IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 29, 2021, accepted August 11, 2021, date of publication August 23, 2021, date of current version August 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106736

Joint Identification and Channel Estimation for Fault Detection in Industrial IoT With Correlated Sensors

LÉLIO CHETOT, MALCOLM EGAN, AND JEAN-MARIE GORCE, (Senior Member, IEEE)
CITI Laboratory, EA3720, Institut National de Recherche en Informatique et en Automatique (INRIA), Institut National des Sciences Appliquées (INSA) Lyon, Université de Lyon, 69621 Villeurbanne, France

Corresponding authors: Lélio Chetot (lelio.chetot@inria.fr), Malcolm Egan (malcom.egan@inria.fr), and Jean-Marie Gorce (jean-marie.gorce@inria.fr).

This work was supported by INRIA Nokia Bell Labs ADR "Network Information Theory."

ABSTRACT As industrial plants increase the number of wirelessly connected sensors for fault detection, a key problem is to identify and obtain data from the sensors. Due to the large number of sensors, random access protocols exploiting non-orthogonal multiple access (NOMA) are a natural approach. In this paper, we develop new algorithms based on approximate message passing for sensor identification and channel estimation accounting for correlation in the activity probability of each sensor and observations of physical variables (e.g., temperature) by the access point. These algorithms form the basis for data decoding, while also identifying faulty machines and estimating local values of the temperature, which can lead to a reduction in the amount of data required to be transmitted. Numerical results show that for an expected activity probability of 0.35, our algorithms improve the normalized mean-square error of the channel estimate by up to 5dB and reduce the rate of sensor identification errors by a factor of four.

INDEX TERMS Maximum likelihood detection, channel estimation, fault detection, correlation, Internet of Things, belief propagation, approximation algorithms, Bayesian methods, message passing, massive machine-type communications.

I. INTRODUCTION

To improve efficiency and minimize economic losses, a key challenge in large-scale industrial plants is rapid identification of faulty or degraded machines. An increasingly popular approach is to place sensors with wireless communication capability on as many machines as is economically feasible [1]. When a centralized access point sends a sync signal, all sensors which detect faulty behavior of their machine then transmit data to the access point.

An important challenge lies in the fact that each sensor only transmits when it detects a fault. For applications with a relatively small number of sensors, it would be reasonable to allocate distinct subcarriers to each sensor, making the identification trivial since detecting a signal in a subcarrier would indicate which sensor is active. However, it will not be the case when the—potentially very large—number of sensors that transmit within a given frame can vary dramatically. It is therefore highly inefficient to allocate orthogonal resources for each sensor's transmission especially when sensors are inactive with a high probability. Doing so would lead to high resource requirements, most of which is not utilized. As a consequence, this scenario—known as random access—requires non-orthogonal resource allocation (e.g., the same time-slots or subcarriers are utilized by multiple sensors), often known as non-orthogonal multiple access (NOMA) [2].

Another, less often accounted for, feature is that sensors on the same or nearby machines often observe physical variables, such as the temperature, that are similar. In particular, the observations by such sensors are statistically correlated [3].

In order to transmit data, each sensor first transmits a pilot sequence which provides a means of identifying active sensors (and hence faulty machines) and estimating channels between the sensors and the access point. There is presently a large body of work on sensor identification and channel estimation in random access for *uncorrelated* sensors. For example, a Bayesian estimation framework has been developed for the user identification problem in [4], which attempts

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

to identify the active subset of sensors based on the posterior distribution informed by observations at the access point.

Due to the nature of random access, active sensors typically form a sparse subset of all sensors. As a consequence, the problem of channel estimation has recently been attacked using compressed sensing, with algorithms based on least absolute shrinkage and selection operator (LASSO) and orthogonal matching pursuit [5]–[7].

The problem of *joint* sensor identification and channel estimation in random access based on NOMA has recently seen significant attention. In [8], channel estimation is performed via a low complexity, but accurate, variant of belief propagation (BP) known as generalised approximate message passing (GAMP), with sensor identification obtained through an *ad hoc* thresholding scheme. In [9], a GAMP-based sensor identification and channel estimation algorithm has been proposed for multi carrier communication systems, with activity probabilities estimated via expectation-maximization. In [10] and [11] expectation propagation is used to address the joint active user detection and channel estimation problem in a multi-user setup with inter-symbol interference introduced by a faster-than-Nyquist signaling. While the correlation between the symbols is considered, the activity of the users is assumed to be statistically independent. In [12]–[14], a systematic approach for the joint identification and channel estimation problem exploited a general framework, known as the group-sparse model, where sensor activity—common to all subcarriers—is treated as a latent variable. A GAMP-type method, known as hybrid GAMP (HGAMP), was then applied by exploiting the group-sparse hybrid GAMP (GS-HGAMP) algorithm in [15] tailored for the group-sparse model.

This existing work on joint identification and channel estimation has largely focused on generic random access systems. For sensor networks tailored to industrial fault detection, the probability a fault occurs is also often dependent on physical variables, such as the temperature, for which a noisy estimate may also be obtained at the access point. Indeed, when the temperature deviates from standard operating levels, the probability of a machine faults or degradation can increase. Such a scenario arises in the context of semiconductor manufacturing [16]. As the access point can locally measure the ambient temperature, this provides additional information, which can potentially improve the performance of algorithms for sensor identification and channel estimation.

In this paper, we develop algorithms for sensor identification and channel estimation in narrowband communication systems in presence of fault probability, which depend on physical variables (such as the temperature, which we will focus on in the remainder of the paper) and may be statistically correlated. The first step is to introduce a statistical model relating observations at the access point (i.e., the ambient temperature and received signal) to the channel, activity of each sensor, and the probability each machine is faulty.

Based on our new model, we derive an identification and channel estimation algorithm by exploiting GAMP. In particular, the model falls into the framework of hybrid GAMP [15]. The algorithm is obtained by developing a loopy BP (LBP) algorithm for the model, and then applying GAMP for the variables associated with the communication channel.

A key feature of the algorithm is that it explicitly accounts for uncertainty and correlation in the probability sensors are active, as opposed to existing approaches where the activity probability is fixed and sensor transmissions are uncorrelated. In addition, our model accounts for the impact of physical variables (such as temperature) on the probability of a fault. We model the probability of a fault conditioned on temperature observations at the access point via the beta distribution, a highly flexible family of models. As such, we call the algorithm β-HGAMP.

Numerical results demonstrate that β-HGAMP outperforms existing algorithms based on GAMP [17] and GS-HGAMP [13], [15]. In particular, β-HGAMP outperforms these approaches by up to 5dB in terms of the normalized mean-square error (NMSE) for channel estimation and for sensor identification, the user error rate (UER) is approximately four times lower when the expected activity probability is 0.35. Finally, unlike existing GAMP and GS-HGAMP algorithms, β-HGAMP yields a posteriori estimates of activity probabilities. These estimates provide insight into the physical variables (e.g., temperature) at each machine that may be useful to detect degradation without including measurements within the data transmission.

A. MAIN CONTRIBUTIONS

The main contributions in this work are summarized as follows:

1) We develop a new framework for wireless sensor networks for fault detection, which incorporates knowledge of physical variables (e.g., temperature).
2) We introduce a new statistical model for device activity that allows for correlated sensors, which generalizes the group-sparse model and allows for features such as error-prone sensors.
3) We develop an algorithm within the HGAMP framework to estimate the channel and identify active sensors.
4) We show via Monte Carlo simulations that significant performance improvements can be obtained over existing algorithms in terms of NMSE and UER.

B. ORGANIZATION

This paper is organized as follows. In Sec. II, we introduce the system model. We then derive in Sec. III new loopy belief propagation and HGAMP algorithms tailored to our model. We assess and discuss the performance of the new β-HGAMP in Sec. IV. Finally, we conclude and provide insights for future work in Sec. V.

Algorithm 1: Transmission Protocol

1 **Step** *0* (Downlink): Sync signal sent by the access point to indicate the beginning of a frame.
2 **Step** *1* (Local at Devices): The sensor on each machine detects whether or not the machine is faulty.
3 **Step** *2* (Uplink): Sensors on faulty machines transmit pilot sequences.
4 **Step** *3* (Local at Access Point): The access point locally measures the room temperature, then performs identification of active sensors and estimates of channel coefficients.
5 **Step** *4* (Uplink): Sensors on faulty machines transmit data.

C. NOTATION

Deterministic scalars are denoted by x or X whereas deterministic vectors and matrices are respectively denoted by $\boldsymbol{x}$ and $\boldsymbol{X}$. Random scalars are denoted by x or X whereas random vectors and matrices are respectively denoted by $\mathbf{x}$ and $\mathbf{X}$. The transpose operator is denoted by $(\cdot)^\mathsf{T}$. Identity, 1-only and 0-only matrices are respectively denoted by $\boldsymbol{I}_N$, $\mathbf{1}_{M\times N}$ and $\mathbf{0}_{M\times N}$. The Bernoulli, beta and complex gaussian distributions are denoted by $\mathscr{B}(\cdot)$, Beta$(\cdot,\cdot)$ and $\mathscr{CN}(\cdot,\cdot)$, respectively. The probability density and mass functions of a random variable x are denoted by $f_\mathsf{x}(\cdot)$ and $\mathbb{P}_\mathsf{x}(\cdot)$. The expectation and variance are denoted by $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$. For $N \in \mathbb{N}$, $[N]$ denotes the set $\{1,\dots,N\}$. The symbol $\propto$ indicates equality up to a multiplicative constant.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider an industrial plant consisting of B machines, each monitored via S sensors equipped with a single antenna and utilizing a single subcarrier. We denote by $[B]$ the index set of machines and by $N = BS$ the total number of sensors. For $b \in [B]$, $\mathbb{M}_b$ denotes the index subset $\{(b-1)S+1,\dots,bS\} \subset [N]$ of the sensors observing the b-th machine and $\bigcup_{b=1}^{B}\mathbb{M}_b = [N]$. The sensors seek to transmit information about the state of the machines they are watching, e.g. faulty behavior, to an access point equipped with K antennas.

A. COMMUNICATION PROTOCOL AND RECEIVED SIGNAL

A typical communication protocol between the sensors and the access point is given in algorithm 1. In **Step 0**, the access point broadcasts a sync signal to all sensors, which indicates the beginning of a frame. At this time, each sensor detects whether or not the corresponding machine is faulty as detailed in **Step 1**.

In **Step 2**, each active sensor n transmits its pilot sequence $\boldsymbol{a}_n \in \mathbb{C}^K$. In **Step 3**, the access point observes the signal

$$\mathbf{Y} = \boldsymbol{A}\mathbf{X} + \mathbf{W}, \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{N\times K}$ is the channel random matrix, $\boldsymbol{A} = [\boldsymbol{a}_1,\dots,\boldsymbol{a}_N] \in \mathbb{C}^{M\times N}$ is the matrix with all the pilot sequences with norm $\sqrt{M}$ and $\mathbf{W} \in \mathbb{C}^{M\times K}$ is the noise random matrix. The entries of $\mathbf{W}$ are assumed to be i.i.d. $\mathscr{CN}(0,\tau_w)$.

TABLE 1. List of the system variables.

Variable	Description
	Dimensions
B	Number of independent machines.
S	Number of sensors per machine.
N	Total number of sensors.
M	Pilot length.
K	Number of receiving antennas at the AP.
	System variables
$\boldsymbol{A}$	Pilot matrix.
T_b	Temperature of machine b.
T_0	Temperature at the AP.
$\mathbf{q}$	Vector of sensor activity probabilities.
$\mathbf{s}$	Vector of sensor states.
$\mathbf{X}$	Channel matrix.
$\mathbf{W}$	Noise matrix.
$\mathbf{Y}$	Received signal matrix.
	System Parameters
$\alpha_b(T_0), \beta_b(T_0)$	Beta distribution parameters of machine b.
μ_0, τ_0	Channel mean and variance.
τ_w	Noise variance.

B. TEMPERATURE-DEPENDENT ACTIVITY PROBABILITY AND CORRELATION

The activation of the sensors is indicated by the random vector $\mathbf{s} = [\mathsf{s}_1,\dots,\mathsf{s}_N]^\mathsf{T} \in \{0,1\}^N$, where $\mathsf{s}_n = 0$ and $\mathsf{s}_n = 1$ correspond to the cases where sensor n is inactive and active, respectively. Given the sensor n is active, we assume that the fading coefficients x_{nk} between this sensor and the k-th access point's antenna is Gaussian distributed with mean μ_0 and variance τ_0. When a sensor is inactive, the fading coefficient is set to zero with probability one. Formally, we then have

$$\forall (n,k) \in [N]\times[K], \begin{cases} \mathsf{x}_{nk}|\mathsf{s}_n = 0 & \sim \text{Dirac}(0) \\ \mathsf{x}_{nk}|\mathsf{s}_n = 1 & \sim \mathscr{CN}(\mu_0,\tau_0)\,. \end{cases} \tag{2}$$

A common assumption is that the probability a machine is faulty, and hence the corresponding sensors are active, is constant [13]. However, many industrial processes are *temperature dependent*; that is, the probability of a fault depends on the temperature of the machine. This arises, for example, in semiconductor manufacturing [16]. Other physical variables, such as pressure or light intensity, may also have a similar impact on faults. For the purpose of exposition, we focus in the remainder of the paper on temperature although our framework can also be applied to these alternative variables.

Let $\mathsf{T}_b \in]-273.15, +\infty[$ be the temperature of machine $b \in [B]$ in degrees Celsius. We assume that probability that machine b is faulty, denoted by q_b, is defined by the function $g_b : \mathsf{T}_b \mapsto \mathsf{q}_b$. Hence, each sensor on this machine will share the same activity probability, inducing correlation in the activity of sensors on the same machine. In practice, the relationship between T_b and q_b is obtained via empirical tests at the design phase of machine b. As a consequence,

$$\forall n \in \mathbb{M}_b, \quad \mathsf{s}_n|\mathsf{T}_b \sim \mathscr{B}(q_b)\,. \tag{3}$$

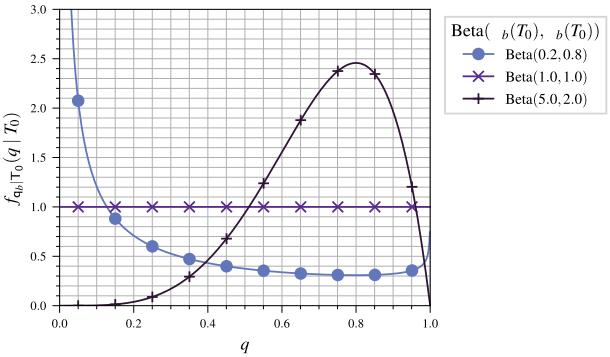


FIGURE 1. Examples of probability density functions for the Beta distribution.

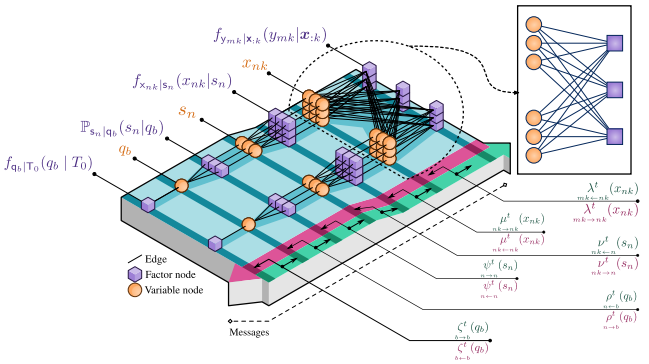


FIGURE 2. Factor graph induced by the joint density (12) with $B, S, N, M, K = (2, 3, 6, 3, 3)$.

where $\mathbb{M}_b = \{(b-1)S+1, \ldots, b\,S\}$ and q_b is a realization of $g_b(T_b)$.

On the other hand, the access point does not have direct access to measurements of the temperature within machine b. Instead, as detailed in **Step 3**, the access point can only measure the room temperature locally, denoted by T_0. The (possibly random) temperature, T_0, measured at the access point and the temperature, T_b, at machine b are not in general the same, but are statistically dependent.

Observe that $\mathsf{q}_b = g_b(\mathsf{T}_b) \in [0, 1]$, which means that we require a general statistical model with support on $[0, 1]$ for the conditional distribution of q_b. A good candidate is

$$\mathsf{q}_b|\mathsf{T}_0 \sim \text{Beta}(\alpha_b(\mathsf{T}_0), \beta_b(\mathsf{T}_0)), \tag{4}$$

corresponding to a beta distribution with parameters $\alpha_b(\mathsf{T}_0), \beta_b(\mathsf{T}_0)$, which depend on the temperature T_0 observed by the access point. Indeed, the beta distribution is both tractable and highly flexible model, including both the uniform distribution and cases where the probability mass is concentrated at zero and one (see Fig. 1).

In particular, the probability mass function of the beta distribution with parameters $\alpha_b(\mathsf{T}_0), \beta_b(\mathsf{T}_0)$ is given by

$$f_{\mathsf{q}_b|\mathsf{T}_0}(q \mid T_0) = \frac{q^{\alpha_b(T_0)-1}(1-q)^{\beta_b(T_0)-1}}{\text{B}(\alpha_b(T_0), \beta_b(T_0))}, \tag{5}$$

where

$$\forall(\alpha, \beta) \in]0, \quad +\infty[^2, \ \text{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{6}$$

and $\Gamma(\cdot)$ is the Gamma function. Note that the precise functional form of $\alpha_b(\mathsf{T}_0), \beta_b(\mathsf{T}_0)$ depends on the function g_b and the statistical dependence between T_0 and T_b, which is established during the design phase of the plant.

C. RELATED WORK ON GROUP SPARSE MODELING

Recent models for the joint sensor detection and channel estimation assume that the activity of each sensor on a machine to be the same or the activity of each sensor is independent. In the former case, when a fault occurs on a machine, the sensors monitoring it all turn on and transmit their identification sequence. Hence, an activity variable s_b describing the state of the group of sensors can be used instead of considering an activity variable s_n per sensor (see [15]).

This *machine-based* group sparse model can be seen as a special case of our model. Choosing the density

$$f_{\mathsf{q}_b|\mathsf{T}_0}(q \mid T_0) = p_b(T_0)\delta(q) + (1 - p_b(T_0))\delta(q-1) \tag{7}$$

leads to the activity probabilities $\{0, 1\}$ where the weight $p_b(T_0) \in [0, 1]$ depends on the temperature. As a consequence, this models the group sparsity since all sensors in a group would be either all active with probability $p_b(T_0)$ or inactive with probability $1 - p_b(T_0)$, without any possibility for intermediate states (i.e., sensors on the same machine that do not agree on whether a fault has occured). In the rest of this paper, it will be assumed that

$$p_b(T_0) = \frac{\alpha_b(T_0)}{\alpha_b(T_0) + \beta_b(T_0)} \tag{8}$$

which corresponds to the expected value of q_b when it is beta distributed according to (5).

However, such a machine-based model is limiting when sensors on the same machine may not all be active at the same time, with correlated activity probabilities. A simplification of the group sparse model, that we call *sensor-based*, is to ignore the groups i.e. assuming that the size of the groups is $S = 1$ such that no distinction is made between the groups and the sensors. Again, we will consider the activity probability of each sensor to be given by (8). This sensor-based model then allows to treat each sensor's activity independently from the other sensors' activity, as in [13].

In summary, the machine-based and sensor-based models are not able to capture the correlation in the activity probability of the sensors. Moreover, these models do not capture the uncertainty arising from the dependence on the temperature.

D. PROBLEM STATEMENT

The focus of the remainder of this paper is to develop algorithms to identify active sensors and estimate channel coefficients in **Step 3** accounting for the local temperature observations at the access point. Aside from learning which machines are faulty, active sensor identification and channel

estimation forms the basis for data decoding in **Step 4**; e.g. via successive interference cancellation.

Formally, we are interested in solving the minimum mean squared error (MMSE) estimation problem

$$\begin{aligned} \boldsymbol{X}^\star &= \underset{\boldsymbol{X}\in\mathbb{C}^{N\times K}}{\arg\min}\, \mathbb{E}\Big[\|\boldsymbol{X}-\mathbf{X}\|_2^2 \mid \mathbf{Y}=\boldsymbol{Y}, \mathsf{T}_0=T_0\Big] \\ &= \mathbb{E}\big[\mathbf{X} \mid \mathbf{Y}=\boldsymbol{Y}, \mathsf{T}_0=T_0\big] \\ &= \int_{\mathbb{C}^{N\times K}} \boldsymbol{X} f_{\mathbf{X}|\mathbf{Y},\mathsf{T}_0}(\boldsymbol{X} \mid \boldsymbol{Y}, T_0)\mathrm{d}\boldsymbol{X} \end{aligned} \tag{9}$$

where $\boldsymbol{Y}$ and T_0 are observed. At the same time, we seek to obtain the posterior distributions $\mathbb{P}_{\mathsf{s}_n|\mathbf{Y},\mathsf{T}_0}(s \mid \boldsymbol{Y}, T_0)$ and $f_{\mathsf{q}_b|\mathbf{Y},\mathsf{T}_0}(q \mid \boldsymbol{Y}, T_0)$. This provides a basis to also obtain the estimates

$$\forall b \in [B],\ q_b^\star = \mathbb{E}\big[\mathsf{q}_b \mid \mathbf{Y}=\boldsymbol{Y}, \mathsf{T}_0=T_0\big] \tag{10a}$$

$$\forall n \in [N],\ s_n^\star = \underset{s\in\{0,1\}}{\arg\max}\, \mathbb{P}_{\mathsf{s}_n|\mathbf{Y},\mathsf{T}_0}(s \mid \boldsymbol{Y}, T_0) \tag{10b}$$

However, each of these three estimators are intractable to compute so it is necessary to resort to algorithms that approximate the optimal solutions. The algorithm for channel estimation and user identification is detailed in Sec. III.

III. DEVICE IDENTIFICATION AND CHANNEL ESTIMATION ALGORITHMS

In this section, we derive algorithms for the sensor identification (10b) and channel estimation (9) problems. We first start by writing the loopy belief propagation (LBP) algorithm for the system model derived in Sec. II. Despite few works granting the convergence for general models [18], [19], this approach has been successfully and widely used in graphical inference problems. We then develop an algorithm within the HGAMP framework called β-HGAMP from LBP to reduce the complexity of the algorithm with limited loss in performance.

A. LOOPY BELIEF PROPAGATION APPROACH

LBP is an algorithm which aims at solving inference problems in a systematic fashion by exploiting the structure of the joint density of the system variables [20]. The idea is to decompose (factorize) this joint density into multiple subfactors, each depending on a subset of the system variables. It is then possible to build a graphical representation of the dependency between the factors and the variables, namely a *factor graph*, where the graph's nodes are split into *factor nodes* and *variable nodes*. An edge between a variable node and a factor node exists if and only if the corresponding variable is an argument of the corresponding factor.

From Sec. II, the system variables are T_0 and the entries of $\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{X}, \boldsymbol{Y}$. Noting that their associated random counterparts form the following Markov chain

$$\mathsf{T}_0 \to \mathbf{q} \to \mathbf{s} \to \mathbf{X} \to \mathbf{Y}, \tag{11}$$

we can write their joint density as

$$\begin{aligned} f_{\mathbf{Y},\mathbf{X},\mathbf{s},\mathbf{q},\mathsf{T}_0}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{s},\boldsymbol{q},T_0) &= f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{Y} \mid \boldsymbol{X}) f_{\mathbf{X}|\mathbf{s}}(\boldsymbol{X} \mid \boldsymbol{s}) \\ &\quad \times \mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s} \mid \boldsymbol{q}) f_{\mathbf{q}|\mathsf{T}_0}(\boldsymbol{q} \mid T_0) \end{aligned} \tag{12}$$

where each term can be factorized as

$$f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{k=1}^{K}\prod_{m=1}^{N} f_{\mathsf{y}_{mk}|\mathbf{x}_n}(y_{mk} \mid \boldsymbol{x}_n), \tag{13a}$$

$$f_{\mathbf{X}|\mathbf{s}}(\boldsymbol{X} \mid \boldsymbol{s}) = \prod_{k=1}^{K}\prod_{n=1}^{N} f_{\mathsf{x}_{nk}|\mathsf{s}_n}(x_{nk} \mid s_n), \tag{13b}$$

$$\mathbb{P}_{\mathbf{s}|\mathbf{q}}(\boldsymbol{s} \mid \boldsymbol{q}) = \prod_{b=1}^{B}\prod_{n\in\mathbb{M}_b} \mathbb{P}_{\mathsf{s}_n|\mathsf{q}_b}(s_n \mid q_b), \tag{13c}$$

$$f_{\mathbf{q}|\mathsf{T}_0}(\boldsymbol{q} \mid T_0) = \prod_{b=1}^{B} f_{\mathsf{q}_b|\mathsf{T}_0}(q_b \mid T_0), \tag{13d}$$

which follow from the description in Sec. II. The factor graph is detailed in Fig. 2 where the factor nodes, the variable nodes and the edges between them are deduced from (12) and (13).

LBP provides a means of solving (9) and (10). In order to develop the algorithm for the model in Sec. II, observe that

$$f_{\mathbf{X}|\mathbf{Y},\mathsf{T}_0}(\boldsymbol{X} \mid \boldsymbol{Y}, T_0) = \frac{\sum_{\boldsymbol{s}} \int f_{\mathbf{Y},\mathbf{X},\mathbf{s},\mathbf{q},\mathsf{T}_0}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{s},\boldsymbol{q},T_0)\mathrm{d}\boldsymbol{q}}{f_{\mathbf{Y}}(\boldsymbol{Y})}, \tag{14a}$$

$$\mathbb{P}_{\mathbf{s}|\mathbf{Y},\mathsf{T}_0}(\boldsymbol{s} \mid \boldsymbol{Y}, T_0) = \frac{\int\int f_{\mathbf{Y},\mathbf{X},\mathbf{s},\mathbf{q},\mathsf{T}_0}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{s},\boldsymbol{q},T_0)\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{X}}{f_{\mathbf{Y}}(\boldsymbol{Y})}, \tag{14b}$$

$$f_{\mathbf{q}|\mathbf{Y},\mathsf{T}_0}(\boldsymbol{q} \mid \boldsymbol{Y}, T_0) = \frac{\int \sum_{\boldsymbol{s}} f_{\mathbf{Y},\mathbf{X},\mathbf{s},\mathbf{q},\mathsf{T}_0}(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{s},\boldsymbol{q},T_0)\mathrm{d}\boldsymbol{X}}{f_{\mathbf{Y}}(\boldsymbol{Y})}. \tag{14c}$$

The messages exchanged by LBP tailored to (9) and (10) are described in Table 2 and lead to the beliefs[1]

$$\text{(FV.2)}\cdot\text{(VF.2)} \to \Upsilon_{\mathsf{x}_{nk}}(x) = \underset{nk\to nk}{\mu^t}(x)\ \underset{nk\leftarrow nk}{\mu^t}(x), \tag{15a}$$

$$\text{(FV.4)}\cdot\text{(VF.4)} \to \Upsilon_{\mathsf{s}_n}(s) = \underset{n\to n}{\psi^t}(s)\underset{n\leftarrow n}{\psi^t}(s), \tag{15b}$$

$$\text{(FV.6)}\cdot\text{(VF.6)} \to \Upsilon_{\mathsf{q}_b}(x) = \underset{b\to b}{\zeta^t}(x)\ \underset{b\leftarrow b}{\zeta^t}(x) \tag{15c}$$

and LBP estimates

$$\hat{x}_{nk}^i = \mathbb{E}_{\Upsilon_{\mathsf{x}_{nk}}}[\mathsf{x}_{nk}], \tag{16a}$$

$$\hat{s}_n^i = \mathbb{1}\left(\log\frac{\Upsilon_{\mathsf{s}_n}(1)}{\Upsilon_{\mathsf{s}_n}(0)} > 0\right), \tag{16b}$$

$$\hat{q}_b^i = \mathbb{E}_{\Upsilon_{\mathsf{q}_b}}[\mathsf{q}_b]. \tag{16c}$$

[1]The beliefs $\Upsilon_{\mathsf{x}_{nk}}$, Υ_{s_n} and Υ_{q_b} estimate the marginal posterior distributions $f_{\mathsf{x}_{nk}|\mathbf{Y},\mathsf{T}_0}$, $\mathbb{P}_{\mathsf{s}_n|\mathbf{Y},\mathsf{T}_0}$ and $f_{\mathsf{q}_b|\mathbf{Y},\mathsf{T}_0}$.

TABLE 2. Belief propagation messages for the factor graph induced by the joint density (12) and the factorizations (13).

Factor	Variable	Factor → Variable		Factor ← Variable	
$f_{\mathsf{y}_{mk}\|\mathbf{x}_k}$	x_{nk}	$\lambda^i_{mk\to n'k}(x_{nk}) \propto \int_{\mathbb{C}^{N-1}} f_{\mathsf{y}_{mk}\|\mathbf{x}_k}(y_{mk} \mid \boldsymbol{x}_{:k}) \left[\prod_{\substack{n'=1\\n'\neq n}}^{N} \lambda^{t-1}_{mk\leftarrow n'k}(x_{n'k})\right] \mathrm{d}_{\setminus n}\boldsymbol{x}_{:k}$	(FV.1)	$\lambda^i_{mk\leftarrow nk}(x_{nk}) \propto \mu^i_{nk\to nk}(x_{nk}) \prod_{\substack{m'=1\\m'\neq m}}^{M} \lambda^i_{m'k\to nk}(x_{nk})$	(VF.1)
$f_{\mathsf{x}_{nk}\|\mathsf{s}_n}$	x_{nk}	$\mu^i_{nk\to nk}(x_{nk}) \propto \sum_{s=0}^{1} f_{\mathsf{x}_{nk}\|\mathsf{s}_n}(x_{nk} \mid s)\, \nu^i_{nk\leftarrow n}(s)$	(FV.2)	$\mu^i_{nk\leftarrow nk}(x_{nk}) \propto \prod_{m=1}^{M} \lambda^i_{mk\to nk}(x_{nk})$	(VF.2)
$f_{\mathsf{x}_n\|\mathsf{s}_n}$	s_n	$\nu^i_{nk\to n}(s_n) \propto \int_{\mathbb{C}} f_{\mathsf{x}_{nk}\|\mathsf{s}_n}(x \mid s_n)\, \mu^i_{nk\leftarrow nk}(x)\mathrm{d}x$	(FV.3)	$\nu^i_{nk\leftarrow n}(s_n) \propto \psi^i_{n\to n}(s_n) \prod_{\substack{k'=1\\k'\neq k}}^{K} \nu^i_{nk'\leftarrow n}(s_n)$	(VF.3)
$\mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_b}$	s_n	$\psi^i_{n\to n}(s_n) \propto \int_0^1 \mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_b}(s_n \mid q)\, \rho^i_{n\leftarrow b}(q)\mathrm{d}q$	(FV.4)	$\psi^i_{n\leftarrow n}(s_n) \propto \prod_{k=1}^{K} \nu^i_{nk\leftarrow n}(s_n)$	(VF.4)
$\mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_b}$	q_b	$\rho^i_{n\to b}(q_b) \propto \sum_{s=0}^{1} \mathbb{P}_{\mathsf{s}_n\|\mathsf{q}_b}(s \mid q_b)\, \psi^i_{n\leftarrow b}(s)$	(FV.5)	$\rho^i_{n\leftarrow b}(q_n) \propto \zeta^i_{b\to b}(q_b) \prod_{n'\in\mathbb{M}_b\setminus\{n\}} \rho^i_{n'\to b}(q_b)$	(VF.5)
$f_{\mathsf{q}_n\|\mathsf{T}_0}$	q_b	$\zeta^i_{b\to b}(q_b) \propto f_{\mathsf{q}_b\|\mathsf{T}_0}(q_b \mid T_0)$	(FV.6)	$\zeta^i_{b\leftarrow b}(q_b) \propto \prod_{n\in\mathbb{M}_b} \rho^i_{n\to b}(q_b)$	(VF.6)

B. β-HGAMP ALGORITHM

LBP still suffers from the intractability of the computation of the messages (FV.1) exchanged in the dense part of the factor graph. Indeed, one can see that the integrals required to compute the messages $\lambda^i_{mk\to nk}$ are high dimensional, which means that LBP still has a very high complexity.

Such an issue was addressed in [17] with the development of the GAMP framework leading to the key gaussian approximation

$$\mu^i_{nk\to nk}(x) \simeq \mathcal{CN}(x; \hat{r}^i_{nk}, v^{r,i}_{nk}) \tag{17}$$

where $\hat{r}^i_{nk}$ and $v^{r,i}_{nk}$ are mean and variance variables iteratively updated by GAMP. This approximation is made possible by using the underlying linear mixing structure $\boldsymbol{A}\mathbf{X}$ under a large system limit assumptions i.e. when both M and N are very large and the coefficients of $\boldsymbol{A}$ are scaling as $O(1/N)$.

Following the approach in [13] and [15], this approximation is then propagated successively in the messages of Table 2, leading to the derivation of an hybrid GAMP algorithm, namely β-HGAMP where the estimates $\hat{x}^i_{nk}$, $\hat{s}^i_n$ and $\hat{q}^i_b$ are computed from 16. The complete algorithm is finally given in algorithm 2. In more detail:

- *Lines* 8 – 15*:* the updates of the variables $\hat{p}^i_{mk}$, $\hat{z}^i_{mk}$ and $\hat{u}^i_{mk}$ may be seen as messages sent by the factor nodes associated to $f_{\mathsf{y}_{mk}|\mathbf{x}_{:k}}$. These updates are obtained via the derivation in [17].
- *Lines* 24 – 32*:* the estimates $\hat{q}^i_b$ and $\hat{s}^i_n$ are computed in these lines, which corresponds to the portion of the factor graph where BP is applied; namely, between the factor nodes $f_{\mathsf{q}_b|\mathsf{T}_0}$ and $f_{\mathsf{x}_{nk}|\mathsf{s}_n}$. The computation is performed using intermediate variables from lines 16 to 23, where the variables $\hat{r}^i_{nk}$ and $v^{r,i}_{nk}$ arise from the Gaussian approximation (17).
- *Lines* 33 – 38*:* the estimates of the channel coefficient $\hat{x}^i_{nk}$ are updated here, which corresponds to the layer of variable nodes x_{nk}. Note that the *final* estimates $\hat{x}^i_{nk}$ will be set to 0 if the corresponding $\hat{s}^i_n$ is 0 and left untouched otherwise.

TABLE 3. Complexity comparisons.

Method	Complexity
GS-HGAMP [15]	$O(MNS)$
β-HGAMP	$O((2M+1)NK + NS(1+S))$

To the best of our knowledge, Algorithm 2 is the first HGAMP algorithm relying on the activity probability vector $\boldsymbol{q}$. By viewing $\boldsymbol{q}$ as latent variables, it is possible to incorporate correlation in the activity of the sensors.

C. COMPLEXITY ANALYSIS

We briefly address the complexity of the proposed β-HGAMP algorithm. It is clear that lines 8 to 23 requires $O(2MNK)$ floating operations (multiplications and divisions) and lines 33 to 38 $O(NK)$ operations.

The complexity bottleneck appears from lines 24 to 32 with the computation of the estimates $\hat{q}^i_b$ and $\hat{q}^i_{n,b}$, which requires the evaluation of a high-dimensional integral. To reduce the complexity of the integration, observe that the integrands are polynomials in the variable q weighted by the density $f_{\mathsf{q}_b}(q)$. It is thus possible to write

$$\hat{q}^i_b = \frac{\sum_{j=0}^{S+1} \pi_j \mathbb{E}\left[\mathsf{q}_b^j\right]}{\sum_{j=0}^{S} \varpi_j \mathbb{E}\left[\mathsf{q}_b^j\right]} \quad \text{and} \quad \hat{q}^i_{n,b} = \frac{\sum_{j=0}^{S} \pi_{n,j} \mathbb{E}\left[\mathsf{q}_b^j\right]}{\sum_{j=0}^{S-1} \varpi_{n,j} \mathbb{E}\left[\mathsf{q}_b^j\right]} \tag{18}$$

where

$$P_b(q) = q \prod_{n\in\mathbb{M}_b} \left[(1-q)\psi_{0,n} + q\psi_{1,n}\right] = \sum_{j=0}^{S+1} \pi_j q^j \tag{19a}$$

$$Q_b(q) = \prod_{n\in\mathbb{M}_b} \left[(1-q)\psi_{0,n} + q\psi_{1,n}\right] = \sum_{j=0}^{S} \varpi_j q^j \tag{19b}$$

$$P_{n,b}(q) = q \prod_{n' \in \mathbb{M}_b \setminus \{n\}} \left[(1-q)\psi_{0,n'} + q\psi_{1,n'}\right] = \sum_{j=0}^{S} \pi_{n,j} q^j \tag{19c}$$

$$Q_{n,b}(q) = \prod_{n' \in \mathbb{M}_b \setminus \{n\}} \left[(1-q)\psi_{0,n'} + q\psi_{1,n'}\right] = \sum_{j=0}^{S-1} \varpi_{n,j} q^j \tag{19d}$$

Since the polynomials' coefficients $\{\pi_j\}$, $\{\varpi_j\}$, $\{\pi_{n,j}\}$, $\{\varpi_{n,j}\}$ depends on the intermediate variables $\{\psi_{0,n}\}$ and $\{\psi_{1,n}\}$, they cannot be computed offline. Each set of coefficients are computed using a recursive algorithm[2] with at most $O(S^2)$ operations from the corresponding roots of the polynomials. The computation of the moments $\mathbb{E}\left[\mathsf{q}_b^j\right]$ may be computed offline and so are assumed to be of constant complexity $O(1)$. Using this method grants the computational complexity of lines 24 to 32 to be $O(NS(1+S))$.

Finally, summing over each block leads to an overall complexity of $O((2M+1)NK + NS(1+S))$. A complexity comparison with respect to GS-HGAMP is provided in Table 3 and one can compare with the methods within [15, table 1]. Though the complexity of β-HGAMP scales with S^2 instead of S for GS-HGAMP, it is worth noting that S would remain relatively small in the context of large-scale industrial plant where it is the number B of machines that would increase.

IV. NUMERICAL RESULTS

A. SETTINGS

We assess the performances of GAMP, GS-HGAMP and β-HGAMP through extensive Monte-Carlo simulations. As a baseline, GS-HGAMP is evaluated with the machine-based and sensor-based modes described in Sec. II-C and quickly summarized here:

1) the machine-based mode performs active sensor detection taking into account the underlying group structure induced by the machines;
2) the sensor-based mode performs individual detection ignoring the group structure; i.e., each sensor is assumed to be independent from the other sensors on the same machine. In practice, this means that the number of groups B is assumed to be equal to the number of sensors N (and so $S = 1$).

Performance is assessed in terms of the normalized mean-square error (NMSE):

$$\mathsf{NMSE}[\hat{\boldsymbol{X}}^i, \boldsymbol{X}]_{\mathrm{dB}} = 10\log_{10}\mathbb{E}\left[\frac{\left\|\hat{\boldsymbol{X}}^i - \boldsymbol{X}\right\|_2^2}{\|\boldsymbol{X}\|_2^2}\right] \tag{20}$$

[2]For a polynomial $P(X) = \sum_{i=0}^{S} a_i X^i = \prod_{i=1}^{S}(X - r_i)$, identify, for $I \in [S]$, the coefficients of each side of the recursion $P_{I+1}(X) = P_I(X)(X - r_{I+1})$ where $P_I(X) = \prod_{i=1}^{I}(X - r_j)$.

Algorithm 2: β-HGAMP

Input: $T_0, \boldsymbol{Y}, \boldsymbol{A}, \mu_0, \tau_0, \tau_w, \epsilon, I_{\max}$

1 **Initialization**
2 $\quad i = 0$
3 $\quad \hat{\boldsymbol{X}}^i = \mu_0 \mathbf{1}_{N\times K}$, $\boldsymbol{V}^{\boldsymbol{X},i} = \tau_0 \mathbf{1}_{N\times K}$
4 $\quad \hat{\boldsymbol{U}}^i = \mathbf{0}_{M\times K}$
5 **end**
6 **while** $\Delta > \epsilon$ **and** $i < I_{max}$ **do**
7 $\quad i = i + 1$
$\quad$ // Updates of system output variables.
8 $\quad$ **for** $(m,k) \in [M] \times [K]$ **do**
9 $\quad\quad v_{mk}^{p,i} = \sum_{n=1}^{N} |a_{mn}|^2 v_{nk}^{x,i-1}$
10 $\quad\quad \hat{p}_{mk}^{i} = \sum_{n=1}^{N} a_{mn}\hat{x}_{nk}^{i-1} - v_{mk}^{p,i}\hat{u}_{mk}^{i-1}$
11 $\quad\quad v_{mk}^{z,i} = \tau_w v_{mk}^{p,i}/(v_{mk}^{p,i} + \tau_w)$
12 $\quad\quad \hat{z}_{mk}^{i} = \hat{p}_{mk}^{i} + v_{mk}^{p,i}(y_{mk} - \hat{p}_{mk}^{i})/(v_{mk}^{p,i} + \tau_w)$
13 $\quad\quad v_{mk}^{u,i} = (1 - v_{mk}^{z,i})/(v_{mk}^{u,i})^2$
14 $\quad\quad \hat{u}_{mk}^{i} = (\hat{z}_{mk}^{i} - \hat{p}_{mk}^{i})/v_{mk}^{p,i}$
15 $\quad$ **end**
$\quad$ // Updates of intermediate variables.
16 $\quad$ **for** $n \in [N]$ **do**
17 $\quad\quad$ **for** $k \in [K]$ **do**
18 $\quad\quad\quad v_{nk}^{r,i} = 1/\sum_{m=1}^{M} |a|_{mn}^2 v_{mk}^{u,i}$
19 $\quad\quad\quad \hat{r}_{nk}^{i} = \hat{x}_{nk}^{i-1} + v_{nk}^{r,i}\sum_{m=1}^{M} a_{mn}\hat{u}_{mk}^{i}$
20 $\quad\quad$ **end**
21 $\quad\quad \psi_{0,n} = \prod_{k=1}^{K} \mathcal{CN}(0; \hat{r}_{nk}^{i}, v_{nk}^{r,i})$
22 $\quad\quad \psi_{1,n} = \prod_{k=1}^{K} \mathcal{CN}(0; \hat{r}_{nk}^{i} - \mu_0, v_{nk}^{r,i} + \tau_0)$
23 $\quad$ **end**
$\quad$ // Updates of activity-related variables.
24 $\quad$ **for** $b \in [B]$ **do**
25 $\quad\quad \hat{q}_b^i = \frac{\int_{[0,1]} q f_{\mathsf{q}_b}(q) \prod_{n\in\mathbb{M}_b}[(1-q)\psi_{0,n} + q\psi_{1,n}]\mathrm{d}q}{\int_{[0,1]} f_{\mathsf{q}_b}(q) \prod_{n\in\mathbb{M}_b}[(1-q)\psi_{0,n} + q\psi_{1,n}]\mathrm{d}q}$
26 $\quad\quad$ **for** $n \in \mathbb{M}_b$ **do**
27 $\quad\quad\quad \hat{q}_{n,b}^i = \frac{\int_{[0,1]} q f_{\mathsf{q}_b}(q) \prod_{n'\in\mathbb{M}_b\setminus\{n\}}[(1-q)\psi_{0,n} + q\psi_{1,n}]\mathrm{d}q}{\int_{[0,1]} f_{\mathsf{q}_b}(q) \prod_{n'\in\mathbb{M}_b\setminus\{n\}}[(1-q)\psi_{0,n} + q\psi_{1,n}]\mathrm{d}q}$
28 $\quad\quad\quad \mathsf{LLR}_n = \log\left(\frac{\hat{q}_{n,b}^i}{(1-\hat{q}_{n,b}^i)}\frac{\psi_{1,n}}{\psi_{0,n}}\right)$
29 $\quad\quad\quad \hat{s}_n^i = \mathbb{1}(\mathsf{LLR}_n > 0)$
30 $\quad\quad\quad \gamma_n = 1/(1 + \exp(-\mathsf{LLR}_n))$
31 $\quad\quad$ **end**
32 $\quad$ **end**
$\quad$ // Updates of channel estimates.
33 $\quad$ **for** $(n,k) \in [N] \times [K]$ **do**
34 $\quad\quad \kappa_{nk} = 1/(1/\tau_0 + 1/v_{nk}^{r,i})$
35 $\quad\quad \nu_{nk} = \mu_0/\tau_0 + \hat{r}_{nk}^{i}/v_{nk}^{r,i}$
36 $\quad\quad \hat{x}_{nk}^{i} = \gamma_n \kappa_{nk} \nu_{nk}$
37 $\quad\quad v_{nk}^{x,i} = \gamma_n(\kappa_{nk} + |\kappa_{nk}\nu_{nk}|^2) - |\hat{x}_{nk}^{i}|^2$
38 $\quad$ **end**
39 **end**

Output: $\hat{\boldsymbol{q}}^i, \hat{\boldsymbol{s}}^i, \hat{\boldsymbol{X}}^i$

which measures the quality of the channel estimates and the user error rate (UER)

$$\mathsf{UER}[\hat{\boldsymbol{s}}^i, \boldsymbol{s}] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}(\hat{s}_n^i \neq s_n)\right]. \tag{21}$$

TABLE 4. Simulation settings.

Setting	Value
N	256 sensors
B	64 machines
S	4 sensors per machines
M	128-long pilot sequences
K	2 antennas
$\boldsymbol{A}$	Spherical distribution[2]
# transmissions	1000
T_0	$\{26.5, 27.5, 28.5, 29.5\}$°C
$\alpha(T_0)$	$\{0.15, 0.25, 0.35, 0.45\}$, see (22)

[2]Each column $\boldsymbol{a}_n$ of the matrix $\boldsymbol{A}$ is drawn from $\mathscr{CN}(\boldsymbol{0}_N, \boldsymbol{I}_N)$, normalized and scaled by $\sqrt{M}$.

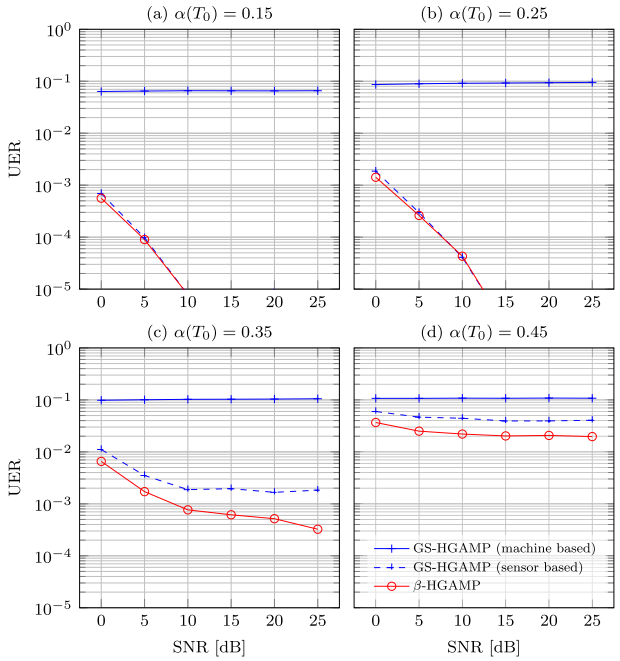


FIGURE 3. UER for the activity detection of the devices.

which measures the probability that a sensor is incorrectly detected. In particular, the UER provides a means to investigate the impact of the communication system on the ability for the access point to correctly identify faults.

The corresponding results are given in Figs. 4 and 3. Fig. 5 shows the estimates of the *a posteriori* activity probabilities $\hat{q}_b^i$ against the true activity probability q_b. These figures have been obtained using the settings described in Table 4 where the parameters of the beta distribution are computed using

$$\alpha(T_0) = \frac{|T_0 - 25|}{10} \quad \text{and} \quad \beta(T_0) = 1 - \alpha(T_0) \tag{22}$$

meaning that the more T_0 deviates from 25° C, the larger the value of $\alpha(T_0)$, the larger the fault probability, and the number of active sensors during the transmission time slot.

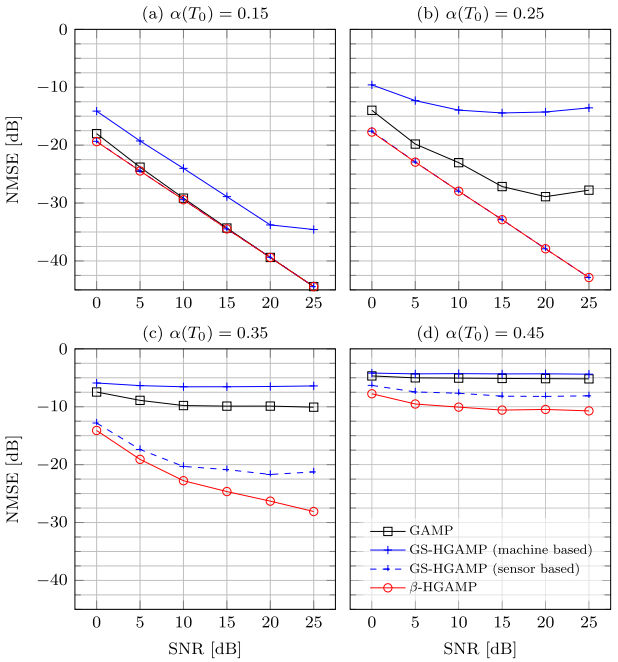


FIGURE 4. NMSE for the channel estimates.

B. DISCUSSION

As noted in Sec. III, convergence guarantees of LBP and HGAMP are not necessarily available. With the aid of damped updates for the GAMP and HGAMP algorithms, the results obtained in Figs. 3 and 4 were stable without observing divergence after extensive simulations.

1) UER AND NMSE

Observe from Fig. 4 that β-HGAMP outperforms GAMP and the two modes of GS-HGAMP, with the sensor-based GS-HGAMP the second best performing algorithm followed by GAMP and then machine-based GS-HGAMP. The same order is observed for the UER on Fig. 3. A UER curve for GAMP is not present since GAMP only provides a channel estimate and not sensor identification. As the average activity probability $\alpha(T_0)$ increases, the performance differences between the algorithms also increase. In more detail:

- Figs. 3 and 4 (a), $\alpha(T_0) = 0.15$: GAMP, GS-HGAMP (sensor-based) and β-HGAMP have a similar NMSE. In terms of the UER, β-HGAMP and sensor-based GS-HGAMP have nearly the same performance.
- Figs. 3 and 4 (b), $\alpha(T_0) = 0.25$: β-HGAMP and sensor-based GS-HGAMP have the lowest NMSE, while GAMP and machine-based GS-HGAMP have a significant performance degradation with a 10 and 30dB gap, respectively. The UER for β-HGAMP and the sensor-based GS-HGAMP are comparable.
- Figs. 3 and 4 (c), $\alpha(T_0) = 0.35$: each algorithm has a reduction of approximately 10dB at a SNR of 25dB in the NMSE for the channel estimate compared with the case of $\alpha(T_0) = 0.25$. The biggest loss is observed

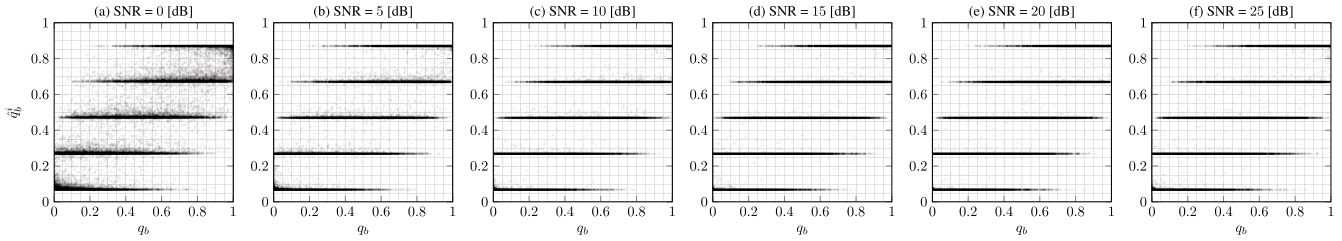


FIGURE 5. Estimates of the *a posteriori* activity probabilities vs. the true activity probabilities for $\alpha(T_0) = 0.35$.

for sensor-based GS-HGAMP with a decrease of around 14dB leaving β-HGAMP to have the best performance. For the UER, two orders of magnitude have been lost for β-HGAMP and sensor-based HGAMP with about a factor 2 to 3 difference between them.

- Figs. 3 and 4 (d), $\alpha(T_0) = 0.45$: there is a significant degradation in the NMSE and UER for all algorithms due to, on average, a larger number of active devices.

From these observations, we see that β-HGAMP outperforms existing methods for the model described in Sec. II. We also observe that despite incorporating some information about the positions of the sensors, machine-based GS-HGAMP has poor performance even compared with sensor-based GS-HGAMP. This is due to the fact that the individual behavior of each sensor cannot be accounted for.

However, sensor-based HGAMP does not exploit the prior information of shared activity probability among groups, unlike β-HGAMP. Hence, for larger average activity probability $\alpha(T_0)$, the probability to have multiple sensors of the same group to be active at the same time is also larger. This information is leveraged by β-HGAMP in order to perform a joint detection for sensors belonging to the same group. In contrast, sensor-based HGAMP does not use this information and performs independent detection for those sensors.

These explanations justify the UER of each algorithm and also the values of the NMSE. Indeed, detection errors have a direct consequence on the channel estimates. When a sensor is detected to be inactive, the corresponding channel estimates will be set to zero. Since machine-based GS-HGAMP is subject to a large number of detection errors because of the block detection, a large number of channel estimates will be zero, significantly increasing the errors in channel estimates and, at the same time, degrading the NMSE. The same reasoning applies to sensor-based GS-HGAMP and β-HGAMP, with a weaker impact of the NMSE. We also mention that GAMP does not leverage the diversity effect offered by the multiple antennas because of the underlying structure of its factor graph, limiting the quality of its estimates.

Finally, the performance degradation observed at $\alpha(T_0) = 0.45$ is explained by the potentially large number of devices which might be active relatively to the length of the pilot sequences. As a rule of thumb, an average of 45% of the sensors are active over each transmission meaning that some transmissions may have a more sensors active than the length of the pilot sequences. Note that considering larger values of M (i.e., longer pilot sequences) will prevent such an issue.

2) ESTIMATES OF THE A POSTERIORI ACTIVITY PROBABILITIES

As detailed in Sec. II, the activity probability of each machine is dependent on the local temperature. As a consequence, the *a posteriori* activity probability estimate $\hat{q}_b^i$ for the b-th machine can be used to obtain an estimate of the local temperature. This information may be useful for diagnosis of machine degradation without sending temperature information within the data packet. We remark that this information is not available using the GAMP and GS-HGAMP algorithms.

The estimates of $\hat{q}_b^i$ are plotted in Fig. 5. In particular, the subfigures correspond to the estimates from each transmission for different SNRs. Observe that the estimates largely form 5 levels.

An explanation for this behavior is as follows. First, recall that each machine in this example is equipped with 4 sensors. For each transmission, each machine can be described by 5 states corresponding to the number of active sensors; i.e., 0, 1, 2, 3, 4. That is, the distribution of the estimates is determined by how many sensors are active. The estimates lying between the levels are due to channel noise, most evident for lower values of the SNR.

V. CONCLUSION

This paper investigates the joint sensor identification and channel estimation problem for fault detection, where multiple sensors are placed on each machine. We derive an algorithm based on the HGAMP framwework, called β-HGAMP, which accounts for uncertainty in activity probabilities in order to improve both the detection and estimation capabilities compared with previous GS-HGAMP algorithm. As a side feature, β-HGAMP estimates the *a posteriori* activity probabilities which can reduce the quantity of data to be transmitted.

Motivated by these results, the investigation of more general models incorporating correlation is an interesting open question. This would provide a flexible and systematic approach for the joint detection and channel estimation problem for massive random access systems when the activity of devices may be correlated beyond the group structure considered in this paper.

Finally, this work provides a basis for joint detection, channel estimation and data recovery in the presence of correlated sensor activity. Combining our approach with the work in [12] appears to be a promising avenue of future work.

REFERENCES

[1] M. Erdelj, N. Mitton, and E. Natalizio, *Applications of Industrial Wireless Sensor Networks*. Boca Raton, FL, USA: CRC Press, 2013, pp. 1–22.

[2] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 1st Quart., 2020.

[3] C. Zheng, M. Egan, L. Clavier, A. E. Kalor, and P. Popovski, "Stochastic resource optimization of random access for transmitters with correlated activation," *IEEE Commun. Lett.*, early access, Jun. 17, 2021, doi: 10.1109/LCOMM.2021.3090110.

[4] 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.

[5] A. Pramanik, S. P. Maity, and Z. Farheen, "Compressed sensing channel estimation in massive MIMO," *IET Commun.*, vol. 13, no. 19, pp. 3145–3152, Dec. 2019.

[6] Y. Huang, Y. He, L. Shi, T. Cheng, Y. Sui, and W. He, "A sparsity-based adaptive channel estimation algorithm for massive MIMO wireless powered communication networks," *IEEE Access*, vol. 7, pp. 124106–124115, 2019.

[7] 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 J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–14, Dec. 2018.

[8] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Proc. Inf. Theory Appl. Workshop*, 2008, pp. 326–333.

[9] 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.

[10] W. Yuan, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, "Iterative receiver design for FTN signaling aided sparse code multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 915–928, Feb. 2020.

[11] W. Yuan, N. Wu, Q. Guo, D. W. K. Ng, J. Yuan, and L. Hanzo, "Iterative joint channel estimation, user activity tracking, and data detection for FTN-NOMA systems supporting random access," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2963–2977, May 2020.

[12] 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.

[13] 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.

[14] Q. Zou, H. Zhang, and H. Yang, "Expectation-maximization-aided hybrid generalized expectation consistent for sparse signal reconstruction," *IEEE Signal Process. Lett.*, vol. 28, pp. 648–652, 2021.

[15] S. Rangan, A. K. Fletcher, V. K. Goyal, and P. Schniter, "Hybrid generalized approximate message passing with applications to structured sparsity," in *Proc. IEEE Int. Symp. Inf. Theory Process.*, Jul. 2012, pp. 1236–1240.

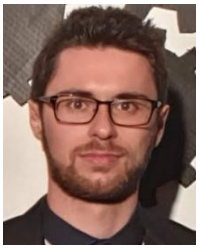
[16] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[17] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory Process.*, Jul. 2011, pp. 2168–2172.

[18] J. Mooij and H. Kappen, "Sufficient conditions for convergence of loopy belief propagation," 2012, *arXiv:1207.1405*. [Online]. Available: http://arxiv.org/abs/1207.1405

[19] A. T. Ihler, J. W. F. Iii, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, no. 5, p. 32, May 2005.

[20] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.



LÉLIO CHETOT received the B.E. degree in telecommunication engineering from the Institut National des Sciences Appliquées (INSA) de Lyon, Villeurbanne, France, in 2018. He is currently pursuing the Ph.D. degree with the CITI Laboratory, Institut National de la Recherche en Informatique et Automatique (INRIA), Villeurbanne, France. His research interest includes graphical Bayesian inference applied to massive machine-type communication networks.



MALCOLM EGAN received the Ph.D. degree in electrical engineering from the University of Sydney, Australia, in 2014. Previously he was an Assistant Professor at INSA Lyon, and a Postdoctoral Researcher at the Laboratoire de Mathématiques, Université Blaise Pascal, France, and the Department of Computer Science, Czech Technical University in Prague, Czech Republic. He is currently a Chargé de Recherche (Permanent Research Staff) at INRIA hosted by CITI, a joint Laboratory between INRIA, INSA Lyon, and Université de Lyon, France. His research interests include information theory and statistical signal processing with applications in wireless and molecular communications. He is currently an Associate Editor for IEEE COMMUNICATIONS LETTERS and previously a Guest Editor of IEEE ACCESS.



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 the Co-Founder of the Centre for Innovation, Telecommunications and Integration of Services (CITI Lab), in 2001. He was a Visiting Scholar at Princeton University, 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 currently the Scientific Coordinator for the Experimental Facility FIT-CorteXlab. He is also a Professor at INSA, Université de Lyon, a Research Member of the Institut National de Recherche en Informatique et en Automatique (INRIA), and the Vice Delegate for Research at INRIA Grenoble. He has co-published more than 150 conference and journal articles. His research interests include wireless networking and communication theory, focusing on realistic modeling, wireless system optimization, and performance assessment considering both infrastructure-based and *ad-hoc* networks. He holds the Internet of Things Industrial and Research Chair of INSA Lyon sponsored by SPIE ICS. He is an Associate Editor of the *EURASIP Journal on Wireless Communications and Networking* (Springer).