Hybrid Generalized Approximate Message Passing for Active User Detection and Channel Estimation with Correlated Group-Heterogeneous Activity

Lélio Chetot, Malcolm Egan, Jean-Marie Gorce

Hybrid Generalized Approximate Message Passing for Active User Detection and Channel Estimation with Correlated Group-Heterogeneous Activity

Lélio CHETOT*, Malcolm EGAN*, Jean-Marie GORCE*
email: {lelio.chetot,malcom.egan,jean-marie.gorce}@inria.fr

***Abstract*—The random access procedure is a bottleneck to the development of wireless networks supporting the use cases of massive machine-type communication and ultra reliable and low-latency communication. Such networks are densely and massively popupated and must meet stringent latency and reliability requirements. Due to these characteristics, grant-free random access is envisioned to alleviate the control overhead generated by the classical random access procedure. However, active user detection and channel estimation algorithms are required. Existing algorithms assume that the activity of each device is homogeneous and independent, which is not the case in many applications (e.g., due to sensors observing a common phenomenon). In order to address this problem, we introduce a new flexible model taking into account a group-heterogeneous activity, using the framework of copula theory. It is then leveraged by a hybrid generalized approximate message passing algorithm to solve the active user detection and channel estimation problem. Our numerical results show that the user detection and channel estimation are both improved with this new algorithm w.r.t. state-of-the-art Bayesian algorithms, with gains up to 10 times fewer detection errors and $10\,\mathrm{dB}$ less channel estimation error.**

***Index Terms*—active user detection, channel estimation, bayesian compressed sensing, approximate message passing, correlated activity, internet of things**

I. INTRODUCTION

Industrial IoT (IIoT) is an important pillar to sustain smart manufacturing and Industry 4.0 visions [2]. At the heart of IIoT is a large collection of sensors that are capable of observing manufacturing equipment and an edge or cloud-based data analytics system able to apply artificial intelligence methods to identify anomalies and improve production.

A key challenge is to develop infrastructure to support communication between the sensors. Historically, communication in manufacturing plants has been supported by wired links, which are able to support a massive number of ultra reliable and low-latency communication (uRLLC) links. Unfortunately, wired connections are both expensive and time-consuming to reconfigure. An alternative approach is to exploit wireless links, as in [3], between plant sensors and the cloud. While uRLLC links are now an integral part of 5G, existing systems are still of a small size. A massive scaling up of the quantity of uRLLC connections has been proposed as a motivating use case for 6G. Such a massive scaling is critical for the automation of large-scale industrial plants.

*L. CHETOT, M. EGAN and J.-M. GORCE are with Univ Lyon, Inria, INSA Lyon, CITI, EA3720 in Villeurbanne (69621), France. This work appeared in part in IEEE VTC Spring 2023, see [1].

Supporting massive scale uRLLC wireless communication systems relies on minimizing the amount of control signaling. To this end, an important innovation from 5G is the grant-free random access (GFRA) protocol [4], where active user detection is performed without a dedicated authorization signal from an access point connected to the cloud. Moreover, control signaling constraints mean that active user detection signals must also be kept short, leading to non-orthogonal multiple access (NOMA). A further complication is that devices in manufacturing plants may send event-driven signals, leading to random and, typically, sporadic transmissions from each device. As multiple devices may be placed on the same machine or production line, nearby devices may observe correlated events. As a consequence, in addition to being both random and sporadic, device activity may also be *correlated*. In order to ultimately decode data, a key challenge for IIoT systems exploiting GFRA is then to identify which devices are active and to estimate their channels.

A. Related Work

A popular framework for performing active user detection and channel estimation (AUDaCE) for random and sporadic activity is compressed sensing (CS) [5]. In the literature, there exist two key families of CS algorithms. The first family uses classical CS techniques such as least absolute shrinkage and selection operator (LASSO) optimization [6], [7], matching pursuit (MP) algorithms [8]–[12] or iterative thresholding algorithms [13]–[15]. The second family is Bayesian CS using message-passing algorithms based on belief propagation (BP), expectation propagation (EP) or approximate message passing (AMP). In [16], the simplest version of AMP has been used to perform AUDaCE. In [17], a generalized approximate message passing (GAMP) algorithm was suggested, empowered by expectation maximization (EM) to estimate hyper parameters with an ad-hoc active user detection. In [18], [19], hybrid generalized approximate message passing (HGAMP) is used to perform AUDaCE with low-precision analog-to-digital converter, incorporating independent activity prior information for the active user detection. In [20], AUDaCE is accomplished using bilinear generalized vector approximate message passing (BiGVAMP) in the context of massive unsourced random access.

The compressed sensing algorithms in [16]–[20] rely on the assumption that the random activity of each device is

independent of the activity of other devices in the system. However, when sensors are placed on the same or nearby machines, events may be detected by multiple sensors, leading to *correlated activity*. While compressed sensing algorithms for independent activity may be applied, this can lead to significant performance degradations [21]. To the best of our knowledge, the only work addressing active user detection and channel estimation for devices with correlated activity is [21]. This work exploited the Bayesian compressed sensing framework, where the prior distribution for the device activity vector is correlated. The correlated prior is then leveraged with an HGAMP algorithm. A limitation of the work in [21] is that only a limited form of correlation was accounted for, reducing its useability for more complex systems.

Other works considered the impact of correlated activity on different communications aspects, such as throughput maximization [22], resource optimization [23] and detection errors [24].

A promising method to account for more general forms of correlation is to exploit copula methods. Originally used in the finance and economic fields [25], copula were sporadically considered for communications as in [26]–[28]. In [26], copula are used for MIMO radar waveforms analysis and synthesis. In [27], copula-aided CS of correlated heterogeneous data measurements is performed. In [28], copula-based interference models for internet of things (IoT) networks are proposed to tractably compute achievable rates.

B. Contributions

In this paper, we consider the AUDaCE problem in GFRA with random, sporadic and correlated device activity. Our work differs from [21] as we consider a general correlation model based on copula theory. Our main contributions are as follows:

1) Group-heterogeneous activity model: A key challenge in developing algorithms to solve the AUDaCE algorithm is to utilize appropriate statistical user activity models. These models should capture the random, sporadic and *correlated* aspects of realistic user activity distributions. The most general model is the multivariate Bernoulli model [29]; however, this model has a large number of parameters (exponential in the number of active devices) and is difficult to sample from. To resolve these problems, we propose a new statistical model for user activity exploiting copula theory. In our model, statistical dependence between the activity of different devices is captured via the copula model, which supports a range of correlation structures generalizing existing user activity models currently utilized in the development of AUDaCE algorithms.

2) Hybrid generalized approximate message passing algorithm: With each user activity model, a new AUDaCE algorithm is required. We first develop a loopy BP (LBP) algorithm to solve the AUDaCE problem under the group-heterogeneous activity (GHetA) user activity model. We then approximate the LBP algorithm by exploiting the principles of GAMP. Our approach leads to a novel algorithm within the family of HGAMP algorithms originally developed in the context of Bayesian CS.

3) Numerical evaluation: We validate our GHetA-HGAMP algorithm via a comprehensive numerical study. We first show that our GHetA-HGAMP algorithm either outperforms or achieves the same performance of state-of-the-art AUDaCE algorithms under the GHetA activity model, even in the special cases tailored to existing algorithms. This shows that the GHetA-HGAMP algorithm is applicable in all correlation regimes that have been previously studied and in the general case only captured by our GHetA model. We then study the impact of estimation error in parameters of the GHetA model. We show that the GHetA-HGAMP algorithm is robust to mismatched correlation when the true correlation is lower than the estimated correlation. This provides useful insight for the design of estimators of the GHetA model parameters.

C. Organization

In Sec. II, the system model as well as the new GHetA activity pattern are described for this chapter. The next Sec. IV tailors the HGAMP framework to the new GHetA-based AUDaCE problem. The performances are analyzed in Sec. V and compared to others GAMP-based methods. The paper is concluded in Sec. VI.

D. Notations

The sets $\mathbb{R}, \mathbb{R}^M, \mathbb{R}^{M\times N}$ (resp. $\mathbb{C}, \mathbb{C}^M, \mathbb{C}^{M\times N}$) respectively denote the sets of real (resp. complex) scalars, vectors of dimension M and matrices of dimensions (M, N). For $(a, b) \in \mathbb{N}^2$ with $a \leq b$, the integer set $[\![a, b]\!]$ is a shorthand notation for $\{k \in \mathbb{N} \mid a \leq k \leq b\}$. We respectively denote by x (or X), $\boldsymbol{x}$, $\boldsymbol{X}$ a scalar, vector and matrix deterministic variable and by x (or X), $\mathbf{x}$, $\mathbf{X}$ their random counterparts. The operators $\boldsymbol{X}^\mathsf{T}, \boldsymbol{X}^*, \boldsymbol{X}^\dagger$ denotes the transpose, conjugate and conjugate transpose of a matrix. The probability density function (pdf), cumulative distribution function (cdf) and probability mass function (pmf) of $\mathbf{X}$ are respectively denoted by $f_\mathbf{X}$, $F_\mathbf{X}$ and $\mathbb{P}_\mathbf{X}$. Expectation and variance are denoted by $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$. The real (resp. complex) Gaussian distribution with matrix mean $\boldsymbol{M}$ and covariance $\boldsymbol{C}$ is denoted by $\mathrm{Norm}(\boldsymbol{M}, \boldsymbol{C})$ and its pdf is $\mathcal{N}(\cdot; \boldsymbol{M}, \boldsymbol{C})$ (resp. $\mathrm{CNorm}(\boldsymbol{M}, \boldsymbol{C})$ and $\mathcal{CN}(\cdot; \boldsymbol{M}, \boldsymbol{C})$).

II. System model for grant-free random access

In this section, we formalize the uplink transmission of user equipments (UEs)' preambles in the context of GFRA and shortly review existing activity models.

A. Equivalent baseband uplink transmission

In GFRA, the primary task is for a subset of N UEs to reliably transmit their preamble to the access point (AP). When the transmission of the preamble is performed over a time window corresponding to M orthogonal frequency division multiplexing (OFDM) symbols, the nth UE may transmit a (possibly random) complex signal denoted by $\mathbf{x}_n = [\mathsf{x}_{mn}]_{m\in[\![1,M]\!]}^\mathsf{T} \in \mathbb{C}^M$. In what follows, all UEs are assumed to transmit on the same subcarrier, and M corresponds to the number of resource elements. This assumption is relevant in the context of IoT, such as in NB-IoT [30]

where UEs may transmit on a single carrier with narrowband transmissions [31]. The matrix of all the transmitted signals is then $\mathbf{X} = [\mathbf{x}_n]_{n\in[\![1,N]\!]} \in \mathbb{C}^{M\times N}$.

The signal matrix $\mathbf{X}$ is observed by the kth AP's antenna after passing through a narrowband flat-fading channel. We denote by h_{nk} the equivalent digital baseband channel random coefficient between UE n and antenna k and by $\mathbf{h}_k = [\mathsf{h}_{nk}]^{\mathsf{T}}_{n\in[\![1,N]\!]}$ the channel random vector, where we assume that the coherence time of the channel spans the duration of the signal. The equivalent baseband received signal thus consists in the linear combination of UEs' signals

$$\mathbf{y}_k = \mathbf{z}_k + \mathbf{w}_k \qquad \text{where} \qquad \mathbf{z}_k = \mathbf{X}\mathbf{h}_k \tag{1}$$

where $\mathbf{w}_k = [\mathsf{w}_{mk}]^{\mathsf{T}}_{m\in[\![1,M]\!]} \in \mathbb{C}^M$ is the random vector of baseband noise coefficients. Finally the random signal received over the K AP's antenna is the concatenation of the signals given in (1), leading to

$$\mathbf{Y} = \mathbf{Z} + \mathbf{W} \qquad \text{where} \qquad \mathbf{Z} = \mathbf{X}\mathbf{H} \tag{2}$$

where $\mathbf{H} = [\mathbf{h}_k]_{k\in[\![1,K]\!]} \in \mathbb{C}^{N\times K}$, $\mathbf{W} = [\mathbf{w}_k]_{k\in[\![1,K]\!]} \in \mathbb{C}^{M\times K}$, $\mathbf{Z} = [\mathbf{z}_k]_{k\in[\![1,K]\!]} \in \mathbb{C}^{M\times K}$ and $\mathbf{Y} = [\mathbf{y}_k]_{k\in[\![1,K]\!]} \in \mathbb{C}^{M\times K}$.

B. Activity pattern

During the random access (RA) window, the N UEs belong either to the sets of *active* or *inactive* UEs, depending on whether some UEs initiate a RA procedure before transmitting data payloads.

The random *state*, active or inactive, of the UE $n \in [\![1, N]\!]$ is denoted by a binary random variable $\mathsf{s}_n \in \{0, 1\}$ where

$$\mathsf{s}_n = \begin{cases} 0 \Leftrightarrow \text{UE } n \text{ is inactive} \\ 1 \Leftrightarrow \text{UE } n \text{ is active} \end{cases} \tag{3}$$

For $k \in [\![1, K]\!]$, we assume the distribution of the channel coefficients h_{nk} to depend on the realization of the states s_n. If $\mathsf{s}_n = 0$, then $\mathsf{h}_{nk} \sim \mathrm{Dirac}(0)$. Otherwise, $\mathsf{s}_n = 1$ and $\mathsf{h}_{nk} \sim \mathrm{Dist}$ where Dist is some distribution, e.g. $\mathrm{CNorm}(\mu_{\mathsf{h}}, \tau_{\mathsf{h}})$.

The random activity *pattern* that consists of all the random states during the RA window, is denoted by the binary random vector $\mathbf{s} = [\mathsf{s}_n]^{\mathsf{T}}_{n\in[\![1,N]\!]} \in \{0, 1\}^N$. The probability distribution of $\mathbf{s}$ is described by its pmf $\mathbb{P}_{\mathbf{s}}$. The characterization of the pmf is important for GFRA since the probability that a UE succeeds its RA depends on the joint activity of all UEs in the cell. A number of different models for the pmf $\mathbb{P}_{\mathbf{s}}$ have been considered in the literature. While the most common model is independent activity, this can fail to capture realistic behavior of UEs in the context of IIoT and in uRLLC and massive machine-type communication (mMTC) more generally. In the following, we summarize existing probabilistic models for UE activity.

1) Independent sparse activity pattern: An independent sparse activity (ISA) pattern $\mathbf{s}$ consists of a collection of mutually independent Bernoulli states with activity probabilities $\boldsymbol{q} = [q_n]_{n\in[\![1,N]\!]} \in [0, 1]^N$. The corresponding pmf factorizes as

$$\mathbb{P}_{\mathbf{s}}(\boldsymbol{s}) = \prod_{n=1}^{N} q_n^{s_n}(1 - q_n)^{1-s_n} \tag{4}$$

and the corresponding correlation between any two states $(n, n') \in [\![1, N]\!]^2$ is

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 & \text{if } n = n' \\ 0 & \text{if } n \neq n' \end{cases}. \tag{5}$$

2) Group-sparse activity pattern: The independent group-sparse activity (GSA) pattern [7], [32] is similar to the independent pattern except that the states are split into G non-overlapping groups as suggested by the following partition

$$[\![1, N]\!] = \bigcup_{g=1}^{G} [\![n_{g-1}, n_g - 1]\!] \tag{6}$$

where $n_0 = 1 < \cdots < n_g < \cdots < n_G = G$ are $G + 1$ integers denoting the groups' boundaries. Denote each group size by $S_g \triangleq n_g - n_{g-1}$ UEs in the group $g \in [\![1, G]\!]$ are assumed to have the same random state $\tilde{\mathsf{s}}_g \sim \mathrm{Bern}(q_g)$. The corresponding pmf then factorizes as

$$\mathbb{P}_{\mathbf{s}}(\boldsymbol{s}) = \prod_{g=1}^{G} q_g^{\prod_{n=n_{g-1}}^{n_g-1} s_n}(1 - q_g)^{\prod_{n=n_{g-1}}^{n_g-1}(1-s_n)} \mathbb{1}\left(\prod_{n=n_{g-1}}^{n_g-1} s_n + \prod_{n=n_{g-1}}^{n_g-1}(1 - s_n)\right). \tag{7}$$

The corresponding correlation between any two states is

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 \text{ if } \exists g \in [\![1, G]\!], (n, n') \in [\![n_{g-1}, n_g - 1]\!]^2, \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

When $G = N$, the GSA pattern becomes an independent pattern. We mention that GSA patterns with overlapping groups exist, but have not yet been applied in the context of active user detection and channel estimation.

3) Group-homogeneous activity pattern: In [21], we introduced a new pattern model that accounts for correlated activities in groups of UEs. Assuming the independent group structure of (6), each UE in group $g \in [\![1, G]\!]$ share the same random activity probability $\mathsf{q}_g \in [0, 1]$ which is assumed, for $(\alpha_g, \beta_g) \in \mathbb{R}^2_{+*}$, to be Beta distributed as

$$\mathsf{q}_g \sim \mathrm{Beta}(\alpha_g, \beta_g). \tag{9}$$

Hence, the activity states are distributed according to

$$\forall n \in [\![n_{g-1}, n_g - 1]\!], \ \mathsf{s}_n \mid \mathsf{q}_g = q_g \sim \mathrm{Bern}(q_g) \tag{10}$$

leading to a group-homogeneous activity (GHomA) pattern with corresponding pmf

$$\mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s} \mid \mathbf{q}) = \prod_{g=1}^{G} \prod_{n=n_{g-1}}^{n_g-1} q_g^{s_n}(1 - q_g)^{1-s_n}. \tag{11}$$

Note that it differs from (7) since it is allowed to have different activity states in the same group. The corresponding correlation, for which we give a proof in Appendix B, is the following

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 & \text{if } n = n', \\ \frac{1}{1+\alpha_g+\beta_g} & \text{if } (n, n') \in [\![n_{g-1}, n_g - 1]\!], \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

4) Multivariate Bernoulli activity pattern: A pattern following the multivariate Bernoulli distribution [29] can describe any structure of joint activity states. However, the number of parameters required to achieve such universality in pattern description grows exponentially with the size of the random vector. As a consequence, this model is not suited to model correlated device activities for massive networks.

III. Correlated group-heterogeneous activity with copula

As highlighted in Sec. I, the activity of distinct UEs may be correlated. For example, the source of correlated activity may be the observation of a common event (e.g., a fault) by multiple UEs. Moreover, existing activity models typically either model activity as independent Sec. II-B1 or consider the general multivariate Bernoulli model Sec. II-B4. As a consequence, the models do not capture correlated activity, account for a limited form of correlated activity or have a large number of parameters which leads to difficulties in parameter estimation and simulation. In between the independent and multivariate Bernoulli models are the group-sparse Sec. II-B2 and group-homogeneous activity models Sec. II-B3. These models are suitable when UE activities are highly correlated. However, the performance under these models can significantly degrade when the correlation is low to moderate.

In this section, we introduce a new correlated activity model—called the GHetA model—which captures a wide range of correlation structures. This is achieved via a combination of methods from copula theory—which has seen limited use in wireless communication [28], but has been widely considered in statistics and signal processing [25]–[27]—and the introduction of continuous latent variables. While continuous latent variables were utilized in the GHomA model introduced in [21], the addition of copula methods yields a generalization to a wide range of correlation structures. The proposed model also admits tractable joint user identification and channel estimation algorithms, as developed in Sec. IV.

A. Proposed group-heterogeneous activity pattern

We aim to statistically characterize the activity pattern $\mathbf{s}$ such that it accounts for generic dependence structures between the activity states. We then propose the new GHetA model which consists of a tuple $(\mathbf{c}, \mathbf{q}, \mathbf{s}, \boldsymbol{T})$ satisfying for $n \in [\![1, N]\!]$

$$\mathsf{q}_n = T_n(\mathsf{c}_n) \tag{13}$$

$$\mathsf{s}_n \mid \mathsf{q}_n = q \sim \mathrm{Bern}(q) \tag{14}$$

where $\mathbf{c} = [\mathsf{c}_n]_{n\in[\![1,N]\!]}$ is a correlated random vector with known parameterized distribution, $\mathbf{q} = [\mathsf{q}_n]_{n\in[\![1,N]\!]} \in [0,1]^N$ are latent variables identified with the activity probabilities, $\mathbf{s} = [\mathsf{s}_n]_{n\in[\![1,N]\!]} \in \{0,1\}^N$ is the random activity pattern and $\boldsymbol{T} = [T_n]_{n\in[\![1,N]\!]}$ is a known component-wise bijective mapping.

In the next subsection, we justify the relevance of this model in Sec. III-A1 and describe its construction in Sections III-A2 and III-A3.

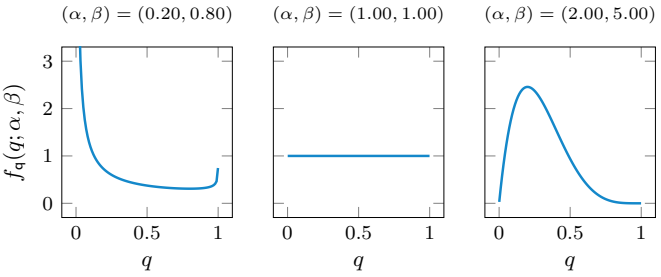

Fig. 1: **Shapes of the Beta probability density function.** The Beta pdf may have a shape morphing from the ∪-shape to the ∩-shape passing by the uniform distribution.

1) Rationale of the model: Since the activity states must be Bernoulli distributed, the key to obtain correlation between them, without using the multivariate Bernoulli distribution [29], is to introduce *continuous latent random variables* $\mathbf{q} = [\mathsf{q}_n]_{n\in[\![1,N]\!]}^{\mathsf{T}} \in [0,1]^N$ such that

$$\mathbb{P}_{\mathbf{s}}(\boldsymbol{s}) = \int_{[0,1]^N} f_{\mathbf{s},\mathbf{q}}(\boldsymbol{s}, \boldsymbol{q}) \mathrm{d}\boldsymbol{q} \tag{15}$$

where the joint pdf factorizes as

$$f_{\mathbf{s},\mathbf{q}}(\boldsymbol{s}, \boldsymbol{q}) = f_{\mathbf{q}}(\boldsymbol{q}) \prod_{n=1}^{N} q_n^{s_n} (1-q_n)^{1-s_n}. \tag{16}$$

Hence, conditional on $\mathbf{q} = \boldsymbol{q}$, the activity variables are independent but correlation in $\mathbf{s}$ still arises when $\mathbf{q}$ is correlated.

The introduction of continuous latent variables provides a means of flexibly controlling the level of correlation between the activity states since, after (95) in Appendix A, their correlation is

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 & \text{if } n = n', \\ \frac{\mathbb{E}[\mathsf{q}_n \mathsf{q}_{n'}] - \mathbb{E}[\mathsf{q}_n]\mathbb{E}[\mathsf{q}_{n'}]}{\sqrt{\mathbb{E}[\mathsf{q}_n](1-\mathbb{E}[\mathsf{q}_n])\mathbb{E}[\mathsf{q}_{n'}](1-\mathbb{E}[\mathsf{q}_{n'}])}} & \text{otherwise} \end{cases} \tag{17}$$

which uniquely depends on the joint distribution of the latent variables.

Motivated by Sklar's theorem [33], a correlated random vector $\mathbf{q}$ may be constructed by *separately* specifying the marginals distributions of $\mathbf{q}$. In particular, we have

$$F_{\mathbf{q}}(\boldsymbol{q}) = F_{\mathbf{u}}(\boldsymbol{u}(\boldsymbol{q})) \quad \text{where} \quad \mathbf{u}(\mathbf{q}) = [F_{\mathsf{q}_n}(\mathsf{q}_n)]_{n\in[\![1,N]\!]}^{\mathsf{T}} \in [0,1]^N \tag{18}$$

where $\mathbf{u}$ is a random vector with joint cdf $F_{\mathbf{u}}$, known as the *copula* function and with the following uniform marginal distributions

$$\forall n \in [\![1, N]\!], \ \mathsf{u}_n \triangleq F_{\mathsf{q}_n}(\mathsf{q}_n) \sim \mathrm{Unif}([0,1]). \tag{19}$$

In order to describe the activity pattern, it is therefore necessary to specify the marginal distributions of $\mathbf{q}$ and the copula function $F_{\mathbf{u}}$.

2) Marginal distributions: Let $n \in [\![1, N]\!]$. As the latent variable q_n plays the role of an activity probability for the Bernoulli random state s_n, it is necessary that q_n must have support on $[0,1]$. Even if there are many possible distributions

that meet this requirement, similar to [21], we choose the marginal of q_n to be Beta distributed according to

$$\mathsf{q}_n \sim \mathrm{Beta}(\alpha_n, \beta_n) \qquad \text{with} \qquad (\alpha_n, \beta_n) \in \mathbb{R}^2_{+*} \tag{20}$$

It is a family of distributions that is known to be very flexible for modeling a wide range of random systems, with the help of only two parameters as suggested by Fig. 1.

Another reason for selecting the Beta distribution is that it can form an approximation for a Bernoulli random variable. Looking at the Beta pdf

$$f_{\mathsf{q}_n}(q; \alpha_n, \beta_n) = \frac{x^{\alpha_n - 1}(1-x)^{\beta_n - 1}}{\mathrm{B}(\alpha_n, \beta_n)} \tag{21}$$

it is clear that the parameters (α_n, β_n) allow to control the skew. With such a choice of marginal distributions, we show in Appendix C that the state correlation is

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 & \text{if } n = n' \\ \dfrac{\mathbb{E}[\mathsf{q}_n \mathsf{q}_{n'}] - \frac{\alpha_n}{\alpha_n + \beta_n} \frac{\alpha_{n'}}{\alpha_{n'} + \beta_{n'}}}{\sqrt{\frac{\alpha_n \beta_n}{(\alpha_n + \beta_n)^2} \frac{\alpha_{n'} \beta_{n'}}{(\alpha_{n'} + \beta_{n'})^2}}} & \text{otherwise} \end{cases} \tag{22}$$

3) Copula function: Since we aim at modeling a correlated activity pattern, we must choose the copula function to be that of a correlated random vector $\mathbf{u}$ with uniform marginals. Such a distribution does not admit a tractable statistical representation by the means of its pdf. So, we reuse Sklar's theorem in order to assume $\mathbf{u}$ is constructed based on

$$F_{\mathbf{c}}(\boldsymbol{c}) = F_{\mathbf{u}}(\boldsymbol{u}(\boldsymbol{c})) \quad \text{where} \quad \mathbf{u}(\mathbf{c}) = [F_{\mathsf{c}_n}(\mathsf{c}_n)]^{\mathsf{T}}_{n \in [\![1,N]\!]} \tag{23}$$

and $\mathbf{c}$ is a correlated random vector.

The choice of the distribution of $\mathbf{c}$ should be motivated by its tractability and capability to describe the targeted dependence structure that we seek for the activity pattern. Indeed, given (18) and (23), we highlight the direct connection between $\mathbf{c}$ and $\mathbf{q}$ with

$$F_{\mathbf{c}}(\boldsymbol{c}) = F_{\mathbf{q}}(\boldsymbol{q}(\boldsymbol{c})) \quad \text{where} \quad \mathbf{q}(\mathbf{c}) = [T_n(\mathsf{c}_n)]^{\mathsf{T}}_{n \in [\![1,N]\!]} \tag{24}$$

and

$$\forall n \in [\![1, N]\!], \; T_n = F^{-1}_{\mathsf{q}_n} \circ F_{\mathsf{c}_n} \tag{25}$$

where $\circ$ denotes the standard function composition. In (24), the connection between the joint pdfs of $\mathbf{c}$ and $\mathbf{q}$ is made thanks to the mappings of (25). If $\mathbf{c}$ has a specific dependence structure, it will be preserved in the latent variables $\mathbf{q}$ and also the activity pattern $\mathbf{s}$ since

$$\mathbf{c} \to \mathbf{q} \to \mathbf{s}. \tag{26}$$

There exist many practical choices of correlated distributions for $\mathbf{c}$ (see [27] and references therein) that can all be considered when applied to the GHetA model. A sampling procedure is given in Sec. III-B and examples are given in Sec. III-C.

B. Sampling from the group-heterogeneous activity model

Sampling from the GHetA model is important to numerically compute expectations of the form

$$\mathbb{E}[g(\mathbf{q})] = \int_{[0,1]^N} g(\boldsymbol{q}) f_{\mathbf{q}}(\boldsymbol{q}) \mathrm{d}\boldsymbol{q} \tag{27}$$

Algorithm 1 Sampling from a copula model.

input:
- N: sample size
- $\mathrm{Dist}(\boldsymbol{\theta})$: correlated distribution of $\mathbf{c}$
- $\{F_{\mathsf{c}_n}\}_{n \in [\![1,N]\!]}$: marginal cdfs
- $\left\{F^{-1}_{\mathsf{q}_n}\right\}_{n \in [\![1,N]\!]}$: inverse marginal cdfs

end

$\boldsymbol{c} \sim \mathrm{Dist}(\boldsymbol{\theta})$

$\boldsymbol{u} = [F_{\mathsf{c}_n}(c_n)]^{\mathsf{T}}_{n \in [\![1,N]\!]}$

$\boldsymbol{q} = [F^{-1}_{\mathsf{q}_n}(u_n)]^{\mathsf{T}}_{n \in [\![1,N]\!]}$

return: $\boldsymbol{q}$

for a function g. For instance, one can approximate (27) using Monte-Carlo integration such that

$$\mathbb{E}[g(\mathbf{q})] \simeq \frac{1}{I} \sum_{i=1}^{I} g(\boldsymbol{q}_i) \tag{28}$$

where the collection $\{\boldsymbol{q}_i\}_{i=1}^I \in [0,1]^{N,I}$ is sampled using Algo. 1. A direct application of such approximation allows to approximate the following expected outer product

$$\mathbb{E}[\mathbf{q}\mathbf{q}^{\mathsf{T}}] = \int_{[0,1]^N} \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} f_{\mathbf{q}}(\boldsymbol{q}) \mathrm{d}\boldsymbol{q} \simeq \frac{1}{I} \sum_{i=1}^{I} \boldsymbol{q}_i \boldsymbol{q}_i^{\mathsf{T}} \tag{29}$$

which is required to compute the correlation matrix of $\mathbf{q}$.

C. Examples of group-heterogeneous activity model

A practical example of the GHetA model is presented in this section. We consider the following distributions

$$\mathbf{c} \sim \mathrm{Norm}(\mathbf{0}_{N,1}, \boldsymbol{K}_{\mathbf{c}}) \text{ where } \begin{cases} \boldsymbol{K}_{\boldsymbol{c}} \in [-1,1]^{N \times N} \\ \boldsymbol{K}_{\boldsymbol{c}} \succeq 0 \end{cases}, \tag{30}$$

$$\forall n \in [\![1, N]\!], \; \mathsf{q}_n \sim \mathrm{Beta}(\alpha_n, \beta_n) \text{ where } (\alpha_n, \beta_n) \in \mathbb{R}^2_{+*}, \tag{31}$$

$$\forall n \in [\![1, N]\!], \; \mathsf{s}_n \sim \mathrm{Bern}(\mathsf{q}_n). \tag{32}$$

Note that the covariance matrix $\boldsymbol{K}_{\mathbf{c}}$ is restricted to be a correlation matrix but can be semi-definite positive, i.e. possibly not invertible. The reason is to allow some particular structures of the correlation matrix.

For instance, one can consider a correlation matrix of the form

$$\boldsymbol{K}_{\boldsymbol{c}} = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{c},1} & & \\ & \ddots & \\ & & \boldsymbol{K}_{\boldsymbol{c},G} \end{bmatrix} \in \mathbb{R}^{N \times N} \tag{33}$$

which is block diagonal where the block matrix $g \in [\![1, G]\!]$ is given by

$$\boldsymbol{K}_{\boldsymbol{c},g} = \rho_g \mathbf{1}_{|\mathbb{G}_g|,|\mathbb{G}_g|} + (1 - \rho_g) \boldsymbol{I}_{|\mathbb{G}_g|} \quad \text{and} \quad \rho_g \in [0,1]. \tag{34}$$

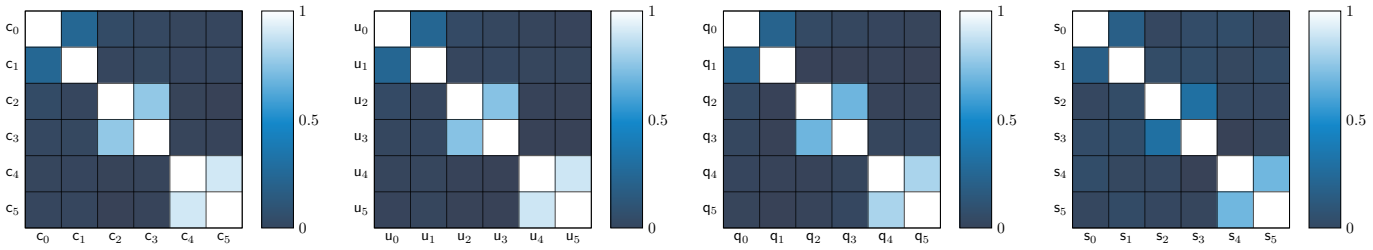

Fig. 2: **Correlation matrices.** Example of correlation matrices for the random vectors $\mathbf{c} \to \mathbf{u} \to \mathbf{q} \to \mathbf{s}$ where $\mathbf{c}$ is correlated Gaussian and $\mathbf{q}$ has Beta marginals. The same three groups of sizes $[2, 2, 2]$ can be distinguished for each vector. The Gaussian correlation coefficients are $[\rho_g]_{g\in[\![1,3]\!]} = [0.25, 0.75, 0.90]$. The Beta parameters are $[\alpha_n]_{n\in[\![1,6]\!]} = [0.25, 0.25, 0.2, 0.2, 0.1, 0.1]$ and $[\beta_n]_{n\in[\![1,6]\!]} = [0.75, 0.75, 1.2, 1.2, 0.1, 0.1]$.

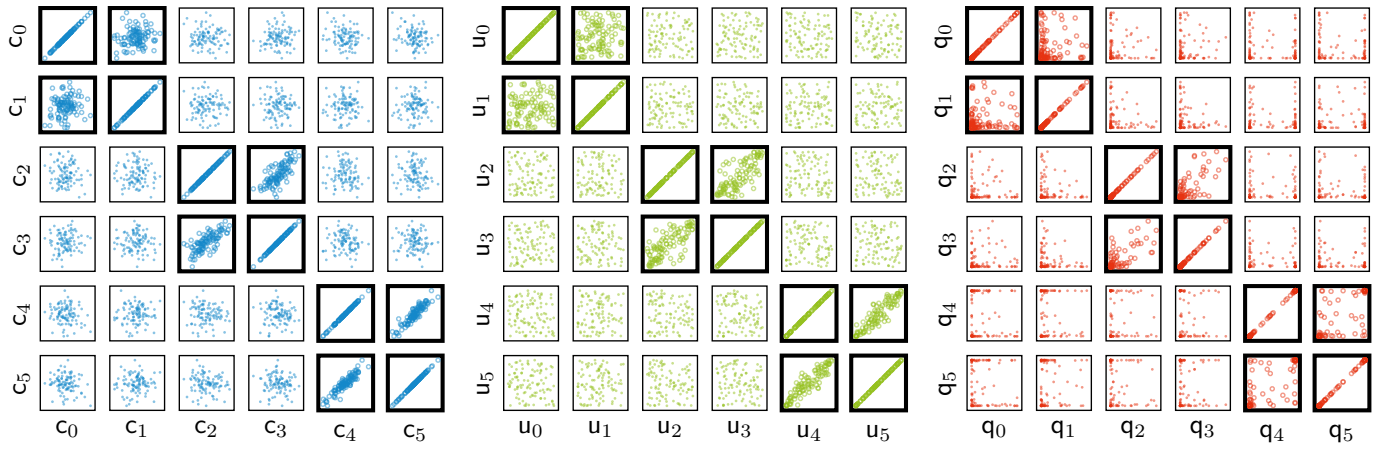

Fig. 3: **Scatter plots.** Pairwise scatter plots between the coordinates of multiple samples of the random vectors $\mathbf{c} \to \mathbf{u} \to \mathbf{q}$. Each subplot depicts how two components of the vectors spread in the plane, depending on their correlation and the distribution parameters. The samples are obtained using Algo. 1 with the same parameters as in Fig. 2. The scatter plots between variables from the same group are drawn with thicker axes.

A consequence of the structure (33) is that the marginals are standard normal distributions with cdf

$$F_{\mathsf{c}_n}(c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2}} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u. \tag{35}$$

Next, we assume that the activity probability vector $\mathbf{q}$ has beta marginal distributions with cdfs

$$\forall n \in [\![1, N]\!],\ F_{\mathsf{q}_n}(q) = \int_0^q \frac{u^{\alpha_n - 1}(1-u)^{\beta_n - 1}}{\mathrm{B}(\alpha_n, \beta_n)} \mathrm{d}u \tag{36}$$

We then obtain the correlated random vector $\mathbf{q}$ with the component-wise transform

$$\mathbf{q} = \left[(F_{\mathsf{q}_n}^{-1} \circ F_{\mathsf{c}_n})(\mathsf{c}_n)\right]_{n\in[\![1,N]\!]}^{\mathsf{T}}. \tag{37}$$

The random activity states are finally drawn based on

$$\forall n \in [\![1, N]\!],\ \mathsf{s}_n \sim \mathrm{Bern}((F_{\mathsf{q}_n}^{-1} \circ F_{\mathsf{c}_n})(\mathsf{c}_n)). \tag{38}$$

The empirical correlation matrices of each random vector are given in Fig. 2, showing that the correlation structure (not the correlation values) is saved from $\mathbf{c}$ to $\mathbf{s}$. This is verified in Fig. 3 that shows pairwise scatter plots of coordinate samples using Algo. 1.

D. Connection to other activity models

1) Generalization of GHomA: The GHetA model boils down to the GHomA model in Sec. II-B3 for a particular choice of the correlation structures since one can formulates the states correlation (see (96) in Appendix A) as

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \begin{cases} 1 & \text{if } n = n' \\ \frac{\mathrm{Cor}[\mathsf{q}_n, \mathsf{q}_{n'}]}{\sqrt{(\alpha_n+\beta_n+1)(\alpha_{n'}+\beta_{n'}+1)}} & \text{otherwise} \end{cases}. \tag{39}$$

This last formulation leads to two insights.

First the state correlation will necessarily be smaller than the activity probability correlation. Indeed, the denominator is strictly greater than 1 since $(\alpha_n, \beta_n, \alpha_{n'}, \beta_{n'}) \in \mathbb{R}_{+,*}^4$ and

$$\sqrt{(\alpha_n + \beta_n + 1)(\alpha_{n'} + \beta_{n'} + 1)} > 1. \tag{40}$$

In the particular case $\alpha_n = \alpha_{n'}$ and $\alpha_n + \beta_n = 1$, the correlation $\mathrm{Cor}[\mathsf{q}_n, \mathsf{q}_{n'}]$ is halved.

Second, it explicitly generalizes the one in (12). Indeed when q_n and $\mathsf{q}_{n'}$ are fully correlated, i.e. when $\mathrm{Cor}[\mathsf{q}_n, \mathsf{q}_{n'}] = 1$, and $(\alpha_n, \beta_n) = (\alpha_{n'}, \beta_{n'}) = (\alpha_g, \beta_g)$ there exist underlying, fully correlated, c_n and $\mathsf{c}_{n'}$ that belong to a common group g. Hence the states correlation coefficient equals that of GHomA since we obtain

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \frac{1}{\alpha_g + \beta_g + 1} \tag{41}$$

as in (12), which shows the generalization.

2) Generalization of GSA: Let $g \in [\![1, G]\!]$. The GSA model from Sec. II-B2 describes activity patterns where states are identical within a group g and are active with probability $\tilde{q}_g$:

$$\forall n \in [\![n_{g-1}, n_g - 1]\!], \mathsf{s}_n \sim \mathrm{Bern}(\mathbb{E}[\mathsf{s}_n] = \bar{q}_g). \tag{42}$$

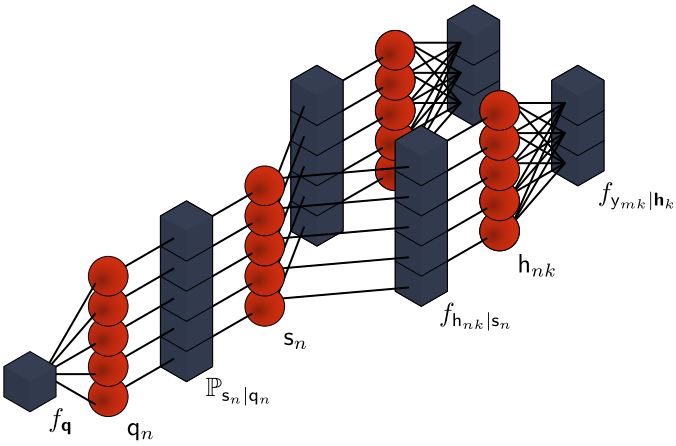

Fig. 4: **Factor graph.** Graphical model of the system presented in Sec. II with $N = 5$, $M = 3$ and $K = 2$. Different from the factor graph of ISA-HGAMP, GSA-HGAMP and GHomA-HGAMP, this factor graph considers the correlation between the variables $\{q_n\}_{n\in[N]}$.

To describe this model with GHetA, assume the Beta distribution parameters (α_n, β_n) to be identical in the same group

$$\alpha_g \triangleq \alpha_{n_{g-1}} = \cdots = \alpha_{n_g-1} \tag{43}$$
$$\beta_g \triangleq \beta_{n_{g-1}} = \cdots = \beta_{n_g-1}. \tag{44}$$

Then assume that the correlation between activity probabilities within the same group is maximal ($\mathrm{Cor}[q_n, q_{n'}] = 1$) leading to the following equality

$$\forall g \in [\![1, G]\!], \mathsf{q}_g \triangleq \mathsf{q}_{n_{g-1}} = \cdots = \mathsf{q}_{n_g-1}. \tag{45}$$

Consider that (α_g, β_g) tends to $(0, 0)$ while keeping constant the average activity probability such that

$$\mathbb{E}[\mathsf{q}_g] = \alpha_g/(\alpha_g + \beta_g) = \bar{q}_g. \tag{46}$$

Hence, the beta pdf which is $\cup$-shaped, will have sharper asymptotes, concentrating the pdf around the abscissa $q \in \{0, 1\}$, leading to a Bernoulli pmf. This is equivalent to say that q_g converges in distribution to a Bernoulli distribution $\mathrm{Bern}(\bar{q}_g)$. The corresponding states in this group, will then have activity probability $\mathsf{q}_g \in \{0, 1\}$, and so be either all active or inactive.

Since $\mathsf{s}_n \sim \mathrm{Bern}(\mathbb{E}[\mathsf{s}_n])$, we compute the expected value of s_n and find that

$$\mathbb{E}[\mathsf{s}_n] = \mathbb{E}_{\mathsf{q}_g}[\mathbb{E}_{\mathsf{s}_n|\mathsf{q}_g}[\mathsf{s}_n \mid \mathsf{q}_g]] = \mathbb{E}[\tilde{\mathsf{q}}_g] = \bar{q}_g \tag{47}$$

which corresponds to the GSA model. Note that the exact generalization is only true in the asymptotic case (α_g, β_g) tends to $(0, 0)$.

3) Generalization of ISA: GSA generalizes ISA from Sec. II-B1 when $G = N$ and so does GHetA.

IV. Active user detection and channel estimation

A. General problem formulation

The key of a successful GFRA procedure lies in its capability to correctly perform AUDaCE. Given the received signal $\mathbf{Y} = \boldsymbol{Y}$ and an activity model, the active user detection consists in identifying the set of active UEs transmitting their identification sequences. This is equivalent to identify the support $\mathbf{s} = \boldsymbol{s}^\star$ of the rows of $\mathbf{H} = \boldsymbol{H}^\star$ where

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{H}^\star + \boldsymbol{W}. \tag{48}$$

The channel estimation consists in estimating the channel coefficients of $\boldsymbol{H}^\star$ outside of the support. The joint AUDaCE consequently formalizes as

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}\in\{0,1\}^N}{\arg\max}\, \mathbb{P}_{\mathbf{s}|\mathbf{Y}}(\boldsymbol{s} \mid \boldsymbol{Y}) \quad \text{and} \quad \hat{\boldsymbol{H}} = \mathbb{E}[\boldsymbol{H} \mid \boldsymbol{Y}] \tag{49}$$

The joint formulation of this inference problem comes from the fact that both estimators require the joint density $f_{\mathbf{s},\mathbf{H},\mathbf{Y}}$ since

$$\mathbb{P}_{\mathbf{s}|\mathbf{Y}}(\boldsymbol{s} \mid \boldsymbol{Y}) = \frac{\int_{\mathbb{C}^{N\times K}} f_{\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y})\mathrm{d}\boldsymbol{H}}{\mathcal{Z}(\boldsymbol{Y})} \tag{50}$$

$$f_{\mathbf{H}|\mathbf{Y}}(\boldsymbol{H} \mid \boldsymbol{Y}) = \frac{\sum_{\boldsymbol{s}\in\{0,1\}^N} f_{\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y})}{\mathcal{Z}(\boldsymbol{Y})} \tag{51}$$

where

$$\mathcal{Z}(\boldsymbol{Y}) = \int_{\mathbb{C}^{N\times K}} \sum_{\boldsymbol{s}\in\{0,1\}^N} f_{\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y})\mathrm{d}\boldsymbol{H} \tag{52}$$

is the system log-partition function. For the GHomA and newly introduced GHetA models, the latent variables $\mathbf{q}$ are introduced in the joint pdf with

$$f_{\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y}) = \int_{[0,1]^N} f_{\mathbf{q},\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y})\mathrm{d}\boldsymbol{q} \tag{53}$$

and the log-partition function can be further written as

$$\mathcal{Z}(\boldsymbol{Y}) = \int_{\mathbb{C}^{N\times K}} \sum_{\boldsymbol{s}\in\{0,1\}^N} \int_{[0,1]^N} f_{\mathbf{q},\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{H}, \boldsymbol{Y})\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{H}. \tag{54}$$

To perform AUDaCE, one needs to know the posteriors of (50) and (51) which is impossible in general because of the intractable marginalization they require. We then develop in the next section a BP algorithm that addresses this issue.

B. Approximation by belief propagation

The posteriors (50) and (51) being intractable, we seek to approximate them. BP is a handy algorithmic framework that provides a systematic means for computing such approximates. To do so, note the system's variables form the following Markov chain

$$\mathbf{q} \to \mathbf{s} \to \mathbf{H} \to \mathbf{Y} \tag{55}$$

which allows to write their joint density as

$$f_{\mathbf{H},\mathbf{s},\mathbf{q}|\mathbf{Y}}(\boldsymbol{H}, \boldsymbol{s}, \boldsymbol{q} \mid \boldsymbol{Y}) = f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y} \mid \boldsymbol{H}) f_{\mathbf{H}|\mathbf{s}}(\boldsymbol{H} \mid \boldsymbol{s}) \mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s} \mid \boldsymbol{q}) f_{\mathbf{q}}(\boldsymbol{q}). \tag{56}$$

Based on Sec. II-A, we can factorize the joint density's factors as

$$f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y} \mid \boldsymbol{H}) = \prod_{k=1}^{K}\prod_{m=1}^{M} f_{\mathsf{y}_{mk}|\mathbf{h}_k}(y_{mk} \mid \boldsymbol{h}_k) \tag{57}$$

$$f_{\mathbf{H}|\mathbf{s}}(\boldsymbol{H} \mid \boldsymbol{s}) = \prod_{k=1}^{K}\prod_{n=1}^{N} f_{\mathsf{h}_{nk}|\mathsf{s}_n}(h_{nk} \mid s_n) \tag{58}$$

$$\mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s} \mid \boldsymbol{q}) = \prod_{n=1}^{N} f_{\mathsf{s}_n|\mathsf{q}_n}(s_n \mid q_n) \tag{59}$$

Algorithm 2 GHetA-HGAMP

Description: GHetA-HGAMP consists into two parts. The **black** lines corresponds to the updates of the GAMP variables for estimating the channel and the blue lines corresponds to the updates of the pattern variables with BP. Estimates of the system variables are colored in red. **Note:** All variance updates $\tau_\cdot$ and associated variables should be understood as **damped** updates, meaning that a new value is computed as $x^{i+1} = (1-d)x^i + d\tilde{x}^i$, where $d \in [0,1]$ is the damping factor. It is set globally to $d = 0.9$

input: $\boldsymbol{Y}, \boldsymbol{X}, \mu_{\mathsf{h}}, \tau_{\mathsf{h}}, \tau_{\mathsf{w}}, I_{\max}$
init:
  $i = 0$
  $\forall (n,k) \in [\![1,N]\!] \times [\![1,K]\!]\ \ \hat{h}^i_{nk} = \mu_{\mathsf{h}},\ \ \hat{\tau}^i_{\mathsf{h},nk} = \tau_{\mathsf{h}}$
  $\forall (m,k) \in [\![1,N]\!] \times [\![1,K]\!]\ \ \hat{u}^i_{mk} = 0$
end
for $i \in [\![1, I_{\max}]\!]$ **do:**
  for $(m,k) \in [\![1,M]\!] \times [\![1,K]\!]$ **do:**
    $\hat{\tau}^i_{\mathsf{p},mk} = \sum_{n=1}^N |x_{mn}|^2 \hat{\tau}^{i-1}_{\mathsf{h},nk}$
    $\hat{p}^i_{mk} = \sum_{n=1}^N x_{mn}\hat{h}^{i-1}_{nk} - \hat{\tau}^i_{\mathsf{p},mk}\hat{u}^{i-1}_{mk}$
    $\hat{\tau}^i_{\mathsf{z},mk} = \tau_{\mathsf{w}}\hat{\tau}^i_{\mathsf{p},mk}/(\hat{\tau}^i_{\mathsf{p},mk} + \tau_{\mathsf{w}})$
    $\hat{z}^i_{mk} = \hat{p}^i_{mk} + \hat{\tau}^i_{\mathsf{p},mk}(y_{mk} - \hat{p}^i_{mk})/(\hat{\tau}^i_{\mathsf{p},mk} + \tau_{\mathsf{w}})$
    $\hat{\tau}^i_{\mathsf{u},mk} = (1 - \hat{\tau}^i_{\mathsf{z},mk})/(\hat{\tau}^i_{\mathsf{p},mk})^2$
    $\hat{u}^i_{mk} = (\hat{z}^i_{mk} - \hat{p}^i_{mk})/\hat{\tau}^i_{\mathsf{p},mk}$
  end
  for $(n,k) \in [\![1,N]\!] \times [\![1,K]\!]$ **do:**
    $\hat{\tau}^i_{\mathsf{r},nk} = (\sum_{m=1}^M |x_{mn}|^2 \hat{\tau}^i_{\mathsf{u},mk})^{-1}$
    $\hat{r}^i_{nk} = \hat{h}^{i-1}_{nk} + \hat{\tau}^i_{\mathsf{r},nk} \sum_{m=1}^M x_{mn}\hat{u}^i_{mk}$
  end
  for $n \in [\![1,N]\!]$ **do:**
    $\phi_{0,n} = \prod_{k=1}^K \mathcal{CN}(0; \hat{r}^i_{nk}, \hat{\tau}^i_{\mathsf{r},nk})$
    $\phi_{1,n} = \prod_{k=1}^K \mathcal{CN}(0; \hat{r}^i_{nk} - \mu_{\mathsf{h}}, \hat{\tau}^i_{\mathsf{r},nk} + \tau_{\mathsf{h}})$
    $\hat{q}^i_{n,n} = \frac{\int_{[0,1]^N} q_n f[\mathbf{q}](\boldsymbol{q}) \prod_{n' \in [\![1,N]\!]\setminus\{n\}} [(1-q_{n'})\phi_{0,n'} + q_{n'}\phi_{1,n'}]}{\int_{[0,1]^N} f[\mathbf{q}](\boldsymbol{q}) \prod_{n' \in [\![1,N]\!]\setminus\{n\}} [(1-q_{n'})\phi_{0,n'} + q_{n'}\phi_{1,n'}]}$
    $\mathsf{LLR}_n = \log\left(\frac{\hat{q}^i_{n,g}}{(1-\hat{q}^i_{n,g})} \frac{\phi_{1,n}}{\phi_{0,n}}\right)$
    $\hat{s}^i_n = \mathbb{1}(\mathsf{LLR}_n > 0)$
    $\gamma_n = (1 + \exp(-\mathsf{LLR}_n))^{-1}$
    $\hat{q}^i_n = \frac{\int_{[0,1]^N} q_n f[\mathbf{q}](\boldsymbol{q}) \prod_{n'=1}^N [(1-q_{n'})\phi_{0,n'} + q_{n'}\phi_{1,n'}] \mathrm{d}q_n}{\int_{[0,1]^N} f[\mathbf{q}](\boldsymbol{q}) \prod_{n'=1}^N [(1-q_{n'})\phi_{0,n'} + q_{n'}\phi_{1,n'}] \mathrm{d}q_n}$
  end
  for $(n,k) \in [\![1,N]\!] \times [\![1,K]\!]$ **do:**
    $\kappa_{nk} = (1/\tau_{\mathsf{h}} + 1/\hat{\tau}^i_{\mathsf{r},nk})^{-1}$
    $\nu_{nk} = \mu_{\mathsf{h}}/\tau_{\mathsf{h}} + \hat{r}^i_{nk}/\hat{\tau}^i_{\mathsf{r},nk}$
    $\hat{h}^i_{nk} = \gamma_n \kappa_{nk} \nu_{nk}$
    $\hat{\tau}^i_{\mathsf{h},nk} = \gamma_n(\kappa_{nk} + |\kappa_{nk}\nu_{nk}|^2) - \left|\hat{h}^i_{nk}\right|^2$
  end
end
return: $\hat{\boldsymbol{s}}^{I_{max}}$, $\hat{\boldsymbol{H}}^{I_{max}}$

and we keep $f_{\mathbf{q}}(\boldsymbol{q})$ unfactorized, which corresponds to a generally correlated model.

This leads to the factor graph depicted in Fig. 4, from which we derive an instance of BP for which the messages are given in Tab. I and the variable beliefs, for $(n,k) \in [\![1,N]\!] \times [\![1,K]\!]$, are

$$\underset{\mathsf{s}_n}{\mathfrak{B}} = \underset{\mathbb{P}_{\mathsf{s}_n|\mathsf{q}_n} \to \mathsf{s}_n}{\mathfrak{M}} \underset{\mathbb{P}_{\mathsf{s}_n|\mathsf{q}_n} \leftarrow \mathsf{s}_n}{\mathfrak{M}} \quad \text{and} \quad \underset{\mathsf{h}_{nk}}{\mathfrak{B}} = \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \to \mathsf{h}_{nk}}{\mathfrak{M}} \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}}. \tag{60}$$

The beliefs allow to approximate the posterior pdfs (50) and (51) by

$$\mathbb{P}_{\mathbf{s}|\mathbf{Y}} \simeq \prod_{n=1}^N \underset{\mathsf{s}_n}{\mathfrak{B}} \quad \text{and} \quad f_{\mathbf{H}|\mathbf{Y}} \simeq \prod_{k=1}^K \prod_{n=1}^N \underset{\mathsf{h}_{nk}}{\mathfrak{B}} \tag{61}$$

and produce the estimates $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{H}}$ by independently solving

$$\hat{s}_n = \mathbb{1}(\underset{s_n}{\mathfrak{B}}(0) < \underset{s_n}{\mathfrak{B}}(1)), \quad \text{and} \quad \hat{h}_{nk} = \mathbb{E}[\mathsf{h}_{nk}; \underset{h_{nk}}{\mathfrak{B}}]. \tag{62}$$

However, there are major downsides in using BP for large and densely connected factor graphs.

1) the messages are functions, each requiring to be discretized in order to be represented on a computer. This is very demanding in terms of storage capability since there are $O(KMN)$ messages.
2) Second, the computation of the message

$$\underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}} = \prod_{m=1}^M \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}} \tag{63}$$

requires the computation of the messages $\underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}}$ which are in general intractable high dimensional integrals (cf. Tab. I).

Hence, we build in the next section a HGAMP algorithm to address this issue.

C. GHetA-HGAMP algorithm

Following the works in [32], [34], [35], we can partially approximate the BP algorithm in Tab. I using the framework of HGAMP. For large systems when $N/M \underset{N \to +\infty}{=} O(1)$, the following approximation of (63) can be made

$$\underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}}{\mathfrak{M}}(h) \approx \mathcal{CN}(h; \hat{r}_{nk}, \hat{\tau}_{\mathsf{r},nk}). \tag{64}$$

where $(\hat{r}_{nk}, \hat{\tau}_{\mathsf{r},nk})$ are mean and variance variables that are iteratively updated by Algo. 2. Propagating this approximation in the BP messages in Tab. I, we obtain the HGAMP instance in Algo. 2 tailored to the GHetA model, namely GHetA-HGAMP.

From Tab. I and Algo. 2, one can see that the original BP algorithm has been turned into an iterative algorithm where the updated quantities are not functions anymore but scalars.

TABLE I
LOOPY BELIEF PROPAGATION MESSAGES FOR THE FACTOR GRAPH OF FIG. 4

Factor $\longrightarrow$ Variable		Factor $\longleftarrow$ Variable
$\mathfrak{M}_{f_{y_{mk}\|\mathbf{h}_k}\to\mathbf{h}_{nk}}(h_{nk})$	$\propto$	$\int_{\mathbb{C}^{N-1}} f_{\mathsf{y}_{mk}\|\mathbf{h}_k}(y_{mk}\mid \boldsymbol{h}_k)\left[\prod_{n'\in[\![1,N]\!]\setminus\{n\}} \mathfrak{M}_{f_{y_{mk}\|\mathbf{h}_k}\leftarrow\mathbf{h}_{n'k}}(h_{n'k})\mathrm{d}h_{n'k}\right]$
$\mathfrak{M}_{f_{y_{mk}\|\mathbf{h}_k}\leftarrow\mathbf{h}_{nk}}(h_{nk})$	$\propto$	$\mathfrak{M}_{f_{h_{nk}\|s_n}\to\mathbf{h}_{nk}}(h_{nk}) \prod_{m'\in[\![1,M]\!]\setminus\{m\}} \mathfrak{M}_{f_{y_{m'k}\|\mathbf{h}_k}\to\mathbf{h}_{nk}}(h_{nk})$
$\mathfrak{M}_{f_{\mathsf{h}_{nk}\|\mathsf{s}_n}\to\mathsf{h}_{nk}}(h_{nk})$	$\propto$	$\sum_{s=0}^{1} f_{\mathsf{h}_{nk}\|\mathsf{s}_n}(h_{nk}\mid s)\,\mathfrak{M}_{f_{h_{nk}\|s_n}\leftarrow\mathsf{s}_n}(s)$
$\mathfrak{M}_{f_{\mathsf{h}_{nk}\|\mathsf{s}_n}\leftarrow\mathsf{h}_{nk}}(h_{nk})$	$\propto$	$\prod_{m=1}^{M} \mathfrak{M}_{f_{y_{mk}\|\mathbf{h}_k}\to\mathbf{h}_{nk}}(h_{nk})$
$\mathfrak{M}_{f_{\mathsf{h}_{nk}\|\mathsf{s}_n}\to\mathsf{s}_n}(s_n)$	$\propto$	$\int_{\mathbb{C}} f_{\mathsf{h}_{nk}\|\mathsf{s}_n}(h\mid s_n)\,\mathfrak{M}_{f_{h_{nk}\|s_n}\leftarrow\mathbf{h}_{nk}}(h)\mathrm{d}h$
$\mathfrak{M}_{f_{h_{nk}\|s_n}\leftarrow\mathsf{s}_n}(s_n)$	$\propto$	$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\to\mathsf{s}_n}(s_n) \prod_{k'\in[\![1,K]\!]\setminus\{k\}} \mathfrak{M}_{f_{h_{nk'}\|s_n}\to\mathsf{s}_n}(s_n)$
$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\to\mathsf{s}_n}(s_n)$	$\propto$	$\int_0^1 \mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_n}(s_n\mid q)\,\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\leftarrow\mathsf{q}_n}(q)\mathrm{d}q$
$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\leftarrow\mathsf{s}_n}(s_n)$	$\propto$	$\prod_{k=1}^{K} \mathfrak{M}_{f_{h_{nk}\|s_n}\to\mathsf{s}_n}(s_n)$
$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\to\mathsf{q}_n}(q_n)$	$\propto$	$\sum_{s=0}^{1} \mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_n}(s\mid q_n)\,\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\leftarrow\mathsf{s}_n}(s)$
$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\leftarrow\mathsf{q}_n}(q_n)$	$\propto$	$\mathfrak{M}_{f_{\mathsf{q}}\to\mathsf{q}_n}(q_n)$
$\mathfrak{M}_{f_{\mathsf{q}}\to\mathsf{q}_n}(q_n)$	$\propto$	$\int_{[0,1]^{N-1}} f_{\mathsf{q}}(q)\left[\prod_{n'\in[\![1,N]\!]\setminus\{n\}} \mathfrak{M}_{f_{\mathsf{q}}\leftarrow\mathsf{q}_{n'}}(q_{n'})\mathrm{d}q_{n'}\right]$
$\mathfrak{M}_{f_{\mathsf{q}}\leftarrow\mathsf{q}_n}(q_n)$	$\propto$	$\mathfrak{M}_{\mathbb{P}_{s_n\|q_n}\to\mathsf{q}_n}(q_n)$

D. Complexity analysis

A short analysis is given for the computational complexity of Algo. 2.

1) GAMP updates: The lines 8 to 19 and 29 to 34 corresponds to the classical GAMP updates. From line 8 to 19, there are $O(KMN)$ operations, coming from the for loops and the sum required to compute line 9, 10, 17 and 18. From line 29 to 34, the number of operations is $O(KN)$. Hence, the complexity of GAMP is of the order of

$$C_{\text{GAMP}} = O(KMN). \tag{65}$$

2) BP updates: The lines 20 to 28 corresponds to the BP updates remaining after the GAMP approximation. First, note that the updates at line 23 and 27 in Algo. 2 require numerical integrations. They are handled using Monte-Carlo integration with I_{samples} samples $\{\boldsymbol{q}_i\}_{i\in[\![1,I_{\text{samples}}]\!]}$ drawn with Algo. 1. Hence the integrals can be approximated using

$$\hat{q}_{n,n} \approx \frac{\sum_{j=1}^{I_{\text{samples}}} q_{n,j}\varpi_{n,j}}{\sum_{j=1}^{I_{\text{samples}}} \varpi_{n,j}} \quad \text{and} \quad \hat{q}_n \approx \frac{\sum_{j=1}^{I_{\text{samples}}} q_{n,j}\varpi_j}{\sum_{j=1}^{I_{\text{samples}}} \varpi_j} \tag{66}$$

where

$$\varpi_j = \prod_{n'=1}^{N} \Big((1-q_{n',j})\phi_{0,n'} + q_{n',j}\phi_{1,n'}\Big) \tag{67}$$

$$\varpi_{n,j} = \frac{\varpi_j}{(1-q_{n,j})\phi_{0,n} + q_{n,j}\phi_{1,n}} \tag{68}$$

TABLE II
COMPUTATIONAL COMPLEXITIES OF RELATED HGAMP ALGORITHMS.

Algorithm	Computational complexity	References
GAMP, ISA-HGAMP	$O(KMN)$	[18], [34]
GSA-HGAMP	$O(KMN)$	[32]
GHomA-HGAMP	$O(KMN + KN + NS^2)$	[21]
GHetA-HGAMP	$O(KMN + KN + N^2 I_{\text{samples}})$	see (70)

which require $O(NI_{\text{samples}})$ operations to be computed. Accounting for the remaining lines and the for loop, the computational complexity of the BP part is

$$C_{\text{BP}} = O(KN + N^2 I_{\text{samples}}). \tag{69}$$

3) GHetA-HGAMP complexity: Combining (65) and (69), the computational complexity of GHetA-HGAMP is

$$C_{\text{HGAMP}} = C_{\text{GAMP}} + C_{\text{BP}} = O(KMN + KN + N^2 I_{\text{samples}}) \tag{70}$$

The BP part is the computational bottleneck of this algorithm, because of the estimates $\hat{q}^i_{n,n}$ and $\hat{q}^i_n$. The reader can refer to Tab. II for a comparison with related HGAMP algorithms.

V. NUMERICAL RESULTS

The goal of this section is the following:

- assess the performance gain of using GHetA-HGAMP for correlated scenarios against state-of-the-art methods that were not designed to deal with heterogeneous correlation;
- assess the robustness of GHetA-HGAMP to parameter mismatches.

The performance of the new GHetA-HGAMP are numerically studied based on extensive Monte-Carlo simulations. For a collection of I_{MC} ground truth values $\{(\boldsymbol{s}_i^\star, \boldsymbol{H}_i^\star)\}_{i\in[\![1,I_{\text{MC}}]\!]}$ and their respective estimates $\{(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{H}}_i)\}_{i\in[\![1,I_{\text{MC}}]\!]}$, the capability of GHetA-HGAMP to perform active user detection is assessed by the false alarm rate (FAR)

$$\mathsf{FAR}(\{\boldsymbol{s}_i^\star, \hat{\boldsymbol{s}}_i\}_{i\in[\![1,I_{\text{MC}}]\!]}) = \frac{1}{NI_{\text{MC}}} \sum_{i=1}^{I_{\text{MC}}} \sum_{n=1}^{N} \mathbb{1}(s_{n,i} < \hat{s}_{n,i}), \tag{71}$$

the missed detection rate (MDR)

$$\mathsf{MDR}(\{\boldsymbol{s}_i^\star, \hat{\boldsymbol{s}}_i\}_{i\in[\![1,I_{\text{MC}}]\!]}) = \frac{1}{NI_{\text{MC}}} \sum_{i=1}^{I_{\text{MC}}} \sum_{n=1}^{N} \mathbb{1}(s_{n,i} > \hat{s}_{n,i}), \tag{72}$$

and the user error rate (UER)

$$\mathsf{UER}(\{\boldsymbol{s}_i^\star, \hat{\boldsymbol{s}}_i\}_{i\in[\![1,I_{\text{MC}}]\!]}) = \frac{1}{NI_{\text{MC}}} \sum_{i=1}^{I_{\text{MC}}} \sum_{n=1}^{N} \mathbb{1}(s_{n,i} \neq \hat{s}_{n,i}). \tag{73}$$

The quality of its channel estimation is assessed by the normalized mean squared error (NMSE)

$$\mathsf{NMSE}(\{\boldsymbol{H}_i^\star, \hat{\boldsymbol{H}}_i\}_{i\in[\![1,I_{\text{MC}}]\!]}) = \frac{\sum_{i=1}^{I_{\text{MC}}} \|\boldsymbol{H}_i^\star - \hat{\boldsymbol{H}}_i\|_2^2}{\sum_{i=1}^{I_{\text{MC}}} \|\boldsymbol{H}_i^\star\|_2^2}. \tag{74}$$

In order to compare the performances of our GHetA-HGAMP, we consider the following benchmark of GAMP-based algorithms, forming the state-of-the-art methods for AUDaCE with Bayesian compressed sensing (BCS).

A. Benchmark

a) ISA-HGAMP (based on [18], [34]): This instance of HGAMP corresponds to the algorithm originally described in [34] but with a slight modification to allow activity detection as in [18]. To do so, we have considered, as in [17], a Bernoulli-Gaussian prior for the channel coefficients so that

$$f_{\mathsf{H}}(\boldsymbol{H}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\Big((1-q_n)\delta(h_{nk}) + q_n\mathcal{CN}(h_{nk};\mu_{\mathsf{h}},\tau_{\mathsf{h}})\Big) \tag{75}$$

where

$$\forall n \in [\![1,N]\!],\ q_n = \alpha_n/(\alpha_n+\beta_n). \tag{76}$$

The log-likelihood ratios for this HGAMP become

$$\mathsf{LLR}_n = \log\frac{\alpha_n}{\alpha_n+\beta_n} + \log\phi_{0,n} - \log\frac{\beta_n}{\alpha_n+\beta_n} - \log\phi_{1,n} \tag{77}$$

that will be used for activity detection. This ISA-HGAMP does not account for any activity correlation.

b) State Evolution for ISA-HGAMP (see [34], [36], [37]): Assume $K=1$. State evolution (SE) accurately tracks the asymptotic mean squared error (MSE) of a (H)GAMP estimator $\hat{\boldsymbol{h}}(\mathbf{y})$ at iteration t with the following recursion

$$\tau_r(t+1) = \tau_w + \frac{N}{M}\,\mathrm{MSE}(\mathsf{r}(t),\mathsf{h}) \tag{78}$$

where

$$\mathsf{r}(t) = \mathsf{h} + \mathsf{w}_r(t) \quad \text{given} \quad \begin{cases}\mathsf{h}\sim\mathrm{CNorm}(0,\tau_h),\\ \mathsf{w}_r(t)\sim\mathrm{CNorm}(0,\tau_r(t))\end{cases} \tag{79}$$

and

$$\mathrm{MSE}(\mathsf{r},\mathsf{h}) = \mathbb{E}[(\hat{h}(\mathsf{r})-\mathsf{h})^2]. \tag{80}$$

This SE can only be applied to (H)GAMP-based algorithms with separable prior $f_{\mathsf{h}}(\boldsymbol{h}) = \prod_{n=1}^{N} f_{\mathsf{h}}(h_n)$. Hence, it is not applicable to GHetA-HGAMP since $\mathbf{h}$ does not have a separable prior because of the correlated activity induced by the copula structure (see Sec. III). The derivation of SE is a rather involved task, making not straightforward its extension to GHetA-HGAMP.

However, we choose to add SE for ISA-HGAMP to our benchmark in order to demonstrate that it is not adapted to GHetA. Hence, the NMSE produced by SE is going to be compared with the NMSE of the other algorithms.

More precisely, we have implemented the following variant of SE taken from [37]

$$\tau_r(t+1) = \tau_w + \frac{N}{M}\,\mathrm{MSE}(\mathbf{r}(t),\mathbf{h}) \tag{81}$$

with

$$\mathbf{r}(t) = \mathbf{h} + \mathbf{w}_r(t) \quad \text{given} \quad \begin{cases}\mathbf{h}\sim\mathrm{CNorm}(\mathbf{0}_K,\tau_h\boldsymbol{I}_K),\\ \mathbf{w}_r(t)\sim\mathrm{CNorm}(\mathbf{0}_K,\tau_r(t)\boldsymbol{I}_K)\end{cases} \tag{82}$$

and

$$\mathrm{MSE}(\mathbf{r},\mathbf{h}) = \frac{1}{K}\mathbb{E}[\|\hat{\boldsymbol{h}}(\mathbf{r})-\mathbf{h}\|_2^2]. \tag{83}$$

TABLE III
SIMULATION PARAMETERS FOR THE BENCHMARK

Parameter = Value	Description
$I_{MC} = 10000$	# of Monte-Carlo iterations
$I_{\max} = 100$	# of HGAMP iterations
$I_{\text{samples}} = 100$	# of samples for the integration
$G = 64$	# of groups
$S = 4$	# of sensors per group
$N = 256$	# of sensors
$M = 128$	# of OFDM symbols
$K = \{1,2,4,8\}$	# of antennas at the AP
$\boldsymbol{X} = \boldsymbol{x}_n \sim \mathrm{CNorm}(\mathbf{0}_{N\times 1}, \frac{1}{M}\boldsymbol{I}_N)$	identification sequences
$(\alpha_n,\beta_n) = (0.3, 0.7)$	Beta distribution's parameters
$\rho_g = \{0.0, 0.1, \ldots, 1.0\}$	correlation within each group
$\mathrm{SNR} = 1/(M\tau_w) = \{0,4,8,12,16,20\}$ dB	signal-to-noise ratio

The main difference between (81) and (78) is that (81) accounts for scenarios with multiple antennas, as in (2). Hence, this SE (81) is the particular case of the one given in [37] when the antennas are independent.

The computation of $\mathbb{E}[\|\hat{\boldsymbol{h}}(\mathbf{r})-\mathbf{h}\|_2^2]$ is done in the following way. First, sample from the distribution of $\mathbf{r}$ and $\mathbf{h}$. Then compute the minimum mean squared error (MMSE) estimator using

$$\hat{\boldsymbol{h}}(\boldsymbol{r}_i(t)) \triangleq \frac{1}{1+\exp(-\mathsf{LLR}_n)}\left(\frac{\mu_{\mathsf{h}}\tau_{\mathsf{r}}(t)\mathbf{1}_K + \tau_{\mathsf{h}}\boldsymbol{r}_i(t)}{\tau_{\mathsf{h}}+\tau_{\mathsf{r}}(t)}\right). \tag{84}$$

Finally, average over the samples to approximate the expectation as

$$\mathbb{E}[\|\hat{\boldsymbol{h}}(\mathbf{r})-\mathbf{h}\|_2^2] \simeq \frac{1}{I_{\text{samples, SE}}}\sum_{i=1}^{I_{\text{samples, SE}}}\|\hat{\boldsymbol{h}}(\boldsymbol{r}_i)-\boldsymbol{h}_i\|_2^2. \tag{85}$$

c) GSA-HGAMP [32]: For the simulations, the activity probability of the group g will be identified to

$$q_g = \frac{1}{S_g}\sum_{n=1}^{N}\frac{\alpha_n}{\alpha_n+\beta_n} \tag{86}$$

which consists in computing the average of the ratios as a proxy for the group activity probability.

d) GHomA-HGAMP [21]: The average activity probability of sensors belonging to the same group is assumed to be the same and defined as

$$\tilde{\alpha}_g = \frac{1}{S_g}\sum_{n=1}^{N}\frac{\alpha_n}{\alpha_n+\beta_n} \qquad \text{and} \qquad \tilde{\beta}_g = 1-\tilde{\alpha}_g. \tag{87}$$

e) GHetA-HGAMP: It is the algorithm introduced in this paper. Different from GAMP, GSA-HGAMP and GHomA-HGAMP, it can deal with more general dependence and correlation structures, still introduced at the level of the activity probability that affects the state correlation as explained in Sec. III.

B. Relative performances of GHetA-HGAMP

a) Description of scenarios: The scenarios consist in simulated transmissions based on Sec. II with the GHetA model described in Sec. III. The settings for the simulations are given in Tab. III and suggests a GHetA model with underlying Gaussian correlation as described in Sec. III-C. The group

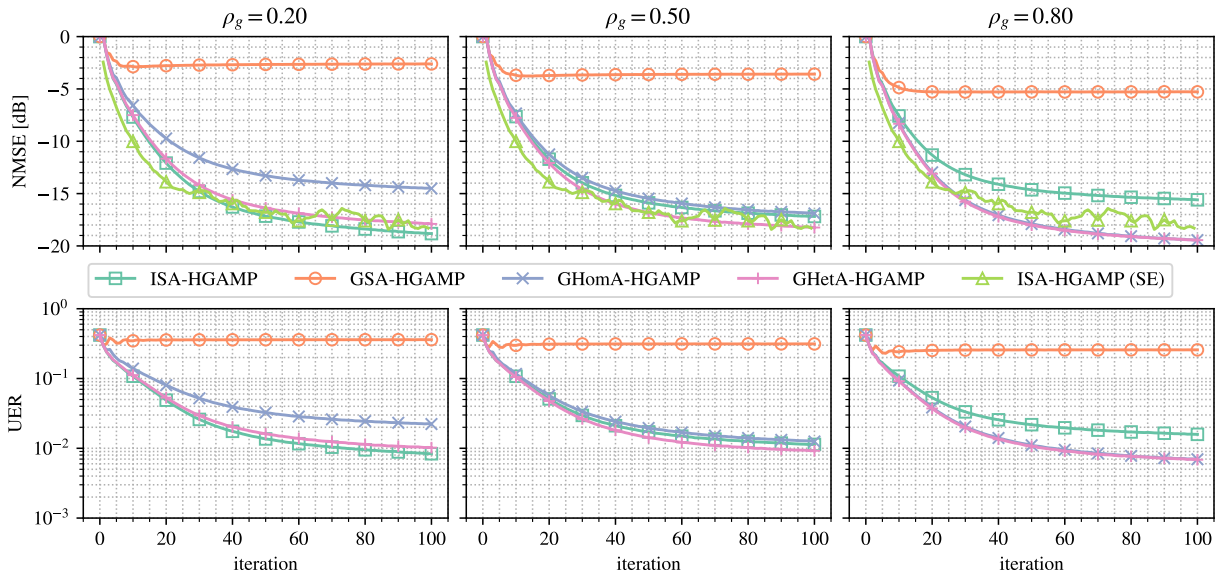

Fig. 5: NMSE and UER of the HGAMP algorithms w.r.t. the number of iterations with different correlation regimes.

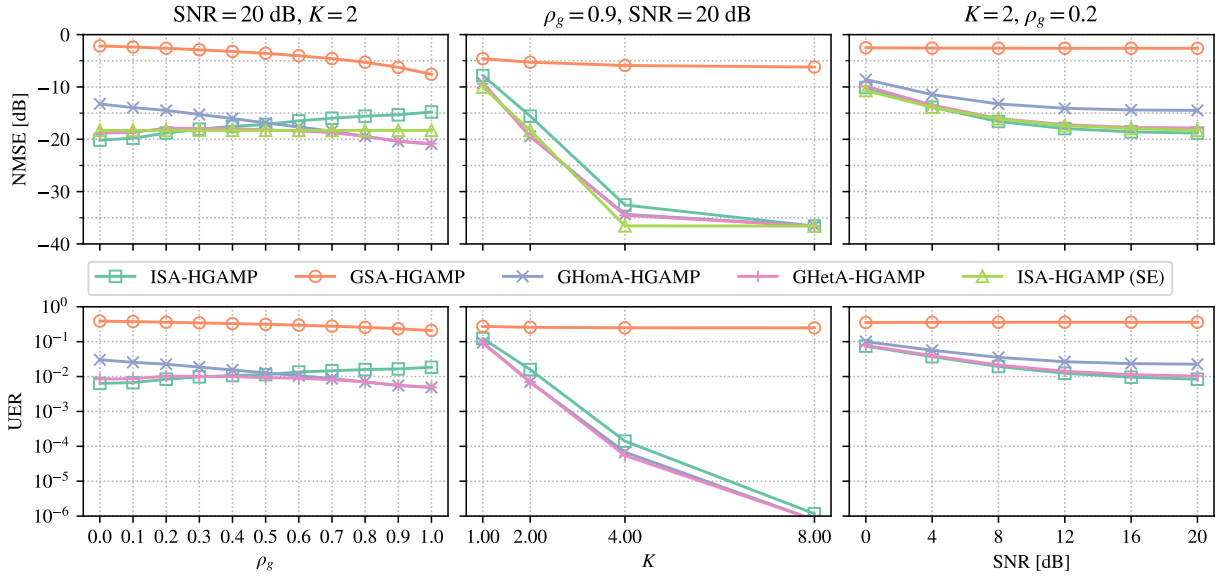

Fig. 6: NMSE and UER of the HGAMP algorithms w.r.t. the SNR, the number of antennas and the group correlation.

correlations $\{\rho_g\}_{g=1}^{G}$ are identical across the groups and will be denoted by ρ. The beta distribution parameters have all been chosen to be $(\alpha, \beta) = (0.3, 0.7)$, leading to an average activity rate per UE of 30%. Given a simulated $\mathbf{Y} = \boldsymbol{Y}$, each benchmark algorithm performs AUDaCE. This process is repeated I_{MC} times.

The goal of the simulations is to assess the gains of the new GHetA-HGAMP w.r.t. algorithms tailored for limited forms of correlated states (independent, identical, homogeneous). Henceforth, we vary the activity correlation regime through the parameter ρ from a low correlation regime ($\rho = 0$) to a high correlation regime ($\rho = 1$).

b) General comments on the results: The obtained results can be read from Figs. 5 and 6.

The Fig. 5 gives the general trend of the UER and NMSE for each benchmarked algorithms. For these two metrics, we see that in a rather low ($\rho = 0.2$) and high ($\rho = 0.8$) correlation regimes, the worst performing algorithm is GSA-HGAMP. On the opposite, GHetA-HGAMP performs the best in all regimes. In between, we respectively find the remaining GHomA-HGAMP, ISA-HGAMP and its SE.

The Fig. 6 plots the UER and NMSE against the SNR, the number of antennas K and the group correlation ρ_g for a fixed set of system parameters.

The plots on the left of Fig. 6 show the impact of the correlation on the algorithm performances. For low correlation regimes ($\rho \in [0, 0.5]$), the dominant algorithms are ISA-HGAMP and GHetA-HGAMP; for the intermediate correlation ($\rho = 0.5$), ISA-HGAMP, GHomA-HGAMP and GHetA-HGAMP have similar NMSE with less than 2 dB difference and quasi-identical UER (and hence FAR and MDR) difference; for high correlation regimes ($\rho \in [0.5, 1]$), the dom-

inant algorithms are GHomA-GAMP and GHetA-HGAMP.

The central plots of Fig. 6 illustrates the impact of the number of antennas on the performances in a very correlation regime. If there is a visible gap for $K \in \{1, 2, 4\}$ between ISA-HGAMP and the pair GHomA/GHetA-HGAMP, it vanishes at $K = 8$. This suggests that having a sufficiently high number of antennas can compensate for the lack of information on the correlation structure. It is explained by the fact that having independent antennas leverages the channel spatial diversity to reduce the uncertainty on the users' states leading to a better channel estimation.

Finally the right plots of Fig. 6 shows the classical performances evolution with the SNR. It confirms that the noise level does not affect the previous analysis regarding the relative ranking of each algorithm's performances.

c) Analysis: First, the results validate the capability of GHetA-HGAMP to perform AUDaCE for the whole range of activity correlation by achieving the best (or near the best) performances. The reasons the other algorithms cannot achieve such universality are the following.

For GSA-HGAMP, its poor performances for the simulation model comes from the fact it enforces binary group detection, i.e. that all UEs belonging to the same group will be jointly detected all active or inactive, with no possibility of intermediate decision. Such behavior results in a dramatic degradation of the FAR and MDR. For this particular scenario, each group of UEs is assumed by GSA-HGAMP to be active with probability 0.3, and so 0.7 to be inactive. This gives GSA-HGAMP more opportunities to incorrectly detect inactive users. This is confirmed by looking at the UER of Figs. 5 and 6.

For ISA-HGAMP, the algorithm only accounts for independent activity by design. Hence, when it performs AUDaCE on a received signal $\boldsymbol{Y}$ generated by independent UEs, the estimation and detection assumptions are fully met, leading to the lowest FAR, MDR, UER and, as a direct consequence, NMSE. On the other hand, when estimation and detection are performed based on a received signal issued by correlated UEs, the algorithm does not exploit this side information and still assumes the UEs to be independent. In the full correlation regime ($\rho = 1$), Fig. 6 shows that the UER of ISA-HGAMP is distant by an order of magnitude to the UER of GHomA-HGAMP and GHetA-HGAMP. It follows a severe degradation of the NMSE with a loss of $10\,\mathrm{dB}$.

In the meantime, GHomA-HGAMP suffers from the reverse problem. It was designed assuming homogeneity within groups of UEs; in other words, UEs belonging to the same group will have the exact same state distribution. Hence, for the chosen simulation model (which is based on (33) and (34)) and for $\rho > 0$, GHomA-HGAMP will treat the received signal as if it were issued with $\rho = 1$. This justifies its dominant performances alongside GHetA-HGAMP when ρ increases.

Note that if the states are independent ($\rho = 0$), GHomA-HGAMP, still perform worse than ISA-HGAMP and GHetA-HGAMP since the group structure (number of groups and group sizes) is still known to GHomA-HGAMP. Similarly to GSA-HGAMP, it enforces the group structure to be preserved while detecting users, with the difference that only their posterior activity probability will be identical, allowing for different states within the groups.

TABLE IV
SIMULATION PARAMETERS FOR THE ROBUSTNESS STUDY OF GHETA-HGAMP TO BIASED CORRELATION

Parameter = Value	**Description**
$I_{MC} = 1000$	# of Monte-Carlo iterations
$I_{\max} = 100$	# of GAMP or HGAMP iterations
$I_{\text{samples}} = 100$	# of samples for the integration
$G = 32$	# of groups
$\forall g \in [G],\ \lvert\mathbb{G}_g\rvert = 8$	# of sensors per group
$N = 256$	# of sensors
$M = 128$	# of OFDM symbols
$K = 2$	# of antennas at the AP
$\boldsymbol{X} = \boldsymbol{x}_n \sim \mathrm{CNorm}(\boldsymbol{0}_{N\times 1}, \frac{1}{M}\boldsymbol{I}_N)$	identification sequences
$(\alpha_n, \beta_n) = (0.3, 0.7)$	Beta distribution's parameters
ρ_g and $\tilde{\rho}_g = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$	model and mismatched group correlations

Considering SE for ISA-HGAMP, we must be careful in the interpretation of the results. One can see that its performances are globally close to GHetA-HGAMP, no matter the correlation regime. When the correlation is small, ISA-HGAMP and its SE follow the same trajectories as GHetA-HGAMP. This is an expected behavior since SE has been derived assuming separable, and so independent, channel priors. However when the correlation becomes high, it can be seen from the central and right plots of Fig. 5 and the left plots of Fig. 6 that SE has comparable performances to GHetA-HGAMP. This is in fact a consequence of the simulation framework. Since it is not possible to run SE on correlated data, the performance were computed on an ISA model with similar activity probility than the correlated one. The ISA model being favorable to SE, it is then normal to observe good performances for high correlation regime. In the meantime, SE is supposed to track the NMSE of ISA-glshgamp. This is the case for the low correlation regime, but not for the high one. Hence, SE fails in its NMSE prediction, demonstrating that the importance of the hypothesis of separable channel prior.

We thus see that accounting for general form of correlation structure with GHetA-HGAMP leads to significant improvements, not only in the estimation and detection performances for single scenarios, but also in the capability to flexibly and efficiently address AUDaCE with all possible correlation regimes.

C. Robustness of GHetA-HGAMP to biased correlation

a) Description of scenarios: Another interesting aspect to study is the robustness of GHetA- HGAMP to mismatched or biased correlation coefficient. With the parameters described in Tab. IV, we simulate different instances of the GHetA model for different activity correlations ρ and AUDaCE is performed with GHetA-HGAMP. However, the correlation given to the algorithm is replaced by a different one, denoted by $\tilde{\rho}$. AUDaCE is thus performed with partial knowledge of the ground truth activity structure, since the groups remain known, as well as the Beta distribution parameters.

b) General comments of the results: The obtained results can be read from Tab. V. For all the scenarios, we see that choosing the algorithm's correlation identical to the ground truth one, always allows to achieve near the best performance.

TABLE V
PERFORMANCES OF GHETA-HGAMP FOR DIFFERENT PAIRS OF TRUE AND MISMATCHED CORRELATIONS.

(A) FAR

FAR(ρ, $\tilde{\rho}$)	0.0	0.2	0.4	0.6	0.8	1.0
0.0	$\mathbf{6.97e-03}$	$7.44e-03$	$8.61e-03$	$1.11e-02$	$1.43e-02$	$2.01e-02$
0.2	$1.10e-02$	$9.04e-03$	$\mathbf{8.93e-03}$	$9.93e-03$	$1.17e-02$	$1.55e-02$
0.4	$1.48e-02$	$9.98e-03$	$\mathbf{8.60e-03}$	$8.64e-03$	$9.48e-03$	$1.14e-02$
0.6	$1.80e-02$	$1.05e-02$	$8.09e-03$	$\mathbf{7.36e-03}$	$7.54e-03$	$8.15e-03$
0.8	$2.08e-02$	$1.08e-02$	$7.11e-03$	$5.92e-03$	$\mathbf{5.58e-03}$	$5.78e-03$
1.0	$2.32e-02$	$1.02e-02$	$6.13e-03$	$4.49e-03$	$3.84e-03$	$\mathbf{3.73e-03}$

(B) MDR

MDR(ρ, $\tilde{\rho}$)	0.0	0.2	0.4	0.6	0.8	1.0
0.0	$\mathbf{2.43e-02}$	$2.49e-02$	$2.81e-02$	$3.40e-02$	$4.04e-02$	$5.20e-02$
0.2	$4.06e-02$	$3.34e-02$	$\mathbf{3.13e-02}$	$3.26e-02$	$3.59e-02$	$4.17e-02$
0.4	$5.74e-02$	$3.85e-02$	$3.24e-02$	$\mathbf{3.05e-02}$	$3.06e-02$	$3.23e-02$
0.6	$7.28e-02$	$4.39e-02$	$3.26e-02$	$2.79e-02$	$2.60e-02$	$\mathbf{2.49e-02}$
0.8	$8.67e-02$	$4.73e-02$	$3.18e-02$	$2.45e-02$	$2.11e-02$	$\mathbf{1.90e-02}$
1.0	$1.01e-01$	$4.87e-02$	$2.99e-02$	$2.14e-02$	$1.68e-02$	$\mathbf{1.39e-02}$

(C) UER

UER(ρ, $\tilde{\rho}$)	0.0	0.2	0.4	0.6	0.8	1.0
0.0	$\mathbf{1.22e-02}$	$1.27e-02$	$1.45e-02$	$1.80e-02$	$2.21e-02$	$2.97e-02$
0.2	$1.99e-02$	$1.64e-02$	$\mathbf{1.56e-02}$	$1.67e-02$	$1.89e-02$	$2.33e-02$
0.4	$2.75e-02$	$1.85e-02$	$1.57e-02$	$\mathbf{1.52e-02}$	$1.58e-02$	$1.77e-02$
0.6	$3.44e-02$	$2.05e-02$	$1.54e-02$	$1.35e-02$	$\mathbf{1.31e-02}$	$1.32e-02$
0.8	$4.06e-02$	$2.17e-02$	$1.45e-02$	$1.15e-02$	$1.02e-02$	$\mathbf{9.74e-03}$
1.0	$4.65e-02$	$2.17e-02$	$1.33e-02$	$9.53e-03$	$7.72e-03$	$\mathbf{6.76e-03}$

(D) NMSE [DB]

NMSE(ρ, $\tilde{\rho}$)	0.0	0.2	0.4	0.6	0.8	1.0
0.0	$\mathbf{-1.72e+01}$	$-1.71e+01$	$-1.65e+01$	$-1.55e+01$	$-1.45e+01$	$-1.32e+01$
0.2	$-1.47e+01$	$-1.57e+01$	$\mathbf{-1.59e+01}$	$-1.56e+01$	$-1.50e+01$	$-1.41e+01$
0.4	$-1.31e+01$	$-1.49e+01$	$-1.57e+01$	$\mathbf{-1.58e+01}$	$-1.57e+01$	$-1.52e+01$
0.6	$-1.20e+01$	$-1.43e+01$	$-1.56e+01$	$-1.63e+01$	$\mathbf{-1.64e+01}$	$-1.63e+01$
0.8	$-1.11e+01$	$-1.39e+01$	$-1.57e+01$	$-1.68e+01$	$-1.73e+01$	$\mathbf{-1.75e+01}$
1.0	$-1.04e+01$	$-1.37e+01$	$-1.59e+01$	$-1.74e+01$	$-1.83e+01$	$\mathbf{-1.88e+01}$

On the contrary, the worst performances are systematically achieved when the mismatch is the largest.

c) Analysis: The comments above show that when there is a correlated activity ($\rho > 0$), it is always desirable to have $\tilde{\rho} = \rho$ to achieve the best performance. More interestingly, when the true correlation becomes moderate ($\rho \leq 0.5$), it is also useful to set a non-zero correlation ($\tilde{\rho} \leq 0.25$), even if the value does not match the true correlation. Indeed, we see in Tab. V that when GHetA-HGAMP assumes independent UEs ($\tilde{\rho} = 0$) the obtained metrics are the worst. For instance, the minimum differences in NMSE between $\tilde{\rho} = 0$ and any other value $\tilde{\rho} > 0$ are respectively 1.8 dB, 2.3 dB, 2.8 dB and 3.3 dB for $\rho \in \{0.4, 0.6, 0.8, 1.0\}$. Since ISA-HGAMP is equivalent to GHetA-HGAMP with $\tilde{\rho} = 0$, one can conclude it is better to use GHetA-HGAMP with any $\tilde{\rho} > 0$ rather than ISA-HGAMP, even if $\tilde{\rho} \neq \rho$. This situation would arise, for instance, if only an inaccurate estimation of ρ is available at the AP.

VI. CONCLUSION

This work introduced a GHetA model to describe the activity patterns of UEs with application to IIoT networks. The proposed model leverages the copula theory enabling a powerful framework to describe general dependence structure between the activity states.

The AUDaCE problem has then been studied under the GHetA model for single-carrier NOMA. The study of the corresponding inverse problem is done within the BCS framework. We first proposed a BP algorithm which takes into account the graphical model induced by any GHetA model. The BP algorithm is approximated deriving an instance of the HGAMP algorithm, named GHetA-HGAMP. Even if the algorithm is applied in the context of IIoT, its scope of application includes any inverse problem with correlated signal sources. The numerical comparison with state-of-the-art BCS algorithms for AUDaCE has demonstrated the capacity of GHetA-HGAMP to leverage the dependence structure of the activity states of the UEs with significant improvement in the metrics of interest (NMSE, FAR, MDR, UER). It has also been shown that knowing the true correlation level improves the estimation and detection capabilities. Generally, we showed that for a wide range of parameter combinations (number of antennas, SNR, correlation), GHetA-HGAMP is a suitable choice to perform AUDaCE.

For scenarios when this information is not available, the estimation of the copula dependence structure remains an open problem. Also, understanding and deriving state evolution for GHetA-HGAMP seems to be a quite difficult problem, due to the non-separable and non-identically distrituted channel coefficients.

APPENDIX A
STATE CORRELATION WITH LATENT VARIABLES

Consider the random system

$$\mathbf{q} \in [0,1]^N \to \mathbf{s} \in \mathbb{F}^N \tag{88}$$

where $\forall n \in [N]$, $\mathsf{s}_n \mid \mathsf{q}_n = q_n \sim \mathrm{Bern}(q_n)$. Then, for $(n, n') \in [N]^2$ and $i \in \{n, n'\}$, the following equalities hold for the expectation

$$\mathbb{E}[\mathsf{s}_i] = \mathbb{E}_{\mathsf{q}_i}[\mathbb{E}_{\mathsf{s}_i}[\mathsf{s}_i \mid \mathsf{q}_i]] = \mathbb{E}[\mathsf{q}_i], \tag{89}$$

the variance

$$\mathbb{V}[\mathsf{s}_i] = \mathbb{E}_{\mathsf{q}_i}[\mathbb{E}_{\mathsf{s}_i}[\mathsf{s}_i^2 \mid \mathsf{q}_i]] - \mathbb{E}[\mathsf{s}_i]^2 = \mathbb{E}[\mathsf{q}_i](1 - \mathbb{E}[\mathsf{q}_i]), \tag{90}$$

and the expected cross-product

$$\begin{aligned}\mathbb{E}[\mathsf{s}_n \mathsf{s}_{n'}] &= \mathbb{E}_{\mathsf{q}_n, \mathsf{q}_{n'}}[\mathbb{E}[\mathsf{s}_n \mathsf{s}_{n'} \mid \mathsf{q}_n, \mathsf{q}_{n'}]] &(91)\\ &= \mathbb{E}_{\mathsf{q}_n, \mathsf{q}_{n'}}[\mathbb{E}_{\mathsf{s}_n}[\mathsf{s}_n \mid \mathsf{q}_n]\mathbb{E}_{\mathsf{s}_{n'}}[\mathsf{s}_{n'} \mid \mathsf{q}_{n'}]] &(92)\\ &= \mathbb{E}[\mathsf{q}_n \mathsf{q}_{n'}]. &(93)\end{aligned}$$

Together, they lead to

$$\begin{aligned}\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] &= \frac{\mathbb{E}[\mathsf{s}_n \mathsf{s}_{n'}] - \mathbb{E}[\mathsf{s}_n]\mathbb{E}[\mathsf{s}_{n'}]}{\sqrt{\mathbb{V}[\mathsf{s}_n]\mathbb{V}[\mathsf{s}_{n'}]}} &(94)\\ &= \frac{\mathbb{E}[\mathsf{q}_n \mathsf{q}_{n'}] - \mathbb{E}[\mathsf{q}_n]\mathbb{E}[\mathsf{q}_{n'}]}{\sqrt{\mathbb{E}[\mathsf{q}_n](1 - \mathbb{E}[\mathsf{q}_n])\mathbb{E}[\mathsf{q}_{n'}](1 - \mathbb{E}[\mathsf{q}_{n'}])}} &(95)\end{aligned}$$

which is also equal to

$$\mathrm{Cor}[\mathsf{s}_n, \mathsf{s}_{n'}] = \frac{\mathrm{Cor}[\mathsf{q}_n, \mathsf{q}_{n'}]\sqrt{\mathbb{V}[\mathsf{q}_n]\mathbb{V}[\mathsf{q}_{n'}]}}{\sqrt{\mathbb{E}[\mathsf{q}_n](1 - \mathbb{E}[\mathsf{q}_n])\mathbb{E}[\mathsf{q}_{n'}](1 - \mathbb{E}[\mathsf{q}_{n'}])}}. \tag{96}$$

APPENDIX B
STATE CORRELATION WITH GROUP-HOMOGENEOUS ACTIVITY

Let $g \in [G]$, $(n, n') \in \mathbb{G}_g^2$ and $i \in \{n, n'\}$. Given (89), (90) and (92) and the statistical properties of the Beta distribution, we have

$$\mathbb{E}[\mathsf{s}_i] = \frac{\alpha_g}{\alpha_g + \beta_g}, \tag{97}$$

$$\mathbb{V}[\mathsf{s}_i] = \frac{\alpha_g \beta_g}{(\alpha_g + \beta_g)^2}, \tag{98}$$

$$\mathbb{E}[\mathsf{s}_n \mathsf{s}_{n'}] = \frac{\alpha_g \beta_g}{(\alpha_g + \beta_g)^2} \frac{1}{1 + \alpha_g + \beta_g} + \left(\frac{\alpha_g}{\alpha_g + \beta_g}\right)^2. \tag{99}$$

Substituting (97), (98) and (99) in (95) gives (12) in Sec. II-B3.

APPENDIX C
STATE CORRELATION WITH GROUP-HETEROGENEOUS ACTIVITY

Let $(n, n') \in [N]^2$ and $i \in \{n, n'\}$. Given (89), (90) and (92) and the statistical properties of the Beta distribution, we have

$$\mathbb{E}[\mathsf{s}_i] = \frac{\alpha_i}{\alpha_i + \beta_i}, \tag{100}$$

$$\mathbb{V}[\mathsf{s}_i] = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2} \tag{101}$$

and where $\mathbb{E}[\mathsf{s}_n \mathsf{s}_{n'}]$ depends on the choice of $f_{\mathbf{q}}$. Substituting (100), (101) and (92) in (95) gives (22) in Sec. III-A2.

REFERENCES

[1] L. Chetot, M. Egan, and J.-M. Gorce, "Active user detection and channel estimation for grant-free random access with gaussian correlated activity," in *2023 IEEE 97th Vehicular Technology Conference (VTC)*, IEEE, 2023, pp. 1–6.

[2] M. Erdelj, N. Mitton, and E. Natalizio, "Applications of industrial wireless sensor networks," in *Industrial Wireless Sensor Networks: Applications, Protocols, and Standards*, Boca Raton, FL: CRC Press, 2013, pp. 1–22.

[3] I. Rodriguez, R. S. Mogensen, A. Fink, *et al.*, "An Experimental Framework for 5G Wireless System Integration into Industry 4.0 Applications," *Energies*, vol. 14, no. 15, p. 4444, 15 Jan. 2021, ISSN: 1996-1073, URL: https://doi.org/10.3390/en14154444.

[4] 3GPP, "TR 21.916: Release 16," 3GPP, Sep. 14, 2021, p. 157, URL: https://www.3gpp.org/ftp/Specs/archive/21_series/21.916/.

[5] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 3, pp. 1805–1838, thirdquarter 2020, ISSN: 1553-877X, URL: https://doi.org/10.1109/COMST.2020.2996032.

[6] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, Jan. 1996, ISSN: 00359246, URL: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J Royal Statistical Soc B*, vol. 68, no. 1, pp. 49–67, Feb. 2006, ISSN: 1369-7412, 1467-9868, URL: https://doi.org/10.1111/j.1467-9868.2005.00532.x.

[8] S. Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec./1993, ISSN: 1053587X, URL: https://doi.org/10.1109/78.258082.

[9] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov. 1993, 40–44 vol.1, URL: https://doi.org/10.1109/ACSSC.1993.342465.

[10] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse Solution of Underdetermined Systems of Linear Equations by Stagewise Orthogonal Matching Pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012, ISSN: 1557-9654, URL: https://doi.org/10.1109/TIT.2011.2173241.

[11] Suhyuk Kwon, Jian Wang, and Byonghyo Shim, "Multipath Matching Pursuit," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014, ISSN: 0018-9448, 1557-9654, URL: https://doi.org/10.1109/TIT.2014.2310482.

[12] P. Schniter, L. C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *2008 Information Theory and Applications Workshop*, San Diego, CA, USA: IEEE, Jan. 2008, pp. 326–333, ISBN: 978-1-4244-2670-6, URL: https://doi.org/10.1109/ITA.2008.4601068.

[13] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004, ISSN: 0010-3640, 1097-0312, URL: https://doi.org/10.1002/cpa.20042.

[14] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009, ISSN: 10635203, URL: https://doi.org/10.1016/j.acha.2009.04.002.

[15] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009, ISSN: 1936-4954, URL: https://doi.org/10.1137/080716542.

[16] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018, ISSN: 1558-0792, URL: https://doi.org/10.1109/MSP.2018.2844952.

[17] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive Sensing-Based Adaptive Active User Detection and Channel Estimation: Massive Access Meets Massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020, ISSN: 1941-0476, URL: https://doi.org/10.1109/TSP.2020.2967175.

[18] Q. Zou, H. Zhang, D. Cai, and H. Yang, "Message Passing Based Joint Channel and User Activity Estimation for Uplink Grant-Free Massive MIMO Systems With Low-Precision ADCs," *IEEE Signal Process. Lett.*, vol. 27, pp. 506–510, 2020, ISSN: 1558-2361, URL: https://doi.org/10.1109/LSP.2020.2979534.

[19] Q. Zou, H. Zhang, D. Cai, and H. Yang, "A Low-Complexity Joint User Activity, Channel and Data Estimation for Grant-Free Massive MIMO Systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 1290–1294, 2020, ISSN: 1558-2361, URL: https://doi.org/10.1109/LSP.2020.3008550.

[20] R. Ayachi, M. Akrout, V. Shyianov, F. Bellili, and A. Mezghani, "Massive Unsourced Random Access Based on Bilinear Vector Approximate Message Passing," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 5283–5287, URL: https://doi.org/10.1109/ICASSP43922.2022.9747338.

[21] L. Chetot, M. Egan, and J.-M. Gorce, "Joint Identification and Channel Estimation for Fault Detection in Industrial IoT with Correlated Sensors," *IEEE Access*, pp. 1–1, 2021, ISSN: 2169-3536, URL: https://doi.org/10.1109/ACCESS.2021.3106736.

[22] A. E. Kalor, O. A. Hanna, and P. Popovski, "Random Access Schemes in Wireless Systems with Correlated User Activity," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5, URL: https://doi.org/10.1109/SPAWC.2018.8445866.

[23] C. Zheng, M. Egan, L. Clavier, A. E. Kalør, and P. Popovski, "Stochastic Resource Optimization of Random Access for Transmitters With Correlated Activation," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3055–3059, Sep. 2021, ISSN: 1558-2558, URL: https://doi.org/10.1109/LCOMM.2021.3090110.

[24] A. Jeannerot, M. Egan, L. Chetot, and J.-M. Gorce, "Mitigating user identification errors in resource optimization for grant-free random access," in *IEEE VTC2023-Sping-97th Vehicular Technology Conference*, 2023.

[25] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. John Wiley & Sons, Oct. 22, 2004, 312 pp., ISBN: 978-0-470-86345-9.

[26] R. B. Sinitsyn and F. J. Yanovsky, "MIMO radar copula ambiguity function," in *2012 9th European Radar Conference*, Oct. 2012, pp. 146–149.

[27] N. Deligiannis, J. F. C. Mota, E. Zimos, and M. R. D. Rodrigues, "Heterogeneous Networked Data Recovery From Compressive Measurements Using a Copula Prior," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5333–5347, Dec. 2017, ISSN: 1558-0857, URL: https://doi.org/10.1109/TCOMM.2017.2746099.

[28] C. Zheng, M. Egan, L. Clavier, G. W. Peters, and J.-M. Gorce, "Copula-Based Interference Models for IoT Wireless Networks," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6, URL: https://doi.org/10.1109/ICC.2019.8761783.

[29] B. Dai, S. Ding, and G. Wahba, "Multivariate Bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, Sep. 2013, ISSN: 1350-7265, URL: https://doi.org/10.3150/12-BEJSP10.

[30] 3GPP, "TR 37.824: Coexistence between NB-IoT and NR," 3GPP, Jun. 2020, p. 46, URL: https://www.3gpp.org/ftp/Specs/archive/37_series/37.824/.

[31] M. Chafii, F. Bader, and J. Palicot, "SC-FDMA with index modulation for M2M and IoT uplink applications," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Barcelona: IEEE, Apr. 2018, pp. 1–5, ISBN: 978-1-5386-1734-2, URL: https://doi.org/10.1109/WCNC.2018.8377028.

[32] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne, and P. Schniter, "Hybrid Approximate Message Passing," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4577–4592, Sep. 2017, ISSN: 1941-0476, URL: https://doi.org/10.1109/TSP.2017.2713759.

[33] M. Sklar, *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8, 1959, 3 pp.

[34] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," ser. 2011 IEEE International Symposium on Information Theory Proceedings, 2011-07, 2011, pp. 2168–2172, URL: https://doi.org/10.1109/ISIT.2011.6033942.

[35] Q. Zou, H. Zhang, C.-K. Wen, S. Jin, and R. Yu, "Concise Derivation for Generalized Approximate Message Passing Using Expectation Propagation," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1835–1839, Dec. 2018, ISSN: 1558-2361, URL: https://doi.org/10.1109/LSP.2018.2876806.

[36] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: II. analysis and validation," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan. 2010, pp. 1–5, URL: https://doi.org/10.1109/ITWKSPS.2010.5503228.

[37] Q. Zou and H. Zhang, "High-Dimensional Multiple-Measurement-Vector Problem: Mutual Information and Message Passing Solution," *IEEE Transactions on Signal Processing*, vol. 71, pp. 2819–2832, 2023, ISSN: 1941-0476, URL: https://doi.org/10.1109/TSP.2023.3301137.

Lélio Chetot received in 2018 and 2022 the engineering and PhD degrees from the Institut National des Sciences Appliquées (INSA), Lyon, France. He is currently an assistant professor in INSA Lyon, member of the team "Models and Algorithms for Reliable Communication Systems" at INRIA, hosted by the CITI laboratory. His research interests are in the areas of statistical signal processing, Bayesian inference and communication theory for wireless networks.

Malcolm Egan received the Ph.D. in Electrical Engineering in 2014 from the University of Sydney, Australia. He is currently a Chargé de Recherche (Tenured Research Scientist) in Inria and a member of CITI, a joint laboratory between Inria, INSA Lyon and Université de Lyon, France. Previously he was an assistant professor in INSA Lyon, and a postdoctoral researcher with the Laboratoire de Mathématiques, Université Blaise Pascal, France and the Department of Computer Science, Czech Technical University in Prague, Czech Republic. He has also held visiting positions at Princeton University and the University of Bristol. From 2018-2022 (full term) he was an associate editor for IEEE Communications Letters. His research interests are in the areas of information theory, statistical signal processing and machine learning with applications in communications.

Jean-Marie Gorce (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the Institut National des Sciences Appliquées (INSA), Lyon, France, in 1993 and 1998, respectively. He was a Co-Founder of the Centre for Innovation, Telecommunications and Integration of Services (CITI Lab), in 2001. He was a Visiting Scholar with Princeton University, Princeton, NJ, USA, from 2013 to 2014. He has been the Principal Investigator of several French and European sponsored projects related to wireless networks. He is a Professor with INSA de Lyon and the Head for Science with INRIA Lyon. He is currently the Scientific Coordinator for the Experimental Facility FIT-CorteXlab. He has co-published more than 200 conference and journal articles. His research focuses on multi-user communicating systems, with approaches combining information theory, coding, algorithms, signal processing and implementation.