\mathbf{h}^T \; \mathbf{x}_d^T]^T$, where $\mathbf{h} = \mathrm{vec}(\mathbf{H})$ and $\mathbf{x}_d = \mathrm{vec}(\mathbf{X}_d)$. We partition $\boldsymbol{\theta}$ into groups such that $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}$ and we assume the prior factors at the level of these groups, which we call the variables in the factor graph terminology. Then, based on Bayes' theorem, the true posterior probability distribution is given by

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta}) \prod_i p(\boldsymbol{\theta}_i) \tag{3}$$

where the factors on the RHS are called the factors in a factor graph, $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, and $Z = p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta}) \prod_i p(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}$ is a normalization factor. Given the posterior probability $p(\boldsymbol{\theta}|\mathbf{y})$, the Bayesian MMSE estimate is given by

$$\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} \tag{4}$$

The posterior probability distribution in (3) involves a high-dimensional integral, therefore the Bayesian MMSE estimator in (4) is not computationally tractable. The EP algorithm provides an iterative method to approximate the posterior distribution of the unknown vector $\boldsymbol{\theta}$ from an exponential family $\mathcal{F}$ [21]. Here, we consider Gaussian distribution. In this regard, we propose the following approximation

$$\frac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta}) \approx \prod_i m(\boldsymbol{\theta}_i) \quad \Rightarrow \quad p(\boldsymbol{\theta}|\mathbf{y}) \approx q(\boldsymbol{\theta}) = \prod_i q(\boldsymbol{\theta}_i) \tag{5}$$

where $q(\boldsymbol{\theta}) \in \mathcal{F}$ is the approximate posterior in factored form at the variable level, and $q(\boldsymbol{\theta}_i) = m(\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)$ where the $m(\boldsymbol{\theta}_i)$ are the extrinsic probability density functions (pdf). Additionally, we assume that the prior pdfs $p(\boldsymbol{\theta}_i)$ to be simple (typically Gaussian or other members of the exponential family), so that they do not need approximation. Therefore, only the data pdf $p(\mathbf{y}|\boldsymbol{\theta})$ requires approximation. The EP approach adjusts the approximate posterior by minimizing a KL distance. In the original EP algorithm [29], the approximate posterior factors get approximated alternatively at the factor level, with each factor being optimized completely. In the original EP, the approximate factors are in the exponential family, but not constrained any further. Hence the factors can involve possibly all variables. However, it is possible to introduce constraints in approximate pdfs (e.g. Gaussians with a block diagonal covariance). In the EP variation considered in [26], the approximate factors are also factorizable at variable level. However, we propose to optimize the factors not at factor level but at variable level. Hence, the name variable level EP (VL-EP), as opposed to the classical factor level EP (FL-EP). But the updating follows exactly the EP principle: we optimize a factor $m(\boldsymbol{\theta}_i)$ by minimizing the KL distance

$$\begin{aligned}&\mathrm{KL}\big(\frac{1}{Z}\,p(\mathbf{y}|\boldsymbol{\theta})\,q(\boldsymbol{\theta}_{\bar{i}})\;||\;m(\boldsymbol{\theta}_i)q(\boldsymbol{\theta}_{\bar{i}})\big)\\ &=\frac{1}{Z}\int p(\mathbf{y}|\boldsymbol{\theta})\,q(\boldsymbol{\theta}_{\bar{i}})\,\ln\frac{\frac{1}{Z}p(\mathbf{y}|\boldsymbol{\theta})\,q(\boldsymbol{\theta}_{\bar{i}})}{m(\boldsymbol{\theta}_i)\,q(\boldsymbol{\theta}_{\bar{i}})}\,d\boldsymbol{\theta}\end{aligned}\tag{6}$$

with respect to (w.r.t.) a Gaussian $m(\boldsymbol{\theta}_i)$. Note that $q(\boldsymbol{\theta}) = q(\boldsymbol{\theta}_i)\,q(\boldsymbol{\theta}_{\bar{i}})$. The minimization of the KL distance leads to (see section 2 in [29] which exposes the original EP)

$$\begin{array}{rcl}\widehat{p}(\boldsymbol{\theta}_i) &=& \int q(\boldsymbol{\theta}_{\bar{i}})\;p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}_{\bar{i}}\\ Z_i &=& \int \widehat{p}(\boldsymbol{\theta}_i)\,d\boldsymbol{\theta}_i\\ \boldsymbol{\mu}_i &=& \frac{1}{Z_i}\int \boldsymbol{\theta}_i\,\widehat{p}(\boldsymbol{\theta}_i)\,d\boldsymbol{\theta}_i\\ \boldsymbol{\Sigma}_i &=& \frac{1}{Z_i}\int \boldsymbol{\theta}_i\,\boldsymbol{\theta}_i^H\,\widehat{p}(\boldsymbol{\theta}_i)\,d\boldsymbol{\theta}_i-\boldsymbol{\mu}_i\boldsymbol{\mu}_i^H\end{array}\tag{7}$$

where Z_i is a normalization constant, μ_i and $\boldsymbol{\Sigma}_i$ are the mean and covariance of the Gaussian $m(\boldsymbol{\theta}_i)$. Note that $\widehat{p}(\boldsymbol{\theta}_i)$ integrates out all other variables and produces the (unnormalized) target pdf for $\boldsymbol{\theta}_i$ that we approximate by the Gaussian $m(\boldsymbol{\theta}_i)$. It is this integration which produces the cleaned $\mathbf{y}$, cleaned from the interference of other variables $\boldsymbol{\theta}_{\bar{i}}$. Actually, the proof of (7) is fairly straightforward. Since the KL distance in (6) needs to be minimized w.r.t. (the parameters of) $m(\boldsymbol{\theta}_i)$, we can write

$$\begin{array}{rcl}\mathrm{KL} &=& c^t-\dfrac{1}{Z_i}\int\ln(m(\boldsymbol{\theta}_i))\,\widehat{p}(\boldsymbol{\theta}_i)\,d\boldsymbol{\theta}_i\\ &=& c^t+n_{\boldsymbol{\theta}_i}\ln(\pi)+\ln\det(\boldsymbol{\Sigma}_i)\\ && +\dfrac{1}{Z_i}\mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1}\int(\boldsymbol{\theta}_i-\boldsymbol{\mu}_i)(\boldsymbol{\theta}_i-\boldsymbol{\mu}_i)^H\;\widehat{p}(\boldsymbol{\theta}_i)\,d\boldsymbol{\theta}_i\}\\ &\geq& c^t+n_{\boldsymbol{\theta}_i}\ln(\pi)+\ln\det(\boldsymbol{\Sigma}_i)+\dfrac{n_{\boldsymbol{\theta}_i}}{Z_i}\end{array}\tag{8}$$

where c^t denotes (various) terms that are constant and $n_{\boldsymbol{\theta}_i}$ denotes the dimension of $\boldsymbol{\theta}_i$. The minimization over $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ leads to the solution in (7) and the minimal value in the last line in (8). As this minimal value is decreasing in Z_i, which itself is linear in $\widehat{p}(\boldsymbol{\theta}_i)$, we can majorize the KL distance by replacing $\widehat{p}(\boldsymbol{\theta}_i)$ by a minorizer and still retain a valid KL distance minimization strategy. We follow this strategy in the following section when the moments of $\widehat{p}(\boldsymbol{\theta}_i)$ can not be computed analytically.

IV. Channel VL-EP for Gaussian-Gaussian Semi-Blind

For the application to the Gaussian inputs Gaussian channel semi-blind (GG-SB) channel estimation problem, we shall consider the problem formulation which eliminates the Gaussian $\mathbf{X}_d$. Then, we have the correspondence $\boldsymbol{\theta}=\mathbf{h}$, $\boldsymbol{\theta}_i=\mathbf{h}_i$. We have $\mathbf{y}=[\mathbf{y}_p^T\;\mathbf{y}_d^T]^T$. We shall consider the development of $\mathbf{y}_p$ and $\mathbf{y}_d$ separately. The alternating updating of the posterior factors loops over the K users. The update for user k updates the posterior factor for $\mathbf{h}_k$. We consider first of all the pilot part and let user k use the n^{th} pilot, $k\in\mathcal{G}_n$ so that $\mathbf{x}_{p,k}=\mathbf{x}_p^{(n)}$. Then, the pilot signal model for user k can be written as

$$\begin{array}{rcl}\mathbf{Y}_p &=& \sqrt{\rho}\,\mathbf{h}_k\,\mathbf{x}_{p,k}^T+\sqrt{\rho}\,\sum_{i\neq k}\mathbf{h}_i\,\mathbf{x}_{p,i}^T+\mathbf{V}_p\\ \mathbf{Y}_p\mathbf{x}_p^{(n)*} &=& \sqrt{\rho}\,P\,\mathbf{h}_k+\sqrt{\rho}\,P\displaystyle\sum_{i\in\mathcal{G}_n\setminus\{k\}}(\widehat{\mathbf{h}}_i+\widetilde{\mathbf{h}}_i)+\mathbf{V}_p\mathbf{x}_p^{(n)*}\end{array}\tag{9}$$

where $\mathbf{V}=[\mathbf{V}_p\;\mathbf{V}_d]$ and $\mathbf{h}_i$ has (approximate) posterior pdf $q(\mathbf{h}_i)\sim\mathcal{CN}(\widehat{\mathbf{h}}_i,\mathbf{C}_i)$ and $\mathbf{C}_i=\mathbb{E}\{\widetilde{\mathbf{h}}_i\widetilde{\mathbf{h}}_i^H\}$. All variables whose pdf appears in different factors in the approximate posterior are treated as independent. Hence, $\sqrt{\rho}\,P\,\mathbf{h}_k$ has a Gaussian pdf with mean $\mathbf{Y}_p\mathbf{x}_p^{(n)*}-\sqrt{\rho}\,P\sum_{i\in\mathcal{G}_n\setminus\{k\}}\widehat{\mathbf{h}}_i$ and covariance

$$\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}=P\,\mathbf{I}_M+\rho\,P^2\sum_{i\in\mathcal{G}_n\setminus\{k\}}\mathbf{C}_i\,.\tag{10}$$

The likelihood from the pilot part needs to be combined with the data likelihood, where $\mathbf{x}_{d,i}$ has prior pdf $p_{\mathbf{x}_{d,i}}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_{L-P})$. The signal for user k can be written as

$$\mathbf{Y}_d=\sqrt{\rho}\,\mathbf{h}_k\,\mathbf{x}_{d,k}^T+\widetilde{\mathbf{Y}}_{d,k}\tag{11}$$

where

$$\begin{array}{rcl}\widetilde{\mathbf{Y}}_{d,k} &=& \sqrt{\rho}\,\sum_{i\neq k}\mathbf{h}_i\,\mathbf{x}_{d,i}^T+\mathbf{V}_d\\ &=& \sqrt{\rho}\,\sum_{i\neq k}(\widehat{\mathbf{h}}_i+\widetilde{\mathbf{h}}_i)\,\mathbf{x}_{d,i}^T+\mathbf{V}_d\end{array}\tag{12}$$

which $\mathbf{Y}_d$ has zero mean and covariance $\rho\,\mathbf{h}_k\mathbf{h}_k^H+\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$, where

$$\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}=\mathbf{I}_M+\rho\sum_{i\neq k}(\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^H+\mathbf{C}_i)\,.\tag{13}$$

We shall model $\widetilde{\mathbf{h}}_i\mathbf{x}_{d,i}^T$ also as Gaussian, since we need to go towards Gaussian approximations $q(\boldsymbol{\theta}_k)$ in any case. So we associate a Gaussian pdf to (11) by moment matching (whereas there are actually products of Gaussian variables). Given $\mathbf{h}_k$ and $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$, the negative log-likelihood leads to

$$-\ln p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})=\|\mathbf{Y}_d\|^2_{(\rho\,\mathbf{h}_k\mathbf{h}_k^H+\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})^{-1}}+c^t\tag{14}$$

with the squared weighted Frobenius norm $\|\mathbf{Y}\|^2_{\mathbf{A}}=\mathrm{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}^H\}$, and where the knowledge of $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ comprises the knowledge of $\widehat{\mathbf{h}}_{\overline{k}}$ and $\mathbf{C}_{\overline{k}}$. We construct a convex majorizer for the negative log-likelihood of (14) as in [30, Section V.A 4)], [31] which can actually also been derived with an expectation maximization (EM) approach. The construction of the majorizer is simply based on first-order Taylor series expansion of concave functions, either w.r.t. $\mathbf{H}$ directly or w.r.t. a covariance type expression (which is then quadratic in $\mathbf{H}$). Then, we can majorize the negative log-likelihood of (14) by

$$\begin{aligned}&-\ln p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})\\ &=-\ln\mathbb{E}_{\mathbf{x}_{d,k}^p,\mathbf{H}_{\overline{k}}^q\mathbf{X}_{d,\overline{k}}^p|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\,p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{h}_{\overline{k}},\mathbf{x}_d)\\ &=-\ln\mathbb{E}_{\mathbf{x}_{d,k}^p|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\,\mathbb{E}_{\mathbf{H}_{\overline{k}}^q\mathbf{X}_{d,\overline{k}}^p|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\frac{p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{h}_{\overline{k}},\mathbf{x}_d)\,p(\mathbf{x}_{d,k})}{p(\mathbf{x}_{d,k}|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})}\\ &=-\ln\mathbb{E}_{\mathbf{x}_{d,k}^p|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\,\mathbb{E}_{\mathbf{H}_{\overline{k}}^q\mathbf{X}_{d,\overline{k}}^p|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\,p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{h}_{\overline{k}},\mathbf{x}_d)+c^t\\ &\leq\mathbb{E}_{\mathbf{x}_{d,k}^p|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\{-\ln\mathbb{E}_{\mathbf{H}_{\overline{k}}^q\mathbf{X}_{d,\overline{k}}^p|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\,p(\mathbf{Y}_d|\mathbf{h}_k,\mathbf{h}_{\overline{k}},\mathbf{x}_d)\}+c^t\\ &=\mathbb{E}_{\mathbf{x}_{d,k}^p|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}\|\mathbf{Y}_d-\sqrt{\rho}\,\mathbf{h}_k\,\mathbf{x}_{d,k}^T\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}}+c^t\\ &=\|\mathbf{Y}_d-\sqrt{\rho}\,\mathbf{h}_k\,\widehat{\mathbf{x}}_{d,k}^T\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}}+\rho\,\mathrm{tr}\{\mathbf{R}_k\}\,\mathbf{h}_k^H\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}\mathbf{h}_k+c^t\\ &=-\ln\widehat{p}(\mathbf{h}_k)\end{aligned}\tag{15}$$

where the inequality follows from Jensen's inequality and the convexity of $-\ln(.)$. The column vector $\widehat{\mathbf{x}}_{d,k}$ and the matrix $\mathbf{R}_k$ which respectively denote the linear MMSE (LMMSE) estimate of $\mathbf{x}_{d,k}$ and associated error covariance matrix based on the current estimate $\widehat{\mathbf{h}}_k$, are given by

$$\begin{aligned} \widehat{\mathbf{x}}_{d,k}^T &= \sqrt{\rho}\,(1+\rho\,\widehat{\mathbf{h}}_k^H \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \widehat{\mathbf{h}}_k)^{-1}\,\widehat{\mathbf{h}}_k^H \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{Y}_d \\ \mathbf{R}_k &= \mathbf{C}_{\widetilde{\mathbf{x}}_{d,k}\widetilde{\mathbf{x}}_{d,k}} = \sigma^2_{\widetilde{x}_{d,k}} \mathbf{I}_{L-P} \end{aligned} \tag{16}$$

where $\sigma^2_{\widetilde{x}_{d,k}} = (1+\rho\widehat{\mathbf{h}}_k^H \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \widehat{\mathbf{h}}_k)^{-1}$. Note that $\mathbb{E}_{\mathbf{H}^q_{\overline{k}} \mathbf{X}^p_{d,\overline{k}} | \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}$ in (15) uses a Gaussian distribution for $\mathbf{H}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$ which is based on moment matching from $p(\mathbf{X}_{d,\overline{k}})$ and $q(\mathbf{H}_{\overline{k}})$ (as VL-EP requires). Note also that because of the Gaussian approximation of $\mathbf{H}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$ and the EM majorization step, the target pdf $\widehat{p}(\mathbf{h}_k)$ is Gaussian. This Gaussian blind information pdf needs to be combined with the Gaussian pilot part and the Gaussian prior to yield

$$\begin{aligned} -\ln q(\mathbf{h}_k) &= \\ &\|\mathbf{Y}_p\, \mathbf{x}_p^{(n)*} - \sqrt{\rho}\,P \textstyle\sum_{i\in\mathcal{G}_n\setminus\{k\}} \widehat{\mathbf{h}}_i - \sqrt{\rho}\,P\,\mathbf{h}_k\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}}} \\ &+ \|\mathbf{Y}_d - \sqrt{\rho}\,\mathbf{h}_k\,\widehat{\mathbf{x}}^T_{d,k}\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}} + c^t \\ &+ \rho\,(\|\widehat{\mathbf{x}}_{d,k}\|^2 + (L-P)\,\sigma^2_{\widetilde{x}_{d,k}})\,\mathbf{h}_k^H \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{h}_k + \mathbf{h}_k^H \mathbf{C}^{-1}_{o,k}\mathbf{h}_k \\ &= -2\,\Re\{\widehat{\mathbf{h}}_k^H \mathbf{C}_k^{-1}\mathbf{h}_k\} + \mathbf{h}_k^H \mathbf{C}_k^{-1}\mathbf{h}_k + c^t \end{aligned} \tag{17}$$

which is Gaussian with mean and covariance

$$\begin{aligned} \widehat{\mathbf{h}}_k &= \sqrt{\rho}\,\mathbf{C}_k\,[P\,\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}} \big(\mathbf{Y}_p\,\mathbf{x}_p^{(n)*} - \sqrt{\rho}\,P \textstyle\sum_{i\in\mathcal{G}_n\setminus\{k\}} \widehat{\mathbf{h}}_i\big) \\ &\qquad + \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{Y}_d\, \widehat{\mathbf{x}}^*_{d,k}] \\ \mathbf{C}_k &= \\ &\big(\rho\,[P^2\,\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}} + (\|\widehat{\mathbf{x}}_{d,k}\|^2 + (L-P)\,\sigma^2_{\widetilde{x}_{d,k}})\,\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}] + \mathbf{C}^{-1}_{o,k}\big)^{-1} \end{aligned} \tag{18}$$

where now $\widehat{\mathbf{h}}_k$ is the new estimate, and $\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}$, $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ are defined in (10), (13), and $\mathbf{C}_{o,k} = \operatorname{diag}(\beta_{1k}, \ldots, \beta_{Mk})$. This channel VL-EP is an iterative procedure that cycles through the $\mathbf{h}_k, k = 1, \ldots K$ and can be considered as an iterative version for the channel maximum a posteriori (MAP) estimation if one puts the $\mathbf{C}_i = \mathbf{0}$ in $\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}$, $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$. The channel MAP estimation instead of alternatingly optimizing the $\mathbf{h}_k$ tries to optimize w.r.t. all of $\mathbf{H}$ at once. The channel VL-EP can be initialized with $\widehat{\mathbf{h}}_k^{(-1)} = \mathbf{0}$ or with the channel MAP estimate for $\widehat{\mathbf{h}}_k$ with associated

$$\mathbf{C}_k = \big(\rho\,[P\,\mathbf{I}_M + (L-P)\,(\mathbf{I}_M + \rho\sum_{i\neq k}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^H)^{-1}] + \mathbf{C}^{-1}_{o,k}\big)^{-1}$$

where $\|\mathbf{x}_{p,k}\|^2 = P$ and $\|\widehat{\mathbf{x}}_{d,k}\|^2 + (L-P)\,\sigma^2_{\widetilde{x}_{d,k}}) \approx (L-P)\,\sigma^2_{x_{d,k}} = L-P$. Note that if parallel updating of the users is performed, one can reduce complexity in the computation of sums of the form $\sum_{i\neq k} A_i = \sum_{i=1}^K A_i - A_k$, so by computing a sum only once and then performing single term corrections.

V. Performance Evaluation

In this section, we provide numerical results verifying the analytical derivations and the performance of the proposed algorithm. The $M = 100$ APs and $K = 24$ users are uniformly distributed at random over a square area of size 1000×1000. The large-scale fading coefficient β_{mk} in (2) models the path loss and shadow fading as follows

$$\beta_{mk} = 10^{\frac{\mathrm{PL}_{mk}}{10}}\;10^{\frac{\sigma_{sh} z_{mk}}{10}} \tag{19}$$

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 $z_{mk} \sim \mathcal{N}(0,1)$, i.e., we assume uncorrelated shadow fading. The three-slope model in [32] is adopted for the path loss. The performance of the different channel estimators is assessed by the normalized mean square error (NMSE) versus SNR. The NMSE is defined as $\mathrm{NMSE} = \frac{avg\|\mathbf{h}-\widehat{\mathbf{h}}\|^2}{avg\|\mathbf{h}\|^2}$ where avg stands for average. Fig. 1 compares the performance of the proposed channel estimation VL-EP for GG-SB and channel MAP estimation and presents NMSE [dB] versus SNR [dB]. The proposed algorithm outperforms the channel MAP estimation and a joint channel and data MAP algorithm, termed Bayesian semi-blind approach in [16]. The joint channel and data MAP alternatingly estimates the channel or data as if the estimate for the other quantity is perfect, whereas the channel MAP estimation takes into account the data error covariance matrix $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$. Therefore the channel MAP estimation outperforms the Bayesian semi-blind iterative algorithm. The performance of these three different semi-blind channel estimation algorithms is compared to the different Cramer-Rao bounds (CRBs). For the semi-blind approaches one can consider the genie-aided scenario in which the data $\mathbf{X}_d$ would be detected exactly, hence becoming also pilots for the channel estimation, leading to the genie-aided Bayesian semi-blind (B-SB) CRB. For our VL-EP or channel MAP scenario, we consider Gaussian channels with the Gaussian input symbols eliminated, leading to the Gaussian inputs B-SB CRB. The deterministic CRB curve in the figure corresponds to a deterministic framework in which both data signal and channel coefficients are modeled as unknown deterministic quantities [16]. The performance of the different CRBs is evaluated by $\mathrm{NMSE} = \mathrm{tr}\{\mathrm{CRB}\}/\mathrm{tr}\{\mathbf{C}_{\mathbf{hh}}\}$, where $\mathbf{C}_{\mathbf{hh}} = \operatorname{diag}(\beta_{11}, \ldots, \beta_{M1} \ldots \beta_{1K}, \ldots, \beta_{MK})$. The simulations show

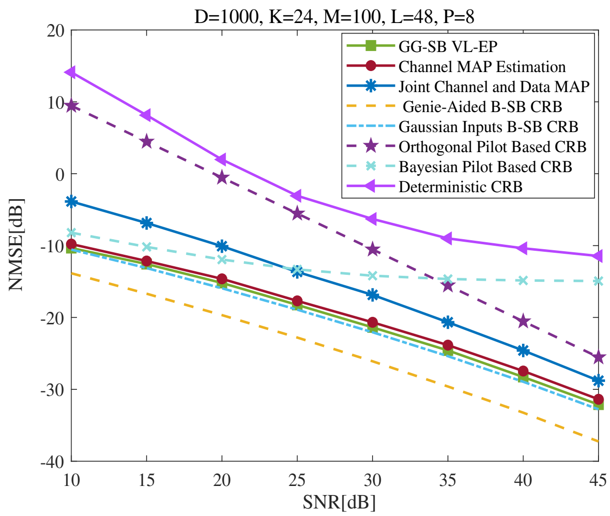

Fig. 1. NMSE [dB] versus SNR [dB]

that exploiting prior information gives significant performance gains. Compared to a fictitious scenario of just orthogonal pilot based channel estimation (pilots still of length P), deterministic semi-blind does not do as well whereas Bayesian semi-blind still does much better. On the other hand, the Bayesian pilot based CRB shows that just adding channel prior information to the contaminating pilots allows already to significantly improve MSE at low to moderate SNR, but floors at higher SNR. Adding the blind channel information from the data second-order statistics breaks this flooring, and both channel MAP and especially VL-EP allow to get performance close to the corresponding CRB, which behaves with just an SNR offset compared to the genie-aided CRB.

VI. Conclusion

In this paper, we considered semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data, exploiting prior channel information to mitigate the pilot contamination which originates from reusing pilot sequences, in CF massive MIMO systems. We proposed a VL-EP algorithm for semi-blind channel estimation which provides an approximate MMSE channel estimator. Numerical simulations corroborated the analytical derivations and the proposed algorithm.

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] 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.

[3] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5g," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1637–1660, 2020.

[4] 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.

[5] 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.

[6] T. L. Marzetta, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[7] R. Gholami, L. Cottatellucci, and D. Slock, "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.

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

[9] 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.

[10] ——, "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.

[11] A. Ashikhmin, H. Q. Ngo, T. L. Marzetta, and H. Yang, "Pilot assignment in cell free massive mimo wireless systems," Apr. 4 2017, uS Patent 9,615,384.

[12] 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.

[13] W. H. Hmida, V. Meghdadi, A. Bouallegue, and J.-P. Cances, "Graph coloring based pilot reuse among interfering users in cell-free massive mimo," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.

[14] H. Liu, J. Zhang, X. Zhang, A. Kurniawan, T. Juhana, and B. Ai, "Tabu-search-based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2286–2290, 2019.

[15] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive mimo based on the hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, 2020.

[16] R. Gholami, L. Cottatellucci, and D. Slock, "Tackling pilot contamination in cell-free massive mimo by joint channel estimation and linear multi-user detection," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 2828–2833.

[17] T. P. Minka, "Expectation propagation for approximate bayesian inference," *arXiv preprint arXiv:1301.2294*, 2013.

[18] M. Seeger, "Bayesian gaussian process models: Pac-bayesian generalisation error bounds and sparse approximations," University of Edinburgh, Tech. Rep., 2003.

[19] K. Ghavami and M. Naraghi-Pour, "Blind channel estimation and symbol detection for multi-cell massive mimo systems by expectation propagation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 943–954, 2017.

[20] J. Cespedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional mimo systems," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, 2014.

[21] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[22] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate inference techniques with expectation constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, p. P11015, 2005.

[23] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, "Expectation propagation detector for extra-large scale massive mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2036–2051, 2020.

[24] K. Ghavami and M. Naraghi-Pour, "Noncoherent massive mimo detection by expectation propagation," in *GLOBECOM 2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[25] X. Tan, Y.-L. Ueng, Z. Zhang, X. You, and C. Zhang, "A low-complexity massive mimo detection based on approximate expectation propagation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7260–7272, 2019.

[26] K.-H. Ngo, M. Guillaud, A. Decurninge, S. Yang, and P. Schniter, "Multi-user detection based on expectation propagation for the noncoherent simo multiple access channel," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6145–6161, 2020.

[27] H. He, H. Wang, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Distributed expectation propagation detection for cell-free massive mimo," *arXiv preprint arXiv:2108.07498*, 2021.

[28] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive mu-mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4418–4432, 2019.

[29] S. Barthelmé and N. Chopin, "Expectation propagation for likelihood-free inference," *Journal of the American Statistical Association*, vol. 109, no. 505, pp. 315–333, 2014.

[30] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[31] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, 2015.

[32] 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.