Active User Detection and Channel Estimation for Grant-Free Random Access with Gaussian Correlated Activity

Lélio Chetot, Malcolm Egan, Jean-Marie Gorce

Active User Detection and Channel Estimation for Grant-Free Random Access with Gaussian Correlated Activity

Lélio CHETOT
lelio.chetot@insa-lyon.fr
Univ Lyon, Inria, INSA Lyon, CITI, EA3720, Villeurbanne (69621), France

Malcolm EGAN
malcom.egan@inria.fr
Univ Lyon, Inria, INSA Lyon, CITI, EA3720, Villeurbanne (69621), France

Jean-Marie GORCE
jean-marie.gorce@insa-lyon.fr
Univ Lyon, Inria, INSA Lyon, CITI, EA3720, Villeurbanne (69621), France

***Abstract*—Industrial IoT (IIoT) is one of the major verticals targeted by the next generations of wireless networks. In order to provide industrial plants with features relying on wireless communications, the grant-free RA (GFRA) protocol appears to be a promising means for supporting massive ultra-reliable connectivity; at the same time, it is a critical bottleneck that requires an access point (AP) to be able to jointly perform active user detection and channel estimation (AUDaCE) to fulfill its main mission of allowing industrial wireless devices to access the core network. This mission is even harder when the GFRA requests are correlated because of event-driven activity triggers. This paper proposes a new tractable gaussian correlated activity (GCA) model for this scenario. The corresponding AUDaCE problem is then studied in the Bayesian compressed sensing (BCS) framework. An hybrid instance of the generalized AMP (GAMP) algorithm is derived and its capability to perform AUDaCE is numerically assessed by extensive Monte-Carlo simulations. The numerical results show gains of 2.5dB in channel estimation gain for twice less detection errors w.r.t. state-of-the-art algorithms.**

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

I. INTRODUCTION

One of the major target verticals of sixth generation (6G) is IIoT consisting of a large, densely populated collection of wireless sensors interfaced via wireless links with cloud-based computing systems. IIoT systems are characterized by their large scale, strict reliability and latency constraints. An open challenge is to develop for the facility sensors transmission strategies dealing with all these requirements. Since the sensors cannot stay in a connected mode with dedicated resources because of their number, each time they transmit, they must complete a random access (RA) procedure. To this end, GFRA was introduced in 5G [1], [2], to reduce the latency and control overhead. In order for the AP to reliably acknowledge the RA requests, estimating the channel state in the uplink based on the RA signaling simultaneously to the active user detection is also required. Performing AUDaCE is then at the center of a successful GFRA procedure.

A. Related work

As suggested in [3], [4], leveraging BCS-non-orthogonal multiple access (NOMA) appears to be relevant for GFRA in IIoT by exploiting the sparse sensor activities.

BCS relies on message-passing algorithms based on belief propagation (BP), expectation propagation (EP) or approximate message passing (AMP). In the case of independent sensors activity, a number of methods have been developed [5]–[10]. In [5], the simplest version of AMP has been used to perform AUDaCE. In [11], a GAMP algorithm was suggested, empowered by expectation maximization (EM) to estimate hyper parameters with an ad-hoc active user detection. In [7], [8], hybrid GAMP (HGAMP) is used to perform AUDaCE with low-precision analog-to-digital converter, incorporating independent activity prior information for the active user detection. In [10], AUDaCE is accomplished using bilinear vector AMP in the context of massive unsourced RA.

The compressed sensing (CS) algorithms in [5], [7], [8], [10], [11] rely on the assumption that the random sensor activities are mutually independent. However, when sensors monitor the same machines, events may be detected by multiple sensors, leading to *correlated activity*. While CS algorithms for independent activity may be applied, this can lead to significant performance degradations [12]. To the best of our knowledge, the only work addressing AUDaCE for sensors with correlated activity is [12]. However, correlation in activity was restricted to be large and homogeneous among groups of users.

B. Main Contributions

Our main contributions are as follows:

- **Model of GCA:** we propose a new tractable statistical model for user activity exploiting the correlated Gaussian distribution. Existing activity models either assume independence (e.g. [7]) or high levels of correlation ([13], [14]). In contrast, the GCA model can support heterogeneous correlation between devices, ranging from independence to high correlation.
- **GCA-HGAMP algorithm:** with each user activity model, a new AUDaCE algorithm is required. In the BCS framework, we first develop a BP algorithm to solve the AUDaCE problem under the GCA model. We then approximate BP by exploiting the principles of HGAMP.

Algorithm 1 Sampling from the GCA model.

INPUT:
  Correlation matrix $\boldsymbol{K} \in [-1,1]^N$.
  Parameters of the Beta distributions $\{\alpha_n, \beta_n\}_{n\in[N]}$.
END
Sample $\boldsymbol{c} \sim \mathsf{Norm}(\boldsymbol{0}_{N\times 1}, \boldsymbol{K})$.
Compute the vector $\boldsymbol{u} = [F_{\mathsf{c}_n}(c_n)]^{\mathsf{T}}_{n\in[N]}$.
Compute the vector $\boldsymbol{q} = [F^{-1}_{\mathsf{q}_n}(u_n;\alpha_n,\beta_n)]^{\mathsf{T}}_{n\in[N]}$.
Sample $\boldsymbol{s} = [s_n]^{\mathsf{T}}$ where $\forall n \in [N],\ s_n \sim \mathsf{Bern}(q_n)$.
RETURN: Desired sample $\boldsymbol{s}$.

- **Numerical Evaluation:** we validate our GCA-HGAMP algorithm via a comprehensive numerical study. We show that our GCA-HGAMP algorithm either outperforms or does not suffer from performances losses in all correlation regimes that have been previously considered by state-of-the-art algorithms with an average gain of 2.5dB in the channel estimation error and twice less detection errors.

C. 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). The integer set $[N]$ is a shorthand notation for $\{1,\dots,N\}$ (resp. $\emptyset$) if $N \geq 1$ (resp. $N < 1$).

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 $\mathsf{Norm}(\boldsymbol{M}, \boldsymbol{C})$ and its pdf is $\mathcal{N}(\cdot;\boldsymbol{M},\boldsymbol{C})$ (resp. $\mathsf{CNorm}(\boldsymbol{M},\boldsymbol{C})$ and $\mathcal{CN}(\cdot;\boldsymbol{M},\boldsymbol{C})$).

II. System Model for Gaussian Correlated Activity

A. Transmission of the idenfication sequences

We consider a single-carrier orthogonal frequency division multiplexing (OFDM) communication system with N single-antenna sensors and 1 K-antenna AP. The digital baseband signal received at the AP is

$$\mathbf{Y} = \boldsymbol{X}\mathbf{H}(\mathbf{s}) + \mathbf{W} \tag{1}$$

where $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_N\end{bmatrix} \in \mathbb{C}^{M\times N}$ is the known matrix of the N concatenated sensors identification sequences spanning M OFDM symbols, $\mathbf{s} \in \{0,1\}^N$ is the random vector of the N sensor activities, $\mathbf{H}(\mathbf{s}) = [\mathsf{h}_{nk}(\mathsf{s}_n)]_{(n,k)\in[N,K]}$ is the random matrix of the channel coefficients between the N sensors and the K AP's antennas, $\mathbf{W} = [\mathsf{w}_{mk}]_{(m,k)\in[M,K]}$ is the random matrix of the additive noise coefficients and $\mathbf{Y} = [\mathsf{y}_{mk}]_{(m,k)\in[M,K]}$ is the random matrix of the output signal coefficients.

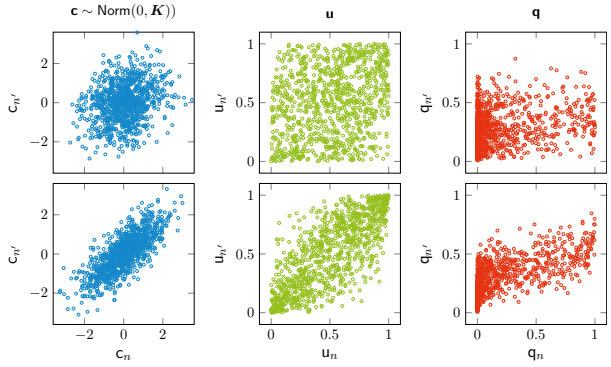


Fig. 1: Scatter plots of the vectors $\mathbf{c}$, $\mathbf{u}$ and $\mathbf{q}$ based on (7). The correlation matrix is $\boldsymbol{K} = \rho\boldsymbol{1}_{2\times 2} + (1-\rho)\boldsymbol{I}_2$ with $\rho = 0.25$ for the top row and $\rho = 0.75$ for the bottom row. The parameters of the Beta distributions are $(\alpha_n, \alpha_{n'}, \beta_n, \beta_{n'}) = (0.2, 2, 0.8, 5)$.

For $(n,m,k) \in [N]\times[M]\times[K]$, $\mu_h \in \mathbb{C}$ and $(\tau_w, \tau_h) \in \mathbb{R}^2_{+,*}$, we assume the noise coefficients to be i.i.d. as

$$\mathsf{w}_{mk} \sim \mathsf{CNorm}(0, \tau_w) \tag{2}$$

and the channel coefficients to be i.i.d. as

$$\mathsf{h}_{nk}(\mathsf{s}_n) \mid \mathsf{s}_n = s \sim \begin{cases}\mathsf{Dirac}(0) & \text{if } s = 0\\ \mathsf{CNorm}(\mu_h, \tau_h) & \text{if } s = 1\end{cases}. \tag{3}$$

Note that the channel coefficients are only assumed to have conditional independence given we know all the sensors activities $\mathbf{s} = \boldsymbol{s}$.

B. Gaussian Correlated Activity Model

For $n \in [N]$, denote by $\mathsf{q}_n \in [0,1]$ the random activity probability of the sensor state s_n such that

$$\begin{cases}\mathsf{q}_n & \sim \mathsf{Beta}(\alpha_n, \beta_n)\\ \mathsf{s}_n \mid \mathsf{q}_n = q_n & \sim \mathsf{Bern}(q_n)\end{cases} \tag{4}$$

for $(\alpha_n, \beta_n) \in \mathbb{R}^2_{+*}$. The motivation for q_n to be Beta distributed is similar to [12]: this distribution describes a wide range of probability profiles with only two parameters, as suggested by its pdf

$$f_{\mathsf{q}_n}(q) = \frac{q^{\alpha_n - 1}(1-q)^{\beta_n - 1}}{\mathrm{B}(\alpha_n, \beta_n)} \tag{5}$$

where $\mathrm{B}(\alpha,\beta)$ denotes the beta function evaluated at (α,β). In order to model the state correlation, we seek to impose a dependence structure at the level of the activity probability random vector $\mathbf{q}$, targeting Beta marginal distributions.

Let $\mathbf{c} \sim \mathsf{Norm}(\boldsymbol{0}_{N\times 1}, \boldsymbol{K})$ be a correlated Gaussian random vector parameterized by $\boldsymbol{K} \in [-1,1]^{N,N}$, a, potentially degenerated, correlation matrix. Assume that

$$\forall n \in [N],\ q_n = T_n(c_n) \quad \text{with} \quad T_n = F^{-1}_{\mathsf{q}_n} \circ F_{\mathsf{c}_n} \tag{6}$$

where F_{q_n} and F_{c_n} respectively denote the marginal cdfs of q_n and c_n. Since $\boldsymbol{K}$ is a correlation matrix, the entries of $\mathbf{c}$

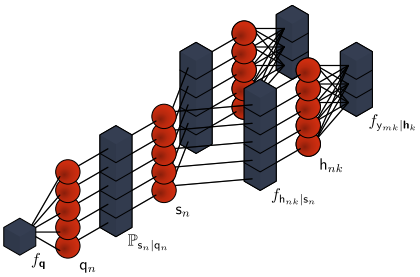


Fig. 2: Factor graph of (19).

TABLE I
BELIEF PROPAGATION MESSAGES CORRESPONDING TO FIG. 2.

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

identically follow the standard normal distribution $\mathsf{Norm}(0,1)$. Hence

$$\forall n \in [N],\ \begin{cases} \mathsf{u}_n &= F_{\mathsf{c}_n}(\mathsf{c}_n) \sim \mathsf{Unif}([0,1]) \\ \mathsf{q}_n &= F^{-1}_{\mathsf{q}_n}(\mathsf{u}_n) \sim \mathsf{Beta}(\alpha_n, \beta_n) \end{cases} \tag{7}$$

or, in a more compact form,

$$\mathbf{q} = \boldsymbol{T}(\mathbf{c}) = [T_n(c_n)]^{\mathsf{T}}_{n\in[N]} \tag{8}$$

The element-wise transform $\boldsymbol{T}$ conserves the targeted Beta marginals of $\mathbf{q}$ and transports the dependence structure from $\mathbf{c}$ to $\mathbf{q}$.

The interest of this model w.r.t. the one of [12] is its greater flexibility. First, the parameters of the beta distributions are not shared anymore between subgroups of activity probabilities. Second, the correlation matrix $\boldsymbol{K}$ adds freedom in capturing custom pairwise state dependence structures. Together, they provide a tractable model for GCA in sensor networks at the expanse of meaningful additional parameters.

For instance, a system where the N sensors are split into G non-overlapping groups, with S_g sensors in group $g \in [G]$ with Gaussian correlation $\rho_g \in [0,1]$, can be described by a correlation matrix of the form

$$\boldsymbol{K} = \mathrm{diag}(\boldsymbol{K}_1, \dots, \boldsymbol{K}_G) \tag{9}$$

with

$$\forall g \in [G],\ \boldsymbol{K}_g = (1-\rho_g)\boldsymbol{I}_{S_g} + \rho_g \mathbf{1}_{S_g\times S_g}. \tag{10}$$

If the sensor n has Beta parameters $\alpha_n = 1-\beta_n$, its average activity probability would be

$$\mathbb{E}[\mathsf{q}_n] = \frac{\alpha_n}{\alpha_n+\beta_n} = \alpha_n. \tag{11}$$

III. Active User Detection and Channel Estimation

A. Problem formulation

Based on the model of Sec. II, the AUDaCE problem consists in jointly estimating an activity pattern $\mathbf{s} = \boldsymbol{s}^\star$ and the corresponding channel matrix $\mathbf{H}(\mathbf{s}) = \boldsymbol{H}^\star$. To do so, we want to compute estimates $\hat{\boldsymbol{s}} \approx \boldsymbol{s}^\star$ and $\hat{\boldsymbol{H}} \approx \boldsymbol{H}^\star$ given the AP observed the output signal $\mathbf{Y} = \boldsymbol{Y}$ with

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}\in\{0,1\}^N}{\arg\max}\, f_{\mathbf{s}|\mathbf{Y}}(\boldsymbol{s}\mid \boldsymbol{Y}), \quad \text{(Active User Detection)} \tag{12}$$

$$\hat{\boldsymbol{H}} = \mathbb{E}[\mathbf{H}\mid \mathbf{Y} = \boldsymbol{Y}]. \quad \text{(Channel Estimation)} \tag{13}$$

The estimators require the computation of the following marginal posteriors

$$\mathbb{P}_{\mathbf{s}|\mathbf{Y}}(\boldsymbol{s}\mid \boldsymbol{Y}) = \int_{\mathbb{C}^{M\times N}} f_{\mathbf{s},\mathbf{H}(\mathbf{s})|\mathbf{Y}}(\boldsymbol{s},\boldsymbol{H}\mid \boldsymbol{Y})\mathrm{d}\boldsymbol{H} \tag{14}$$

$$f_{\mathbf{H}|\mathbf{Y}}(\boldsymbol{H}\mid \boldsymbol{Y}) = \sum_{\boldsymbol{s}\in\{0,1\}^N} f_{\mathbf{s},\mathbf{H}(\mathbf{s})|\mathbf{Y}}(\boldsymbol{s},\boldsymbol{H}\mid \boldsymbol{Y}) \tag{15}$$

where each marginalization involves the posterior density

$$f_{\mathbf{s},\mathbf{H}(\mathbf{s})|\mathbf{Y}}(\boldsymbol{s},\boldsymbol{H}\mid \boldsymbol{Y}) = \int_{[0,1]^N} f_{\mathbf{q},\mathbf{s},\mathbf{H}(\mathbf{s})|\mathbf{Y}}(\boldsymbol{q},\boldsymbol{s},\boldsymbol{H}\mid \boldsymbol{Y})\mathrm{d}\boldsymbol{q} \tag{16}$$

$$= \frac{f_{\mathbf{q},\mathbf{s},\mathbf{H}(\mathbf{s}),\mathbf{Y}}(\boldsymbol{q},\boldsymbol{s},\boldsymbol{H},\boldsymbol{Y})}{\displaystyle\int_{[0,1]^N} \sum_{\boldsymbol{s}\in\{0,1\}^N} \int_{\mathbb{C}^{M\times N}} f_{\mathbf{q},\mathbf{s},\mathbf{H}(\mathbf{s}),\mathbf{Y}}(\boldsymbol{q},\boldsymbol{s},\boldsymbol{H},\boldsymbol{Y})\mathrm{d}\boldsymbol{H}\mathrm{d}\boldsymbol{q}}. \tag{17}$$

Since the system variables form the following Markov chain

$$\mathbf{q} \to \mathbf{s} \to \mathbf{H} \to \mathbf{Y} \tag{18}$$

after Bayes' theorem, the full joint density which appears in the numerator and denominator of (17) factorizes as

$$f_{\mathbf{q},\mathbf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{q},\boldsymbol{s},\boldsymbol{H},\mathbf{Y}) = f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y}\mid \boldsymbol{H}) f_{\mathbf{H}|\mathbf{s}}(\boldsymbol{H}\mid \boldsymbol{s}) \mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s}\mid \boldsymbol{q}) f_{\mathbf{q}}(\boldsymbol{q}) \tag{19}$$

with

$$f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y}\mid \boldsymbol{H}) = \prod_{(m,k)\in[M]\times[K]} f_{\mathsf{y}_{mk}|\mathbf{h}_k}(y_{mk}\mid \boldsymbol{h}_k), \tag{20}$$

$$f_{\mathbf{H}|\mathbf{s}}(\boldsymbol{H}\mid \boldsymbol{s}) = \prod_{(n,k)\in[N]\times[K]} f_{\mathsf{h}_{nk}|\mathsf{s}_n}(h_{nk}\mid s_n), \tag{21}$$

$$\mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s}\mid \boldsymbol{q}) = \prod_{n\in[N]} \mathbb{P}_{\mathsf{s}_n|\mathsf{q}_n}(s_n\mid q_n). \tag{22}$$

Note that a factorization of $f_{\mathbf{q}}(\boldsymbol{q})$ is not provided since $\mathbf{q}$ is assumed to be correlated without any specific structure other

Algorithm 2 GCA-HGAMP

Description: GCA-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.

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

than (8) in the general case. The factor graph corresponding to such a factorization is depicted in Fig. 2.

From (17), it is clear that computing (14) and (15) is intractable because of the high-dimensional sum and integrals. Hence, we will first resort to the BP algorithm to approximate the posterior distributions (14) and (15) and then derive a HGAMP algorithm to make the approximations and estimations tractable.

B. HGAMP for AUDACE

The derivation of the HGAMP algorithm to address the AUDaCE problem needs first the construction of a BP algorithm based on the factor graph in Fig. 2. The corresponding BP messages are given in Tab. I and the corresponding beliefs are

$$\forall n \in [N],\ \mathfrak{B}_{\mathsf{s}_n} = \underset{\mathbb{P}_{\mathsf{s}_n|\mathsf{q}_n} \to \mathsf{s}_n}{\mathfrak{M}} \underset{\mathbb{P}_{\mathsf{s}_n|\mathsf{q}_n} \leftarrow \mathsf{s}_n}{\mathfrak{M}} \tag{23}$$

$$\forall (n,k) \in [N] \times [K],\ \mathfrak{B}_{\mathsf{h}_{nk}} = \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \to \mathsf{h}_{nk}}{\mathfrak{M}} \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}} \tag{24}$$

We then build the following posterior approximates as

$$\mathbb{P}_{\mathsf{s}|\mathsf{Y}}(\boldsymbol{s} \mid \boldsymbol{Y}) \approx \prod_{n \in [N]} \mathfrak{B}_{\mathsf{s}_n}(s_n) \tag{25}$$

$$f_{\mathsf{H}|\mathsf{Y}}(\boldsymbol{H} \mid \boldsymbol{Y}) \approx \prod_{(n,k) \in [N] \times [K]} \mathfrak{B}_{\mathsf{h}_{nk}}(h_{nk}) \tag{26}$$

Substituting these approximates in (12) and (13) allows to perform AUDaCE using

$$\forall n \in [N],\ \hat{s}_n \approx \underset{s \in \{0,1\}}{\arg\max}\, \mathfrak{B}_{\mathsf{s}_n}(s) \tag{27}$$

$$\forall (n,k) \in [N] \times [K],\ \hat{h}_{nk} \approx \hat{s}_n \int_{\mathbb{C}} h \mathfrak{B}_{\mathsf{h}_{nk}}(h) \mathrm{d}h \tag{28}$$

However, for $(n,k) \in [N] \times [K]$, the computation of the message

$$\underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}} = \prod_{m=1}^{M} \underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}_{nk}}{\mathfrak{M}} \tag{29}$$

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.

Following the works in [13], [15], [16], 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 (29) can be made

$$\underset{f_{\mathsf{h}_{nk}|\mathsf{s}_n} \leftarrow \mathsf{h}}{\mathfrak{M}}(h) \approx \mathcal{CN}(h; \hat{r}_{nk}, \hat{\tau}_{\mathsf{r},nk}). \tag{30}$$

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 GCA model, namely GCA-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. Also, note that the updates at line 23 and 27 in Algo. 2 require numerical integration. It is handled using Monte-Carlo integration with I_{samples} samples $\{\boldsymbol{q}_i\}_{i \in [I_{\text{samples}}]}$ drawn with Algo. 1. As a whole, we can show that the algorithm complexity is

$$C_{\text{GCA}} = O(NMK + NK + NI_{\text{samples}}) \tag{31}$$

which is comparable to the state-of-the-art algorithm (see [13] and [12]).

IV. Numerical results

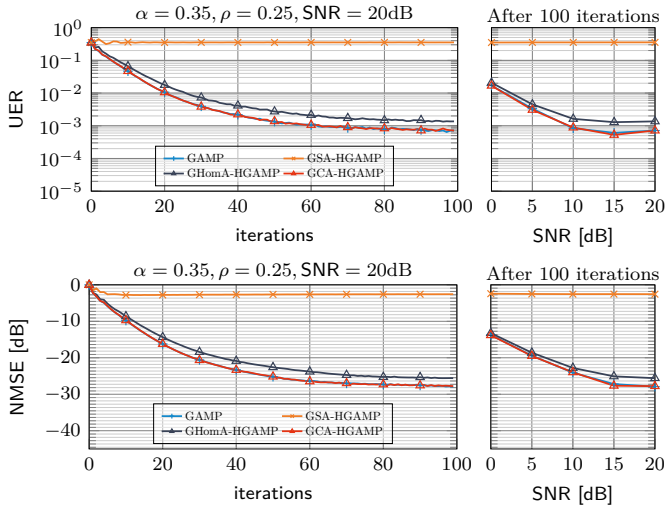
The performances of the new GCA-HGAMP are numerically studied based on extensive Monte-Carlo simulations. For a collection of I_{MC} ground truth values $\{(\boldsymbol{s}^\star_i, \boldsymbol{H}^\star_i)\}_{i \in [I_{\text{MC}}]}$ and their respective estimates $\{(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{H}}_i)\}_{i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

$$\mathsf{UER}(\{\boldsymbol{s}^\star_i, \hat{\boldsymbol{s}}_i\}_{i \in [I_{\text{MC}}]}) = \frac{1}{NI_{\text{MC}}} \sum_{i \in [I_{\text{MC}}]} (N - (\boldsymbol{s}^\star_i)^{\mathsf{T}} \hat{\boldsymbol{s}}_i) \tag{32}$$

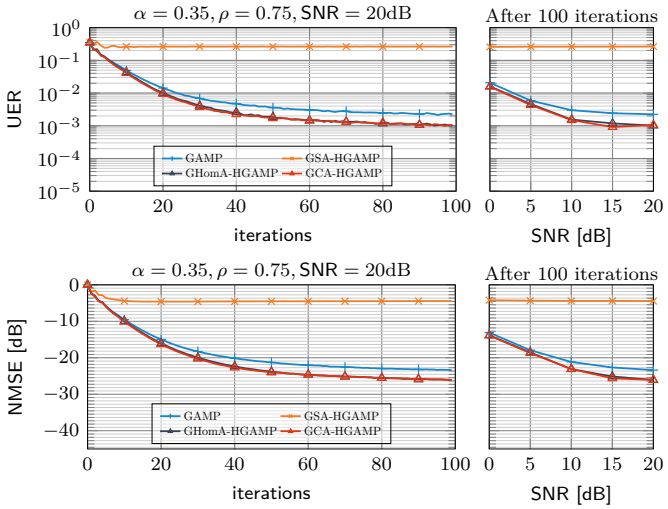
and the quality of its channel estimation is assessed by the NMSE

$$\mathsf{NMSE}(\{\boldsymbol{H}^\star_i, \hat{\boldsymbol{H}}_i\}_{i \in [I_{\text{MC}}]}) = \frac{1}{I_{\text{MC}}} \sum_{i \in [I_{\text{MC}}]} \frac{\|\boldsymbol{H}^\star_i - \hat{\boldsymbol{H}}_i\|_2^2}{\|\boldsymbol{H}^\star_i\|_2^2}. \tag{33}$$

In order to compare the performances of our GCA-HGAMP, we also provide the performances for the following algorithms.



(a) Low correlation regime.



(b) High correlation regime.

Fig. 3: Performances in terms of user error rate (UER) and normalized mean squared error (NMSE) (see (32) and (33)).

A. Benchmark

a) Modified GAMP (based on [15]): This modified GAMP corresponds to the algorithm originally described in [15] but with a slight modification to allow activity detection. To do so, we have considered, as in [11], a Bernoulli-Gaussian prior for the channel coefficients so that

$$f_{\mathsf{H}}(\boldsymbol{H}) = \prod_{(n,k)\in[N]\times[K]} \Big((1-q_n)\delta(h_{nk}) + q_n \mathcal{CN}(h_{nk}; \mu_{\mathsf{h}}, \tau_{\mathsf{h}}) \Big) \quad (34)$$

where

$$\forall n \in [N],\ q_n = \alpha_n/(\alpha_n + \beta_n). \quad (35)$$

The log-likelihood ratios for this GAMP become

$$\mathsf{LLR}_n = \log \frac{\alpha_n}{\alpha_n + \beta_n} + \log \phi_{0,n} - \log \frac{\beta_n}{\alpha_n + \beta_n} - \log \phi_{1,n} \quad (36)$$

that will be used for activity detection. This modified GAMP does not account for any activity correlation.

b) Group-sparse activity (GSA)-HGAMP [13]: For the simulations, the activity probability of the group g will be identified to

$$q_g = \frac{1}{S_g} \sum_{n\in\mathbb{G}_g} \frac{\alpha_n}{\alpha_n + \beta_n} \quad (37)$$

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

c) Group-homogeneous activity (GHomA)-HGAMP [12]: The average activity probability of sensors belonging to the same group is assumed to be the same and defined as

$$\tilde{\alpha}_g = \frac{1}{S_g} \sum_{n\in\mathbb{G}_g} \frac{\alpha_n}{\alpha_n + \beta_n} \quad \text{and} \quad \tilde{\beta}_g = 1 - \tilde{\alpha}_g. \quad (38)$$

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
$G = 64$	# of groups
$\forall g \in [G],\ \lvert\mathbb{G}_g\rvert = 4$	# of sensors per group
$N = 256$	# of sensors
$M = 128$	# of OFDM symbols
$K = 2$	# of antennas at the AP
$\boldsymbol{X} = [\tilde{\boldsymbol{x}}_n/\lVert\tilde{\boldsymbol{x}}_n\rVert_2]_{n\in[N]},\ \tilde{\boldsymbol{x}}_n \sim \mathsf{CNorm}(\boldsymbol{0}_{N\times 1}, \boldsymbol{I}_N)$	identification sequences
$\forall n \in [N],\ (\alpha_n, \beta_n) = (0.35, 0.65)$	Beta distribution's parameters
$\forall g \in [G],\ \rho_g = \{0.25, 0.75\}$	correlation within each group

TABLE II
SIMULATION PARAMETERS FOR THE COMPARATIVE STUDY OF GCA AGAINST MODIFIED GAMP, GS-HGAMP AND GHOMA-HGAMP

d) GCA-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. II-B.

B. Relative performances of GCA to the benchmark

The numerical study focuses on the relative performances of GCA-HGAMP against modified GAMP, GSA-HGAMP and GHomA-HGAMP. The simulation parameters are given in Tab. II and the corresponding NMSE and UER have been drawn in Fig. 3. Note that the simulation scenarios ($\rho = 0.25$ and $\rho = 0.75$) are tailored to GCA-HGAMP. Hence, the main goal of the simulations is to assess the gains of the proposed GCA-HGAMP algorithm w.r.t. to the benchmark algorithms that do not account for the more general Gaussian correlated activity.

C. Discussion of the results

The important indicator are the correlation coefficients $\{\rho_g\}_{g\in[G]}$ that fully describe the correlation matrix $\boldsymbol{K}$ of the underlying Gaussian random vector $\mathbf{c}$. For a particular group

$g \in [G]$, when ρ_g is close to 0 the matrix is close to $\boldsymbol{I}_{S_g}$ and when ρ_g is close to 1 the matrix is close to $\mathbf{1}_{S_g \times S_g}$, making $\boldsymbol{K}$ block diagonal. Hence, two correlation regimes, low $\rho_g = 0.25$ and high $\rho_g = 0.75$, have been considered to assess the performances, for which we give a review hereafter.

In all the correlation regimes, GSA-HGAMP always performs the worst and GCA-HGAMP always performs the best.

In the low correlation regime, the activity of the sensors is close to be independent ($\rho = 0$) so that the modified GAMP algorithm is as good as GCA-HGAMP over the whole SNR range. GAMP and GCA reduce the NMSE by ~ 2.5dB for the channel estimation w.r.t. GHomA (cf. Fig. 3a) below . The reason is that GHomA-HGAMP still relies on the underlying group structure of the sensors, which is responsible for a higher number of false alarms and missed detections for GHomA-HGAMP, leading to an overall twice worse UER (see Fig. 3a). Indeed, when the activity correlation is low, the activity behavior of sensors within the same group is very unlikely to be the same.

In the high correlation regime, the trend observed in the low correlation regime is reversed. Both GHomA and GCA-HGAMP have the best performance, reducing the NMSE by ~ 2.5dB w.r.t. the modified GAMP algorithm in Fig. 3b; the UER is again halved in favor of GHomA and GCA-HGAMP. The reason is that GAMP cannot leverage the correlation structure which is favorable to similar activity probabilities within each group of sensors to improve its detection capability.

An additional remark is that GCA-HGAMP is able to perform the best in all scenarios where GHomA-HGAMP and GAMP only perform best when the correlation regime respectively approaches the high and low correlation ones, which are their reference scenarios. As a whole, the results show the ability of GCA-HGAMP to adapt to various correlation structures with significant gains in terms of NMSE and UER.

V. Conclusion

This paper addresses the problem of AUDaCE in the context of massive GFRA where the activity of transmitting devices is correlated. A tractable GCA model is proposed leveraging the correlated distribution of the device activity probabilities. Based on this new model, the AUDaCE problem is formulated within the BCS framework. A GCA-HGAMP algorithm with polynomial complexity is then derived to approximate the optimal solutions of the AUDaCE problem. Finally, numerical simulations show significant gains in terms of UER and NMSE for the proposed algorithm to account for more general correlation against state-of-the-art algorithms that account for a limited form of correlated activity, or even no correlation at all. For future works, more general correlated activity models could be considered using the framework of the copula theory.

References

[1] 3GPP, "TR 21.915: Release 15," 3GPP, Oct. 1, 2019, p. 118, URL: https://www.3gpp.org/ftp/Specs/archive/21_series/21.915/.

[2] 3GPP, "TR 21.916: Release 16," 3GPP, Sep. 14, 2021, p. 157, URL: https://www.3gpp.org/ftp/Specs/archive/21_series/21.916/.

[3] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 3, pp. 2294–2323, 2018, ISSN: 1553-877X, URL: https://doi.org/10.1109/COMST.2018.2835558.

[4] 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.

[5] 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.

[6] K. He, Y. Li, C. Yin, and Y. Zhang, "A novel compressed sensing-based non-orthogonal multiple access scheme for massive MTC in 5G systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 81, Apr. 19, 2018, ISSN: 1687-1499, URL: https://doi.org/10.1186/s13638-018-1079-4.

[7] 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.

[8] 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.

[9] Y. Mei, Z. Gao, Y. Wu, *et al.*, "Compressive Sensing Based Joint Activity and Data Detection for Grant-Free Massive IoT Access," *IEEE Trans. Wirel. Commun.*, pp. 1–1, 2021, ISSN: 1558-2248, URL: https://doi.org/10.1109/TWC.2021.3107576.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.