

Bayesian Receiver Design via Bilinear Inference for Cell-Free Massive MIMO With Low-Resolution ADCs

Takumi Takahashi¹, Member, IEEE, Hiroki Iimori², Member, IEEE,
Kengo Ando³, Graduate Student Member, IEEE, Koji Ishibashi⁴, Senior Member, IEEE,
Shinsuke Ibi⁵, Member, IEEE, and Giuseppe Thadeu Freitas de Abreu⁶, Senior Member, IEEE

Abstract—We propose a novel joint channel and data estimation (JCDE) scheme to combat the rate limitation in fronthaul links of cell-free massive MIMO (CF-mMIMO) systems introduced by the use of analog-to-digital converters (ADCs) at access points (APs), which makes channel estimation and multi-user detection at the central AP (CAP) challenging. The latter problem is solved here via the new JCDE scheme which differs from state-of-the-art (SotA) alternatives due to two contributions. The first is the design and incorporation of de-quantization (DQ) step which relies only on scalar Gaussian approximation (SGA) assumptions in conformity with mild central limit theorem (CLT), in contrast to the much harder asymptotic conditions required by the classic bilinear generalized approximate message passing (BiGAMP) algorithm. The second is a modification of bilinear Gaussian belief propagation (BiGaBP), whereby quantized outputs are linearized via the Bussgang decomposition enabling tractable signal processing. The resulting DQ-aided JCDE method achieves both low-complexity and high-accuracy by exploiting both the spatial degrees of freedom (DoF) obtained from, and the observations at the CAP to compensate for the low-resolution distortion introduced by, the distributed APs. The efficacy of the proposed method over the SotA is confirmed via computer simulations.

Index Terms—Cell-free massive MIMO, receiver design, limited feedback, Bayesian bilinear inference, low-resolution ADC.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (mMIMO) has been adopted as a key technology to address the

Manuscript received 11 July 2022; accepted 3 December 2022. Date of publication 19 December 2022; date of current version 12 July 2023. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP22H01483 and Grant JP20K14734. The associate editor coordinating the review of this article and approving it for publication was S. K. Mohammed. (*Corresponding author: Takumi Takahashi.*)

Takumi Takahashi is with the Graduate School of Engineering, Osaka University, Suita 565-0871, Japan (e-mail: takahashi@comm.eng.osaka-u.ac.jp).

Hiroki Iimori was with the School of Computer Science and Engineering, Jacobs University Bremen, 28759 Bremen, Germany. He is now with Ericsson Research, Yokohama Blue Avenue, Nishi-ku, Yokohama-shi, Kanagawa 220-0012, Japan (e-mail: h.iimori@ieee.org).

Kengo Ando and Giuseppe Thadeu Freitas de Abreu are with the School of Computer Science and Engineering, Jacobs University Bremen, 28759 Bremen, Germany (e-mail: K.Ando@jacobs-university.de; g.abreu@jacobs-university.de).

Koji Ishibashi is with the Advanced Wireless & Communication Research Center (AWCC), The University of Electro-Communications, Chofu-shi, Tokyo 182-8585, Japan (e-mail: koji@ieee.org).

Shinsuke Ibi is with the Faculty of Science and Engineering, Doshisha University, Kyotanabe 610-0394, Japan (e-mail: sibi@mail.doshisha.ac.jp).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2022.3228326>.

Digital Object Identifier 10.1109/TWC.2022.3228326

increasing demand in data rates at initial standardization stages of fifth-generation (5G) systems, which aim to achieve best-effort services focusing on faster downlink speeds. In beyond 5G and future sixth-generation (6G) networks, however, massive multiple-input multiple-output (mMIMO) is also expected to be exploited in the realization of *user-centric* communication systems, which seek to always satisfy wireless link quality both in the downlink (service) and the uplink (access) connectivities.

Cell-free massive MIMO (CF-mMIMO) has been recently proposed [1] as a promising new concept of mMIMO architectures that enables achieving this goal. The CF-mMIMO architecture virtually configures a mMIMO system using a large number of single- or multiple-antenna APs that are geographically distributed and connected through wired fronthaul links to a common high-performance central processing unit (CPU), while simultaneously serving multiple user equipment (UE). This renders the benefits of lower average path-loss, smaller spatial correlation, and larger degrees of freedom (DoFs) simultaneously possible. In addition, the centralized processing at the CPU of signals collected from local APs enables the alleviation of the cell boundary problem typical of cellular architectures by *effective* interference suppression, resulting in a more *user-centric* system, whereby the system can dynamically adjust to serve the needs of users, whether than users having to conditions of the system.

Early CF-mMIMO systems proposed in seminal works such as [1], [2], and [3] are, however, not scalable in practice due to the assumption that all APs are ideally connected to one CPU, responsible for coordinating and processing the signals of all UEs. In order to address this issue, a *user-centric* decentralized approach was proposed in [4], [5], and [6], which suggested the formation of local clusters [7], [8] consisting of a local CPU – also referred to as CAP – and associated APs, such that each UE is not served by all APs simultaneously, but by a subset of the latter that provides the best channel conditions. In particular, in [5] and [6], the dynamic cooperation clustering (DCC) method of [9] was applied to dynamically change the CAP-AP connections so as to achieve a scalable CF-mMIMO system, considering the time variability of the network structure due to the mobility of UEs.

Next to scalability, the energy consumption and deployment cost of the network are further challenges of the CF-mMIMO

architecture, as these parameters increase rapidly with the number of APs in the system. These challenges are mitigated by the use of low-cost hardware components, which in turn translates to the requirement of robustness against hardware impairments [10], [11], [12], [13]. In particular, rate limitations in wired fronthaul links and the constrained computational capabilities of APs are two factors now considered unavoidable in practical CF-mMIMO system [14], [15], which require *appropriate* data compression and demodulation techniques at the CAPs and APs to be circumvented.

In light of the above, we contribute in this article with an appropriate receiver design to address the aforementioned inherent problems of scalable CF-mMIMO system, including channel estimation (CE) and multi-user detection (MUD) in the uplink. We highlight that while preceding work such as [10], [11], [12], [13], [14], and [15] focused on metrics such as achievable data rate, spectral efficiency, and/or energy efficiency in their system performance evaluation, our contribution is fundamentally *algorithmic*, such that our numerical results are presented in terms of normalized mean square error (NMSE) of estimated quantities and bit error rate (BER), computed from actual constellation points. Besides the contribution of the newly proposed receiver, the performance evaluation here presented is therefore complementary to the aforementioned work in demonstrating the practical feasibility and the relevance of robust receiver designs for CF-mMIMO systems.

When considering the SotA literature on techniques to combat the rate-limited fronthaul link problem, one finds that a common approach is to perform data compression at the APs, which amounts to quantizing the signals using a small number of bits in order to reduce the load on the fronthaul links [10], [11], [12], [13], [14], [15]. A practical solution to that end is to employ low-resolution ADCs with 1–3 bits instead of the typical 10+ bits at APs, which has the further advantages of enabling low power consumption and reduced hardware cost. Other possible solutions are compute-and-forward schemes based on the cardinality reduction of symbols [16], [17], and singular value decomposition (SVD)-based latent semantic analysis (LSA) techniques [18], [19], which can efficiently reduce the fronthaul load, but require higher performance of the APs.

We highlight that among the many low-resolution ADC solutions to the rate-limited fronthaul problem are also clever estimate-compress-forward (ECF) methods such as those proposed *e.g.* in [14] and [15], in which the APs performs, besides data compression, other signal processing operations such as channel estimation and data filtering. Indeed, such an approach has been shown via numerical results to outperform the simpler “compress-only” counterparts, but their performance and implementation cost are in a trade-off relationship, which ultimately compromises the scalability of the system.

Keeping our aim sharply at the objective of reducing deployment costs in order to add feasible scalability to CF-mMIMO systems, we therefore focus our contribution on the previously mentioned paradigm, whereby APs only compress their received signals using low-resolution ADCs,

before forwarding the resulting quantized signals to the CAP. Under such a premise, it is the CAP that is required to employ sophisticated (and low-complexity) signal demodulation techniques in order to mitigate the severe distortion in the signals forwarded by the APs as a consequence of the nonlinear quantization noise induced by low-bit ADCs at fronthaul links. In summary, we assume that only down-conversion and quantization are performed at the APs, while baseband CE and MUD are performed at the CAP.

Having clarified our basic assumption, we also note that several methods have been proposed for recovering and estimating signals from quantized observations, including the optimal maximum likelihood (ML) estimation, as well as sub-optimal variation of the latter based on sphere decoding (SD) [20], whose exponential complexities are too high to be feasible for scaled CF-mMIMO systems. In comparison, the near-ML estimator based on convex optimization [21] and spatial filtering based on Bussgang’s theorem [22], [23], [23], [24], [25], can be leveraged as the basis of effective low-complexity estimation alternatives, which can achieve high estimation performance by exploiting the abundant DoF provided by CF-mMIMO systems.

The challenge in the latter approach is, however, the high communication overhead implied by the underlying requirement that channel state information (CSI) acquisition is performed via conventional coherent two-stage estimation, where CE is based on predefined pilot sequences, followed by MUD based on the estimated CSI. Indeed, very long orthogonal pilot sequences must be used in order to obtain accurate CSI knowledge from quantized observations, because the use of coarse quantization not only breaks the orthogonality of pilot sequences, but also greatly reduces the number of *effective* measurements available for CSI estimation [26], [27], [28], [29].

A solution to the latter challenge is JCDE via Bayesian bilinear inference, which exploits estimated data symbols as a soft pilot sequence, thus enabling a significant reduction in the length of pilot sequences without sacrificing estimation performance. Taking into account the inherent large-system setting of the CF-mMIMO architecture, whereby the number of APs is assumed to be larger than that of UEs, the BiGAMP algorithm proposed in [30] emerges as a particularly attractive framework for the design of JCDE schemes for CF-mMIMO systems, as it enables low-complexity bilinear inference by exploiting the law of large numbers.

Indeed, the BiGAMP methodology has been extended to bilinear inference based on quantized observations by means of a unified generalized approximate message passing (GAMP) framework [31], [32], [33], and it has been empirically shown [27], [28] that Bayesian JCDE schemes outperform the conventional two-step estimation and can asymptotically approach the Genie-aided ideal performance provided a massive number of observation dimensions and sufficient time resources can be assumed. The working premise of BiGAMP is, however, heavily dependent on the large-system limit approximation based on zero-mean independent and identically distributed (i.i.d.) random observations [34], [35], which renders the approach vulnerable to the lack of DoF caused

by spatial correlation among fading coefficients for instance. In addition, the convergence behavior of iterative estimation based on BiGAMP is not stable, which may result in serious performance degradation, especially in the presence of non-orthogonal pilot sequences [36].

A recently-proposed approach to tackle the latter issue is the BiGaBP algorithm [37], which aims to generalize BiGAMP by relaxing its dependence on the large-system limit approximation and leveraging the belief propagation (BP) concept [38], which ultimately allows JCDE to be performed robustly even with non-orthogonal pilot sequences. Subsequently to [37], the BiGaBP inference framework was recently employed [39] in the receiver design of grant-free (GF) access schemes, where the approach was shown to outperform earlier linear Bayesian counterparts. In light of all the above, higher-complexity alternatives based on vector AMP (VAMP) [40], [41], [42], [43] and the expectation propagation (EP) algorithm [44] will not be considered in this article, as they do not offer the same scalability potential of the technique proposed here.

The contributions of the article are summarized as follows:

- **New Scalable CF-mMIMO Receiver:** A novel JCDE algorithm, dubbed DQ-aided BiGaBP, is presented, wherein message passing DQ (MPDQ) is performed via bilinear inference based on a spatial-temporal representation of the received signal given through a linearization using the Bussgang decomposition. A major challenge overcome by the proposed receiver design is extending BiGaBP to multi-dimensional observations with different quantization resolutions under consistent statistical assumptions and approximation accuracy.¹ To the best of our knowledge, such an extension of the BiGaBP algorithm to incorporate arbitrary distributions on the output has not yet been reported in the literature. The proposed extension is accomplished by integrating a novel DQ step here-derived, and a new purpose-designed JCDE step here obtained via a modification of BiGaBP which incorporates a linearization of quantized outputs by means of the Bussgang decomposition. We further emphasize that in addition to these contributions, both of which are novelties compared to the scheme proposed in [37], our proposed design is suitable for low-resolution CF-mMIMO system, as it is easily applicable to complex inference problems in which signals with different resolutions are observed at different intensities.
- **New Insight on DoF-ADCs Resolution Trade-off:** Our receiver design takes maximum advantage of spatial DoF via a Bayesian framework that enables performance improvement in terms of both CE and MUD when processing signals quantized by low-resolution ADCs at APs and forwarded to a CAP. Thanks to the feature, the simulation results here presented reveal an original trade-off relationship between increased DoF due to spatially distributed APs and reduced signal resolution due to low-resolution ADCs, which contributes new insight on

the design of scalable CF-mMIMO system, by indicating that the ratio of the number of antennas concentrated in the CAP to the number of antennas distributed at the APs should be designed depending on ADC quality, in order for system performance to be optimized. Counter-intuitively, it is also shown that the performance of spatial filtering may actually be severely degraded by equipping APs with multiple antennas, although the proposed DQ-aided BiGaBP method proves somewhat robust to the up-scaling of AP antennas, thanks to the JCDE mechanism.

In addition to the aforementioned contributions, we highlight that, to the best of our knowledge, there is no previous work in the literature offering a systematic bilinear Bayesian receiver design for scalable CF-mMIMO systems for highly correlated channels and considering realistic constraints such as limited-capacity fronthaul links to the CAP and quantized signal distortion induced by the use of low-resolution ADCs.

Notation: Throughout the article, the following notation is used unless otherwise specified. Vectors and matrices are denoted by lower- and upper-case bold-face letters, respectively. The conjugate, transpose, and conjugate transpose operators are denoted by $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$, respectively. The probability density function (PDF) of the outcome a of a random variable a given the occurrence b of a random variable b is denoted by $p_{a|b}(a|b)$. The expectation of a given random variable is denoted by $\mathbb{E}[\cdot]$, while $\mathbb{E}_{a|b}[\cdot|b]$ denotes the conditional expectation of the variable a , given the outcome b of the variable b . The real and imaginary parts of a complex quantity are respectively denoted by $\Re\{\cdot\}$ and $\Im\{\cdot\}$. The imaginary unit is denoted by $j \triangleq \sqrt{-1}$. The $a \times a$ square identity matrix is denoted by \mathbf{I}_a , while the (i, j) -th element of a matrix \mathbf{A} is denoted by a_{ij} . The diagonal matrix constructed by placing the elements of the vector \mathbf{a} on its main diagonal is denoted by $\text{diag}[\mathbf{a}]$, while $\text{diag}[\mathbf{A}]$ denotes a diagonal matrix obtained by replacing the off-diagonal elements of \mathbf{A} with 0's. The block diagonal matrix created by aligning the matrices $\mathbf{A}_1, \dots, \mathbf{A}_D$ along the main diagonal is denoted by $\text{blkdiag}[\mathbf{A}_1 \cdots \mathbf{A}_D]$. The real and complex Gaussian distributions with mean a and variance b are respectively denoted by $\mathcal{N}(a, b)$ and $\mathcal{CN}(a, b)$, while the PDF and the cumulative distribution function (CDF) of the standard normal distribution are denoted by $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp[-x^2/2]$ and $\Phi(x) = \int_{-\infty}^x \phi(v)dv$, respectively. The notation $a \sim p_a(a)$ indicates that a variable a obeys $p_a(a)$. Finally, the symbol \propto is used to denote proportionality.

II. SYSTEM MODEL

Consider an uplink CE and MUD problem in a scalable CF-mMIMO system that adopts *user-centric* DCC [5], where the CPU uses the knowledge of the large-scale fading coefficients of all UEs to cluster a group of APs in order to achieve scalability and mitigate inter-cluster interference. As a consequence of such an architecture, each local cluster consisting of a CAP, its surrounding APs, and the corresponding UEs, as illustrated in Fig. 1, can be considered individually for the purpose of system design, without loss of generality, as it is free of interference from other clusters.

¹To this end, we take advantage of the fact that the BiGaBP framework relies on the SGA in conformity with CLT, whose underlying assumptions are much softer than the large-system limit assumption on which the BiGAMP algorithm heavily relies.

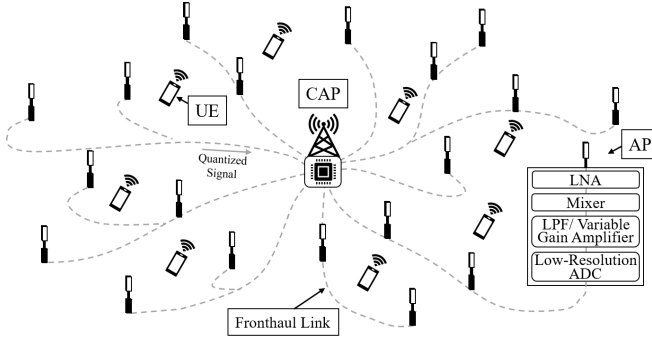


Fig. 1. Illustration of a cluster of a scalable CF-mMIMO system, where a CAP and its surrounding APs serve multiple UEs within a given service area. Each of the APs possesses only the minimum functionality required to perform down-conversion to baseband signals followed by quantization with low-resolution ADCs, before forwarding their signals to the CAP.

In what follows, it shall be assumed that the cluster contains U spatially distributed single- or multiple-antenna APs connected by wired fronthaul links to a common CAP, serving M synchronized single-antenna UEs. We shall also assume that the signal received at the APs is individually quantized using low-resolution ADCs and subsequently forwarded to the CAP through fronthaul links. In contrast, the signal received at the CAP is completely forwarded to the CPU free of distortion.

For future convenience, let $\mathcal{U} \triangleq \{0, 1, \dots, U\}$ be a set of receiver indices, reserving the index $u = 0$ to the CAP, such that the other indices $u \in \mathcal{U}$ with $u \geq 1$ correspond to the remaining APs. Each u -th AP is assumed to have N_u receive antennas, while the M UEs served by the system, with $m \in \mathcal{M} \triangleq \{1, \dots, M\}$, are assumed to be single-antenna devices, distributed randomly following a Poisson point process (PPP) with intensity μ_{UE} within the coverage area of the cluster.

Following [5], [19], the mean power of channels between every UE and a receiver (*i.e.*, CAP or AP) will be determined according to the urban micro cell path-loss model of [45], namely

$$\delta(d_m(u)) = 22.7 + 36.7 \log_{10}(d_m(u)) + 26 \log_{10}(f_c), \quad (1)$$

where f_c [GHz] denotes the carrier frequency, and $d_m(u)$ [m] is the distance between the u -th receiver and the m -th UE, with $u \in \mathcal{U}$ and $m \in \mathcal{M}$.

Using equation (1), and considering that spatial correlation exists among the N_u receive (RX) antennas of the u -th AP, the associated channel vector can be expressed as [46]

$$\mathbf{h}_m(u) \triangleq \sqrt{\rho} \sqrt{\gamma_m(u)} \boldsymbol{\Theta}_m(u)^{1/2} \mathbf{h}, \quad (2)$$

where ρ is the transmit (TX) power assumed identical to all UEs, $\gamma_m(u) = 10^{-\delta(d_m(u))/10}$ is the path-loss of the channel between the m -th UE and the u -th AP, $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{I}_{N_u})$ is a complex standard normal vector corresponding to N_u uncorrelated small-scale channels, and the RX fading correlation matrix $\boldsymbol{\Theta}_m(u)$ is expressed as

$$\begin{aligned} \boldsymbol{\Theta}_m(u) &\triangleq \mathbf{a}(\omega_m(u)) \mathbf{a}(\omega_m(u))^H \\ &= \boldsymbol{\Theta}_m(u)^{1/2} \boldsymbol{\Theta}_m(u)^{1/2} \in \mathbb{C}^{N_u \times N_u}, \end{aligned} \quad (3)$$

with $\mathbf{a}(\omega_m(u)) \triangleq [1, e^{j\pi \cos(\omega_m(u))}, \dots, e^{j\pi \cos(\omega_m(u))(N_u-1)}]$ denoting the antenna steering vector at the u -th AP, pointing to the angle of arrival $\omega_m(u)$ corresponding to the m -th UE.

At the k -th discrete time instance, with $k \in \mathcal{K} \triangleq \{1, \dots, K\}$, the complex baseband signal $\mathbf{y}_k(u) \in \mathbb{C}^{N_u \times 1}$ at the u -th receiver can be written as the following linear regression model [5]

$$\begin{aligned} \mathbf{y}_k(u) &\triangleq [y_{1k}(u), \dots, y_{n_{uk}}(u), \dots, y_{N_{uk}}(u)]^T \\ &= \mathbf{H}(u) \mathbf{x}_k + \mathbf{z}_k(u), \end{aligned} \quad (4)$$

where $\mathbf{H}(u) \triangleq [\mathbf{h}_1(u), \dots, \mathbf{h}_M(u)] \in \mathbb{C}^{N_u \times M}$ is a matrix concatenating the channel vectors of all M single-antenna UEs, $\mathbf{x}_k \triangleq [x_{1k}, \dots, x_{Mk}]^T \in \mathbb{C}^{M \times 1}$ is a TX symbol vector stacking the scalar symbols from all UEs, and $\mathbf{z}(u) \triangleq [z_1(u), \dots, z_{N_u}(u)]^T \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_u})$ is an additive white Gaussian noise (AWGN) vector at the u -th receiver.

Assuming that the channel matrix $\mathbf{H}(u)$ is constant during K successive transmissions, and concatenating K successive RX vectors as given by equation (4) yields the following compact spatial-temporal RX signal representation

$$\mathbf{Y}(u) = \mathbf{H}(u) \mathbf{X} + \mathbf{Z}(u), \quad (5)$$

where $\mathbf{X} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_K] \in \mathbb{C}^{M \times K}$, $\mathbf{Y}(u) \triangleq [\mathbf{y}_1(u), \dots, \mathbf{y}_k(u), \dots, \mathbf{y}_K(u)] \in \mathbb{C}^{N_u \times K}$, and $\mathbf{Z}(u) \triangleq [\mathbf{z}_1(u), \dots, \mathbf{z}_k(u), \dots, \mathbf{z}_K(u)] \in \mathbb{C}^{N_u \times K}$.

We highlight that each row of the TX symbol matrix \mathbf{X} corresponds to a given UE, carrying K_p pilot symbols with indices $k \in \mathcal{K}_p \triangleq \{1, \dots, K_p\}$ followed by K_d data symbols with indices $k \in \mathcal{K}_d \triangleq \{K_p + 1, \dots, K\}$, with $K = K_p + K_d$ and where we assume $K_p \ll K_d < K$. Consequently, the spatial-temporal matrices can be sub-divided into pilot and data parts as follows: $\mathbf{X} \triangleq [\mathbf{X}_p, \mathbf{X}_d]$, with $\mathbf{X}_p \in \mathbb{C}^{M \times K_p}$ and $\mathbf{X}_d \in \mathbb{C}^{M \times K_d}$; $\mathbf{Y}(u) \triangleq [\mathbf{Y}_p(u), \mathbf{Y}_d(u)]$, with $\mathbf{Y}_p(u) \in \mathbb{C}^{N_u \times K_p}$ and $\mathbf{Y}_d(u) \in \mathbb{C}^{N_u \times K_d}$; and $\mathbf{Z}(u) \triangleq [\mathbf{Z}_p(u), \mathbf{Z}_d(u)]$ with $\mathbf{Z}_p(u) \in \mathbb{C}^{N_u \times K_p}$ and $\mathbf{Z}_d(u) \in \mathbb{C}^{N_u \times K_d}$.

A. Quantization With Low-Resolution ADCs at APs

Let the in-phase and quadrature components of the signal received at each antenna of each AP be quantized separately using an ADC with b -bit resolution according to the sets of $2^b + 1$ quantization thresholds $\mathcal{I}_b(u) \triangleq \{\tau_0(u), \tau_1(u), \dots, \tau_{2^b}(u)\}$, with $-\infty = \tau_0(u) < \tau_1(u) < \dots < \tau_{2^b}(u) = \infty$, coupled with the sets of 2^b quantization labels $\mathcal{L}_b(u) \triangleq \{l_0(u), l_1(u), \dots, l_{2^b-1}(u)\}$ where $l_i(u) \in (\tau_i(u), \tau_{i+1}(u)]$. The joint operation of the $2N_u$ b -bit ADCs at the u -th receiver can be represented by the quantization function $Q_{u,b}^{N_u}(\cdot) : \mathbb{R}^{N_u} \rightarrow \mathcal{L}_b^{N_u}(u)$ that maps the RX vector $\mathbf{y}_k(u) \in \mathbb{C}^{N_u \times 1}$ onto the corresponding quantized output vector $\mathbf{r}_k(u)$, defined by

$$\begin{aligned} \mathbf{r}_k(u) &\triangleq Q_{u,b}^{N_u}(\Re\{\mathbf{y}_k(u)\}) + jQ_{u,b}^{N_u}(\Im\{\mathbf{y}_k(u)\}) \\ &= [r_{1k}(u), \dots, r_{n_{uk}}(u), \dots, r_{N_{uk}}(u)]^T, \end{aligned} \quad (6)$$

where the n_u -th output is obtained from the corresponding quantization label as [25]

$$\begin{aligned} \Re\{r_{n_{uk}}(u)\} &= Q_{u,b}^1(\Re\{y_{n_{uk}}(u)\}) = l_i(u) \\ &\quad \text{if } \Re\{y_{n_{uk}}(u)\} \in (\tau_i(u), \tau_{i+1}(u)], \end{aligned} \quad (7a)$$

$$\begin{aligned} \Im\{r_{n_{uk}}(u)\} &= Q_{u,b}^1(\Im\{y_{n_{uk}}(u)\}) = l_j(u) \\ &\quad \text{if } \Im\{y_{n_{uk}}(u)\} \in (\tau_j(u), \tau_{j+1}(u)]. \end{aligned} \quad (7b)$$

It will be assumed hereafter that the optimal quantization labels are determined via the Lloyd-Max algorithm [47]. To that end, let the RX symbols from all users at the u -th AP be modeled as a complex Gaussian random variable with mean 0 and variance $\psi^y(u)$ approximated by

$$\psi^y(u) \triangleq \mathbb{E} [|y_{n_u k}(u)|^2] = \rho \sum_{m=1}^M \gamma_m(u) + \sigma^2, \quad (8)$$

where σ^2 is the noise variance and $\gamma_m(u)$ are path-losses already defined after equation (2).

Then, the labels employed by the u -th AP are given by [25]

$$\mathcal{L}_b(u) = \alpha(u) \tilde{\mathcal{L}}_b(u), \quad (9)$$

where $\tilde{\mathcal{L}}_b(u) = \{\tilde{l}_0(u), \tilde{l}_1(u), \dots, \tilde{l}_{2^b-1}(u)\}$ is a set of labels that minimize the mean square error (MSE) between $\mathbf{y}_k(u)$ and $\mathbf{r}_k(u)$, while and the coefficients $\alpha(u)$ are computed so as to ensure that $\mathbb{E} [|r_{n_u k}(u)|^2] = \psi^y(u)$, namely

$$\begin{aligned} \alpha(u) &= \sqrt{\frac{\psi^y(u)}{2 \sum_{i=0}^{2^b-1} \tilde{l}_i(u)^2 \left(\Phi \left(\sqrt{\frac{2\tau_{i+1}(u)^2}{\psi^y(u)}} \right) - \Phi \left(\sqrt{\frac{2\tau_i(u)^2}{\psi^y(u)}} \right) \right)}}, \end{aligned} \quad (10)$$

with $\Phi(x)$ denoting the standard normal CDF, as defined at the end of Section I.

B. Linearization Based on Bussgang's Theorem

The correlated distortion introduced by the finite-resolution ADC process makes subsequent stochastic signal processing of quantized signals very challenging. Under the assumption that the quantizer input follows a Gaussian distribution, however, the Bussgang's theorem [22] can be leveraged to circumvent this issue. To that end, let the quantized output $\mathbf{r}_k(u)$ be modeled [22] by the following linear relationship with the received signal $\mathbf{y}_k(u)$,

$$\mathbf{r}_k(u) = \mathbf{G}(u) \mathbf{y}_k(u) + \mathbf{d}_k(u), \quad \forall u \in \mathcal{U} \setminus \{0\}, \quad (11)$$

where $\mathbf{G}(u) \in \mathbb{C}^{N_u \times N_u}$ is the Bussgang's gain, and $\mathbf{d}_k(u) \in \mathbb{C}^{N_u \times 1}$ is a distortion that is uncorrelated with $\mathbf{y}_k(u)$, such that $\mathbb{E} [\mathbf{y}_k(u) \mathbf{d}_k(u)^H] = \mathbb{E} [\mathbf{y}_k(u)] \mathbb{E} [\mathbf{d}_k(u)^H]$.

The uncorrelated nature of $\mathbf{d}_k(u)$ with respect to $\mathbf{y}_k(u)$ enables to address the quantization distortion of the ADC output, and the input signal itself, separately. Consequently, the generation of spatial filters and the statistical processing described in Section III can be significantly simplified. For further details, we refer the reader to [23] and [25].

Regarding the RX vector $\mathbf{y}_k(u) \sim \mathcal{CN}(\mathbf{0}_{N_u \times 1}, \mathbf{C}_{yy}(u))$, it follows from equation (11) that

$$\mathbf{C}_{ry}(u) \triangleq \mathbb{E} [\mathbf{r}_k(u) \mathbf{y}_k(u)^H] = \mathbf{G}(u) \mathbf{C}_{yy}(u), \quad (12)$$

with

$$\mathbf{G}(u) \triangleq \text{diag} [g_1(u), \dots, g_{N_u}(u)], \quad (13a)$$

$$\mathbf{C}_{yy}(u) \triangleq \mathbb{E} [\mathbf{y}_k(u) \mathbf{y}_k(u)^H], \quad (13b)$$

where the n_u -th diagonal element of $\mathbf{G}(u)$ is given by $\text{diag} [\mathbf{C}_{yy}(u)] = \psi^y(u) \mathbf{I}_{N_u}$, yielding [25]

$$\begin{aligned} g_{n_u}(u) &= \frac{\mathbb{E} [Q_{u,b}^1(y_{n_u k}(u)) y_{n_u k}(u)^*]}{\mathbb{E} [|y_{n_u k}(u)|^2]} \\ &= \sum_{i=0}^{2^b-1} \frac{l_i(u)}{\sqrt{\pi \psi^y(u)}} \left(e^{-\frac{\tau_i(u)^2}{\psi^y(u)}} - e^{-\frac{\tau_{i+1}(u)^2}{\psi^y(u)}} \right) \\ &\triangleq g(u), \quad \forall n_u, \end{aligned} \quad (14)$$

such that we have $\mathbf{G}(u) = g(u) \mathbf{I}_{N_u}$.

Notice that in the ideal case of ADCs with infinite resolution ($b = \infty$), equation (14) yields $g(u) = 1$, and consequently $\mathbf{G}(u) = \mathbf{I}_{N_u}$, which in turn if substituted into equation (11) results in $\mathbf{r}_k(u) = \mathbf{y}_k(u)$, indicating the consistency of the Bussgang's decomposition.

C. Effective Signal Model at a CAP

Thanks to the Bussgang's linearization described by equation (11) and detailed above, the uncompressed signals received directly by the CAP and the low-resolution compressed signals gathered from the surrounding APs can be combined into a single effective RX vector, which can be compactly expressed as (15), shown at the bottom of the next page, where $N = \sum_{u=0}^U N_u$, and we have implicitly defined the vectors \mathbf{r}_k , \mathbf{y}_k , \mathbf{d}_k and \mathbf{z}_k , and matrices \mathbf{G} and \mathbf{H} .

In view of equation (15), it will prove convenient to define a set of indices related directly to the entries of the vector \mathbf{r}_k , namely, $\mathcal{N} \triangleq \{1, \dots, N\}$, with $\mathcal{N} = \bigcup_{u=0}^U \mathcal{N}_u$, $\mathcal{N}_{\text{AP}} \triangleq \mathcal{N} \setminus \mathcal{N}_0$,

$$\mathcal{N}_u \triangleq \left\{ n \in \mathcal{N} \mid \sum_{i=0}^{u-1} N_i + 1 \leq n \leq \sum_{i=0}^u N_i \right\}, \quad (16)$$

where $|\mathcal{N}_u| = N_u$.

Finally, a complete spatial-temporal effective receive signal model concatenating K consecutive instances of \mathbf{r}_k can be built by extending equation (15) in similarity to equation (5), yielding

$$\begin{aligned} \mathbf{R} &\triangleq [\mathbf{r}_1 \quad \dots \quad \mathbf{r}_K] \\ &= \mathbf{G} \mathbf{H} \mathbf{X} + \mathbf{G} \underbrace{[\mathbf{z}_1 \quad \dots \quad \mathbf{z}_K]}_{\triangleq \mathbf{Z} \in \mathbb{C}^{N \times K}} + \underbrace{[\mathbf{d}_1 \quad \dots \quad \mathbf{d}_K]}_{\triangleq \mathbf{D} \in \mathbb{C}^{N \times K}} \\ &= \mathbf{G} \mathbf{H} \mathbf{X} + \mathbf{G} \mathbf{Z} + \mathbf{D}. \end{aligned} \quad (17)$$

The goal of the article can now be stated concisely: design a novel JCDE algorithm that enables the CAP to simultaneously detect the intended symbol matrix \mathbf{X}_d and accurately estimate the effective channel matrix \mathbf{H} , out of the spatial-temporal effective receive signal matrix \mathbf{R} given in equation (17), which thanks to the Bussgang linearization approach utilized contains both its own undistorted RX signal matrix $\mathbf{Y}_d(0)$ and the signals obtained from the APs, that are subjected to distortion due to the *coarse* quantization performed by low-resolution ADCs.

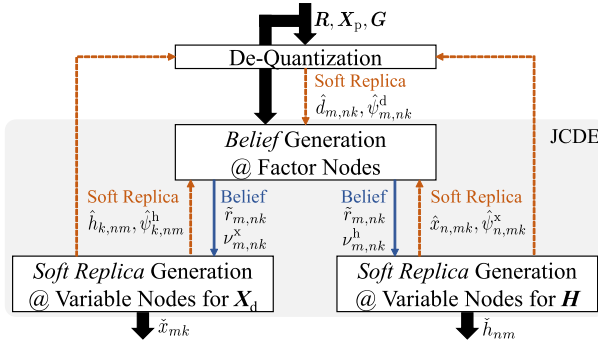


Fig. 2. Schematic of the belief propagation process employed in the proposed DQ-aided BiGaBP algorithm.

III. JOINT CHANNEL AND DATA ESTIMATION VIA DQ-AIDED BIGABP

In this section, a MPDQ process based on a DQ-aided BiGaBP mechanism to design Bayesian JCDE receivers for scalable CF-mMIMO architectures is described. As illustrated in Fig. 2, the estimation procedure of the proposed method is divided into two steps: the DQ process and the JCDE procedure. By iterating between these steps alternately, the estimation accuracy of the unknown data matrix \mathbf{X}_d and effective channel matrix \mathbf{H} can be gradually improved, given the knowledge of the effective signal matrix \mathbf{R} and pilot symbols \mathbf{X}_p .

To elaborate, in the DQ step, the RX signal $\mathbf{Y} \triangleq [\mathbf{Y}(0)^T, \dots, \mathbf{Y}(U)^T]^T$ is first recovered and, subsequently, estimates of the signal distortion \mathbf{D} are obtained using the knowledge of \mathbf{R} , \mathbf{X}_p , and of the tentative estimates of \mathbf{X}_d and \mathbf{H} obtained in the JCDE step. In turn, in the JCDE step, estimates \mathbf{X}_d and \mathbf{H} are computed employing the BiGaBP approach, using the knowledge of \mathbf{R} , \mathbf{X}_p , and of the tentative estimates of \mathbf{D} obtained in the DQ step. We highlight that the bilinear inference in the JCDE step is performed by exchanging *beliefs* (i.e., likelihood information reflecting estimation reliability) and *soft replicas* (i.e., tentative estimates) on the tripartite factor graph (FG) consisting of factor nodes (FNs) and two variable nodes (VNs), which correspond to the channel coefficients and TX symbols, respectively.

In what follows, we provide detailed descriptions of the DQ and JCDE steps of the proposed technique briefly explained above, leaving tedious derivations of certain equations to Appendices A and B for the sake of readability. To that end,

and for future convenience, let us define the soft replicas of x_{mk} and h_{nm} as $\hat{x}_{n,mk}, \forall n$, and $\hat{h}_{k,nm}, \forall k$, respectively, such that their MSEs can be respectively expressed as

$$\hat{\psi}_{n,mk}^x \triangleq \mathbb{E} \left[|\tilde{x}_{n,mk}|^2 \right] \quad \text{and} \quad \hat{\psi}_{k,nm}^h \triangleq \mathbb{E} \left[|\tilde{h}_{k,nm}|^2 \right], \quad (18)$$

where the quantities $\tilde{x}_{n,mk}, \forall n$ and $\tilde{h}_{k,nm}, \forall k$, denote the estimation errors corresponding to $\hat{x}_{n,mk} \triangleq x_{mk} - \hat{x}_{n,mk}, \forall n$, and $\tilde{h}_{k,nm} \triangleq h_{nm} - \hat{h}_{k,nm}, \forall k$, respectively.

For simplicity, in what follows we assume that quadrature phase-shift keying (QPSK) modulation with unit power is employed, where the constellation set is expressed as

$$\mathcal{X} = \left\{ \pm \frac{1}{\sqrt{2}} \pm j \frac{1}{\sqrt{2}} \right\}.$$

A. DQ Step

Let us start with the DQ step, which processes the RX symbols with indices $\forall n \in \mathcal{N}_{AP}$, forwarded from the u -th AP via the fronthaul link at the k -th discrete time, outputting therefore $r_{nk}, \forall n \in \mathcal{N}_{AP}$. In the sequel, we refrain from explicitly including the qualifier “ $\forall n \in \mathcal{N}_{AP}$ ” after each equation, in order to avoid being repetitive. In addition, in this subsection the real and imaginary parts of a complex number c will be abbreviated as $c^{(0)} = \Re \{c\}$ and $c^{(1)} = \Im \{c\}$, respectively, i.e., $c = c^{(0)} + j c^{(1)}$.

With that clarified, since soft replicas are not available at the first iteration ($t = 1$), the initial conditional expectation of y_{nk} , given r_{nk} , is calculated as

$$\begin{aligned} \hat{y}_{m,nk} &= \int y_{nk} p_{y_{nk}|r_{nk}}(y_{nk}|r_{nk}) dy_{nk} \\ &= \int y_{nk}^{(0)} p_{y_{nk}^{(0)}|r_{nk}^{(0)}}(y_{nk}^{(0)}|r_{nk}^{(0)}) dy_{nk}^{(0)} \\ &\quad + j \int y_{nk}^{(1)} p_{y_{nk}^{(1)}|r_{nk}^{(1)}}(y_{nk}^{(1)}|r_{nk}^{(1)}) dy_{nk}^{(1)} \\ &= \sum_{\theta \in \{0,1\}} j^\theta \int y_{nk}^{(\theta)} p_{y_{nk}^{(\theta)}|r_{nk}^{(\theta)}}(y_{nk}^{(\theta)}|r_{nk}^{(\theta)}) dy_{nk}^{(\theta)}, \quad \forall m, \end{aligned} \quad (19)$$

where $\theta \in \{0,1\}$ indicates a binary index that distinguishes between real and imaginary parts, and we highlight that $y_{nk}^{(0)}$ and $y_{nk}^{(1)}$ are independent of each other.

Owing to the CLT, valid for the large-system regime ($M \gg 1$), and making use of the variance approximation given in equation (8), it follows that $y_{nk}^{(\theta)}$ obeys $\mathcal{N}(0, \frac{\psi^{y(u)}}{2})$, such

$$\begin{aligned} \underbrace{\begin{bmatrix} \mathbf{y}_k(0) \\ \mathbf{r}_k(1) \\ \vdots \\ \mathbf{r}_k(U) \end{bmatrix}}_{\triangleq \mathbf{r}_k \in \mathbb{C}^{N \times 1}} &= \underbrace{\text{blkdiag} [\mathbf{I}_{N_0} \ g(1)\mathbf{I}_{N_1} \ \cdots \ g(U)\mathbf{I}_{N_U}]}_{\triangleq \mathbf{G} \in \mathbb{C}^{N \times N}} \cdot \underbrace{\begin{bmatrix} \mathbf{y}_k(0) \\ \mathbf{y}_k(1) \\ \vdots \\ \mathbf{y}_k(U) \end{bmatrix}}_{\triangleq \mathbf{y}_k \in \mathbb{C}^{N \times 1}} + \underbrace{\begin{bmatrix} \mathbf{0}_{N_0 \times 1} \\ \mathbf{d}_k(1) \\ \vdots \\ \mathbf{d}_k(U) \end{bmatrix}}_{\triangleq \mathbf{d}_k \in \mathbb{C}^{N \times 1}} = \mathbf{G} \cdot \left(\underbrace{\begin{bmatrix} \mathbf{H}(0) \\ \mathbf{H}(1) \\ \vdots \\ \mathbf{H}(U) \end{bmatrix}}_{\triangleq \mathbf{H} \in \mathbb{C}^{N \times M}} \mathbf{x}_k + \underbrace{\begin{bmatrix} \mathbf{z}_k(0) \\ \mathbf{z}_k(1) \\ \vdots \\ \mathbf{z}_k(U) \end{bmatrix}}_{\triangleq \mathbf{z}_k \in \mathbb{C}^{N \times 1}} \right) + \mathbf{d}_k \\ &= \mathbf{G} (\mathbf{H} \mathbf{x}_k + \mathbf{z}_k) + \mathbf{d}_k, \end{aligned} \quad (15)$$

that the conditional PDF of $y_{nk}^{(\theta)}$, given $r_{nk}^{(\theta)}$, appearing in the integral on the last line of equation (19) can be expressed via a truncated Gaussian distribution as

$$p_{y_{nk}^{(\theta)}|r_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}|l_i(u)\right) = \frac{p_{y_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}\right)}{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}\right) dy_{nk}^{(\theta)}}, \quad (20)$$

where the denominator is introduced to normalize the integral of the conditional PDF to 1.

The relationship between $p_{y_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}\right) = \mathcal{N}\left(0, \frac{\psi^y(u)}{2}\right)$ and $p_{y_{nk}^{(\theta)}|r_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}|l_i(u)\right)$ is illustrated in Fig. 3a, where we highlight that the prior information of $r_{nk}^{(\theta)} = l_i(u)$ is confined to the interval $(\tau_i(u), \tau_{i+1}(u)]$, of the domain of $y_{nk}^{(\theta)}$ such that the gray area under the normal distribution corresponds to the probability mass of $r_{nk}^{(\theta)}$, which after normalization yields the truncated Gaussian model given in equation (20) and shown with a dotted line in Fig. 3a.

Utilizing the mean and variance obtained from equation (20) for $\theta = \{0, 1\}$, the soft replica of y_{nk} and its corresponding MSE can be obtained as

$$\hat{y}_{m,nk} = \sum_{\theta \in \{0,1\}} j^\theta \eta\left(y_{nk}^{(\theta)}; 0, \sqrt{\frac{\psi^y(u)}{2}}, \tau_i(u), \tau_{i+1}(u)\right), \quad \forall m, \quad (21a)$$

$$\hat{\psi}_{m,nk}^y = \sum_{\theta \in \{0,1\}} \zeta\left(y_{nk}^{(\theta)}; 0, \sqrt{\frac{\psi^y(u)}{2}}, \tau_i(u), \tau_{i+1}(u)\right), \quad \forall m, \quad (21b)$$

where $\eta(s)$ and $\zeta(s)$ are functions yielding, respectively, the mean and variance of a random variable s obeying the truncated Gaussian distribution, which are respectively defined as [48]

$$\eta(s; \mu_s, \sigma_s, a, b) \triangleq \mu_s + \sigma_s \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}, \quad (22a)$$

$$\zeta(s; \mu_s, \sigma_s, a, b) \triangleq \sigma_s^2 \left[1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right], \quad (22b)$$

with $\alpha \triangleq \frac{a-\mu_s}{\sigma_s}$ and $\beta \triangleq \frac{b-\mu_s}{\sigma_s}$.

Owing to the availability of the soft replicas as prior information at the second and subsequent iterations ($t \neq 1$), the conditional PDF of y_{nk} can be expressed as

$$p_{y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}, r_{nk}^{(\theta)}}\left(y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}, l_i(u)\right) \propto \frac{p_{y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}\left(y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}\right)}{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}\left(y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}\right) dy_{nk}^{(\theta)}}, \quad (23)$$

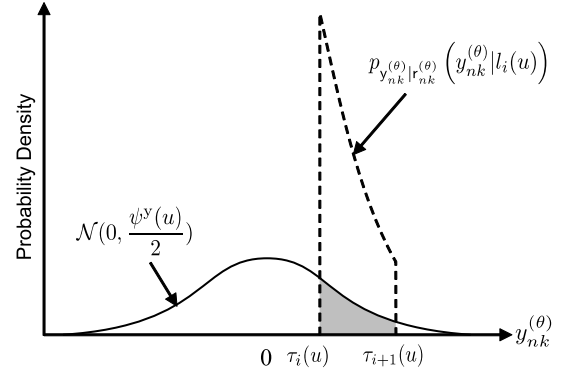
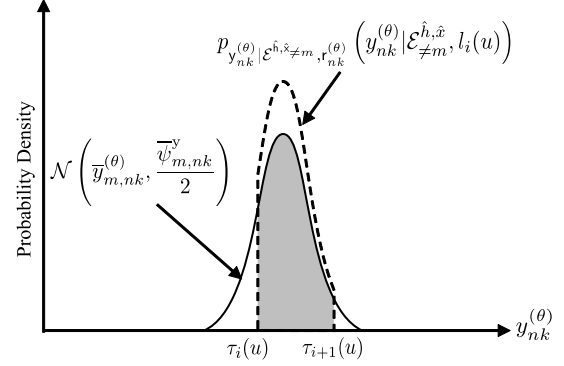
(a) $t = 1$.(b) $t \neq 1$.

Fig. 3. Illustration of the conditional expectation and its corresponding PDF approximated by a truncated and normalized Gaussian distribution.

where for simplicity we have denoted $\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}} \triangleq \{\hat{h}_{k,ni}, \hat{x}_{n,ik}\}_{i \neq m}^M$ and $\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}} \triangleq \{\hat{h}_{k,ni}, \hat{x}_{n,ik}\}_{i \neq m}^M$ and in which we assume that $\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}$ and $r_{nk}^{(\theta)}$ are independent of each other.

For the sake of readability, the detailed derivation of equation (23) is given in Appendix A. Here, suffice it to emphasize that the expression on the right-hand side of equation (23) is simply the truncated Gaussian distribution. In order to determine the original distribution to be truncated in (23), we use the knowledge of the soft replicas to express the RX symbol as equation (24), shown at the bottom of the next page. Under an SGA of the effective noise $\kappa_{m,nk}^y$, we obtain

$$p_{y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}\left(y_{nk}^{(\theta)}|\mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}\right) = \mathcal{N}\left(\bar{y}_{m,nk}^{(\theta)}, \frac{\bar{\psi}_{m,nk}^y}{2}\right), \quad (25)$$

with

$$\bar{y}_{m,nk} = \sum_{i \neq m}^M \hat{h}_{k,ni} \hat{x}_{n,ik}, \quad (26a)$$

$$\bar{\psi}_{m,nk}^y = \sum_{i \neq m}^M \left[|\hat{h}_{k,ni}|^2 \hat{\psi}_{n,ik}^x + \left(|\hat{x}_{n,ik}|^2 + \hat{\psi}_{n,ik}^x \right) \hat{\psi}_{k,ni}^h \right] + \rho \gamma_m(u) + \sigma^2. \quad (26b)$$

Fig. 3b shows the relationship between the conditional PDFs in equations (23) and (25). As illustrated thereby, knowledge

of the soft replicas and their MSEs obtained in the JCDE step enables the correction of the mean and the variance of the conditional PDF of (25), allowing for more accurate DQ processing. In particular, using equations (23) and (25), the soft replica of y_{nk} and its MSE are expressed as

$$\hat{y}_{m,nk} = \sum_{\theta \in \{0,1\}} j^\theta \eta \left(y_{nk}^\theta; \bar{y}_{m,nk}^{(\theta)}, \sqrt{\frac{\psi_{m,nk}^y}{2}}, \tau_i(u), \tau_{i+1}(u) \right), \quad (27a)$$

$$\hat{\psi}_{m,nk}^y = \sum_{\theta \in \{0,1\}} \zeta \left(y_{nk}^\theta; \bar{y}_{m,nk}^{(\theta)}, \sqrt{\frac{\psi_{m,nk}^y}{2}}, \tau_i(u), \tau_{i+1}(u) \right). \quad (27b)$$

Utilizing the estimates given by the equations (21) and (27), and thanks to the relationship obtained from the Busgang decomposition given in equation (11), the soft replica of the signal distortion and its MSE can be calculated as

$$\hat{d}_{m,nk} = r_{nk} - g(u)\hat{y}_{m,nk}, \quad (28a)$$

$$\hat{\psi}_{m,nk}^d = \mathbb{E} \left[|\hat{d}_{m,nk}|^2 \right] = g(u)^2 \hat{\psi}_{m,nk}^y. \quad (28b)$$

with $\tilde{d}_{m,nk} = d_{nk} - \hat{d}_{m,nk}$.

One DQ step is completed when the process described above is carried out for all $u \in \mathcal{U} \setminus \{0\}$, that is, for all signals except the RX signals that are received by the CAP directly from users, namely, $r_{nk}, \forall n \in \mathcal{N}_0$. In light of that, we set both $\hat{d}_{m,nk} = 0$ and $\hat{\psi}_{m,nk}^d = 0, \forall n \in \mathcal{N}_0$.

B. JCDE Step

Next, let us describe the JCDE step for estimating $x_{mk}, \forall m, k \in \mathcal{K}_d$, and $h_{nm}, \forall n, m$. For starters, we emphasize that in the JCDE step, all belief distributions propagated

from FNs to VNs are approximated by a scalar Gaussian distribution based on the CLT, so that only information on the mean and variance of the estimated belief distribution is actually required [38], [49].

The iterative process, which is carried out for all RX indices in parallel, starts at the FNs with soft interference cancellation (soft IC) performed on y_{nk} with utilizing all the soft replicas obtained in the previous JCDE iteration, which for future convenience will be compactly denoted by $\mathcal{E}_{\forall m}^{\hat{h}, \hat{x}} \triangleq \left\{ \hat{h}_{k,nm}, \hat{x}_{n,mk} \right\}_{\forall m}$. At the first iteration ($t = 1$), the soft replicas are appropriately initialized, e.g., $\mathbf{X}_d = \mathbf{0}$. Subsequently, in the detection of an arbitrary TX symbol x_{mk} , the cancellation process is expressed as equation (29), shown at the bottom of the page, where g_n is the n -th diagonal element of \mathbf{G} .

Here it is worth emphasizing that, in contrast to [37], the cancellation process performed according to equation (29) does not resume to the subtraction of inter-UE interference, but instead incorporates the removal of signal distortion as well. Then, under the SGA of the effective noise $\kappa_{m,nk}^r$ whose variance is given by equation (30), shown at the bottom of the page, the conditional PDF of $\tilde{r}_{m,nk}$, given x_{mk} , can be expressed as

$$p_{\tilde{r}_{m,nk}|x_{mk}}(\tilde{r}_{m,nk}|x_{mk}) \propto \exp \left[-\frac{|\tilde{r}_{m,nk} - g_n \hat{h}_{k,nm} x_{mk}|^2}{\nu_{m,nk}^x} \right], \quad (31)$$

where the simplicity of equation (30) owes to the uncorrelated nature of the quantization distortion following the Busgang model of equation (11).

The estimation of channel coefficients h_{nm} can be obtained similarly, such that under the SGA, the conditional PDF of

$$y_{nk} = h_{nm}x_{mk} + \sum_{i \neq m}^M h_{ni}x_{ik} + z_n = \sum_{i \neq m}^M \hat{h}_{k,ni} \hat{x}_{n,ik} + \underbrace{\sum_{i \neq m}^M \left[\tilde{h}_{k,ni} (\hat{x}_{n,ik} + \tilde{x}_{n,ik}) + \hat{h}_{k,ni} \tilde{x}_{n,ik} \right]}_{\triangleq \kappa_{m,nk}^y} + h_{nm}x_{mk} + z_n. \quad (24)$$

$$\tilde{r}_{m,nk} = r_{nk} - g_n \underbrace{\sum_{i \neq m}^M \hat{h}_{k,ni} \hat{x}_{n,ik}}_{\text{Signal distortion cancellation} - \hat{d}_{m,nk}} = g_n h_{nm} x_{mk} + g_n \underbrace{\left(\sum_{i \neq m}^M \left[\tilde{h}_{k,ni} (\hat{x}_{n,ik} + \tilde{x}_{n,ik}) + \hat{h}_{k,ni} \tilde{x}_{n,ik} \right] + z_n \right)}_{\triangleq \kappa_{m,nk}^r} + \tilde{d}_{m,nk}. \quad (29)$$

Inter-UE interference cancellation

$$\begin{aligned} \nu_{m,nk}^x &= \mathbb{E}_{\tilde{r}_{m,nk}|x_{mk}} \left\{ \left| g_n \left(\tilde{h}_{k,nm} x_{mk} + \kappa_{m,nk}^r \right) + \tilde{d}_{m,nk} \right|^2 \middle| x_{mk} \right\} \\ &= g_n^2 \left(\hat{\psi}_{k,nm}^h + \underbrace{\sum_{i \neq m}^M \left\{ |\hat{h}_{k,ni}|^2 \hat{\psi}_{n,ik}^x + (|\hat{x}_{n,ik}|^2 + \hat{\psi}_{n,ik}^x) \hat{\psi}_{k,ni}^h \right\}}_{\triangleq \hat{\psi}_{m,nk}} + \sigma^2 \right) + \hat{\psi}_{m,nk}^d. \end{aligned} \quad (30)$$

$\tilde{r}_{m,nk}$, given h_{nm} , can be expressed as

$$p_{\tilde{r}_{m,nk}|h_{nm}}(\tilde{r}_{m,nk}|h_{nm}) \propto \exp\left[-\frac{|\tilde{r}_{m,nk} - g_n h_{nm} \hat{x}_{n,mk}|^2}{\nu_{m,nk}^h}\right], \quad (32)$$

with

$$\begin{aligned} \nu_{m,nk}^h &= \mathbb{E}_{\tilde{r}_{m,nk}|h_{nm}} \left\{ |g_n (h_{nm} \tilde{x}_{n,mk} + \kappa_{m,nk}^r) + \tilde{d}_{m,nk}|^2 | h_{nm} \right\} \\ &= g_n^2 \left(\xi_{nm} \hat{\psi}_{n,mk}^x + \hat{\psi}_{m,nk} \right) + \hat{\psi}_{m,nk}^d, \end{aligned} \quad (33)$$

where $\xi_{nm} \triangleq \rho\gamma_m(u)$ s.t. $n \in \mathcal{N}_u$.

The subsequent VN process can be performed in a manner similar to the JCDE algorithm described in [37], such that further details are omitted and offered only in a summarized form in the pseudo-code given in the next subsection.

C. Algorithm Description

The pseudo-code of the DQ-aided BiGaBP scheme described above is given in Algorithm 1. Besides the effective RX signal matrix \mathbf{R} and the pilot matrix \mathbf{X}_p , the algorithm requires only system parameters such as the quantization thresholds and labels $\{\mathcal{T}_b(u)\}_{u=1}^U$ and $\{\mathcal{L}_b(u)\}_{u=1}^U$, respectively, the path-losses $\{\gamma_m(u), \forall(u, m)\}$, the Bussgang's gain \mathbf{G} , and the maximum number of iterations t_{\max} , outputting the hard-decision estimates of the TX matrix $\{\tilde{x}_{mk}, \forall(k \in \mathcal{K}^d, m)\}$ and the estimates of the channel coefficients $\{\hat{h}_{nm}, \forall(m, n)\}$. To describe the main blocks, in lines 1 to 4, the soft replicas and their MSE are initialized; lines 6 to 8 correspond to the DQ step; and lines 9 to 15 to the JCDE step. Further, following the repetition of these steps t_{\max} times alternately, in lines 17 to 20 \mathbf{X}_d and \mathbf{H} are estimated.

Notice that although the large-system approximation assumption is relaxed in BiGaBP schemes compared to the conventional BiGAMP, the operating principle of BiGaBP still depends on an accurate SGA of residual components, as can be inferred from Subsections III-A and III-B. In CF-mMIMO systems, adopting single-antenna APs can provide independent fading between distributed antennas; however, the spatial correlation between co-located antennas, mounted on the CAP and multiple-antenna APs, may affect the performance of the proposed algorithm.

In particular, mismatches between the SGA and the stochastic behavior of actual effective noise may result in belief *outliers*, degrading the accuracy of soft replicas and causing performance deterioration due to error propagation [49], especially during earlier iterations of the algorithms and/or in systems with short pilot sequences [37]. In order to mitigate such potential issues, belief damping [37], [50], which prevents the algorithm from converging to local minima and belief scaling [39], [49], which controls convergence speed, are introduced into Algorithm 1. Specifically, belief scaling is applied in line 15 through the parameter $\gamma_s(t) = (\frac{t}{t_{\max}})^2$ [49], which is designed to be a function of the number of iterations, while belief damping is introduced in lines 11, 12, 14, and 15, as described in [50] and summarized as follows.

Algorithm 1 - DQ-Aided BiGaBP

Input: $\{\mathcal{T}_b(u)\}_{u=1}^U, \{\mathcal{L}_b(u)\}_{u=1}^U, \{\gamma_m(u), \forall(u, m)\}, \rho, \mathbf{R}, \mathbf{X}_p, \mathbf{G}, t_{\max}$
Output: $\{\tilde{x}_{mk}, \forall(k \in \mathcal{K}^d, m)\}, \{\hat{h}_{nm}, \forall(m, n)\}$

- 1: $\forall(k \in \mathcal{K}_p, m, n): \hat{x}_{n,mk} = [\mathbf{X}_p]_{mk}$ and $\psi_{n,mk}^x = 0$
- 2: $\forall(k \in \mathcal{K}_d, m, n): \hat{x}_{n,mk} = 0$ and $\psi_{n,mk}^x = 1$
- 3: $\forall(k, m, n): \hat{h}_{k,nm} = 0$ and $\hat{\psi}_{k,nm}^h = \xi_{nm} = \rho\gamma_m(u)$ s.t. $n \in \mathcal{N}_u$
- 4: $\forall(k, m, n \in \mathcal{N}_0): \hat{y}_{m,nk} = r_{nk}, \hat{d}_{m,nk} = 0$, and $\hat{\psi}_{m,nk}^y = \hat{\psi}_{m,nk}^d = 0$
- 5: **for** $t = 1, \dots, t_{\max}$
 $\forall(k, m, n \in \mathcal{N}_{AP}):$ /* DQ step */
- 6: Obtain using eq. (26a) and eq. (26b) $\bar{y}_{m,nk}$ and $\bar{\psi}_{m,nk}^y$
- 7: Obtain using eq. (27a) and eq. (27b) $\hat{y}_{m,nk}$ and $\hat{\psi}_{m,nk}^y$
- 8: Obtain using eq. (28a) and eq. (28b) $\hat{d}_{m,nk}$ and $\hat{\psi}_{m,nk}^d$
/* Begin of JCDE Step: $\forall(m, n)$ */
- 9: Obtain using eq. (29) $\{\tilde{r}_{m,nk}, \forall k\}$ after soft IC
- 10: Obtain using eq. (33) $\{\nu_{m,nk}^h, \forall k\}$ under the SGA
- 11: $\forall k: \bar{h}_{k,nm} = \bar{\psi}_{k,nm}^h \sum_{i \neq k}^K \frac{g_n \hat{x}_{n,mi}^* \bar{y}_{m,ni}}{\nu_{m,ni}^h}$ and $\bar{\psi}_{k,nm}^h = \left[\sum_{i \neq k}^K \frac{g_n^2 |\hat{x}_{n,mi}|^2}{\nu_{m,ni}^h} \right]^{-1}$
- 12: $\forall k: \hat{h}_{k,nm} = \frac{\xi_{nm} \bar{h}_{k,nm}}{\bar{\psi}_{k,nm}^h + \xi_{nm}}$ and $\hat{\psi}_{k,nm}^h = \frac{\xi_{nm} \bar{\psi}_{k,nm}^h}{\bar{\psi}_{k,nm}^h + \xi_{nm}}$
- 13: Obtain using eq. (30) $\{\nu_{m,nk}^x, \forall k \in \mathcal{K}_d\}$ under the SGA
- 14: $\forall k \in \mathcal{K}_d: \bar{x}_{n,mk} = \bar{\psi}_{n,mk}^x \sum_{i \neq n}^N \frac{g_i \hat{r}_{k,im}^* \tilde{r}_{m,ik}}{\nu_{m,ik}^x}$ and $\bar{\psi}_{n,mk}^x = \left[\sum_{i \neq n}^N \frac{g_i^2 |\hat{r}_{k,im}|^2}{\nu_{m,ik}^x} \right]^{-1}$
- 15: $\forall k \in \mathcal{K}_d: \hat{x}_{n,mk} = \frac{1}{\sqrt{2}} \left[\tanh \left[\left(\frac{t}{t_{\max}} \right)^2 \frac{\sqrt{2} \Re\{\bar{x}_{n,mk}\}}{\bar{\psi}_{n,mk}^x} \right] + j \tanh \left[\left(\frac{t}{t_{\max}} \right)^2 \frac{\sqrt{2} \Im\{\bar{x}_{n,mk}\}}{\bar{\psi}_{n,mk}^x} \right] \right]$
and $\hat{\psi}_{n,mk}^x = 1 - |\hat{x}_{n,mk}|^2$
/* End of JCDE Step */
- 16: **end for**
- 17: $\forall(m, n): \bar{h}_{nm} = \bar{\psi}_{nm}^h \sum_{k=1}^K \frac{g_n \hat{x}_{n,mk}^* \bar{y}_{m,nk}}{\nu_{m,nk}^h}$ and $\bar{\psi}_{nm}^h = \left[\sum_{k=1}^K \frac{g_n^2 |\hat{x}_{n,mk}|^2}{\nu_{m,nk}^h} \right]^{-1}$
- 18: $\forall(m, n): \hat{h}_{nm} = \frac{\xi_{nm} \bar{h}_{nm}}{\bar{\psi}_{nm}^h + \xi_{nm}}$
- 19: $\forall(k, m): \bar{\psi}_{mk}^x = \left[\sum_{n=1}^N \frac{g_n^2 |\hat{r}_{k,nm}|^2}{\nu_{m,nk}^x} \right]^{-1}$
- 20: $\forall(k, m): \tilde{x}_{mk} = \underset{\chi_q \in \mathcal{X}}{\operatorname{argmin}} \left| \chi_q - \bar{\psi}_{mk}^x \sum_{n=1}^N \frac{g_n \hat{h}_{k,nm}^* \tilde{r}_{m,nk}}{\nu_{m,nk}^x} \right|$

Let a quantity z be calculated by a function f_z . Then the t -th damped value of z , here denoted $z^{(t)}$, is computed as the weighted average of $z^{(t-1)}$ and f_z , with weights set by the damping factor $\eta_d \in [0, 1]$, i.e.

$$z = f_z \xrightarrow{\text{damping}} z^{(t)} = \eta_d z^{(t-1)} + (1 - \eta_d) f_z. \quad (34)$$

D. Introduction of Local MMSE Filters

Before we move forward to assessing the performance of the proposed DQ-aided BiGaBP algorithm, let us introduce a further modification of the latter aimed at suppressing the harmful effect of spatial correlation between the antennas mounted on the CAP, which can be achieved by introducing a local minimum mean square error (MMSE) filter to whiten the signals $\mathbf{y}_k(0)$ received by the CAP and exploited in the data estimation process described in Section III-B. In particular, to that end, the following processing is performed by the FNs for the signals with indices $\forall n \in \mathcal{N}_0$.

Algorithm 2 - Improved JCDE Step via Local MMSE Filter: $\forall(m, n)$

-
-
- 1: Obtain using eq. (29) $\{\tilde{r}_{m,n,k}, \forall(k, n \in \mathcal{N}_{\text{AP}})\}$ after soft IC
 - 2: Obtain using eq. (35) $\{\tilde{\mathbf{y}}_{m,k}(0), \forall k\}$ after soft IC
 - 3: Obtain using eq. (33) $\{\nu_{m,n,k}^h, \forall k\}$ under SGA
 - 4: $\forall k: \bar{h}_{k,n,m} = \bar{\psi}_{k,n,m}^h \sum_{i \neq k}^K \frac{g_n \hat{x}_{n,m,i} \bar{y}_{m,n,i}}{\nu_{m,n,i}^h}$ and

$$\bar{\psi}_{k,n,m}^h = \left[\sum_{i \neq k}^K \frac{g_n^2 |\hat{x}_{n,m,i}|^2}{\nu_{m,n,i}^h} \right]^{-1}$$
 - 5: $\forall k: \hat{h}_{k,n,m} = \frac{\xi_{nm} \bar{h}_{k,n,m}}{\bar{\psi}_{k,n,m}^h + \xi_{nm}}$ and $\hat{\psi}_{k,n,m}^h = \frac{\xi_{nm} \bar{\psi}_{k,n,m}^h}{\bar{\psi}_{k,n,m}^h + \xi_{nm}}$
 - 6: Obtain using eq. (30) $\{\nu_{m,n,k}^x, \forall(k \in \mathcal{K}_d, n \in \mathcal{N}_{\text{AP}})\}$ under SGA
 - 7: Obtain using eq. (37) $\{\Omega_k, \forall k \in \mathcal{K}_d\}$ under VGA
 - 8: Obtain using eq. (41a) and eq. (41b) $\{\bar{x}_{m,k}, \forall k \in \mathcal{K}_d\}$ and $\{\bar{\psi}_{m,k}^x, \forall k \in \mathcal{K}_d\}$
 - 9: $\forall(k \in \mathcal{K}_d, n \in \mathcal{N}_0): \bar{x}_{n,m,k} = \bar{x}_{m,k}$ and $\bar{\psi}_{n,m,k} = \bar{\psi}_{m,k}^x$
 - 10: Obtain using eq. (41c) and eq. (41d) $\{\bar{x}_{n,m,k}, \forall(k \in \mathcal{K}_d, n \in \mathcal{N}_{\text{AP}})\}$ and $\{\bar{\psi}_{n,m,k}^x, \forall(k \in \mathcal{K}_d, n \in \mathcal{N}_{\text{AP}})\}$
 - 11: $\forall k \in \mathcal{K}_d: \hat{x}_{n,m,k} =$

$$\frac{1}{\sqrt{2}} \left(\tanh \left[\left(\frac{t}{t_{\max}} \right)^2 \frac{\sqrt{2} \Re\{\bar{x}_{n,m,k}\}}{\bar{\psi}_{n,m,k}^x} \right] + j \tanh \left[\left(\frac{t}{t_{\max}} \right)^2 \frac{\sqrt{2} \Im\{\bar{x}_{n,m,k}\}}{\bar{\psi}_{n,m,k}^x} \right] \right)$$
 and $\hat{\psi}_{n,m,k}^x = 1 - |\hat{x}_{n,m,k}|^2$
-
-

First, the soft IC of the RX signal received at the CAP is expressed as

$$\begin{aligned}
 \tilde{\mathbf{y}}_{m,k}(0) &= \mathbf{y}_k(0) - \sum_{i \neq m}^M \hat{\mathbf{h}}_{k,i}(0) \hat{x}_{ik} \\
 &= \mathbf{h}_m(0) x_{mk} \\
 &\quad + \underbrace{\sum_{i \neq m}^M \left\{ \tilde{\mathbf{h}}_{k,i}(0) (\hat{x}_{ik} + \tilde{x}_{ik}) + \hat{\mathbf{h}}_{k,i}(0) \tilde{x}_{ik} \right\}}_{\triangleq \boldsymbol{\kappa}_{k,m}^y} + \mathbf{z}_k(0),
 \end{aligned} \tag{35}$$

where $\mathbf{h}_m(0) = \hat{\mathbf{h}}_{k,m}(0) + \tilde{\mathbf{h}}_{k,m}(0)$ with $\hat{\mathbf{h}}_{k,m}(0) = [\hat{h}_{k,1m}, \dots, \hat{h}_{k,N_0m}]^\top$.

Then, under a vector Gaussian approximation (VGA) of $\boldsymbol{\kappa}_{k,m}^y$, the conditional PDF of (35), given x_{mk} , becomes

$$\begin{aligned}
 p_{\tilde{\mathbf{y}}_{m,k}(0)|x_{mk}}(\tilde{\mathbf{y}}_{m,k}(0)|x_{mk}) &\propto \exp \left[-(\tilde{\mathbf{y}}_{m,k}(0) - \hat{\mathbf{h}}_{k,m}(0)x_{mk})^H \boldsymbol{\Omega}_{m,k}^{-1} \right. \\
 &\quad \left. \times (\tilde{\mathbf{y}}_{m,k}(0) - \hat{\mathbf{h}}_{k,m}(0)x_{mk}) \right],
 \end{aligned} \tag{36}$$

with the covariance matrix given by

$$\begin{aligned}
 \boldsymbol{\Omega}_{m,k} &\triangleq \mathbb{E}_{\tilde{\mathbf{y}}_{m,k}(0)|x_{mk}} \left[|\tilde{\mathbf{h}}_{k,m}(0)x_{mk} + \boldsymbol{\kappa}_{k,m}^y|^2 |x_{mk} \right] \\
 &= \sum_{m=1}^M \underbrace{\left(\hat{\psi}_{i,k}^h + \hat{\mathbf{h}}_{k,i}(0)^H \hat{\mathbf{h}}_{k,i}(0) \hat{\psi}_{ik}^x \right)}_{\triangleq \boldsymbol{\Omega}_k} + \sigma^2 \mathbf{I}_{N_0} \\
 &\quad - \hat{\mathbf{h}}_{k,m}(0)^H \hat{\mathbf{h}}_{k,m}(0) \hat{\psi}_{mk}^x,
 \end{aligned} \tag{37}$$

where $\hat{\boldsymbol{\Psi}}_{m,k}^h \triangleq \mathbb{E}[\tilde{\mathbf{h}}_{k,m}(0)\tilde{\mathbf{h}}_{k,m}(0)^H] = \text{diag}[\hat{\psi}_{k,1m}^h, \dots, \hat{\psi}_{k,N_0m}^h]$.

The subsequent VN process is similar to that of the JCDE algorithm of [37], modified by considering (36) under the VGA, such that only the calculations actually required are

summarized in Algorithm 2, with details on the derivation of the message update rule offered in Appendix B. By replacing the JCDE step of Algorithm 1, *i.e.*, lines 9 to 15, with Algorithm 2, the pseudo-code of the DQ-aided BiGaBP with local MMSE filters is obtained.

IV. PERFORMANCE ASSESSMENT

Computer simulations were conducted to validate the performance of the proposed Bayesian JCDE receiver under various system conditions. The coverage area was set to a square of side 400 [m], with the CAP located at the center and the U APs distributed over an evenly spaced square mesh, while, at each realization, a random number M of UEs were randomly distributed inside the area following a PPP with intensity $\mu_{\text{UE}} = 10$.

Simulated results are offered both in terms of BER and NMSE, averaged over 1000 different random UEs positions and channel realizations. The frame format (*i.e.*, pilot and payload symbol sequences) was set to follow [51], with $K_p = 14$ and $K_d = 126$ such that $K = 140$. Gray-coded QPSK-modulated signals were used for \mathbf{X}_d , but since the number of UEs is random at each realization, with $\mu_{\text{UE}} = 10$, both over- and under-loaded conditions arise. Therefore, in realizations where $K_p \geq M$, orthogonal pilot sequences are used, while in cases when $K_p < M$, the codebook of pilot sequences is given by the rows of a matrix constructed from K_p random columns of an $M \times M$ discrete Fourier transform (DFT) matrix.

The propagation model follows the urban macro cell scenario [52], with the height of the CAP and the APs set to $h_{\text{AP}} = 10$ [m]; carrier frequency, bandwidth, and noise variance at the CAP and APs assumed to be $f_c = 2$ [GHz], $B = 20$ [MHz], $\sigma^2 = -96$ [dBm] [53], respectively; and TX power of UEs ρ in the range of -10 [dBm] to 30 [dBm]. In addition, we defined the antenna decentralized ratio (ADR) as the ratio between the total number of antennas allocated to the APs $|\mathcal{N}_{\text{AP}}| = N - N_0$ and the number of RX antennas N ; thus, $\eta_r \triangleq (N - N_0)/N = 1 - N_0/N$. As for algorithmic parameters, the damping factor η_d was set to 0.5 and the number of iterations was set constant to $t_{\max} = 32$, although other criteria, based on convergence could also be employed.

A. Remarks on Complexity

Before proceeding to the presentation of simulation results on the proposed DQ-aided JCDE method, let us offer a brief analysis of the computational complexities of the algorithms introduced in Section III. Starting with the MPDQ mechanism embedded in Algorithm 1, notice that the corresponding DQ step is performed only over the $|\mathcal{N}_{\text{AP}}|K$ RX symbols collected from APs. The associated complexity is therefore governed by equations (26) and (28), which consist of only scalar-by-scalar calculations, whose number of multiplications, divisions, additions, and subtractions is of order $\mathcal{O}(|\mathcal{N}_{\text{AP}}|MK)$. Similarly, the complexity associated with evaluating the mean and variance of the truncated Gaussian distribution in equation (27) is of order $\mathcal{O}(|\mathcal{N}_{\text{AP}}|MK)$, since such an operation can be handled through a look-up table.

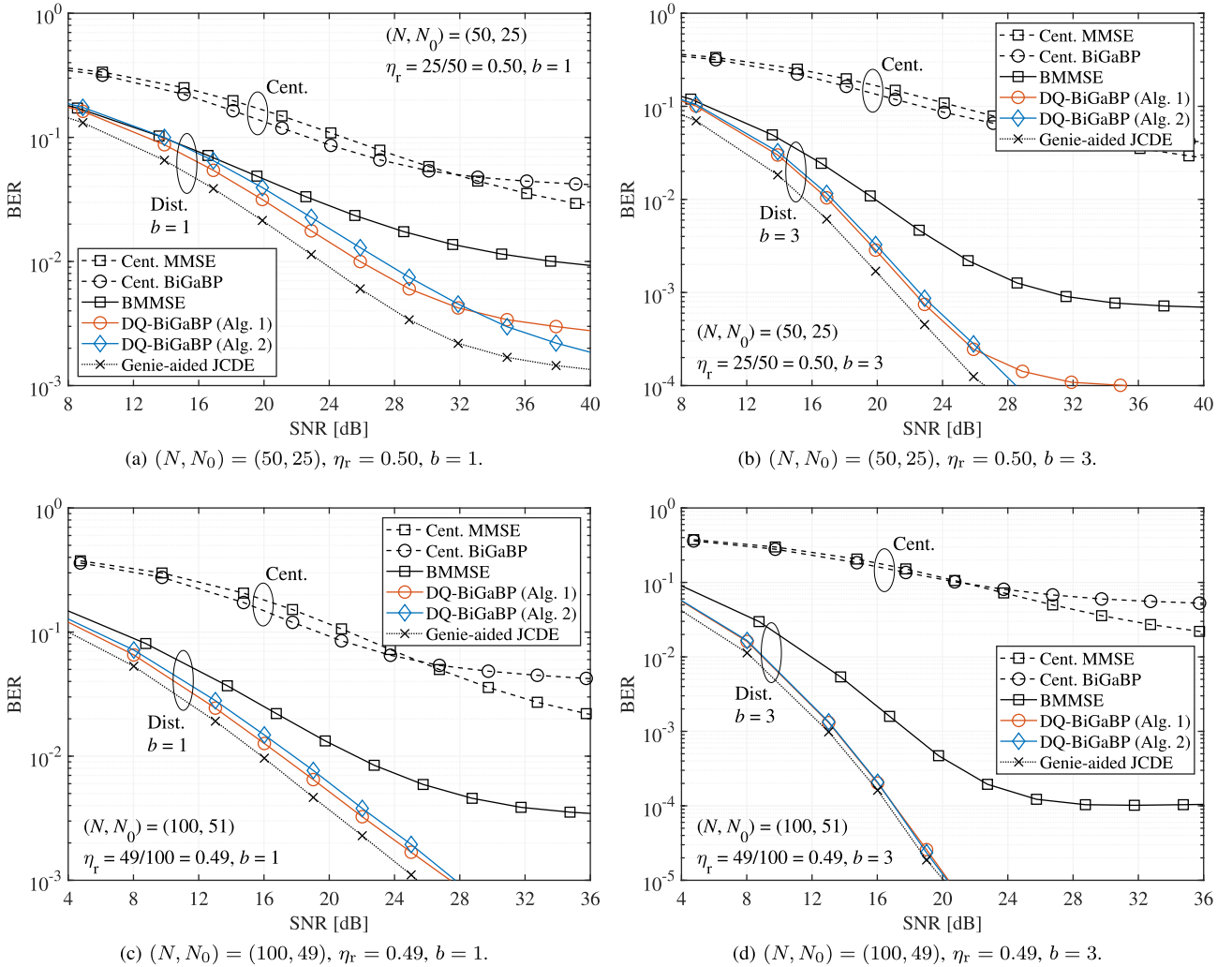


Fig. 4. BER performance of CF-mMIMO systems adopting single-antenna APs with respect to signal-to-noise ratio (SNR) ($N = \{50, 100\}$).

Consequently, the complexity of the DQ step is in order of $\mathcal{O}(|\mathcal{N}_{\text{AP}}|MK)$ per iteration.

Next, in the BiGaBP-based JCDE step, calculations are conducted on NK RX symbols, such that the number of floating-point operations is proportional to $\mathcal{O}(NMK)$, as shown in [39]. Finally, the post-processing operations in lines 17, 18, 19, and 20 have complexity orders $\mathcal{O}(NMK)$, $\mathcal{O}(NM)$, $\mathcal{O}(MK)$, and $\mathcal{O}(MK)$, respectively.

It follows from the above that the complexity of Algorithm 1 is dominated by the JCDE step, which is of order $\mathcal{O}(NMK)$ per iteration and therefore similar to those of existing linear/bilinear Bayesian inference schemes found in literature, e.g. [28], [29], [31], [32], [33], [34], and [39].

In turn, the complexity of Algorithm 2 is dominated by the inverse operation of the $N_0 \times N_0$ matrix Ω_k in equation (37), which is of order $\mathcal{O}(N_0^2 MK + N_0^3 K + |\mathcal{N}_{\text{AP}}|MK)$ per iteration. Given the much higher complexity compared to Algorithm 1, we emphasize that Algorithm 2 is offered for the sake of completeness and as an possible (but not mandatory) add-on to Algorithm 1, at the choice of the system designer, to improve performance. It shall be shown in the sequel,

however, that Algorithm 1 already outperforms the SotA without the inclusion of Algorithm 2².

B. BER Performance

Our first set of results is given in Fig. 4, where the performances in terms of BER as a function of the SNR, with $\text{SNR} \triangleq \frac{\mathbb{E}\{\|\mathbf{H}\mathbf{X}\|^2\}}{KNN_0}$, of the following CF-mMIMO systems adopting single-antenna APs (i.e., $N_u = 1, \forall u \in \mathcal{U} \setminus \{0\}$, and thus $U = |\mathcal{N}_{\text{AP}}| = N - N_0$) are compared:

- *Cent. MMSE*: Baseline linear MMSE receiver with a fully centralized setup; i.e., all N RX antennas are concentrated at the CAP ($\eta_r = 0$), where MUD and CE are sequentially performed using the linear MMSE filters.
- *Cent. BiGaBP*: SotA BiGaBP-based JCDE receiver [37] with a fully centralized set up; i.e., $\eta_r = 0$, providing

²This *robust* Bussgang minimum mean square error (BMMSE)-based linear receiver [19], [23], [25], exploits the heteroscedastic covariance matrix that captures non-uniformity across the received signal vector, which is caused by different statistical properties of the effective noise at each receiver chain in equation (15).

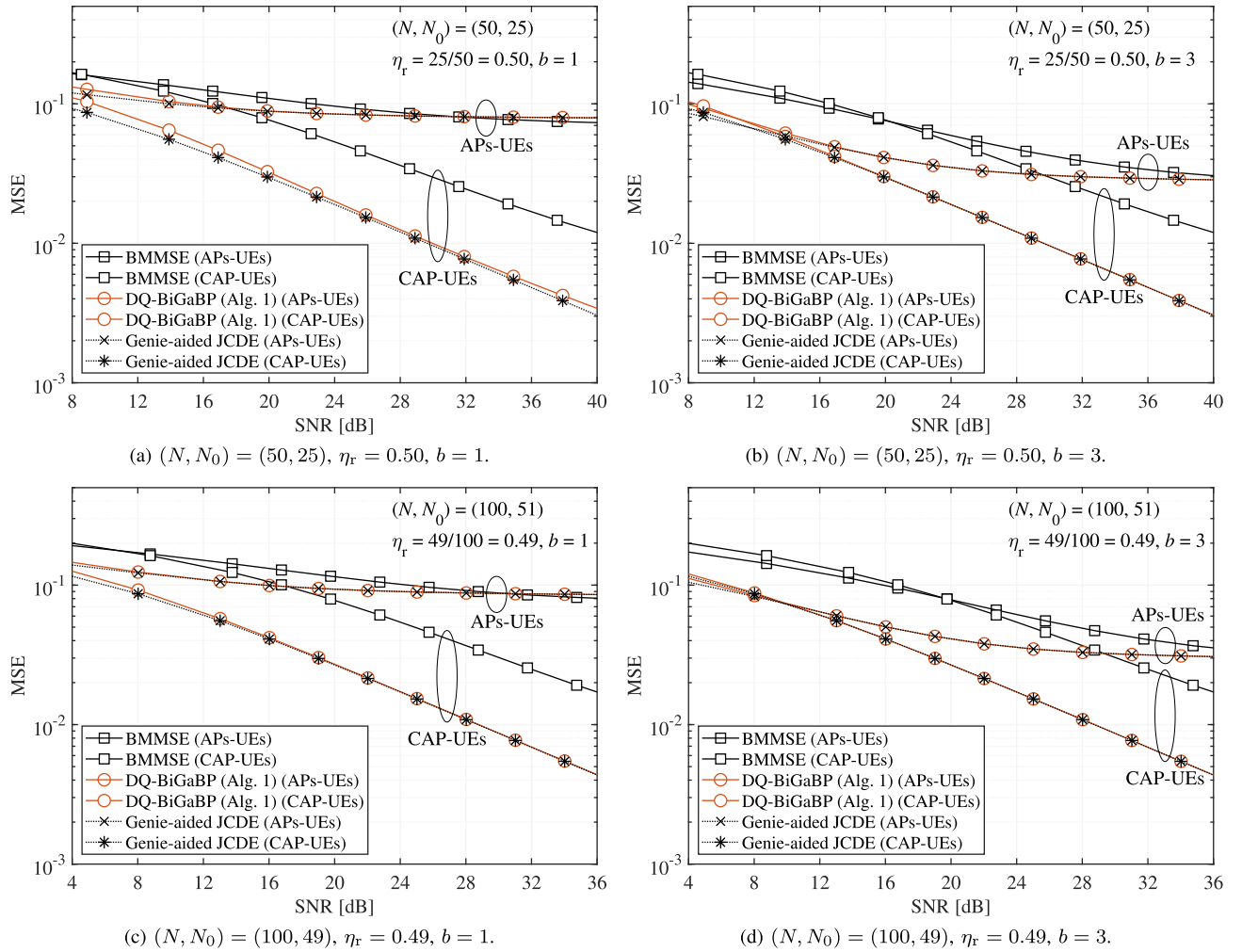


Fig. 5. NMSE performance of CF-mMIMO systems adopting single-antenna APs with respect to SNR ($N = \{50, 100\}$).

a reference performance to verify the gain achieved by distributed antenna placement with APs.

- **BMMSE:** SotA linear receiver based on BMMSE [23], [25], where MUD and CE are sequentially performed. Provides a performance reference to two-step CF-mMIMO estimation schemes, due to the near-optimal MUD performance for a given CE accuracy level, achieved due to the large spatial DoF.
- **DQ-BiGaBP (Alg. 1):** Proposed DQ-aided JCDE receiver presented in Algorithm 1.
- **DQ-BiGaBP (Alg. 2):** Proposed DQ-aided JCDE receiver integrating the local MMSE filters of in Algorithm 2. Provides a performance reference highlighting the effect of spatial correlation between the RX antennas at the CAP.
- **Genie-aided JCDE:** Idealized scheme in which perfect knowledge of \mathbf{X}_d is provided as prior information at each iteration of the bilinear Bayesian JCDE step. Provides an absolute lower bound in terms of the NMSE performance of CE, since perfect knowledge of \mathbf{X}_d can only be obtained when the JCDE algorithm works ideally.

The results in Fig. 4 clearly demonstrate the efficacy of the antenna dispersion strategy of CF-mMIMO systems compared to centralized architectures, even with *coarse* quantization at the APs. It is also confirmed, furthermore, that the proposed methods, *i.e.*, DQ-BiGaBP (Alg. 1) and DQ-BiGaBP (Alg. 2), significantly improve over all SotA methods, which suffer from much higher error floors as a result of their poorer CE performance in presence of non-orthogonal pilots caused either by system overloading or quantization errors. In fact, as shown in Figs. 4a and 4b even in the scenario with $N = 50$, the systems running DQ-BiGaBP (Alg. 1) exhibit error floors due to the lack of spatial DoF. However, it is found that the latter issue is mitigated by the improved DQ-BiGaBP (Alg. 2) scheme, owing to the local MMSE filter which suppresses the harmful effects of strong spatial correlation between antenna elements mounted on the CAP, at the expense of a small performance degradation in the low SNR region caused by noise enhancement when the determinant of the RX covariance matrix is small.

For the results above, it is fair to infer that a simple but effective way to remedy the issue of spatial correlation in CF-mMIMO systems is to boost effective DoF by increasing

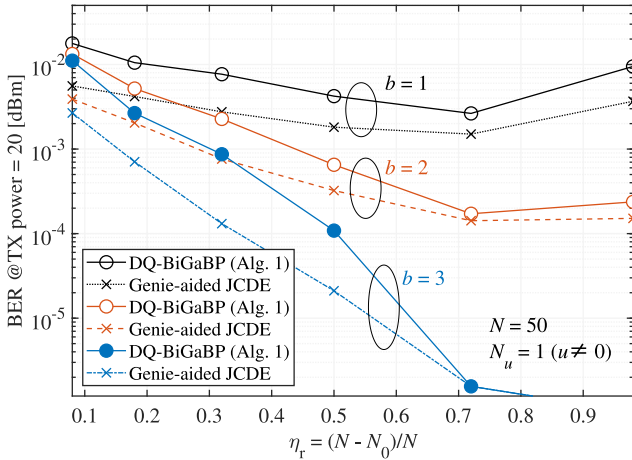
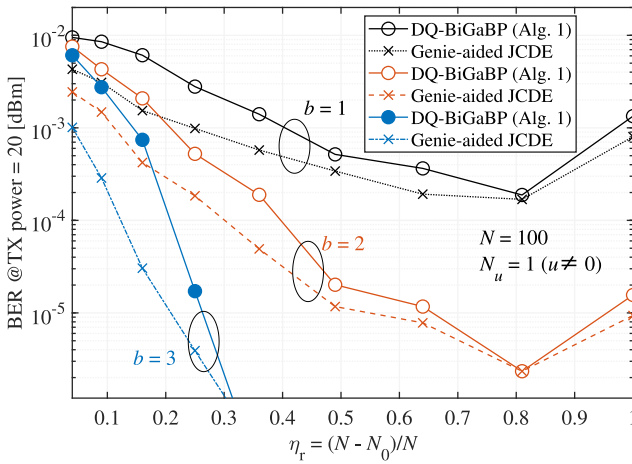
(a) $N = 50$, $b \in \{1, 2, 3\}$.(b) $N = 100$, $b \in \{1, 2, 3\}$.

Fig. 6. BER performances of CF-mMIMO systems adopting single-antenna APs with respect to ADR η_r .

the numbers of RX antennas and/or the APs. In order to verify that hypothesis, we study in Figs. 4c and 4d the performance of CF-mMIMO systems with $N = 100$. In that case, it is found that indeed, the lower-complexity DQ-BiGaBP (Alg. 1) scheme not only already exhibits an error floor-free water-fall BER curve, approaching that of the ideal Genie-aided JCDE reference, but also outperforms DQ-BiGaBP (Alg. 2) in the low SNR region, which is affected by the noise enhancement. From these results, one can conclude that: a) systems with a large number of power-saving and low-cost APs running the low-complexity DQ-BiGaBP (Alg. 1) is generally the setup of choice; but b) if large-scale systems are not feasible due *e.g.* to deployment limitations, then the DQ-BiGaBP (Alg. 2) with a local MMSE filter can be used to compensate for the lack of spatial DoF. Similar discussion can be made for other system configurations except for extremely small or large η_r . The impact of η_r on system performance, including the extreme cases where η_r is close to either 0 or 1, will be studied in Subsection IV-D.

C. MSE Performance

Next, we evaluate the CE performance of the proposed low-complexity method, DQ-BiGaBP (Alg. 1), in terms of the

NMSE of the channel estimates. In view of Fig. 4, results for the Cent. MMSE and Cent. BiGaBP methods are omitted as these SotA alternatives have shown to perform poorly, while results for DQ-BiGaBP (Alg. 2) are omitted since the Genie-aided JCDE method has already provided a lower-bounding reference. With that clarified, the NMSE performances of the remaining three algorithms with the same settings as in Fig. 4 are compared in Fig. 5. To gain deeper insight, the NMSE performance of the estimated channels between APs and UEs, “(APs-UEs),” and between CAPs and UEs, “(CAP-UEs),” were evaluated separately and plotted for each of the methods considered.

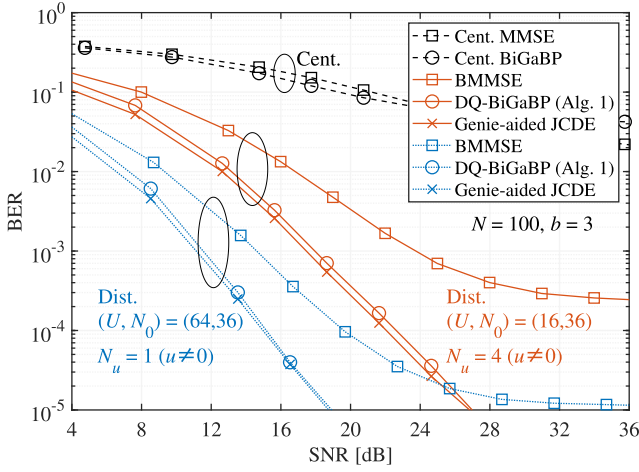
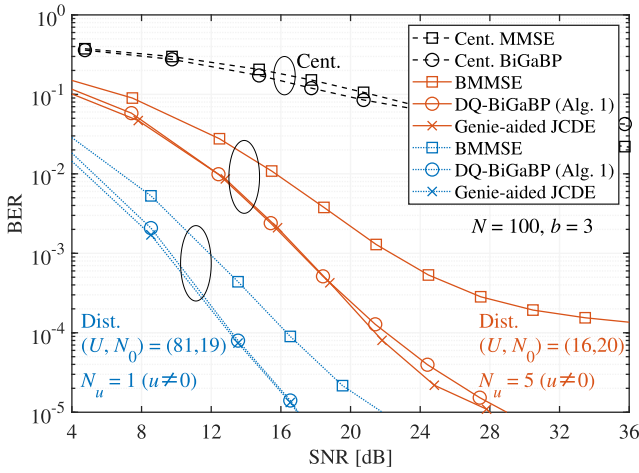
As expected, the performances of “(APs-UEs)” are found to be much worse than those of “(CAP-UEs)” owing to the coarse quantization at the APs. In particular, in the high SNR regime, error floors due to uncertainty caused by quantization noise are unavoidable, and their levels can only be reduced by increasing the number of quantization bits. In contrast, the proposed JCDE mechanism improves SNR, achieving lower NMSEs by exploiting the quasi-orthogonality of the data structure,³ thus significantly outperforming the BMMSE-based scheme and approaching the Genie-aided JCDE method.

In particular, it is found that the gain in SNR achieved by the proposed method increases with the number b of quantization bits, reaching approximately 10 dB for $\text{NMSE} = 10^{-2}$ in the case of “(CAP-UEs),” *i.e.*, $b \rightarrow \infty$. In addition, since in the Genie-aided JCDE scheme CE is performed under perfect knowledge of the TX symbols $\mathbf{X} \triangleq [\mathbf{X}_p, \mathbf{X}_d]$, the asymptotic proximity of the proposed method to that reference indicates that no performance improvement over the proposed DQ-BiGaBP scheme is to be expected from an increase of the pilot sequence length beyond $K_p = 14$ ($\leq K = 140$). In addition, the importance of having a sufficient number of antennas at the CAP is evidenced by the large gap in the performance of “(APs-UEs).”

D. Insights on Optimal Configuration of CF-mMIMO Systems

To gain insight on the optimal setup of CF-mMIMO systems in terms of antenna allocation among the CAP and APs, we compare in Fig. 6 the BER performances of systems with a total number of RX antennas set to $N \in \{50, 100\}$, a TX power of 20 [dBm], and different numbers of quantization bits $b \in \{1, 2, 3\}$, as a function of the ADR η_r , which indicates a more centralized (cellular-like) system for $\eta_r \rightarrow 0$, or a more distributed system for larger η_r , respectively. Again, and as expected, the performance of the proposed DQ-BiGaBP method is found to with growing η_r and N , approaching the Genie-aided performance as η_r and N increases, due to the reduction of correlation between RX antennas and increase of the DoF that follows from larger b and N .

³In this article, the data symbols are generated with an i.i.d. process and transmitted at discrete time slots, implying that the transmitted sequence tends to be orthogonal when K_d is large from a statistical perspective. Mathematically speaking, the mutual coherence between randomly generated data sequences with sufficiently large K_d becomes asymptotically close to zero, *i.e.*, $\frac{1}{K_d} \mathbf{X}_d \mathbf{X}_d^H \approx \mathbf{I}_M$; this is referred to as “quasi-orthogonality of the data structure.”

(a) Single-antenna APs vs. 4-antenna APs ($\eta_r = 0.64$).(b) Single-antenna APs vs. 5-antenna APs ($\eta_r \simeq 0.80$).Fig. 7. BER performance of CF-mMIMO systems adopting multiple-antenna APs with respect to SNR ($N = 100$).

In turn, the performance of all methods, including the ideal Genie-aided, degrades with decreasing b , which emphasizes the crucial role of the double-precision observations at the CAP in mitigating the distortion of the severe information compression carried out at the low-cost APs, in enabling effective information recovery via JCDE. This is also indicated by the poor NMSE performance of “(APs-UEs)” shown in Fig. 5. In other words, it is found that the CF-mMIMO systems should not have all the antennas distributed, and that incorporating both sufficient spatial DoF obtained by the distributed APs, and resolution compensation by double-precision observations at the CAP, is what enables the scaling CF-mMIMO systems, by allowing both low-complexity and high-accuracy estimation performance to be achieved simultaneously.

E. Multiple-Antenna APs

Finally, we study the effect of increasing the number of antennas at APs. In particular, we compare in Fig. 7 the BER performances of CF-mMIMO systems adopting multiple-antenna APs against those of systems with single-antenna APs having the same (or approximately the same) ADR.

Quite counter-intuitively, it is found that the systems with single-antenna APs outperform those with four-antenna (Fig. 7a) and five-antenna (Fig. 7b) APs, respectively. These results can be understood based on those of the preceding Figs. 4 and 5, which indicate that the spatial correlation that occurs at each multiple-antenna AP contributes to reducing effective spatial DoF that causes the observed BER deterioration compared to single-antenna AP systems.

It is also found that the BMMSE-based method suffers from higher error floors in the multiple-antenna case, which can be credited to noise enhancements induced by the required inversion of *ill-conditioned* (i.e., nearly underdetermined) channel matrices, due to spatial correlation among fading coefficients. In contrast, the proposed method approaches the Genie-aided JCDE performance and achieves $\text{BER} = 10^{-5}$, because it exploits the spatial fading independence between the APs in the iterative process to enhance the reliability of the soft IC, thereby enabling the transformation of the problem into an overdetermined condition by digging out the hidden DoF (which is the very aim of the JCDE mechanism). All in all, this result suggests that the Bayesian JCDE receiver is robust to changes in the configuration of the CF-mMIMO system, which is always assumed to enjoy macro-diversity.

V. CONCLUSION

We proposed a novel JCDE receiver based on Bayesian bilinear inference for uplink scalable CF-mMIMO systems adopting power-saving and low-cost APs with low-resolution ADCs. To consider the CE and MUD at a CAP, first, the effective RX signal model observed at the CAP in the form of tractable spatial-temporal linear representation was reformulated using a linearization based on the Bussgang’s theorem. Based on the above signal model, a novel MPDQ algorithm was then proposed considering the BiGaBP framework under consistent assumptions and approximation accuracy. The proposed algorithm possesses linear computational complexity with respect to the number of RX antennas, the number of UEs, and the frame length of transmitted signal, respectively. Finally, computer simulations were conducted to demonstrate the validity of our proposed method in terms of BER and NMSE performances in various CF-mMIMO system configurations. The results revealed the effectiveness of the CF-mMIMO system over the conventional centralized system. Furthermore, comparative study with the SotA alternatives and the case studies covering various scenarios provided an insight into the optimal system design and the robustness to changes in the system configuration.

APPENDIX A

DERIVATION OF EQUATION (23)

Straightforwardly, using Bayes’ rule and the product rule, equation (23) can be rewritten as

$$\begin{aligned} & p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}, r_{nk}^{(\theta)}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}, l_i(u)) \\ &= \frac{p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}})}{p_{y_{nk}^{(\theta)}}(y_{nk}^{(\theta)})} p_{y_{nk}^{(\theta)} | r_{nk}^{(\theta)}}(y_{nk}^{(\theta)} | l_i(u)) \end{aligned}$$

$$\begin{aligned}
&= \frac{p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}})}{p_{y_{nk}^{(\theta)}}(y_{nk}^{(\theta)})} \frac{p_{y_{nk}^{(\theta)}}(y_{nk}^{(\theta)})}{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)}}(y_{nk}^{(\theta)}) dy_{nk}^{(\theta)}} \\
&= C_{m,nk}^{(\theta)} \cdot \frac{p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}})}{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}) dy_{nk}^{(\theta)}}, \quad (38)
\end{aligned}$$

with the proportionality constant given by

$$C_{m,nk}^{(\theta)} = \frac{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}}(y_{nk}^{(\theta)} | \mathcal{E}_{\neq m}^{\hat{h}, \hat{x}}) dy_{nk}^{(\theta)}}{\int_{\tau_i(u)}^{\tau_{i+1}(u)} p_{y_{nk}^{(\theta)}}(y_{nk}^{(\theta)}) dy_{nk}^{(\theta)}}. \quad (39)$$

APPENDIX B

DERIVATION OF THE VN PROCESS IN ALGORITHM 2

Under high-precision SGA and VGA, the beliefs corresponding to x_{mk} are combined over the RX indices, which results in the PDF of an extrinsic belief l_{mk}^x given x_{mk} . This is expressed as

$$\begin{aligned}
&p_{m,k}^x | x_{mk} (l_{mk}^x | x_{mk}) \\
&= p_{\tilde{y}_{m,k}(0) | x_{mk}}(\tilde{y}_{m,k}(0) | x_{mk}) \\
&\quad \times \prod_{n \in \mathcal{N}_{AP}} p_{\tilde{r}_{m,nk} | x_{mk}}(\tilde{r}_{m,nk} | x_{mk}) \\
&\propto \exp \left[-\frac{|\bar{x}_{mk} - x_{mk}|^2}{\bar{\psi}_{mk}^x} \right] \propto \mathcal{CN}(\bar{x}_{mk}, \bar{\psi}_{mk}^x), \quad (40a)
\end{aligned}$$

$$\begin{aligned}
&p_{n,mk}^x | x_{mk} (l_{n,mk}^x | x_{mk}) \\
&= p_{\tilde{y}_{m,k}(0) | x_{mk}}(\tilde{y}_{m,k}(0) | x_{mk}) \\
&\quad \times \prod_{i \in \mathcal{N}_{AP} \setminus \{n\}} p_{\tilde{r}_{m,ik} | x_{mk}}(\tilde{r}_{m,ik} | x_{mk}) \\
&\propto \exp \left[-\frac{|\bar{x}_{n,mk} - x_{mk}|^2}{\bar{\psi}_{n,mk}^x} \right] \\
&\propto \mathcal{CN}(\bar{x}_{n,mk}, \bar{\psi}_{n,mk}^x), \quad n \in \mathcal{N}_{AP}. \quad (40b)
\end{aligned}$$

with

$$\bar{x}_{mk} = \frac{\bar{B}_{mk} \bar{a}_{mk} + \bar{A}_{mk} \bar{b}_{mk}}{\bar{A}_{mk} + \bar{B}_{mk}}, \quad (41a)$$

$$\bar{\psi}_{mk}^x = \frac{\bar{A}_{mk} \bar{B}_{mk}}{\bar{A}_{mk} + \bar{B}_{mk}}, \quad (41b)$$

$$\bar{x}_{n,mk} = \frac{\bar{C}_{n,mk} \bar{a}_{mk} + \bar{A}_{mk} \bar{c}_{n,mk}}{\bar{A}_{mk} + \bar{C}_{n,mk}}, \quad (41c)$$

$$\bar{\psi}_{n,mk}^x = \frac{\bar{A}_{mk} \bar{C}_{n,mk}}{\bar{A}_{mk} + \bar{C}_{n,mk}}, \quad n \in \mathcal{N}_{AP}, \quad (41d)$$

where

$$\bar{a}_{mk} = \bar{A}_{mk} \frac{\hat{\mathbf{h}}_{k,m}(0)^H \boldsymbol{\Omega}_k^{-1} \tilde{\mathbf{y}}_{m,k}(0)}{1 - \lambda_{mk} \hat{\psi}_{mk}^x}, \quad (42a)$$

$$\bar{A}_{mk} = \frac{1 - \lambda_{mk} \hat{\psi}_{mk}^x}{\lambda_{mk}}, \quad (42b)$$

$$\bar{b}_{mk} = \bar{B}_{mk} \sum_{n \in \mathcal{N}_{AP}} \frac{g_n \hat{h}_{k,nm}^* \tilde{r}_{m,nk}}{\nu_{m,nk}^x}, \quad (42c)$$

$$\bar{B}_{mk} = \left(\sum_{n \in \mathcal{N}_{AP}} \frac{g_n^2 |\hat{h}_{k,nm}|^2}{\nu_{m,nk}^x} \right)^{-1}, \quad (42d)$$

$$\bar{c}_{n,mk} = \bar{C}_{n,mk} \sum_{i \in \mathcal{N}_{AP} \setminus \{n\}} \frac{g_i \hat{h}_{k,im}^* \tilde{r}_{m,ik}}{\nu_{m,ik}^x}, \quad (42e)$$

$$\bar{C}_{n,mk} = \left(\sum_{i \in \mathcal{N}_{AP} \setminus \{n\}} \frac{g_i^2 |\hat{h}_{k,im}|^2}{\nu_{m,ik}^x} \right)^{-1}, \quad (42f)$$

with $\lambda_{mk} = \hat{\mathbf{h}}_{k,m}(0)^H \boldsymbol{\Omega}_k^{-1} \hat{\mathbf{h}}_{k,m}(0)$, and where equation (42b) is derived by applying the matrix inversion lemma to eliminate the dependency of $\boldsymbol{\Omega}_{m,k}$ in equation (37) to the index $m \in \mathcal{M}$, enabling us to share the same inverse matrix $\boldsymbol{\Omega}_k^{-1}$ for $\forall m$ [54].

REFERENCES

- [1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [2] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.
- [3] A. Papazafeiropoulos, P. Kourtessis, M. D. Renzo, S. Chatzinotas, and J. M. Senior, "Performance analysis of cell-free massive MIMO systems: A stochastic geometry approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3523–3537, Apr. 2020.
- [4] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.
- [5] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.
- [6] A. K. Papazafeiropoulos, E. Björnson, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Scalable cell-free massive MIMO systems with hardware impairments," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.
- [7] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2112–2127, Oct. 2013.
- [8] A. Yang, Y. Jing, C. Xing, Z. Fei, and J. Kuang, "Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5659–5671, Oct. 2015.
- [9] E. Björnson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011.
- [10] M. Bashar et al., "Exploiting deep learning in limited-fronthaul cell-free massive MIMO uplink," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1678–1697, Aug. 2020.
- [11] X. Hu, C. Zhong, X. Chen, W. Xu, H. Lin, and Z. Zhang, "Cell-free massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6844–6857, Oct. 2019.
- [12] X. Ma, X. Lei, P. T. Mathiopoulos, K. Yu, and X. Tang, "Scalable cell-free massive MIMO systems with finite resolution ADCs/DACs over spatially correlated Rician fading channels," 2021, *arXiv:2109.07061*.
- [13] M. Bashar et al., "Uplink spectral and energy efficiency of cell-free massive MIMO with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 223–245, Jan. 2021.
- [14] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1038–1053, Feb. 2019.
- [15] G. Femenias and F. Riera-Palou, "Fronthaul-constrained cell-free massive MIMO with low resolution ADCs," *IEEE Access*, vol. 8, pp. 116195–116215, 2020.
- [16] B. Nazer, V. R. Cadambe, V. Ntranos, and G. Caire, "Expanding the compute-and-forward framework: Unequal powers, signal levels, and multiple linear combinations," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4879–4909, Sep. 2016.
- [17] Q. Huang and A. Burr, "Compute-and-forward in cell-free massive MIMO: Great performance with low backhaul load," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 601–606.

- [18] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.
- [19] K. Ando, H. Iimori, T. Takahashi, K. Ishibashi, and G. T. F. De Abreu, "Uplink signal detection for scalable cell-free massive MIMO systems with robustness to rate-limited fronthaul," *IEEE Access*, vol. 9, pp. 102770–102782, 2021.
- [20] Y.-S. Jeon, N. Lee, S.-N. Hong, and R. W. Heath, Jr., "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, Jul. 2018.
- [21] J. Choi, J. Mo, and R. W. Heath, Jr., "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, Mar. 2016.
- [22] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Laboratory of Electronics, Cambridge, MA, USA, Tech. Rep., 216, 1952.
- [23] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Apr. 2017.
- [24] J. Mo and R. W. Heath, Jr., "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
- [25] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.
- [26] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," 2014, [arXiv:1404.7736](https://arxiv.org/abs/1404.7736).
- [27] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
- [28] Z. Zhang, X. Cai, C. Li, C. Zhong, and H. Dai, "One-bit quantized massive MIMO detection based on variational approximate message passing," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2358–2373, May 2018.
- [29] I. Atzeni and A. Tolli, "Channel estimation and data detection analysis of massive MIMO with 1-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3850–3867, Jun. 2022.
- [30] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.
- [31] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2168–2172.
- [32] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing dequantization with applications to compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270–6281, Dec. 2012.
- [33] T. C. Zhang, C. K. Wen, S. Jin, and T. Jiang, "Mixed-ADC massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Sep. 2016.
- [34] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [35] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [36] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborova, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 1–6.
- [37] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear Gaussian belief propagation for large MIMO channel and data estimation," in *Proc. IEEE GLOBECOM*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
- [38] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A, Math. General*, vol. 36, no. 43, pp. 11111–11121, 2003.
- [39] 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.
- [40] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.
- [41] X. Meng and J. Zhu, "Bilinear adaptive generalized vector approximate message passing," *IEEE Access*, vol. 7, pp. 4807–4815, 2019.
- [42] S. Sarkar, A. K. Fletcher, S. Rangan, and P. Schniter, "Bilinear recovery using adaptive vector-AMP," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3383–3396, Jul. 2019.
- [43] M. Akrouf, A. Housseini, F. Bellili, and A. Mezghani, "Bilinear generalized vector approximate message passing," 2020, [arXiv:2009.06854](https://arxiv.org/abs/2009.06854).
- [44] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 368–386, Jan. 2020.
- [45] *Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, document TS 36.814, 3GPP, Mar. 2017.
- [46] H. Bolcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.
- [47] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
- [48] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, vol. 1. Hoboken, NJ, USA: Wiley, 1970.
- [49] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.
- [50] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [51] *NR; User Equipment (UE) Radio Transmission and Reception*, 3GPP, document TS 38.101–1, Sep. 2018.
- [52] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, document ITU-R M.2135-1, M Series Mobile, Radiodetermination, Amateur and Related Satellites Services Report, 2009.
- [53] R.-A. Stoica, H. Iimori, and G. T. F. de Abreu, "Multiuser detection for large massively concurrent NOMA systems via fractional programming," in *Proc. CAMSAP*, Guadeloupe, France West Indies, 2019, pp. 1–5.
- [54] L. Hanzo, T. H. Liew, B. L. Yeap, R. Y. S. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. Hoboken, NJ, USA: Wiley, Mar. 2011.



Takumi Takahashi (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 2016, 2017, and 2019, respectively. From 2018 to 2019, he was a Visiting Researcher at the Centre for Wireless Communications, University of Oulu, Finland. In 2019, he joined the Graduate School of Engineering, Osaka University, as an Assistant Professor. His current research interests include belief propagation, compressed sensing, signal processing, and wireless communications.



Hiroki Iimori (Member, IEEE) received the B.Eng. and M.Eng. degrees (Hons.) in electrical and electronic engineering from Ritsumeikan University, Kyoto, Japan, in 2017 and 2019, respectively, and the Ph.D. degree (Hons.) in electrical engineering from Jacobs University Bremen, Germany, in 2022. He was a Visiting Scholar at the University of Toronto, Canada, in 2020. In 2021, he was a Research Intern with the Ericsson Radio S&R Research Laboratory, Yokohama, Japan. Since 2022, he has been with Ericsson Research. His research interests include optimization theory, wireless communications, and signal processing. He was awarded the YKK Doctoral Fellowship by Yoshida Scholarship Foundation and the IEICE Young Researcher of the Year Award by the IEICE Smart Radio Committee in 2020.



Kengo Ando (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Jacobs University Bremen, Germany. His current research interests are wireless communications and signal processing. He was a recipient of the YKK Graduate Fellowship for master's students from the Yoshida Scholarship Foundation, Japan, from 2020 to 2022, and the

Fellowship for overseas study in 2022 from the KDDI Foundation, Japan.



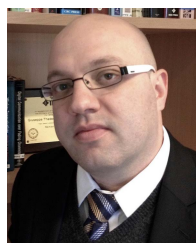
Shinsuke Ibi (Member, IEEE) received the B.E. degree in advanced engineering from the Suzuka College of Technology, Japan, in 2002, and the M.E. and Ph.D. degrees in communication engineering from Osaka University, Japan, in 2004 and 2006, respectively. From 2005 to 2006, he was a Visiting Researcher at the Centre for Wireless Communications, University of Oulu, Finland. In 2006, he joined the Graduate School of Engineering, Osaka University. From 2010 to 2011, he was a Visiting Researcher at the University of Southampton, U.K.

He moved to Doshisha University in 2019. He is currently a Professor with the Faculty of Science and Engineering. His research interests include EXIT-based coding theory, iterative detection, digital signal processing, cognitive radio, and communication theory. He received the 64th and 71st best paper awards from IEICE and the 24th Telecom System Technology Award from the Telecommunication Advancement Foundation.



Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor with the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. Since April 2012, he has been with the Advanced Wireless and Communication Research Center, The Uni-

versity of Electro-Communications, where he is currently a Professor. From 2010 to 2012, he was a Visiting Scholar with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His current research interests are grant-free access, cell-free architecture, millimeter-wave communications, energy harvesting communications, wireless power transfer, channel codes, signal processing, and information theory.



Giuseppe Thadeu Freitas de Abreu (Senior Member, IEEE) received the B.Eng. degree in electrical engineering and the *Latu Sensus* degree in telecommunications engineering from the Universidade Federal da Bahia (UFBa), Salvador, Bahia, Brazil, in 1996 and 1997, respectively, and the M.Eng. and D.Eng. degrees in physics, electrical and computer engineering from Yokohama National University, Japan, in March 2001 and March 2004, respectively. He was a Post-Doctoral Fellow and later an Adjunct Professor (a Docent) in statistical signal

processing and communications theory at the Department of Electrical and Information Engineering, University of Oulu, Finland, from 2004 to 2006 and from 2006 to 2011, respectively. Since 2011, he has been a Professor in electrical engineering at Jacobs University Bremen, Germany. From April 2015 to August 2018, he was a Full Professor at the Department of Computer and Electrical Engineering, Ritsumeikan University, Japan. His research interests include communications and signal processing, including communications theory, estimation theory, statistical modeling, wireless localization, cognitive radio, wireless security, MIMO systems, ultrawideband and millimeter wave communications, full-duplex and cognitive radio, compressive sensing, energy harvesting networks, random networks, connected vehicles networks, and joint communications and sensing. He was a recipient of the Uenohara Award by Tokyo University in 2000 for his master's thesis work. He was a co-recipient of best paper awards at several international conferences and was awarded the JSPS, Heiwa Nakajima, and NICT Fellowships (twice) in 2010, 2013, 2015, and 2018, respectively. He served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2009 to 2014 and the IEEE TRANSACTIONS ON COMMUNICATIONS from 2014 to 2017, and as an Executive Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2018 to 2021.