Low complexity Detector for massive uplink random access with NOMA in IoT LPWA networks

Diane Duchemin, Jean-Marie Gorce, Claire Goursaud

▶ To cite this version:

HAL Id: hal-02146649

https://hal.inria.fr/hal-02146649

Submitted on 4 Jun 2019

Low complexity Detector for massive uplink random access with NOMA in IoT LPWA networks

Diane Duchemin, Jean-Marie Gorce and Claire Goursaud

CITI Laboratory

Univ Lyon, INRIA, INSA Lyon

Lyon, France

{diane.duchemin, jean-marie.gorce, claire.goursaud}@insa-lyon.fr

Abstract—We focus on the random uplink transmissions of an unknown subset of nodes, disseminated in a cell. Under the constraints of massive Machine Type Communication (MTC) in cellular Low Power Wide Area Networks (LPWAN) and Ultra Reliable Low Latency Communications (URLLC), we assume a low coordination with the receiver and the usage of Gaussian coded Non Orthogonal Multiple Access (NOMA). We then target direct data transmission and thus propose a low complexity optimal-based detection of the active users: the It-MAP. This algorithm relies on the Maximum A Posteriori (MAP) detector and, similarly to Orthogonal Matching Pursuit (OMP)-like algorithms, proceeds by iteration to decrease its intrinsic complexity. We also show the gain of employing It-MAP rather than an OMP-based detection and the advantage of exploiting the possibility to tune the algorithm, in order to avoid either Missed Detection or False Alarm, depending on the wished trade-off between the reliability, the latency and the resource usage of the full transmission.

Index Terms—Multiple User Detection, MAP, Random Access, Non Orthogonal Multiple Access

I. INTRODUCTION

The Low Power Wide Area Networks (LPWAN) have recently driven a strong interest due to their advantages towards large scale IoT deployement. Proprietary solutions [1] such as Sigfox or LoRa have been proposed to meet their requirements, and new standards have now been released, such as Cat-M and NB-IoT [2], [3]. All of them agree on the limitation of the radio resources consumed by the signalling part of the communication protocol, especially for the end-device. As such, they mainly adopt a communication protocol based on remaining in idle mode as long as possible for the devices and waking up just for message transmissions since the emission is the most consuming action. LPWAN thus rely on massive random access and derive from a connected mode as usual in cellular systems to communications occurring under unconnected mode. For instance, the recent NB-IoT standards proposed to shift from a 5 steps handshake mechanism before data transmission to an "early" data transmission at the third step [4]. Ideally, a direct transmission would be preferred for small data, especially in the case of Ultra Reliable Low Latency Communications (URLLC) [5], [6], as any additional signalling message makes the overhead much larger than the transmitted data and consequently extends the latency as well.

Besides, Non Orthogonal Multiple Access (NOMA) schemes have a high potential for such massive IoT networks since they allow to share a single radio resource for multiple devices, thus reducing the bottleneck effect in random access, at the cost of introducing intra-interference in the system [6], [7]. However, it brings a more challenging role to the receiving equipment, which thus must detect all active users in spite of the intra-interference of the NOMA scheme [8], [9].

In this work, we propose a new optimal-based low complexity detector for random access of devices using a Gaussian coded NOMA. Section II introduces the system model as well as the detection algorithm, NBOMP [9], to which the proposed detector will be compared. The descriptions of the optimal and our new detection algorithms lie in section III while the numerical results of the comparison with NBOMP are given in section IV. Section V concludes this paper.

II. REFERENCE SCENARIO AND ALGORITHM

A. Scenario's model and assumptions

We focus on the direct and random message transmissions from a set $\mathcal{N}$ of N communicating devices to a single Base Station (BS). The devices have no possibility to self-coordinate their transmissions. The transmissions nevertheless occur in a synchronous and power-constrained manner (by such we mean that they arrive simultaneously at the BS and with a known and fixed received power, thanks to a beacon emitted by the BS for instance). At a given time, there can be 0,1,2,... up to N active users, i.e. the number of active users is not fixed. In the following of the document, the subset of active users is denoted $\mathcal{N}_a$. The nodes have an equal activity probability, which we denote p_n. The communicating devices employ a known set of codewords of length m, whose each is dedicated to only one of them. The attribution of a particular code $C_n \in \mathbb{C}^m$ to its node $n \in \mathcal{N}$ is also known a priori. The nodes own a single antenna whereas the BS is equipped with α antennas. The communications are disturbed by an additive white Gaussian noise $\mathbf{Z} \in \mathbb{C}^{\alpha \times m}$, whose distribution is known ($z_{jk} \sim \mathcal{N}_{\mathbb{C}}(0,1)$ with $j \in [1, \alpha]$, $k \in [1, m]$). For the channel however, not only the distribution but also the exact realisations of the channel $H_n \in \mathbb{C}^{\alpha} \ \forall n \in \mathcal{N}$ are part of the prior knowledge of the receiver. Additional assumptions include a block fading model, i.e. the channel is coherent during the transmission of length m; a spatially independent

channel, i.e. the channel response from a user n to an antenna j of the BS is independent of that associated to any other user and/or antenna ; and finally a Rayleigh fading model; therefore $H_n = [h_{n,1}, ... h_{n,j}, ... h_{n,\alpha}]^t$ with $h_{n,j} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The expression of the signal received at the BS is:

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & ... & y_{1,m} \\ ... & y_{j,k} & ... \\ ... & & y_{\alpha,m} \end{bmatrix} = \sum_{n \in \mathcal{N}_a} \sqrt{\rho_0} H_n C_n^t + \mathbf{Z} \quad (1)$$

where $\mathbf{Y} \in \mathbb{C}^{\alpha \times m}$ is the received signal over m channel use and on the α BS antennas, ρ_0 the received signal power. Vectorizing (1) provides:

$$Y = \sum_{n \in \mathcal{N}_a} \sqrt{\rho_0} \, \mathbf{H_n} \cdot C_n + Z \quad (2)$$

where $\mathbf{H_n} = \mathbf{I_m} \otimes H_a$, with $\mathbf{I_m}$ the identity matrix $m \times m$ and $\otimes$ the Kronecker product.

Notation : $\mathbf{H}$ denotes a matrix; H a vector; h a scalar.

B. NBOMP detection

NBOMP stands for Normalized Block Orthogonal Matching Pursuit. This greedy algorithm has been proposed by Xie et al. [9] in order to jointly detect active users and estimate their transmitted messages under the previously defined hypotheses, this is, based on the a-priori knowledge of the dedicated coding matrix and the channel impulse response for each node and considering perfectly synchronised and power equalised messages at the receiver, with a block Rayleigh fading (in addition to the white Gaussian noise). Moreover, it assumes a low active subset size (i.e. the sparsity assumption), and Gaussian coding matrices, constituted of normalised columns. These two latter points allow to apply algorithms dedicated to Compressed Sensing, such as OMP, to recover the signals.

As defined in [9], NBOMP proceeds by iterations, whose maximum number K is predefined. During each iteration k, based on its a-priori knowledge, it performs both the tasks of detection and estimation: it firstly adds a new detected node to the detected subset λ_k, and secondly estimates the data sent by the assumed active nodes (i.e. belonging to λ_k). At the end of an iteration, the estimated contributions of the detected users are retrieved from the received signal, in order to execute the detection rule of the next iteration without the previously detected communications, thus NBOMP has similarities with a Successive Interference Cancellation (SIC). This detection rule is based on maximizing a normalized correlation :

$$\delta_{NBOMP,k}(Y_k) = \arg\max_{n \in \{\mathcal{N} \setminus \lambda_k\}} \frac{\left\| (\mathbf{H_n} C_n)^t Y_k \right\|_2^2}{\|H_n\|_2^2} \quad (3)$$

where Y_k is the received signal after the interference cancellation of iteration k-1 : $Y_k = Y_{k-1} - \sqrt{\rho_0} \mathbf{H}_{\delta(\mathbf{Y_{k-1}})} C_{\delta(Y_{k-1})}$.

Without loss of generality, we assume that the nodes simply wish to transmit an activity notification to the receiver: the message itself is an "on" notification (which can be triggered by the node application, e.g., due to a measured value reaching a threshold), no measurement value is included. In this case,

the scenario is then equivalent to the "On-Off" framework of [8] where the authors aim at detecting the active set of devices (without considering the actual value of the symbol, sent in combination with the user code). Similarly, we assume that the active nodes only send their dedicated code-words to notify the receiver. Each node thus owns a unique coding Gaussian vector instead of a matrix. In this work, we can thus consider that NBOMP task is limited to the detection problem.

Algorithm 1 Detection-restricted NBOMP algorithm

Input: Y, $\mathcal{N}$, K, ρ_0, (C_n, H_n) $\forall n \in \mathcal{N}$
  $\lambda_0 = \emptyset$; $Y_0 = Y$ // *Initialisation*
  for $k = 0$ to $K - 1$ **do**
   $\delta_{NBOMP,k}(Y_k) = \arg\max_{n \in \{\mathcal{N} \setminus \lambda_k\}} \frac{\left\| (\mathbf{H_n} C_n)^t Y_k \right\|_2^2}{\|H_n\|_2^2}$
   $\lambda_{k+1} = \lambda_k \cup \{\delta_{NBOMP,k}(Y_k)\}$
   $Y_{k+1} = Y_0 - \sqrt{\rho_0} \sum_{n \in \lambda_{k+1}} \mathbf{H_n} C_n$
  end for
  return $\delta_{NBOMP}(Y) = \lambda_K$

III. PROPOSITION FOR RANDOM ACCESS DETECTION

A. Optimal detector

In order to evaluate the performance of NBOMP detection and the further proposed algorithm, we derive the optimal detection rule [10] of our scenario, for reference.

Let $\mathcal{H}_i$ denote the hypothesis that the subset $\Omega_i \subset \{\mathcal{N} \cup \emptyset\}$ is the active nodes set of a given realisation of the received signal Y, and P_i denote the probability that $\mathcal{H}_i$ occurs. Let Ψ denote the set of the $\binom{N}{N_a}$ combinations of N_a elements from $\mathcal{N}$, with $N_a = 0$ to N. Ψ thus contains all the possible active node sets, and $\Omega_i \in \Psi$. Let $\delta : Y \in \mathbb{C}^{\alpha \times m} \to i \in [0, |\Psi| - 1]$ be a decision function, which splits the observation space into $|\Psi|$ non overlapping decision spaces $\mathcal{Y}_i$ and associates each to an hypothesis $\mathcal{H}_i$.

Proposition 1. *The Maximum A Posteriori (MAP) detector is:*

$$\begin{aligned} \delta_{opt}(Y) &= \arg\min_{j \in [0;|\Psi|-1]} \frac{1}{2} \left\| Y - \sqrt{\rho_0} \sum_{n \in \Omega_j} \mathbf{H_n} C_n \right\|^2 - \log P_j \\ &= \arg\max_{j \in [0;|\Psi|-1]} \sqrt{\rho_0} \, \Re \left(\sum_{n \in \Omega_j} C_n^t H_n^t Y \right) \\ &\quad - \frac{\rho_0}{2} \sum_{n \in \Omega_j} \sum_{l \in \Omega_j} \Phi_{n,l} \chi_{n,l} + \log P_j \end{aligned} \quad (4)$$

with $\chi_{n,l} = \sum_{u=1}^m c_{n,u}^* c_{l,u}$; $\Phi_{n,l} = \sum_{a=1}^\alpha h_{n,a}^* h_{l,a}$.

Proposition 2. *The Maximum Likelihood (ML) detector is:*

$$\delta_{\substack{opt \\ ML}}(Y) = \arg\max_{j \in [0;|\Psi|-1]} f(Y|\mathcal{H}_j) \quad (5)$$

$$= \arg\min_{j \in [0;|\Psi|-1]} \left\| Y - \sqrt{\rho_0} \sum_{n \in \Omega_j} \mathbf{H_n} C_n \right\|^2 \quad (6)$$

Proof. The optimal detection rule is the decision function for which the Bayes risk $R(\delta)$ is minimized:

$$\delta_{opt}(Y) = \arg\min_{\delta} R(\delta(Y)) = \arg\min_{\delta} \sum_{i=0}^{|\Psi|-1} R(\delta(Y)|\mathcal{H}_i)P_i \tag{7}$$

where $R(\delta|\mathcal{H}_i)$ is the Bayes risk under hypothesis $\mathcal{H}_i$, i.e.

$$R(\delta(Y)|\mathcal{H}_i) = \sum_{j=0}^{|\Psi|-1} C_{j,i} P(\mathcal{Y}_j|\mathcal{H}_i) \tag{8}$$

with $C_{j,i}$ the cost of deciding that Ω_j is the active nodes subset when $\mathcal{H}_i$ actually holds. As we are looking for the optimal decision function δ_{opt}, the costs are the detection errors such that $C_{i,i}=0 \; \forall i \in [0,|\Psi|-1]$ and $C_{j,i}=1 \; \forall j \neq i$.
$P(\mathcal{Y}_j|\mathcal{H}_i) = \int_{\mathcal{Y}_j} f(Y|\mathcal{H}_i)dY$ is the probability that Y lies in $\mathcal{Y}_j$ when $\mathcal{H}_i$ occurs. The Bayes risk $R(\delta(Y))$ is thus the detection error probability, which we aim to minimize.

$$\delta_{opt}(Y) = \arg\min_{\delta} \sum_{i=0}^{|\Psi|-1} P_i \sum_{j=0}^{|\Psi|-1} C_{j,i} \int_{\mathcal{Y}_j} f(Y|\mathcal{H}_i)dY \tag{9}$$

$$= \arg\min_{\delta} \sum_{j=0}^{|\Psi|-1} \int_{\mathcal{Y}_j} \sum_{i=0}^{|\Psi|-1} C_{j,i} f(Y|\mathcal{H}_i)P_i \; dY \tag{10}$$

Therefore, δ_{opt} is the decision function associated to the spaces $\mathcal{Y}_j$ such that the risk is minimized on each $\mathcal{Y}_j$:

$$\delta_{opt}(Y) = \arg\min_{j \in [0;|\Psi|-1]} \sum_{i=0}^{|\Psi|-1} C_{j,i} f(Y|\mathcal{H}_i)P_i \tag{11}$$

$$= \arg\max_{j \in [0;|\Psi|-1]} f(Y|\mathcal{H}_j)P_j \tag{12}$$

which is the Maximum A Posteriori detector, with

$$f(Y|\mathcal{H}_j) = \frac{1}{\sqrt{(2\pi)^{\alpha m}}} \exp\left(-\frac{\left\|Y - \sqrt{\rho_0}\sum_{n\in\Omega_j}\mathbf{H_n}C_n\right\|^2}{2}\right) \tag{13}$$

In case of uniformly distributed hypotheses ($P_i = \frac{1}{|\Psi|} \; \forall i$), we would obtain the Maximum Likelihood expression:

$$\delta_{opt \atop ML}(Y) = \arg\max_{j \in [0;|\Psi|-1]} f(Y|\mathcal{H}_j)$$

$$= \arg\min_{j \in [0;|\Psi|-1]} \left\|Y - \sqrt{\rho_0}\sum_{n\in\Omega_j}\mathbf{H_n}C_n\right\|^2$$

Combining (12) and (13), we have:

$$\delta_{opt}(Y) = \arg\min_{j \in [0;|\Psi|-1]} \frac{1}{2}\left\|Y - \sqrt{\rho_0}\sum_{n\in\Omega_j}\mathbf{H_n}C_n\right\|^2 - \log P_j$$

$$= \arg\max_{j \in [0;|\Psi|-1]} \sqrt{\rho_0}\,\Re\left(\sum_{n\in\Omega_j} C_n^t H_n^t Y\right)$$

$$- \frac{\rho_0}{2}\sum_{n\in\Omega_j}\sum_{l\in\Omega_j}\Phi_{n,l}\chi_{n,l} + \log P_j \qquad \square$$

$\delta_{opt}(Y)$ can also be expressed as a function of a log likelihood ratio (LLR) as follow:

$$\delta_{opt}(Y) = j \text{ if}$$
$$f(Y|\mathcal{H}_j)P_j \geq f(Y|\mathcal{H}_i)P_i \quad \forall i \neq j \tag{14}$$

$$\Leftrightarrow \log L_i(Y) \geq \log \frac{P_i}{P_j} \quad \forall i \neq j \tag{15}$$

$$\stackrel{[1]}{\Leftrightarrow} \sqrt{\rho_0}\,\Re\left(\sum_{n\in\Omega_j} C_n^t \mathbf{H_n}^t Y - \sum_{n\in\Omega_i} C_n^t \mathbf{H_n}^t Y\right)$$

$$\geq \log \frac{P_i}{P_j} + \frac{\rho_0}{2}\left(\sum_{n\in\Omega_j}\Phi_{n,n}\chi_{n,n} - \sum_{n\in\Omega_i}\Phi_{n,n}\chi_{n,n}\right) \tag{16}$$

$$+ \sum_{n\in\Omega_j}\sum_{\substack{l\in\Omega_j \\ l\neq n}}\Phi_{n,l}\chi_{n,l} - \sum_{n\in\Omega_i}\sum_{\substack{l\in\Omega_i \\ l\neq n}}\Phi_{n,l}\chi_{n,l}\right) \; \forall i \neq j$$

[1] since the LLR is: $\log L_i(Y) = \frac{1}{2}(\left\|Y - \sqrt{\rho_0}\sum_{n\in\Omega_i}\mathbf{H_n}C_n\right\|^2 - \left\|Y - \sqrt{\rho_0}\sum_{n\in\Omega_j}\mathbf{H_n}C_n\right\|^2)$.

The optimal detector thus has to compute $|\Psi|$ correlations of the form (4) (or $|\Psi|-1$ LLR, see (16)) and choose their maximal value to determine the estimated active nodes subset from N nodes. This high complexity is however not tractable for a real implementation. Therefore, we now introduce a MAP-inspired detector, with a lower complexity.

B. It-MAP: Iterative MAP-based detector

The goals of this proposed algorithm are to use an optimal-based detection rule while reducing the complexity of this kind of algorithm -by limiting the considered hypotheses number- and to let the algorithm adapt the iterations number to the active nodes set size. Similarly to NBOMP, this detector is structured by iterative steps, whose each provides a new detected node. Each iteration k is dedicated to the choice of either adding a new detected node to the subset of already detected devices λ_{k-1}, or letting the detected subset as is (which stops the iterative process). Indeed, we employ equation (16) at each iteration with $\Omega_i = \lambda_{k-1}$ and $\Omega_j = \Omega_i \cup \{n\}, n \in \mathcal{N} \setminus \lambda_{k-1}$. Thus, we limit the hypotheses $\mathcal{H}_j$ considered at iteration k to $\Psi_k \subset \Psi$, with Ψ_k the set of $|\mathcal{N} \setminus \lambda_{k-1}|$ subsets, composed of 1 element from the non-detected nodes subset $\mathcal{N} \setminus \lambda_{k-1}$, added to λ_{k-1}: $\Psi_k = \bigcup_{n\in\mathcal{N}\setminus\lambda_{k-1}}\{\lambda_{k-1}, n\}$.

Let p_n denote the activity probability of node n. Let $c_{ItMAP}(k,j)$ be the biased correlator of our iterative MAP-based detector at iteration k, evaluating the likelihood of hypothesis $\mathcal{H}_j$ with respect to hypothesis $\mathcal{H}_i$:

$$c_{ItMAP}(k,j) = \sqrt{\rho_0}\,\Re\left(\sum_{\substack{n\in\Omega_j \\ n\notin\Omega_i}} C_n^t \mathbf{H_n}^t Y\right) - \log \frac{P_i}{P_j}$$

$$- \frac{\rho_0}{2}\left(\sum_{\substack{n\in\Omega_j \\ n\notin\Omega_i}}\Phi_{n,n}\chi_{n,n} + \sum_{\substack{n\in\Omega_j \\ n\notin\Omega_i}}\sum_{l\in\Omega_i}2\Re(\Phi_{n,l}\chi_{n,l})\right) \tag{17}$$

The proposed detection algorithm is It-MAP, algorithm 2, while It-MAP-K (algorithm 3) and It-ML are defined in order to enlighten about some features of the It-MAP detection
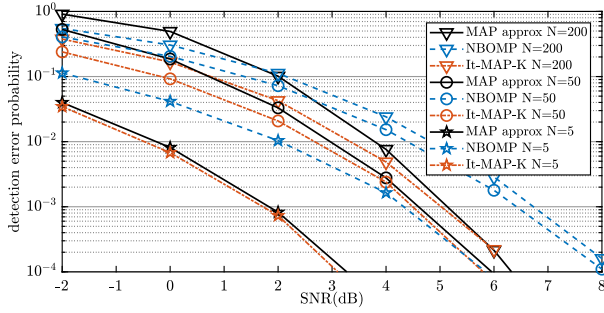
Fig. 1. NBOMP, It-MAP and approximated MAP error rates, for varying N

Fig. 2. NBOMP vs. It-MAP detection error probabilities, for varying N_a

---

Algorithm 2 Iterative MAP-based detector (It-MAP)

Input: Y, $\mathcal{N}$, (C_n, H_n, p_n) $\forall n \in \mathcal{N}$
  $k = 0$; $\lambda_0 = \emptyset$; $\Omega_i = \emptyset$ // *Initialisation*
  while $\exists j \in [0, |\Psi| - 1]$ subject to $(\Omega_j \in \Psi_{k+1} \cap c_{ItMAP}(k+1, j) > 0)$ **do**
    $k = k + 1$
    $\delta_{ItMAP,k}(Y) = \underset{j \in [1, |\Psi|-1] \text{ s.t. } \Omega_j \in \Psi_k}{\arg\max} c_{ItMAP}(k, j)$
    $\lambda_k = \Omega_{\delta_{ItMAP,k}(Y)}$
    $\Omega_i = \lambda_k$
  end while
  return $\delta_{ItMAP}(Y) = j$ s.t. $\Omega_j = \lambda_k$

Algorithm 3 Fixed iterations number It-MAP (It-MAP-K)

Input: Y, $\mathcal{N}$, K, (C_n, H_n, p_n) $\forall n \in \mathcal{N}$
  $k = 0$; $\lambda_0 = \emptyset$; $\Omega_i = \emptyset$ // *Initialisation*
  for $k = 0$ to $K - 1$ **do**
    $\delta_{ItMAPK,k}(Y) = \underset{j \in [1, |\Psi|-1] \text{ s.t. } \Omega_j \in \Psi_k}{\arg\max} c_{ItMAP}(k, j)$
    $\lambda_k = \Omega_{\delta_{ItMAPK,k}(Y)}$
    $\Omega_i = \lambda_k$
  end for
  return $\delta_{ItMAPK}(Y) = j$ s.t. $\Omega_j = \lambda_k$

---

algorithm, and will be used in the numerical results section. It-ML is only different from It-MAP due to its correlator:

$$
\begin{aligned}
c_{ItML}(k, j) = &\sqrt{\rho_0}\, \Re\Big(\sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} C_n^t \mathbf{H_n}^t Y\Big) \\
&- \frac{\rho_0}{2}\Big(\sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \Phi_{n,n}\chi_{n,n} + \sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \sum_{l \in \Omega_i} 2\Re(\Phi_{n,l}\chi_{n,l})\Big)
\end{aligned}
\tag{18}
$$

IV. NUMERICAL RESULTS

This section proposes a simulation based evaluation of the the different detection algorithms performance. We provide numerically computed NBOMP, It-MAP It-ML error probabilities together with the approximated detection error rate of the MAP detector from (25), restricted to the case of $N_a = 1$. The main metric is the detection error probability, and is based on the following definition of a success: only the exact active nodes subset recovery (i.e. no missed detection neither false alarm) corresponds to a successful detection.
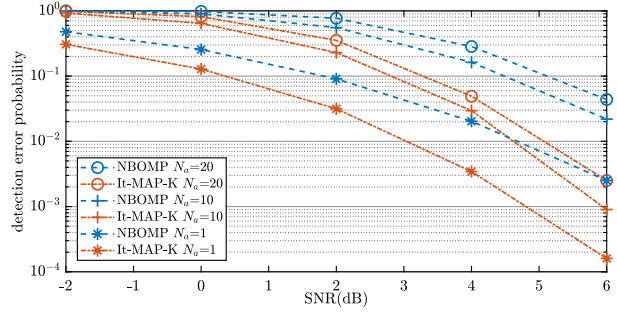
A. Pertinence of the MAP-based detection rule

Our performance evaluation starts with a comparison of the NBOMP detection with the It-MAP detection rule. The It-MAP algorithm used here is the "fixed iterations number" version (algorithm 3). The active node set size N_a is provided as the number of iterations K for both algorithms, and the assumed activity probability $p_n = \frac{N_a}{N}$ is given to It-MAP-K.

Fig. 1 and Fig. 2 show the decrease of the detection error probabilities as a function of the SNR, for varying N and N_a respectively. Fig. 1 shows the error rates for $N_a = 1$, with $N = 5$ to 200 nodes, setting m=32, α=8. The approximated error rate of the optimal detector (MAP) is also depicted to provide an idea of the potential improvement gap (keeping in mind that the line drawn is an upper bound of the actual error probability, not meaningful at low SNR values). One can observe that the larger the SNR, the larger the gap between the NBOMP detection error rate and the approximated MAP error probability. However, the interval between the It-MAP-K and the approximated MAP error rates vanishes as the SNR grows. The upper bound of the MAP error probability, assuming independent sufficient statistics, thus seems to be a good approximation of the error probability of the It-MAP-K.

Moreover, the It-MAP-K outperforms the NBOMP detection, as can be seen in Fig. 2 as well. Indeed, in the latter, the NBOMP and It-MAP-K error rates are drawn for $N = 100$, $m = 32$ and $\alpha = 8$ with N_a equal to 1, 10 or 20. In each case, the number of iterations K realized by the algorithms are set to the respective value of N_a. As a general comment, one can notice that the detection error probability of It-MAP-K is always lower than that of NBOMP. For example, at SNR= 6dB, errors are reduced by at least a factor of 10.

We can thus conclude that, in the cases of unique active user as well as multiple active nodes, the detection rule of our It-MAP algorithm is more efficient than the normalized correlation based detection of NBOMP.

B. Efficiency of the stopping criterion

We now present a second study focusing on the accuracy of the stopping criterion of our iterative MAP-based detector. Indeed, we want to evaluate if the algorithm ends after having detected all the active nodes, and without adding inactive nodes in the detected subset. We respectively refer to Missed Detection (MD) or False Alarm (FA) if it is not the case.
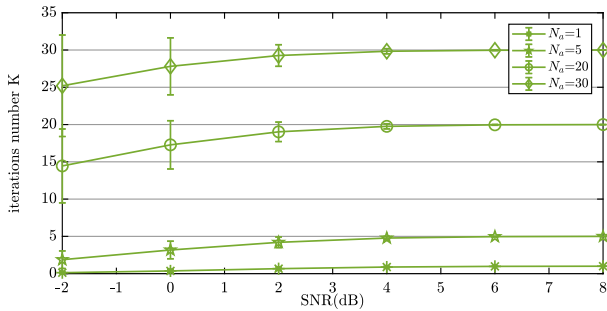
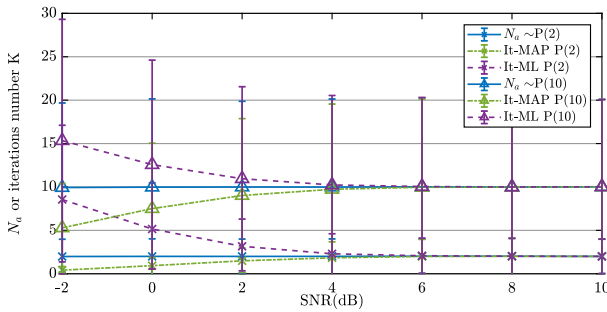Fig. 3. Iterations number (K) mean and variance of It-MAP

Fig. 4. Missed Detection and False Alarm rates of It-MAP

Fig. 5. N_a and K means and variances of It-MAP and It-ML for $N_a \sim \mathcal{P}(\Lambda)$

Fig. 6. Detection error probability for $N_a \sim \mathcal{P}(\Lambda)$

The results have been obtained with the algorithm 2 for It-MAP. Fig. 3 shows the means and variances of the number of iterations K of It-MAP when trying to detect $N = 1$ to 30 actives nodes from a set of size 100, with $m = 32$ and $\alpha = 8$, as a function of the SNR. We can notice that K converges to N_a in any case and that the actual active subset size is even reached from SNR= 4dB. One could thus infer that there are good chances to have a successful detection. One could also observe that the estimated active node number's average is lower than N_a for low SNR values and think that this would imply that the It-MAP algorithm is more subject to MD rather than FA, especially at low SNR. The next figure however brings more details on this point, clearly illustrating that K is only a consequence of the detection process and that $K = N_a$ is *not* synonym of a successful detection. Fig. 4 illustrates the MD and FA rates of It-MAP for SNR ranging from -2dB to 8dB. As a general overview, the 2 groups of curves keep decreasing as the SNR grows so that one metric doesn't decrease at the cost of increasing the other with regard to the SNR increase, and, at SNR= 8dB, the MD rates are lower than 10^{-2} and the FA rates are below 10^{-4}. It thus seems that It-MAP provides a fairly good detection. We can also notice that increasing N_a decreases the MD rate while increasing the FA one and that the MD rate is larger than the FA one, not only at low SNR values. The reason is that It-MAP introduces the term $-\log \frac{P_i}{P_j}$ in the biased correlator, which is a negative term in these simulations. It thus creates a cost on the False Alarms. Said differently, the It-MAP favours MD rather than FA since the "regulation" term, in terms of occurrence probability, of the correlator bias brings a higher threshold onto the correlation part (the threshold being less likely to be overcome, especially
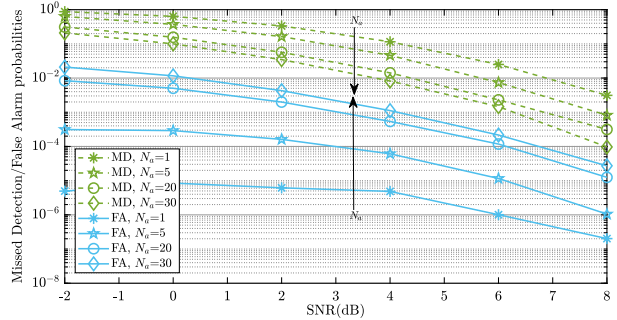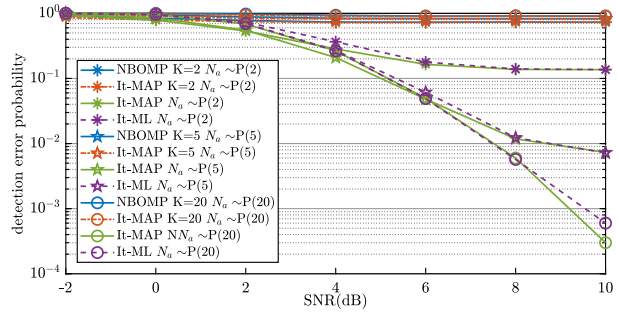
when the noise power is large) thus leading to a higher MD rate. As N_a increases, the nodal activity probability is larger, thus reducing the aforementioned favouritism.

This effect of balancing MD and FA rates can be further exploited by network operators using this detection algorithm, e.g. in order to provide radio resources to the detected active nodes willing to access the network. Indeed, if the uplink activity notifications of the node act as random access requests, depending on the targeted performances, the network providers could choose to preferably allocate some parts of the spectrum to potentially non-active users by being favourable towards FA rather than missing users, thus leading to a higher reliability but also "wasted" resources. Conversely avoiding resource wastage costs a higher latency (since the non-detected users will have to reiterate their requests) and a higher message loss rate (if no reattempt mechanism is employed or expires).

C. Global performance evaluation in a realistic context

Fig. 5, 6 provide the general performance of the proposed detection algorithms in a context mimicking sporadic communications, i.e. with a variable active subset size. Indeed, the simulations were carried out with a Poisson distributed N_a and show the gain of employing the It-MAP detection algorithm rather than NBOMP, or any version with a constraint on the iterations number. The It-ML is also depicted here to show the cost of not knowing the activity rate.

As can be observed in Fig. 5, It-MAP and It-ML easily satisfy the active subset variations, as not only the iterations number means tend to the mean N_a but their variance are also identical. The difference between It-ML and It-MAP rely on the MD/FA favoured feature of the IT-MAP, which doesn't

exist for It-ML. Thus, in presence of high level of noise, the It-ML more often overestimates the active node subset size. Fig. 6 shows the general detection error probability as a function of the SNR. As intended, the iterations-constrained detection algorithms have a very poor accuracy. Even though the It-MAP with a fixed iterations number K acts very slightly better than NBOMP, it is incomparable to It-MAP and It-ML. Indeed, the latter can provide an error probability which is lower than that of NBOMP by up to 3 order of magnitude, for the case of $\Lambda = 20$, at SNR= 10dB. As mentioned above, knowing the activity probability brings a little advantage which corresponds to the gap from It-ML to It-MAP detection error rate, but is clearly negligible at large SNR values.

V. CONCLUSION

In this paper, we have proposed a new detection algorithm, the It-MAP, inspired by the NBOMP and the MAP detectors. It is based on the derivation of the optimal detection rule for multiple access and has a low complexity due to the iterative processing. The It-MAP provides notable performance gains compared to NBOMP, thanks to its detection rule as well as its capability to adapt to unknown active set size. We also have enlighten the possibility to introduce and exploit a regulation term in order to avoid either MD or FA.

As a result, this work provides an efficient tool for users detection in random NOMA, adapted to IoT communication scheme, without requiring the unrealistic prior knowledge of the active users number. The It-MAP detector is thus a first step towards the non-coordinated communications, as ideally required by LPWA IoT networks but also URLLC. Further works on the integration of the channel estimation is however needed in order to release the channel knowledge assumption.

APPENDIX

We derive an approximation of the optimal detection error probability based on the distribution of the sufficient statistic $S_{j,i}(Y)$ which is derived from the LLR expression of the decision rule (16). The sufficient statistic is defined as follow:

$$S_{j,i}(Y) = \sqrt{\rho_0} \, \Re\Big(\sum_{n \in \Omega_j} C_n^t \mathbf{H_n}\,^t Y - \sum_{n \in \Omega_i} C_n^t \mathbf{H_n}\,^t Y \Big) \quad (19)$$

When $\mathcal{H}_j$ holds, no error occurs if $S_{j,i}(Y|\mathcal{H}_j) \geq \eta_{j,i} \; \forall i \neq j$ where the threshold is:

$$\eta_{j,i} = \log \frac{P_i}{P_j} + \frac{\rho_0}{2}\Big(\sum_{n \in \Omega_j} \Phi_{n,n}\chi_{n,n} - \sum_{n \in \Omega_i} \Phi_{n,n}\chi_{n,n}$$
$$+ \sum_{\substack{n \in \Omega_j}} \sum_{\substack{l \in \Omega_j \\ l \neq n}} \Phi_{n,l}\chi_{n,l} - \sum_{\substack{n \in \Omega_i}} \sum_{\substack{l \in \Omega_i \\ l \neq n}} \Phi_{n,l}\chi_{n,l} \Big). \quad (20)$$

Given that $\mathcal{H}_j$ occurs, $S_{j,i}(Y|\mathcal{H}_j) \sim \mathcal{N}_{\mathbb{R}}(M_{j,i}, \sigma_{j,i}^2)$ with:

$$M_{j,i} = \rho_0 Re\Big(\sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \Phi_{n,n}\chi_{n,n} + \sum_{\substack{n \in \Omega_j}} \sum_{\substack{l \in \Omega_j \\ \{n,l\} \not\subset \Omega_i, \; l \neq n}} \Phi_{n,l}\chi_{n,l}$$
$$- \sum_{\substack{n \in \Omega_j}} \sum_{\substack{l \in \Omega_i \\ l \notin \Omega_j}} \Phi_{n,l}\chi_{n,l} \Big) \quad (21)$$

$$\sigma_{j,i}^2 = \frac{\rho_0}{2}\Big(\sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \Phi_{n,n}\chi_{n,n} + \sum_{\substack{n \in \Omega_i \\ n \notin \Omega_j}} \Phi_{n,n}\chi_{n,n}$$
$$+ \sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \sum_{\substack{l \in \Omega_j \\ l \notin \Omega_i, l \neq n}} \Phi_{n,l}\chi_{n,l} + \sum_{\substack{n \in \Omega_i \\ n \cap l \notin \Omega_j, \; l \neq n}} \sum_{l \in \Omega_i} \Phi_{n,l}\chi_{n,l}$$
$$- 2\Re\Big(\sum_{\substack{n \in \Omega_j \\ n \notin \Omega_i}} \sum_{\substack{l \in \Omega_i \\ l \notin \Omega_j}} \Phi_{n,l}\chi_{n,l} \Big)\Big) \quad (22)$$

Let P_e denote the detection error probability, Pe_j be the detection error probability given hypothesis $\mathcal{H}_j$, and $\bar{pe}_{j,i} = P[S_{j,i}(Y|\mathcal{H}_j) \geq \eta_{j,i}]$.

$$Pe = \sum_{j \in [0,|\Psi|-1]} Pe_j . P_j \approx \sum_{j \in [0,|\Psi|-1]} P_j (1 - \prod_{i \neq j} \bar{pe}_{j,i}) \quad (23)$$

where the approximation stands when the $\bar{pe}_{j,i}$ are independent, i.e. only at high SNR when they have a low correlation.

$$\bar{pe}_{j,i} = \int_{\eta_{j,i}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{j,i}^2}} \exp\Big(-\frac{(s_{j,i} - M_{j,i})^2}{2\sigma_{j,i}^2}\Big) ds_{j,i} \quad (24)$$

thus: $$Pe \approx 1 - \sum_{j \in \{0,\Psi-1\}} P_j \prod_{i \neq j} \frac{1}{2} erfc\left(\frac{\eta_{j,i} - M_{j,i}}{\sqrt{2\sigma_{j,i}^2}} \right) \quad (25)$$

REFERENCES

[1] C. Goursaud and J. M. Gorce, "Dedicated networks for IoT: PHY / MAC state of the art and challenges," *EAI Endorsed Transactions on Internet of Things*, vol. 1, p. 150597, oct 2015.
[2] 3GPP, "Further LTE Physical Layer Enhancements for MTC (RP-141865 Revised WI)," 2014.
[3] 3GPP, "Release 13 - description," 2015.
[4] 3GPP, "Further NB-IoT enhancements (RP-171428)," 2017.
[5] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Communications Magazine*, vol. 55, pp. 70–78, feb 2017.
[6] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, C. Stefanovic, P. Popovski, Q. Wang, M. Schellmann, E. Kosmatos, P. Demestichas, M. Raceala-Motoc, P. Jung, S. Stanczak, and A. Dekorsy, "Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks," *IEEE Access*, vol. 6, pp. 28969–28992, 2018.
[7] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive Non-Orthogonal Multiple Access for Cellular IoT: Potentials and Limitations," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 55–61, 2017.
[8] A. K. Fletcher, S. Rangan, and V. K. Goyal, "OnOff Random Access Channels: A Compressed Sensing Framework." https://arxiv.org/pdf/0903.1022v2.pdf, 2009.
[9] R. Xie, H. Yin, X. Chen, and Z. Wang, "Many Access for Small Packets Based on Precoding and Sparsity-Aware Recovery," *IEEE Transactions on Communications*, vol. 64, pp. 4680–4694, nov 2016.
[10] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Boston, MA: Springer US, 2008.