

# Iterative List Detection and Decoding for mMTC

Roberto B. Di Renna, *Student Member, IEEE*, and Rodrigo C. de Lamare, *Senior Member, IEEE*

**Abstract**—The main challenge of massive machine-type communications (mMTC) is the joint activity and signal detection of devices. The mMTC scenario with many devices transmitting data intermittently at low data rates and via very short packets enables its modelling as a sparse signal processing problem. In this work, we consider a grant-free system and propose a detection and decoding scheme that jointly detects activity and signals of devices. The proposed scheme consists of a list detection technique, an  $l_0$ -norm regularized activity-aware recursive least-squares algorithm, and an iterative detection and decoding (IDD) approach that exploits the device activity probability. In particular, the proposed list detection technique uses two candidate-list schemes to enhance the detection performance. We also incorporate the proposed list detection technique into an IDD scheme based on low-density parity-check codes. We derive uplink sum-rate expressions that take into account metadata collisions, interference and a variable activity probability for each user. A computational complexity analysis shows that the proposed list detector does not require a significant additional complexity over existing detectors, whereas a diversity analysis discusses its diversity order. Simulations show that the proposed scheme obtains a performance superior to existing suboptimal detectors and close to the oracle LMMSE detector.

**Index Terms**—Decision feedback receivers, error propagation mitigation, iterative detection and decoding, massive machine-type communication, random access, spatial multiplexing.

## I. INTRODUCTION

MASSIVE machine-type communications (mMTC) has been considered a technology with great potential in future networks. This potential can be widespread across different industries, including healthcare, logistics, manufacturing, process automation, energy, and utilities. Different from the conventional human type communications, mMTC for IoT have unique service features, as transmissions between two MTC devices, low data rates, very short packets and high requirements of energy efficiency and security [1], [2].

In this context, grant-free access is a promising technique to meet the specifications of mMTC. Dividing the short packages in preamble (metadata) and payload (data), the central aggregation node can detect the active devices and estimate their channels in just one transmission. Given the multiple applications for mMTC, each type of device has its own activity behaviour. For example, monitoring devices of high-risk patients in a hospital request access to the network more frequently than gas sensors from a smart home. Thus, it is natural to assume that each device has a different probability of being active ( $\rho_n$ ).

The authors are with the Centre for Telecommunications Studies (CETUC), Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro 22453-900, Brazil (e-mail: robertobrauer@cetuc.puc-rio.br; delamare@cetuc.puc-rio.br). This work was supported in part by the National Council for Scientific and Technological Development (CNPq) and in part by FAPERJ.

As mMTC is a very dense scenario and due to the coherence interval of the channel, even with a small number of active devices (10% of the cell) the reuse of metadata sequences must occur. Naturally, metadata allocation has to adapt with the traffic activity pattern, designating the resources in a random manner [3]. Despite many works [4]–[7] that consider the reuse only in the neighbouring cells, the assumption of intra-cell metadata contamination should be applied [8]. This assumption is directly related to the uplink capacity analysis. For this scenario, the sum-rate expressions besides take metadata contamination into account, must consider the different activity probability of each device.

As envisaged, this kind of network supports a massive number of devices with sporadic activity, this scenario can be interpreted as a sparse signal processing problem [9]. Exploiting the sparsity of the system, detection algorithms should be reformulated to the mMTC scenario. A common approach is to apply a regularization parameter into the cost function, as in [10] wherein Zhu and Giannakis proposed the Sparse Maximum a Posteriori Probability (S-MAP) detection which performs a MAP detection of the new sparse problem, considering a zero-augmented finite alphabet. A variation of the well-known sphere decoder has been proposed in [11] and named as K-Best. In [12], a sparsity constraint has been incorporated in the successive interference cancellation (SA-SIC). In order to avoid matrix inversions, in [13] a version of SA-SIC with sorted QR decomposition and ordered detection based on the activity probability of devices has been reported. Despite its large computational complexity, a solution belonging to the class of Bayesian interference algorithms has been described in [14], which has the advantage of performing the detection without knowing the activity factor  $\rho_n$ .

In the approach proposed in this paper, when a device has to transmit data, it splits the codeword in multiple frames and transmit them in multiple transmission slots. During the same coherence time, the channel is constant and it is estimated from uplink metadata every time the device transmits. In each time slot, each active device selects randomly a metadata sequence from a predetermined codebook and sends the rest of the codeword. As the number of orthogonal metadata sequences is lower than the devices, the system is suitable to frame collisions [3], [9].

In this work, inspired by the joint activity and data detection problem in sparse scenarios, we introduce a detection scheme named activity-aware variable group-list decision feedback (AA-VGL-DF). The proposed AA-VGL-DF scheme consists of a list detector, an  $l_0$ -norm regularized recursive least-squares (RLS) algorithm that exploits the sparsity of the system to adjust the receive filters, and an iterative detection and decoding (IDD) scheme. Inspired by our previous list-detection work, the AA-MF-SIC [15], we employ two list

detection techniques based on the constellation points, thus increasing the accuracy of each symbol detection. In order to reduce the computational complexity, we consider a variable size of each list of constellation points candidates based on the SINR. An IDD scheme based on low-density parity-check (LDPC) codes, which incorporates the  $l_0$ -norm regularized RLS algorithm, the list-detector and takes advantage of the activity probability of each device is also devised for signal detection in mMTC. We then derive uplink achievable sum-rate expressions that take into account metadata collisions, interference and a variable activity probability for each user. An analysis of the computational complexity shows that the AA-VGL-DF detector does not require a significant additional complexity over existing techniques, whereas a diversity analysis discusses the diversity order achieved by the AA-VGL-DF detector. Simulations show that our AA-VGL-DF scheme successfully mitigates the error propagation and approaches the oracle linear minimum mean-squared error (LMMSE) detector performance with a competitive complexity.

The main contributions of this work are:

- 1) The AA-VGL-DF detection scheme with an  $l_0$ -norm regularized RLS algorithm with two list-based strategies, which detects active devices and their symbols;
- 2) An IDD scheme based on LDPC codes modified for mMTC that incorporates the AA-VGL-DF detector;
- 3) A diversity order analysis along with a complexity analysis of AA-VGL-DF and existing approaches based on required floating-point operations (FLOPs);
- 4) A derivation of a closed-form expression for the achievable spectral efficiency which takes metadata collisions, interference and the activity probability of each user;
- 5) A comparative study with simulation results of the AA-VGL-DF and existing techniques.

The organization of this paper is as follows: Section II briefly describes the random access and the channel models. Section III presents the AA-VGL-DF detector, the  $l_0$ -norm regularization and the variable group-list constraint. Section IV introduces the IDD and details the modifications that are suitable for mMTC. Analyses of computational complexity, diversity order and the achievable uplink sum-rate are developed in Section V. Section VI presents the setup for simulations and results while Section VII draws the conclusions.

*Notations:* Matrices and vectors are denoted by bold-faced capital letters and lower-case letters, respectively. The space of complex (real)  $N$ -dimensional vectors is denoted by  $\mathbb{C}^N$  ( $\mathbb{R}^N$ ). The  $i$ -th column of a matrix  $\mathbf{A} \in \mathbb{C}^{M \times N}$  is denoted by  $\mathbf{a}_i \in \mathbb{C}^M$ . The superscripts  $(\cdot)^T$  and  $(\cdot)^H$  stand for the transpose and conjugate transpose, respectively. For a given vector  $\mathbf{x} \in \mathbb{C}^N$ ,  $\|\mathbf{x}\|$  denotes its Euclidean norm.  $\mathbb{E}[\cdot]$  stands for expected value,  $\mathbf{I}$  is the identity matrix and  $\text{diag}[\cdot]$  is to reshape a vector in the main diagonal of a matrix.

## II. SYSTEM MODEL

This section presents the sparse signal model and the system setup considered for the mMTC scenario. The sparse signal model takes into account considerations available in a 3GPP specification and the system setup details the received signal in a given coherence time.

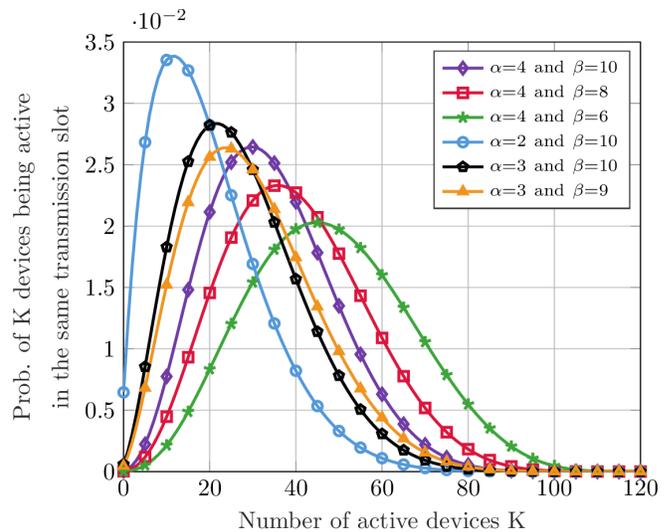


Fig. 1: Model of probability of having  $K$  active devices within a total of  $N = 120$ . All devices become independently active with a probability determined by the random variable  $\rho_n \in [0, 0.1, \dots, 1]$  with a beta distribution.

### A. Sparse signal model

As networks should support many different applications with distinct requirements, naturally each device has its own activity behaviour. In order to model this scenario, we consider the beta-binomial distribution [16] to model the problem as the probability of being active  $\rho_n$  is randomly drawn from a beta distribution, proposed as a traffic model by the 3GPP [17]. Considering  $N$  devices with a single antenna accessing a base station (BS), the probability mass function is given by

$$p(K) = \binom{N}{K} \rho_n^K (1 - \rho_n)^{N-K} \quad (1)$$

where  $\rho_n$  is a random variable with a beta distribution that represents the probability of being active of the  $n$ -th device and  $K$  is the number of active devices out of  $N$  at the same transmission slot. The sparsity of the scenario is modified as soon as each random variable  $\rho_n$  with beta distribution is modelled. Thus, each device has its own activity probability.

Hence, the probability of having  $K$  active devices within a total of  $N$  at the same transmission slot is given by

$$p(K | N, \alpha, \beta) = \frac{\Gamma(N+1)}{\Gamma(K+1)\Gamma(N-K+1)} \frac{\Gamma(N+\alpha)\Gamma(N-K+\beta)}{\Gamma(N+\alpha+\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \quad (2)$$

where  $\Gamma(\cdot)$  is the gamma function,  $\alpha$  and  $\beta$  are real positive parameters that appear as exponents of the random variable  $\rho_n$  and control the shape of the distribution. The average number of active devices is  $\frac{N\alpha}{\alpha+\beta}$  and its variance is  $\frac{N\alpha\beta(\alpha+\beta+N)}{(\alpha+\beta)^2(\alpha+\beta+1)}$ .

In most mMTC applications the devices have low probability of being active. Fig. 1 shows the probability of a specific number of devices to be active at the same time for different values of  $\alpha$  and  $\beta$  in a scenario of  $N = 120$ .

## B. System setup

The massive uplink connectivity scenario illustrated in Fig. 2 is considered, where  $N$  devices with a single antenna access a single base station (BS), equipped with  $M$  antennas. When a device has data to transmit, it splits the codeword in multiple frames and transmit them in multiple transmission slots. In each time slot, each active device selects randomly a metadata sequence from a predetermined codebook and sends the rest of the codeword. Since in practice the BS would have a list of devices that are associated with it, and their unique identifiers, we assume that the metadata sequences are known at the BS. Since these unique identifiers are known to the BS, the metadata sequence is also known at the BS. Given the sporadic activity of devices, they will communicate to the BS only when it is needed, so not all of them will be active during the same coherence time.

As the frame size of mMTC is typically very small (between 10 and 100 bytes) [2], it is possible to assume the devices are synchronized in time. That is, the devices are turned on or turned off in the same transmission slot, as represented in Fig. 2. The duration of a transmission slot ( $\tau = \tau_\phi + \tau_x$ ) is smaller than the coherence time and coherence bandwidth of the channel. The time index  $t$  indicates each transmitted vector in the same transmission slot. As we considered a grant-free random access model, each frame has metadata and data. Thus, the time index indicates how each frame is divided, as in Fig. 2.

The received signal  $\mathbf{y}[t]$  in a given coherence time is organized in a  $M \times 1$  vector that contains the transmitted metadata ( $\phi[t]$ ) or the data ( $\mathbf{x}[t]$ ), as

$$\mathbf{y}[t] = \begin{cases} \mathbf{H} \sqrt{\tau_\phi} \mathbf{B} \phi[t] + \mathbf{v}[t], & \text{if } 1 \leq t \leq \tau_\phi \\ \mathbf{H} \sqrt{\tau_x} \mathbf{B} \mathbf{x}[t] + \mathbf{v}[t], & \text{if } \tau_\phi < t \leq \tau \end{cases} \quad (3)$$

where  $\mathbf{H}$  is the  $M \times N$  channel matrix,  $\mathbf{B}$  is the  $N \times N$  transmission power matrix,  $\mathbf{v}$  is the  $M \times 1$  noise vector, while  $\tau_\phi$  and  $\tau_x$  are the number of metadata and data symbols, respectively. For each time instant  $t$ , the metadata and data are represented by the  $N \times 1$  vectors

$$\begin{aligned} \phi[t] &= \Delta \boldsymbol{\varphi}[t] = [\delta_1 \varphi_1[t], \dots, \delta_N \varphi_N[t]]^T \text{ and} \\ \mathbf{x}[t] &= \Delta \mathbf{s}[t] = [\delta_1 s_1[t], \dots, \delta_N s_N[t]]^T, \end{aligned} \quad (4)$$

where  $\boldsymbol{\varphi}[t]$  and  $\mathbf{s}[t]$  are  $N \times 1$  vectors of symbols from a regular modulation scheme denoted by  $\mathcal{A}$ , as quadrature phase-shift keying (QPSK). The  $N \times N$  diagonal matrix  $\Delta$  controls each device activity in the specific transmission slot, with  $\Pr(\delta_n = 1) = \rho_n$  and  $\Pr(\delta_n = 0) = 1 - \rho_n$ . Thus, each transmitted vector ( $\phi[t]$  or  $\mathbf{x}[t]$ ) is composed by the augmented alphabet  $\mathcal{A}_0$ , where  $\mathcal{A}_0 = \mathcal{A} \cup \{0\}$ .

As in mMTC systems the transmission power  $b$  of each device is different [1], [2], we gather the transmission power component in the  $N \times N$  diagonal matrix  $\mathbf{B} = \text{diag}([\sqrt{b_1}, \sqrt{b_2}, \dots, \sqrt{b_N}])$ . The noise vector  $\mathbf{v}$  is modelled as an independent zero-mean complex-Gaussian  $M \times 1$  vector with variance  $\sigma_v^2$ .

In our work, we consider the block fading model, where a channel realization is constant across a transmission slot

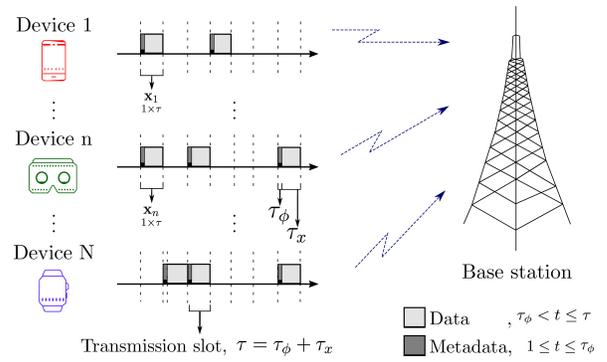


Fig. 2: mMTC single-cell system model. When active, each device transmits  $\tau_\phi$  and  $\tau_x$  symbols of metadata and data, respectively, in the coherence time.

duration and changes independently from slot to slot. The  $M \times N$  channel matrix  $\mathbf{H}$  corresponds to the channel realizations between the BS and devices, modelled as

$$\mathbf{H} = \mathbf{A} \mathcal{N}^{1/2}, \quad (6)$$

where  $\mathbf{H}$  gathers independent fast fading, geometric attenuation and log-normal shadow fading.  $\mathbf{A}$  is the  $M \times N$  matrix of fast fading coefficients circularly symmetric complex Gaussian distributed, with zero mean and unit variance. The  $N \times N$  diagonal matrix  $\mathcal{N}$  models the path loss and shadowing experienced by each device and is modelled as  $10 \log_{10}(\chi) + \omega$ , where  $\chi$  is the signal-to-noise ratio (SNR) and  $\omega$  is a Gaussian random variable with zero mean and variance  $\sigma_\omega^2$  [18]. Thus, each vector  $\mathbf{h}_n$  can be written as

$$\mathbf{h}_n = \mathbf{a}_n \sqrt{\eta_n}, \quad \forall n = 1, \dots, N. \quad (7)$$

The  $\eta_n$  coefficients are assumed to be known at the BS and changes very slowly, reaching a new value just in a new transmission slot. All signal model parameters are described in Table I.

Given the features of mMTC scenarios, the number of devices  $N$  is larger than that of antennas  $M$  at the base station, in a way that it consists of an underdetermined system. However, the transmitted symbols can be detected as their vectors have a sparse structure as the rows corresponding to the inactive users are zero. So, the activity detection problem is reduced to finding the non-zero rows of  $\phi[t]$ .

The motivation of this work is to propose an efficient detection technique for mMTC. Conventional detection techniques are not suitable to deal with the small coherence time and limitation of the orthogonal metadata sequences. Thus, we present an iterative and adaptive detection technique that exploits the sparsity of the system and, using the specifications provided for the mMTC scenario, jointly detects the activity and data of devices, outperforming existing approaches.

## III. VARIABLE GROUP-LIST DECISION FEEDBACK DETECTION

This section details the proposed AA-VGL-DF detection scheme. Unlike our previous work [19], AA-VGL-DF consists of an adaptive receive filter adjusted by  $l_0$ -norm regularized

TABLE I: Description of signal model parameters.

| Parameter       | Description  |
|-----------------|--|
| $M$             | Number of base station antennas;   |
| $N$             | Number of devices;   |
| $K$             | Number of active devices;  |
| $\tau$          | Number of transmitted symbols per trans. slot, given by $\tau = \tau_x + \tau_\phi$ , where $\tau_x$ represents the data and $\tau_\phi$ the metadata; |
| $\rho_n$        | Random variable with a beta distribution that represents the probability of being active of the $n$ -device;   |
| $\mathbf{y}[t]$ | $M \times 1$ received symbol vector of the time instant $t$ ;  |
| $\phi[t]$       | $N \times 1$ metadata vector of the time instant $t$ composed by the augmented alphabet $\mathcal{A}_0$ ;  |
| $\mathbf{x}[t]$ | $N \times 1$ data vector of the time instant $t$ composed by the augmented alphabet $\mathcal{A}_0$ ;  |
| $\Delta$        | $N \times N$ diagonal matrix that controls each device activity in the specific transmission slot;   |
| $\mathbf{B}$    | $N \times N$ diagonal matrix that gathers the transmission power of each device;   |
| $\mathbf{v}[t]$ | noise component, modelled as a independent zero-mean complex-Gaussian $M \times 1$ vector with variance $\sigma_v^2$ ;                                 |
| $\mathbf{H}$    | $M \times N$ channel matrix, where $\mathbf{H} = \mathbf{A}\mathcal{N}^{1/2}$ ;  |
| $\mathbf{A}$    | $M \times N$ matrix of fast fading coefficients;   |
| $\mathcal{N}$   | $N \times N$ diagonal matrix that gathers the path loss and shadowing experienced by each device;  |

RLS with decision feedback and two list detection techniques to reduce the detection error propagation. The detection scheme is illustrated in a block diagram in Fig. 3.

As the transmission slots are separated by metadata and data, AA-VGL-DF has two modes of operation, training mode ( $1 \leq t \leq \tau_\phi$ ) and decision-directed mode ( $\tau_\phi < t \leq \tau$ ). Although the same scheme is used for both cases, in the training mode the focus is to use the metadata to update the RLS algorithm, while the decision-directed mode uses the filter to detect the received data symbols. AA-VGL-DF detects each symbol at a time, per layer. The detection order is updated at each new layer, using the least squares estimation (LSE) criterion. The adaptive receive filter can be decomposed into feedforward and feedback filters. The feedforward one is updated at every new received vector by the  $l_0$ -norm regularized RLS algorithm. The feedback filter is a component that is concatenated to the feedforward filter in order to cancel the interference of the previously detected symbols.

In the first layer, the filter is composed by the feedforward part and obtains a soft estimate of the symbol with the received vector  $\mathbf{y}[t]$ . Inspired by our previous work [15], we apply the shadow area constraints (SAC) criterion to evaluate the reliability of the hard decision of the soft symbol estimate. If the symbol is considered unreliable, a list of candidates drawn from the constellation symbols is generated and the best candidate, chosen by the maximum likelihood criterion, replaces the unreliable symbol. After the first detection, the LSE defines which will be the next layer, the received vector  $\mathbf{y}[t]$  is concatenated with the previous detected symbol and the filter of the selected layer is concatenated with the feedback part. This procedure is repeated until the last layer. At this point, the scheme has three vectors, one with all soft estimates ( $\tilde{\mathbf{d}}_\psi[t]$ ), the second one with the detected symbols ( $\hat{\mathbf{d}}_\psi[t]$ ) and the third one ( $\mathcal{D}_\psi[t]$ ) that keeps the information about the reliability of each soft estimate. Those three vectors are reordered to the original sequence and the second list of

candidates begins. The idea of this list is to perform a group list verification of the most unreliable symbols, after the last detection. With more reliable detected vectors, a more accurate decision of the first filter to be used in the next detection will be taken. In order to define which soft estimates will be rechecked with the second list, we apply the SAC criterion again, but with a larger radius. After the reordering process, if AA-VGL-DF is in the decision-directed mode, the reordered soft estimation vector ( $\tilde{\mathbf{d}}[t]$ ) is converted to LLRs in order to be decoded by the iterative scheme. Otherwise, the reordered vectors of soft estimates and detected symbols ( $\hat{\mathbf{d}}[t]$ ), are used to define the first symbol to be detected in the next received metadata vector.

We will present the proposed AA-VGL-DF scheme in detail in the following subsections. We first detail the adaptive decision feedback structure, the receive filters and the received vector concatenation. Then, we describe the internal verification list scheme and the detection order update. Lastly, we present the external list scheme and the  $l_0$ -norm regularized RLS algorithm.

#### A. Adaptive Decision Feedback Structure

The main idea is to use a feedforward filter to detect the transmitted symbol and a feedback filter to cancel the interference. For each operation mode, both filters are concatenated and written as

$$\mathbf{w}_{\psi_n}[t] = \begin{cases} \mathbf{w}_{\psi_n}^f[t], & n = 1; \\ \left[ \mathbf{w}_{\psi_n}^f{}^T[t], \mathbf{w}_{\psi_n}^b{}^T[t] \right]^T, & n = 2, \dots, N. \end{cases} \quad (8)$$

where  $\mathbf{w}_{\psi_n}[t]$  corresponds to both filters used for the detection of the symbol of the  $n$ -th device (or layer). Both filters update their weights and the detection order  $\psi$  at each new symbol detection. The received vector  $\mathbf{y}[t]$  is concatenated with the  $N \times 1$  vector  $\hat{\mathbf{d}}_{\psi_{n-1}}[t]$  which contains the previously detected symbols as

$$\mathbf{y}_{\psi_n}[t] = \begin{cases} \mathbf{y}[t], & n = 1; \\ \left[ \mathbf{y}^T[t], \hat{\mathbf{d}}_{\psi_{n-1}}^T[t] \right]^T, & n = 2, \dots, N \end{cases} \quad (9)$$

and each soft symbol estimate of the  $n$ -th device is given by

$$\tilde{d}_{\psi_n}[t] = \mathbf{w}_{\psi_n}^H[t] \mathbf{y}_{\psi_n}[t]. \quad (10)$$

Thus, the filter  $\mathbf{w}_{\psi_n}[t]$  and the received vector  $\mathbf{y}_{\psi_n}[t]$  increases in length at each detection. In the last detection,  $\mathbf{w}_{\psi_N}[t]$  and  $\mathbf{y}_{\psi_N}[t]$  are a  $(M + N) \times 1$  vectors, as  $\mathbf{w}_N[t] = \left[ \mathbf{w}_{\psi_{N,1}}^f[t], \dots, \mathbf{w}_{\psi_{N,M}}^f[t], \mathbf{w}_{\psi_{N,M+1}}^b[t], \dots, \mathbf{w}_{\psi_{N,M+N}}^b[t] \right]^T$ .

1) *Internal list:* In order to improve the detection performance, we include a SAC, first presented in our previous work [15], to evaluate the reliability of the soft estimates. As shown in Fig. 3, with the augmented alphabet of a QAM modulation scheme, SAC compares the distance between the soft estimate and all the possible constellation symbols with

$$r = \arg \min_{i \in 1, \dots, (|\mathcal{A}|+1)} \|\mathcal{A}_{0_i} - \tilde{d}_{\psi_n}[t]\|^2, \quad (11)$$

where  $|\mathcal{A}|$  is the modulation order. If the soft estimate falls into the shadow area ( $r > r^{\text{th}}$  or  $r > r^{\text{th}_0}$ ), the estimate

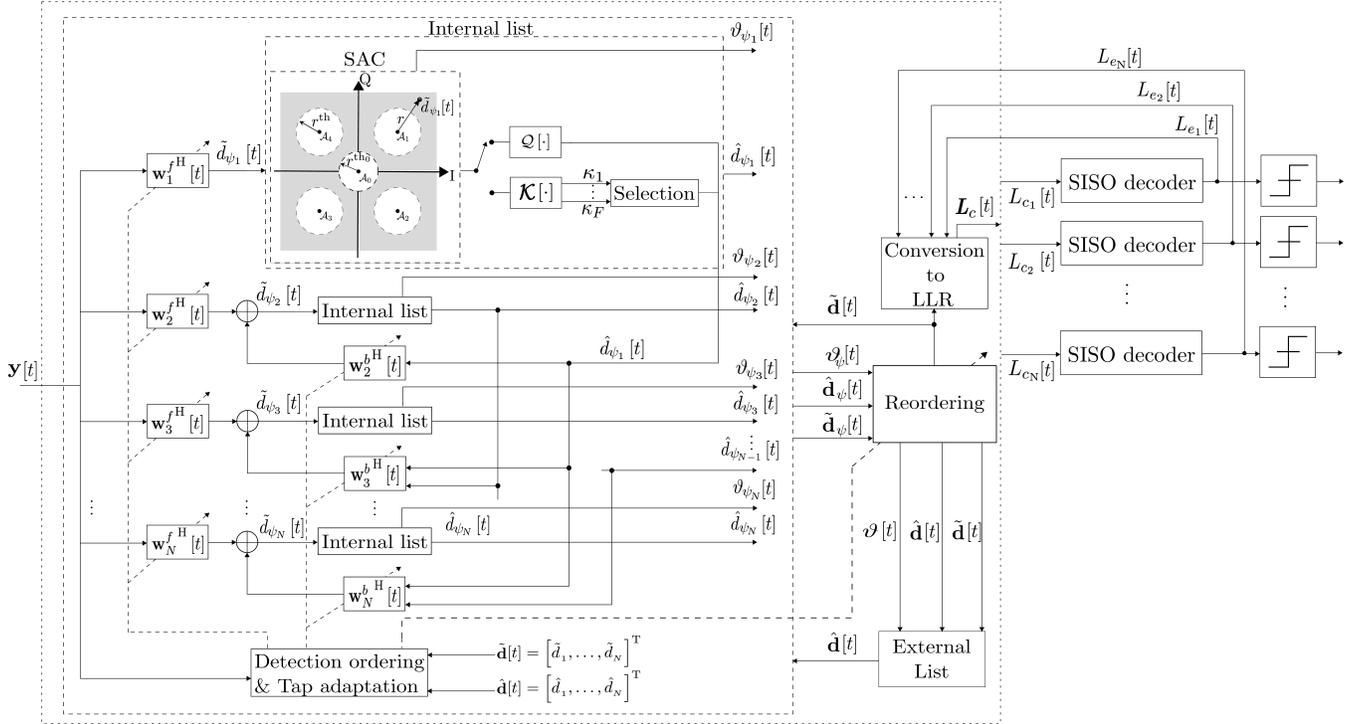


Fig. 3: Detailed structure of the AA-VGL-DF detector and the IDD scheme. To simplify notations, just one received vector is considered in the base station.

is considered unreliable and then  $\tilde{d}_{\psi_n}[t]$  proceeds to the list scheme. Otherwise, it is just quantized to the nearest symbol of the augmented constellation  $\mathcal{A}_0$ , as

$$\hat{d}_{\psi_n}[t] = \mathcal{Q}[\mathbf{w}_{\psi_n}^H[t] \mathbf{y}_{\psi_n}[t]]. \quad (12)$$

The  $r^{\text{th}}$  and  $r^{\text{th}_0}$  radius of each reliability region are defined by the probability of being active of each device [15] and the radius of the region around the zero (inactive device) is the complement of the radius of the regions around the constellation symbols,  $r^{\text{th}_0} = 1 - r^{\text{th}}$ .

The list scheme is a verification of a list of candidates drawn from constellation symbols to the actual detection. The list  $\boldsymbol{\kappa} = [\kappa_1 \cdots \kappa_{(|\mathcal{A}_0|)}]$  is used to select the best candidate according to

$$\boldsymbol{\kappa}_{\text{opt}} = \arg \min_{i \in 1, \dots, (|\mathcal{A}_0|)} \|\mathbf{y}[t] - \hat{\mathbf{h}}_{\psi_n} \kappa_i\|^2, \quad (13)$$

where the vector  $\hat{\mathbf{h}}_{\psi_n}$  contains the estimate of the channel between the device that performs symbol detection and the BS. As the channel estimation is not the focus of this work, we considered the well known linear MMSE (LMMSE) estimation. The estimate of each  $\hat{\mathbf{h}}_{\psi_n}$  is detailed in Section V-B.

The vector with minimum argument  $\boldsymbol{\kappa}_{\text{opt}}$  indicates which candidate  $\kappa$  will replace the quantized version of the unreliable soft symbol estimate  $\tilde{d}_{\psi_n}[t]$ .

2) *Detection order update*: The metric chosen to update the detection order  $\psi$ , is the minimum LSE. At each symbol detection, we compute the  $l_0$ -norm cost function  $\mathcal{J}_j[t]$  for the symbols that were not detected yet. The set  $S_j[t]$  contains the index of the remaining symbols to be detected and is updated with the output  $\psi_j$  of the cost function. Thus, the index of

the chosen filter is stored in the sequence of detection  $\psi$ , represented as

$$\psi_j = \arg \min_{j \in S_j} \mathcal{J}_j[t], \quad (14)$$

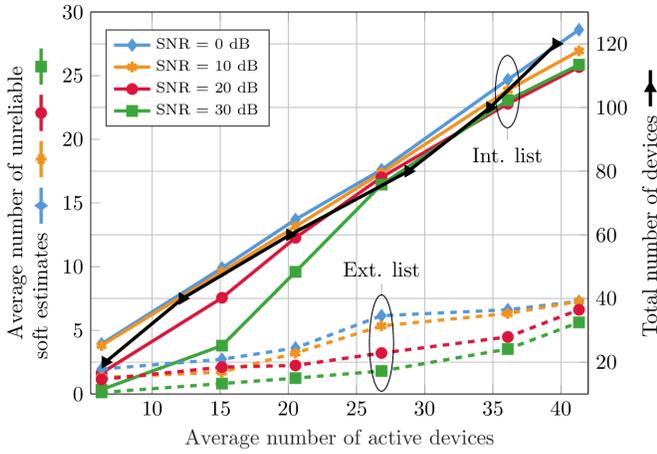
where the indicator  $j$  in (14) belongs to the set  $S_j = \{1, 2, \dots, N\} - \{\psi_1, \psi_2, \dots, \psi_{n-1}\}$ , which contains the index of not yet detected symbols. Hence, the output index of the cost function is removed by the set and this symbol is chosen to be detected. In the next detection, the already detected symbol does not participate in the new cost function computation. Therefore, the  $l_0$ -norm cost functions of each phase, are respectively given by

$$\mathcal{J}_j^\varphi[t] = \sum_{l=0}^t \lambda^{t-l} |\varphi_j[l] - \mathbf{w}_j^H[t] \mathbf{y}_{\psi_n}[t]|^2 + \gamma \|\mathbf{w}_j[t]\|_0, \quad (15)$$

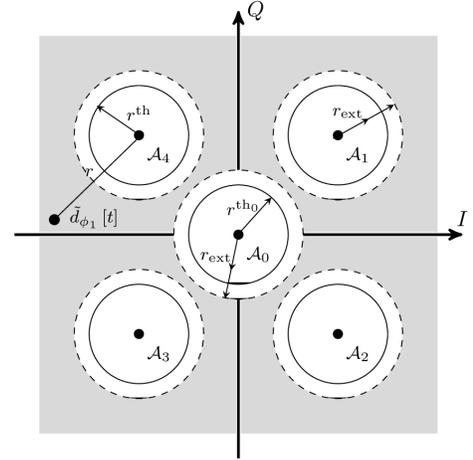
$$\mathcal{J}_j^d[t] = \sum_{l=\tau_\phi+1}^{t-1} \lambda^{t-1-l} |\hat{d}_j[l] - \mathbf{w}_j^H[t] \mathbf{y}_{\psi_n}[t]|^2 + \gamma \|\mathbf{w}_j[t]\|_0 \quad (16)$$

where  $\|\cdot\|_0$  denotes  $l_0$ -norm that counts the number of zero entries in  $\mathbf{w}_j$  and  $\gamma$  is a non-zero positive constant to balance the regularization and, consequently, the estimation error. Moreover,  $0 < \lambda \leq 1$  is the forgetting factor which gives exponentially less weight to older error samples.  $\mathcal{J}_j^\varphi[t]$  represents in (15) the cost function for the metadata mode and  $\mathcal{J}_j^d[t]$  for the data mode.

After the detection of all symbols, the vector with the soft estimates can be reorganized to the original order and integrate the iterative LDPC decoder. However, as for the next received vector AA-VGL-DF will verify which filter should be used first, we included another list detection technique at this point,



(a) The internal list results are the continuous line while the external list values are shown in dashed lines. The probability of being active of each device is randomly drawn from a beta distribution with  $\alpha = 4$  and  $\beta = 8$ .



(b) The radius  $r^{\text{th}}$  and  $r^{\text{th}}$  delimits the reliable regions for the internal list while  $r_{\text{ext}}$  for the external list.

Fig. 4: Parameters of internal and external lists. (a) Number of unreliable soft estimates by the number of active devices, judge by the SAC of the internal and external list scheme and (b) Shadow area constraints for internal and external lists.

in order to provide more refined information to the first cost function of the next received vector.

3) *External list*: The idea of the external list is to carry out a group list verification of the most unreliable symbols, after the last detection. In the internal list block, AA-VGLDF also keeps the information about the reliability of each soft estimate. The  $N \times 1$  binary vector  $\vartheta_\psi[t]$  gathers this information as  $\vartheta_{\psi_n}[t] = 1$  when a soft estimate falls into the shadow area and  $\vartheta_{\psi_n}[t] = 0$ , otherwise. After the reordering process, the external list block receives the estimated symbols  $\tilde{\mathbf{d}}[t]$ , the detected symbols  $\hat{\mathbf{d}}[t]$  and the vector with the reliability information,  $\vartheta[t]$ .

With the knowledge of which symbols had a reliable soft estimate, the external list generates all possible combinations  $\mathcal{G}$  of the symbols of the considered augmented alphabet  $\mathcal{A}_0$  and gathers all these vectors in an  $\nu \times \mathcal{G}$  matrix  $\mathbf{G}$  given by

$$\mathbf{G} = \begin{bmatrix} \mathcal{A}_{0_1} & \mathcal{A}_{0_1} & \mathcal{A}_{0_1} & \cdots & \mathcal{A}_{0_3} & \cdots & \mathcal{A}_{0_{|\mathcal{A}_0|}} \\ \mathcal{A}_{0_1} & \mathcal{A}_{0_2} & \mathcal{A}_{0_1} & \cdots & \mathcal{A}_{0_1} & \cdots & \mathcal{A}_{0_{|\mathcal{A}_0|}} \\ \mathcal{A}_{0_1} & \mathcal{A}_{0_1} & \mathcal{A}_{0_2} & \cdots & \mathcal{A}_{0_2} & \cdots & \mathcal{A}_{0_{|\mathcal{A}_0|}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathcal{A}_{0_1} & \mathcal{A}_{0_1} & \mathcal{A}_{0_1} & \cdots & \mathcal{A}_{0_4} & \cdots & \mathcal{A}_{0_{|\mathcal{A}_0|}} \end{bmatrix}, \quad (17)$$

where  $\nu$  is the number of soft estimates considered unreliable. With the complete candidate vector matrix, the verification of the most appropriate vector occurs as in the internal list, as follows:

$$\mathbf{g}_{\text{opt}} = \arg \min_{i \in 1, \dots, \mathcal{G}} \left\| \mathbf{y}[t] - \sum_{j=1}^{\nu} \hat{\mathbf{h}}_j g_{j,i} \right\|^2, \quad (18)$$

where  $\hat{\mathbf{h}}_j$  is the estimated channel of the unreliable symbol to be verified. The vector candidate  $\mathbf{g}_{\text{opt}}$  is chosen and its values replace the ones considered unreliable in  $\hat{\mathbf{d}}[t]$  in order to proceed to the detection, ordering and parameter estimation of the next received vector.

As an example, let us suppose that we have 4 unreliable estimates (non-zero elements) in vector  $\vartheta_\psi[t]$ . Considering a QPSK modulation, the augmented alphabet  $\mathcal{A}_0$  would have 5 elements and  $\mathcal{G}$  would be a  $4 \times 70$  matrix, considering all the 70 combination possibilities of 5 possible symbols in 4 unreliable estimates.

Since mMTC is a crowded scenario, building the external  $\mathbf{G}$  matrix with all possible combinations would be impractical. Thus, considering the distribution of probability of being active of devices, we notice in Fig. 4a that just a few symbols are considered unreliable by the internal list per received vector and this number is reduced as the SNR grows. Therefore, we considered a constraint in order to reduce the computational complexity of the external list. Instead of verifying all the possibilities of all unreliable soft estimates, we check the worst cases, that is, the soft estimates considered more unreliable. Thus, we define another radius in the SAC in order to designate which unreliable symbols should be or not in the external list. As the number of unreliable estimates varies with the SNR, we choose the radius, represented in Fig. 4b, as  $r_{\text{ext}} = r^{\text{th}} \left[ \left( \frac{M}{\hat{K}} \right) + \left( N \frac{\sigma_x^2}{\sigma_v^2} \right) \right]$  since it follows the increase of the average SNR value. Thus, the reliability of  $\tilde{\mathbf{d}}[t]$  will be rechecked, with the new radius  $r_{\text{ext}}$ . As we have a  $\tilde{\mathbf{d}}[t]$ , we have an estimation of which device is active or not, given by  $\hat{K}$ . So, the number  $\nu$  of considered unreliable soft estimates is reduced as the SNR value grows, as shown in Fig. 4a. In the next subsection, we detail the proposed  $l_0$ -norm regularized RLS algorithm. Other list detection strategies can also be considered [20], [21].

### B. $l_0$ -norm Regularized RLS Algorithm

In order to exploit the sparse activity of devices and compute the parameters of the proposed DF detector without the need for explicit channel estimation, we devise an  $l_0$ -norm regularized RLS algorithm that minimizes the cost function.

Approximating the value of the  $l_0$ -norm [26], the cost function in (15) can be rewritten as

$$\begin{aligned} \mathcal{J}_j^\varphi [t] &= \sum_{l=0}^t \lambda^{t-l} |\varphi_j [l] - \mathbf{w}_j^H [t] \mathbf{y}_{\psi_n} [t]|^2 \\ &+ \gamma \sum_{p=1}^{2M} (1 - \exp(-\xi |w_{j,p} [t]|)), \end{aligned} \quad (19)$$

where the parameter  $\xi$  regulates the range of the attraction to zero on small coefficients of the filter. Thus, taking the partial derivatives for all entries  $t$  of the coefficient vector  $\mathbf{w}_j [t]$  in (19) and setting the results to zero, yields

$$\begin{aligned} \mathbf{w}_j [t] &= \mathbf{w}_j [t-1] + \mathbf{k} [t] \epsilon_n^* [t] \\ &- \gamma \xi \operatorname{sgn} (w_{j,p} [t]) \exp(-\xi |w_{j,p} [t]|) \end{aligned} \quad (20)$$

where  $\mathbf{k} [t]$  is the gain vector and  $\operatorname{sgn}(\cdot)$  is a component-wise sign function defined as

$$\operatorname{sgn} (w_{j,p} [t]) = \begin{cases} w_{j,p} [t] / |w_{j,p} [t]|, & w_{j,p} [t] \neq 0; \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

In order to reduce computational complexity in (20), the exponential function is approximated by the first order of the Taylor series expansion, given by

$$\exp(-\xi |w_{j,p} [t]|) \approx \begin{cases} 1 - \xi |w_{j,p} [t]|, & |w_{j,p} [t]| \leq 1/\xi; \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

As the exponential function is positive, the approximation of (22) is also positive. In this way, (20) becomes

$$\begin{aligned} \mathbf{w}_j [t] &= \mathbf{w}_j [t-1] + \mathbf{k} [t] \epsilon_n^* [t] \\ &- \gamma \xi \operatorname{sgn} (w_{j,p} [t]) f_\xi (w_{j,p} [t]) \end{aligned} \quad (23)$$

where the function  $f_\xi (w_{j,p} [t])$  is given by

$$f_\xi (w_{j,p} [t]) = \begin{cases} \xi^2 (w_{j,p} [t]) + \xi, & -1/\xi \leq w_{j,p} [t] < 0; \\ \xi^2 (w_{j,p} [t]) - \xi, & 0 \leq w_{j,p} [t] \leq 1/\xi; \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

We notice that the function  $f_\xi (w_{j,p} [t])$  in (23) imposes an attraction to zero of small coefficients. So, if the value of  $w_{j,p} [t]$  is not equal or in the range  $[-1/\xi, 1/\xi]$ , no additional attraction is exerted. Thus, the convergence rate of near-zero coefficients of parameters of devices in mMTC applications that exhibit sparsity will be increased [26]. The pseudo-code, which also considers an IDD scheme with AA-VGL-DF, is described in Algorithm 1. Alternatively, a designer can employ other more sophisticated adaptive algorithms for parameter estimation [22]–[25].

#### IV. PROPOSED SOFT INFORMATION PROCESSING AND DECODING

In order to devise an IDD scheme, we incorporate the detected symbols by AA-VGL-DF in an iterative soft information decoding scheme. Unlike existing approaches such as [27], we incorporate the probability of each device being active in the mMTC scenario in the computation of each a priori probability symbol, which avoids the need for channel estimation.

The *a priori* probabilities are computed based on the extrinsic LLRs  $L_{e_n}^z [t]$ , provided by the LDPC decoder. In the first iteration, all  $L_{e_n}^z [t]$  are zero and, assuming the bits are statistically independent of one another, the *a priori* probabilities are calculated as

$$\Pr (x_n [t] = \bar{x}) = \sum_{\bar{x} \in \mathcal{A}_0} \bar{x} \left( \prod_{z=1}^{M_c} [1 + \exp(-\bar{x}^z L_{e_n}^z [t])]^{-1} \right), \quad (25)$$

where  $M_c$  represents the total number of bits of symbol  $\bar{x}$ , the superscript  $z$  indicates the  $z$ -th bit of symbol of  $\bar{x}$ , in  $\bar{x}^z$  (whose value is  $(+1, -1)$ ). As each device has a different activity probability  $\rho_n$ , the *a priori* probabilities should take into account, as

$$\begin{aligned} \rho_n + (1 - \rho_n) \Pr (x_n [t] = \bar{x}), & \text{ if } (\bar{x}^1 \text{ and } \bar{x}^2) = 0, \\ (1 - \rho_n) \Pr (x_n [t] = \bar{x}), & \text{ otherwise.} \end{aligned} \quad (26)$$

where in the next iteration of the scheme, the new *a priori* probabilities incorporates the probability that the  $n$ -th device is active and the extrinsic LLR values. As described in Section II, the probabilities  $\rho_n$  are randomly drawn from a beta distribution.

As the output of the proposed receive filter has a large number of independent variables, we can approximate it as a Gaussian distribution [28]. Hence, we approximate  $\tilde{d}_n [t]$  by the output of an equivalent AWGN channel with  $\tilde{d}_n [t] = \mu_n [t] x_n [t] + b_n [t]$ . Therefore, the likelihood function  $P(\tilde{d}_n [t] | \bar{x})$  is approximated by

$$P(\tilde{d}_n [t] | \bar{x}) \approx \frac{1}{\pi \zeta_n^2 [t]} \exp\left(-\frac{1}{\zeta_n^2 [t]} |\tilde{d}_n [t] - \mu_n [t] \bar{x}|^2\right), \quad (27)$$

where the mean  $\mu_n [t]$  is given by

$$\begin{aligned} \mu_n [t] &= \mathbb{E} \left\{ \tilde{d}_n [t] x_n [t] \right\} = \mathbb{E} \left\{ \mathbf{w}_n^H [t] \mathbf{y}_n [t] x_n [t] \right\} \\ &\approx \mathbf{w}_n^H [t] \left( \sum_{p=1}^{t-1} \lambda^{t-1-p} \mathbf{y}_n [p] x_n [p] \right). \end{aligned} \quad (28)$$

Note that  $x_n$  is the previously detected symbol. In the first case,  $\mu_n [t] = \mathbf{w}_n^H [t] \mathbf{y}_n [t]$ . Each  $b_n [t]$  is a zero-mean complex Gaussian variable with variance  $\zeta_n^2 [t]$  as

$$\begin{aligned} \zeta_n^2 [t] &= \operatorname{var} \left\{ \tilde{d}_n [t] \right\} = \mathbb{E} \left\{ \|\tilde{d}_n [t]\|^2 \right\} - \mu_n^2 [t] \\ &= \mathbf{w}_n^H [t] \mathbb{E} \left\{ \mathbf{y}_n [t] \mathbf{y}_n^H [t] \right\} \mathbf{w}_n [t] - \mu_n^2 [t] \\ &\approx \mathbf{w}_n^H [t] \left( \sum_{p=1}^t \lambda^{t-p} \mathbf{y}_n [p] \mathbf{y}_n^H [p] \right) \mathbf{w}_n [t] - \mu_n^2 [t]. \end{aligned} \quad (29)$$

Then, the extrinsic LLRs computed by the AA-VGL-DF detector for the  $z$ -th bit ( $z \in \{1, \dots, M_c\}$ ) of the symbol  $x_n$  transmitted by the  $n$ -th device are given by

$$L_{c_n}^z [t] = \log \frac{\sum_{\bar{x} \in \mathcal{A}_z^{+1}} \Pr(\tilde{d}_n [t] | \bar{x}) \Pr(\bar{x})}{\sum_{\bar{x} \in \mathcal{A}_z^{-1}} \Pr(\tilde{d}_n [t] | \bar{x}) \Pr(\bar{x})} - L_{e_n}^z [t] \quad (30)$$

where  $\mathcal{A}_z^{+1}$  is the set of  $2^{M_c-1}$  hypotheses of  $\bar{x}$  for which the  $z$ -th bit is  $+1$  (analogously for  $\mathcal{A}_z^{-1}$ ).

---

**Algorithm 1** Proposed IDD with AA-VGL-DF
 

---

1. Initialization:  $M, N, \rho, \xi, \gamma, \lambda, \mathbf{P}_{\psi_n} = \rho \mathbf{I}_M$   
 % For training mode,  
 % For each metadata sequence  $\hat{\mathbf{d}}[t]$  and  $\mathbf{y}_{\psi_n}[t]$ ,
  2. Compute the Kalman gain vector  
 $\mathbf{k}_{\psi_n}[t] = (\mathbf{P}_{\psi_n}[t] \mathbf{y}_{\psi_n}[t]) / (\lambda + \mathbf{y}_{\psi_n}^H[t] \mathbf{P}_{\psi_n}[t] \mathbf{y}_{\psi_n}[t])$ ;
  3. Estimate  $\tilde{\mathbf{d}}_{\psi_n}[t] = \mathbf{w}_n^H[t] \mathbf{y}_n[t]$ ;
  4. Update the error value with  $\epsilon_{\psi_n}[t] = \hat{\mathbf{d}}_{\psi_n}[t] - \tilde{\mathbf{d}}_{\psi_n}[t]$ ;
  5. Update the filters with Eq. (23);
  6. Update the auxiliary matrix  
 $\mathbf{P}_{\psi_n}[t] = \lambda^{-1} (\mathbf{P}_{\psi_n}[t] - \mathbf{k}_{\psi_n}[t] \mathbf{y}_{\psi_n}^H[t] \mathbf{P}_{\psi_n}[t])$ ;
  7. Concatenate  $\mathbf{y}_{\psi_n}[t]$  with  $\hat{\mathbf{d}}_{\psi_n}[t]$ ;
  8. Update the sequence of detection with Eq.(11);
  - % For decision-directed mode,
  9. Compute the *a priori probability* with Eqs. (25) and (26);
  10. Repeat steps 2. to 6.;
  11. Evaluate the reliability of the soft estimation  $\tilde{\mathbf{d}}_{\psi_n}[t]$  with SAC and proceeds with the internal list if it is judged as unreliable;
  12. Update the sequence of detection with the output of 11;
  13. Proceed with the update of  $\boldsymbol{\vartheta}_{\psi}[t]$ ,  $\mathbf{y}_{\psi_n}[t]$  and  $\mathbf{w}_{\psi_n}[t]$ ;
  14. After all detections, update  $\hat{\mathbf{d}}[t]$  with the external list;
  15. Compute  $\mu_{\psi_n}[t]$  and  $\zeta_{\psi_n}^2[t]$  with Eqs. (28) and (29);
  16. Verify the likelihood function  $P(\tilde{\mathbf{d}}_n[t] | \bar{\mathbf{x}})$  with Eq. (27);
  17. Compute the LLR value according to Eq.(30).
- 

## V. ANALYSIS OF THE AA-VGL-DF ALGORITHM

In this section, the computational complexity required by the AA-VGL-DF algorithm is evaluated and both the diversity order achieved by the AA-VGL-DF detector and the achievable rate of the uplink transmission from the  $n$ -th user are discussed.

### A. Computational Complexity

The computational complexity of AA-VGL-DF is analyzed below by counting each required numerical operation in terms of complex FLOPs. In particular, Table II compares the number of required FLOPs, for a different number of devices  $N$ , receive antennas  $M$  and the  $\mathcal{G}$  group size. We consider both well-known algorithms as linear minimum-mean-squared-error (LMMSE) and modifications for mMTC, as SA-SIC [12], SA-SIC with A-SQRD [13] and AA-RLS-DF [19].

Including the internal list scheme in our previous work [19], AA-RLS-DF, as shown in Fig. 5, results in just a slight complexity increase. Recalling that  $\mathcal{G}$  is the number of combinations of the unreliable soft estimates, the upper bound of AA-VGL-DF is a version where there are not constraints in the lists. As verified in Fig. 4a, the number of unreliable soft estimates increase as the number of devices rises. The computational cost of AA-VGL-DF is comparable with a standard DF detector with an RLS algorithm. Since many other considered algorithms have a similar computational cost, AA-VGL-DF has a competitive complexity when compared with other schemes.

### B. Uplink Sum-Rate

As seen that the mMTC has an amount of features that distinguish from the standard massive MIMO communications,

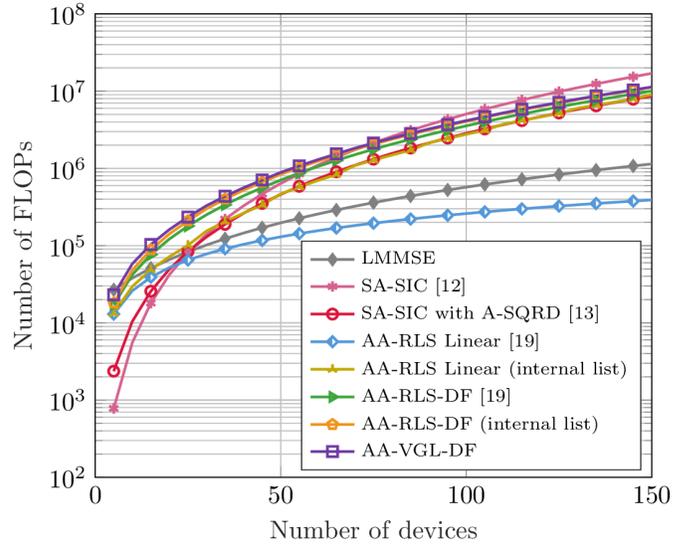


Fig. 5: Comparison of complexity of considered algorithms. The values chosen were  $M = 20$ ,  $\tau_\phi = N/2$  and the variables related to the lists  $\mathcal{G}$  reach the maximum of 5. Just 25% of  $\boldsymbol{\vartheta}$  is equal to 1, following the binomial distribution as  $\alpha = 4$  and  $\beta = 8$ .

we compute the uplink sum-rate considering our detector. Whereas the scenario implies the consideration of a different number of active devices at the same transmission slot and the probability of collision due to reuse of metadata sequences, we compute the achievable rate of each device. We take into account all the possible contamination events from the active devices, as the number of transmission slots are large enough. We also considered that the BS can estimate the number of active devices, as well as the average channel energy  $\eta_n$ . In this way, each device has the knowledge of its channel energy and is able to associate it with its rate, as both parameters are broadcast by the BS.

Differently from the literature, the expressions derived take into account, beyond the filter computed by each approach, the probability of metadata collisions, the probability of having a specific number of active devices and different features of each device, as variable activity probability, transmission power and the path loss and shadowing experienced.

*Theorem 1:* An approximation of a lower bound of the maximal achievable sum-rate (in bits per symbol) is

$$R = \sum_{K=1}^N p(K) K \sum_{c=0}^{K-1} p(c|K) \mathbb{E}_{\{\eta\}} [\mathcal{R}(\mathcal{C}_i, K, \{\eta\})] \quad (31)$$

where  $p(K)$ , given by (1), is the probability of having  $K$  active devices in a total of  $N$  and  $p(c|K)$  is the probability of having  $c$  devices with the same metadata sequence of the  $i$ -th device being observed out of  $K$  active devices and  $\mathcal{C}_i$  refers to a set of those contaminator devices. The expression of  $p(c|K)$  is given by

$$p(c|K) = \binom{K-1}{c} \left( \frac{1}{\tau_\phi} \right)^c \left( 1 - \frac{1}{\tau_\phi} \right)^{K-1-c}. \quad (32)$$

The procedure and the derivation of the first summations are given by [3].  $\mathbb{E}_{\{\eta\}}$  designates the expectation with respect

TABLE II: FLOPs counting of considered techniques in detail.

| Algorithms                    | Required number of FLOPs   |
|-------------------------------|--|
| LMMSE                         | $2M^3 + 4(N+1)M^2 + 2(N^2 + N + 1)M - (N^2 + N)$                           |
| SA-SIC [12]                   | $ \mathcal{A}_0  (N^3 + N^2 + 6)$  |
| SA-SIC A-SQRD [13]            | $2N^3 + 4(M+1)N^2 + (M-1)N$  |
| AA-RLS Linear                 | $(6M^2 + 10M)N$  |
| AA-RLS Linear (internal list) | $[6M^2 + 10M + \vartheta_n(2M \mathcal{A}_0 )]N$                           |
| AA-RLS-DF [19]                | $\sum_{i=1}^N [6(M+i)^2 + 10(M+i)]$  |
| AA-RLS-DF (internal list)     | $\sum_{i=1}^N [6(M+i)^2 + 10(M+i) + \vartheta_n(2M \mathcal{A}_0 )]$       |
| AA-VGL-DF                     | $\sum_{i=1}^N [6(M+i)^2 + 10(M+i) + \vartheta_n(2M \mathcal{A}_0 )] + 2MG$ |

to  $\eta_j, j \in \{i, \mathcal{C}_i\}$  and  $\mathcal{R}(\mathcal{C}_i, K, \{\eta\})$  is a lower bound on the maximal achievable rate of device  $i$  conditioned on a collider set with indices  $\mathcal{C}_i$  within  $K$  active devices is given by

$$\mathcal{R}(\mathcal{C}_i, K, \{\eta\}) = \log_2(1 + \text{SINR}(\mathcal{C}_i, K, \{\eta\})). \quad (33)$$

1) *Perfect Channel Estimation*: We first consider the case when the BS has perfect CSI, i.e., it has the perfect knowledge of  $\mathbf{H}$ . Therefore, the channel capacity  $C$ , recalling  $\tilde{\mathbf{d}}[t]$  and  $\mathbf{y}[t]$  in (10) and suppressing the time index for simplicity, is

$$\begin{aligned} C &= \max_{p_{\tilde{\mathbf{a}}}(\tilde{\mathbf{d}})} I(\mathbf{y}; \tilde{\mathbf{d}}) \\ &= \max_{p_{\tilde{\mathbf{a}}}(\tilde{\mathbf{d}})} H(\mathbf{y}) - H(\tilde{\mathbf{d}}|\mathbf{y}) = \max_{p_{\tilde{\mathbf{a}}}(\tilde{\mathbf{d}})} H(\mathbf{W}^H \mathbf{y}) - H(\mathbf{W}^H \mathbf{v}) \end{aligned} \quad (34)$$

where  $H$  is the differential entropy and  $I$  the mutual information. Thus, as the considered signals are Gaussian, the mutual information is given by [29]

$$\begin{aligned} I(\mathbf{y}; \tilde{\mathbf{d}}) &= \\ \log_2(\det(\mathbb{E}[\mathbf{W}^H \mathbf{y} \mathbf{y}^H \mathbf{W}])) &- \log_2(\det(\mathbb{E}[\mathbf{W}^H \mathbf{v} \mathbf{v}^H \mathbf{W}])). \end{aligned} \quad (35)$$

Thus, computing  $\Omega = \mathbb{E}[\mathbf{W}^H \mathbf{y} \mathbf{y}^H \mathbf{W}]$ , we have to rewrite the following:

$$\begin{aligned} \Omega &= \mathbb{E} \left[ \sum_{j=1}^K \mathbf{w}_j^H y_j y_j^H \mathbf{w}_j \right] \\ &= \mathbb{E} \left[ \sum_{j=1}^K \mathbf{w}_j^H (\mathbf{h}_j \sqrt{b_j} x_j + \mathbf{v}) (\mathbf{h}_j \sqrt{b_j} x_j + \mathbf{v})^H \mathbf{w}_j \right] \\ &= \mathbb{E} \left[ \sum_{j=1}^K \left( \mathbf{w}_j^H \mathbf{h}_j \sqrt{b_j} x_j x_j^H \sqrt{b_j} \mathbf{h}_j^H \mathbf{w}_j \right) + \right. \\ &\quad \left. \left( \mathbf{w}_j^H \mathbf{h}_j \sqrt{b_j} x_j v_j^H \mathbf{w}_j \right) + \right. \\ &\quad \left. \left( \mathbf{w}_j^H v_j x_j^H \sqrt{b_j} \mathbf{h}_j^H \mathbf{w}_j \right) + \left( \mathbf{w}_j^H v_j v_j^H \mathbf{w}_j \right) \right]. \end{aligned} \quad (36)$$

As the main objective is to compute the maximum achievable rate of a device  $i$  out of  $K$  active devices in the same

time instant  $t$  we have,

$$\begin{aligned} \Omega &= \mathbb{E} \left[ \sum_{j=1}^K \left( \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} x_j x_j^H \sqrt{b_j} \mathbf{h}_j^H \mathbf{w}_i \right) + \left( \mathbf{w}_i^H \mathbf{v} \mathbf{v}^H \mathbf{w}_i \right) \right] \\ &= \mathbb{E} \left[ \left| \mathbf{w}_i^H \mathbf{h}_i \sqrt{b_i} x_i \right|^2 + \sum_{j=1, j \neq i}^K \left| \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} x_j \right|^2 + \left| \mathbf{w}_i^H \mathbf{v} \right|^2 \right] \end{aligned} \quad (37)$$

where the first term is the signal of interest and the other additive terms are treated as a Gaussian noise. Thus, substituting (37) in (35), we get the expression of the signal-to-noise-plus-interference for the fixed channel realization  $\mathbf{H}$ , as

$$\text{SINR} = \frac{\left| \mathbf{w}_i^H \mathbf{h}_i \sqrt{b_i} \right|^2}{\sum_{j=1, j \neq i}^K \left| \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} \right|^2 + \|\mathbf{w}_i^H \mathbf{v}\|^2}. \quad (38)$$

2) *Imperfect Channel Estimation*: In practice, the channel matrix  $\mathbf{H}$  has to be estimated at the BS. Thus, we define the computation of the LMMSE estimate of the channel estimate of the  $n$ -th device, as

$$\begin{aligned} \mathbf{y}_n &= \mathbf{Y}_\phi \boldsymbol{\varphi}_n^H = \sum_{n' \in N} \left( \sqrt{\tau_\phi b_{n'}} \mathbf{h}_{n'} \boldsymbol{\varphi}_{n'}^H + \mathbf{V}_\phi \right) \boldsymbol{\varphi}_n^H, \\ &= \sqrt{\tau_\phi b_n} \mathbf{h}_n + \\ &\quad \sum_{n' \neq n} \sqrt{\tau_\phi b_{n'}} \mathbf{h}_{n'} \boldsymbol{\varphi}_{n'}^H \boldsymbol{\varphi}_n^H + \mathbf{V}_\phi \boldsymbol{\varphi}_n^H \end{aligned} \quad (39)$$

where  $\boldsymbol{\varphi}_n$  is the  $1 \times \tau_\phi$  metadata vector of the  $n$ -th device,  $\mathbf{V}_\phi$  is the  $M \times \tau_\phi$  noise matrix and the components of  $(\mathbf{V}_\phi \boldsymbol{\varphi}_n^H)$  are i.i.d., as  $\|\boldsymbol{\varphi}_n\|^2$ . Then, the LMMSE estimate of  $\mathbf{h}_n$ ,  $\hat{\mathbf{h}}_n$  is

$$\begin{aligned} \hat{\mathbf{h}}_n &= \frac{\mathbb{E}\{\mathbf{y}_n^H \mathbf{h}_n\}}{\mathbb{E}\{\mathbf{y}_n \mathbf{y}_n^H\}} \mathbf{y}_n, \\ &= \frac{\eta_n \sqrt{\tau_\phi b_n}}{\sum_{n' \in N} \tau_\phi b_{n'} \eta_{n'} |\boldsymbol{\varphi}_{n'}^H \boldsymbol{\varphi}_n^H|^2 + \sigma_v^2} \mathbf{y}_n. \end{aligned} \quad (40)$$

Thus,  $\hat{\mathbf{H}}$  is the  $N \times M$  matrix of channel estimate. We denote  $\boldsymbol{\mathcal{E}} = \hat{\mathbf{H}} - \mathbf{H}$ , where the elements of  $\boldsymbol{\mathcal{E}} = [\boldsymbol{\mathcal{E}}_1, \boldsymbol{\mathcal{E}}_2, \dots, \boldsymbol{\mathcal{E}}_M]$  are random variables with zero mean and variance  $(\eta_i) / (b_i \eta_i + 1)$ . Furthermore, owing to the properties of LMMSE estimation,  $\boldsymbol{\mathcal{E}}$  is independent of  $\hat{\mathbf{H}}$ . Split-

ting (37) in devices with and without the same metadata sequence assigned to the  $i$ -th device, we have

$$\Omega = \mathbb{E} \left[ \sum_{j \in \{i, C_i\}} \left( \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} x_j x_j^H \sqrt{b_j} \mathbf{h}_j^H \mathbf{w}_i \right) + \sum_{j \notin \{i, C_i\}} \left( \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} x_j x_j^H \sqrt{b_j} \mathbf{h}_j^H \mathbf{w}_i \right) + \left( \mathbf{w}_i^H \mathbf{v} \mathbf{v}^H \mathbf{w}_i \right) \right], \quad (41)$$

Recalling that  $\mathcal{E} = \hat{\mathbf{H}} - \mathbf{H}$  and considering the independence between  $\mathcal{E}$ ,  $\hat{\mathbf{H}}$  and separating the signal of interest we obtain,

$$\Omega = \mathbb{E} \left[ \left| \mathbf{w}_i^H \hat{\mathbf{h}}_i \sqrt{b_i} x_i \right|^2 + \sum_{j \in \{C_i\}} \left| \mathbf{w}_i^H \hat{\mathbf{h}}_j \sqrt{b_j} x_j \right|^2 + \left| \mathbf{w}_i^H \mathbf{v} \right|^2 + \sum_{j \in \{i, C_i\}} \left| \mathbf{w}_i^H \boldsymbol{\varepsilon}_j \sqrt{b_j} x_j \right|^2 + \sum_{j \notin \{i, C_i\}} \left| \mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j} x_j \right|^2 \right] \quad (42)$$

where the first term is the signal of interest and the other additive terms are treated as a Gaussian noise. Thus, substituting (42) in (35), we get the expression, (45) in the top of the next page.

### C. Diversity Order

This section is devoted to present the diversity order achieved by the AA-VGL-DF detector. We adopt the geometrical approach presented in [30] and used in the previous work [31] in order to reach the expression. As for non-ergodic scenarios the error probability is the probability that the signal level is less than the specified value, also known as outage probability, the diversity order, that is, the asymptotic slope of the outage probability curve [32] [33], is given by

$$d \triangleq \lim_{x \rightarrow \infty} \frac{\log \left( P_r \left( R_{k, \text{span}\{\bar{k}\}} \right) \leq x \right)}{\log(x)} \quad (46)$$

where  $R_{k, \text{span}\{\bar{k}\}} = R_{k, \text{span}\{1, 2, \dots, k-1, k+1, \dots, K\}}$  is the squared projection height from the  $k$ th column vector  $\mathbf{h}_K$  of  $\mathbf{H}$ . From the definition in [33],  $R_{k, \text{span}\{\bar{k}\}} = \|\boldsymbol{\Upsilon} \mathbf{h}_K\|^2$  where  $\boldsymbol{\Upsilon} = \mathbf{I} - \mathbf{P}\mathbf{P}^H$  is the projection matrix to the orthogonal space of  $\text{span}\{\bar{k}\}$  and  $\mathbf{P}$  is composed of any orthonormal bases of this subspace. An important point is that only the  $K$  active devices are considered for the computation of the diversity order.

*Theorem 2: The diversity order achieved by the AA-VGL-DF detector is given by*

$$d_{\text{VGL}} = M - K + (\boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + \vartheta^0) + \mathcal{G} \quad (47)$$

where  $\boldsymbol{\vartheta}$  is the  $K \times 1$  binary vector presented in Section III-A that gathers the information about the reliability of the soft estimates.  $\mathbf{m}_{\text{ord}} = [|\mathcal{A}|, |\mathcal{A}|, \dots, |\mathcal{A}|]^T$  is also a  $K \times 1$  vector but each column has the number of symbols of the considered alphabet and  $\mathcal{G}$  is the number of all the possible combinations of the symbols of the considered augmented alphabet generated by the external list.  $\vartheta^0$  is the total number of zeros in the vector  $\boldsymbol{\vartheta}$ .

*Proof:* As the decision feedback scheme applies an interference cancellation at each detection step, as it is common in the literature [34], [35], we can make an analogy to the well-known successive interference cancellation (SIC) scheme. Assuming the channel model described in Section II and making the common assumption that there is no error propagation related to the interference cancellation [30]–[33], the interference nulling out can be expressed as a general matrix form given by

$$\mathbf{y}_{\perp} = \boldsymbol{\Upsilon} \cdot \mathbf{y}, \quad (48)$$

where (48) projects  $\mathbf{y}$  onto the direction orthogonal to the span  $\{\bar{k}\}$ . Following the procedure in [30], to reach the expression of the diversity order of each step of the SIC, the idea is to rotate the set of channels  $\{\mathbf{h}_1, \dots, \mathbf{h}_K\}$  in a way that  $\mathbf{h}_K$  becomes parallel to  $\mathbf{p}_K$ , one of the orthonormal basis of the subspace. Considering the detection of the first step,  $\mathbf{h}_K$  is fixed and position  $\mathbf{h}_{K-1}$  into the  $[\mathbf{p}_{K-1} \mathbf{p}_K]$  plane. In this way, the received signal vector can be written as

$$\mathbf{y}_{\perp} = \sqrt{b_1 \tau_x} x_1 \boldsymbol{\Upsilon} \cdot \mathbf{h}_1 + \mathbf{v}, \quad (49)$$

where the time instants are suppressed to reduce the notation. The rotations happens until the last channel vector,  $\mathbf{h}_K$ , is positioned into the  $[\mathbf{p}_2 \mathbf{p}_3, \dots, \mathbf{p}_K]$  hyper plane. In the well-known SIC scheme, the diversity order after all rotations is  $(M - K + 1)$ , as  $\mathbf{h}_{1\perp}$  has a total of  $(M - K + 1)$  nonzero components. On the other hand, as the AA-VGL-DF scheme has an internal list at each detection step, the diversity gain can be increased.

Assuming that the first soft estimate was considered unreliable by the SAC, the internal list scheme would imply more than one possible received vector to be cancelled. Thus, designating the order of the alphabet of the chosen modulation scheme as  $|\mathcal{A}|$ , for the first step, the diversity order is  $(M - K + |\mathcal{A}|)$ . For the steps that the soft estimation is reliable, the diversity order is the same as the SIC scheme. Thereby, the result can be achieved by induction. For the  $i$ th step, the diversity order can be represented by

$$\begin{cases} M - K + \vartheta_i^0, & \text{if } \tilde{d}_n \text{ is reliable and} \\ M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}}, & \text{if } \tilde{d}_n \text{ is unreliable,} \end{cases} \quad (50)$$

where  $\boldsymbol{\vartheta}$  and  $\mathbf{m}_{\text{ord}}$  vectors are scaled as  $i \times 1$  and  $\vartheta_i^0$  is the total number of zeros in the vector  $\boldsymbol{\vartheta}$  until the  $i$ th step. Therefore, for  $K = 5$  and  $\boldsymbol{\vartheta}^T = [1 \ 0 \ 1 \ 1 \ 0]$ , we have

$$\begin{cases} \text{for } i = 1, & d_{\text{VGL}} = M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} \\ \text{for } i = 2, & d_{\text{VGL}} = M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + 1 \\ \text{for } i = 3, & d_{\text{VGL}} = M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + 1 \\ \text{for } i = 4, & d_{\text{VGL}} = M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + 1 \\ \text{for } i = 5 = K, & d_{\text{VGL}} = M - K + \boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + 2. \end{cases} \quad (51)$$

Thus, considering the internal list, the diversity order achieved by AA-VGL-DF is  $M - K + (\boldsymbol{\vartheta}^T \mathbf{m}_{\text{ord}} + \vartheta^0)$ ,

The external list also contributes to the diversity gain. As the external list is comparable as a low complexity ML detector, the increase gain in the diversity order can follow the same idea. As the ML detector has a diversity gain of  $M$  [32] and the size of the group list is variable, we consider the  $\nu$  number of

$$\begin{aligned}
\text{SINR} &= \frac{|\mathbf{w}_i^H \hat{\mathbf{h}}_i \sqrt{b_i}|^2}{\sum_{j \in \{C_i\}} |\mathbf{w}_i^H \hat{\mathbf{h}}_j \sqrt{b_j}|^2 + \sum_{j \in \{j, C_i\}} |\mathbf{w}_i^H \boldsymbol{\varepsilon}_j \sqrt{b_j}|^2 + \sum_{j \notin \{i, C_i\}} |\mathbf{w}_i^H \mathbf{h}_j \sqrt{b_j}|^2 + \|\mathbf{w}_i^H\|^2} \\
&= \frac{\mathbf{w}_i^H b_i \left( \eta_i - \left( b_j \eta_i^2 \right) / \left( \sqrt{b_j} \eta_i + \sum_{j \in C_i} \eta_j + 1 \right) \right) \mathbf{w}_i}{\mathbf{w}_i^H \left( \sum_{j \in \{C_i\}} b_j \left( \eta_i - \frac{b_j \eta_i^2}{\sqrt{b_j} \eta_i + \sum_{j \in C_i} \eta_j + 1} \right) + \sum_{j \in \{j, C_i\}} b_j \left( \frac{\eta_j}{b_j \eta_j + 1} \right) + \sum_{j \notin \{i, C_i\}} b_j \eta_j \right) \mathbf{w}_i + \left( \frac{(1-\lambda)^2 \sigma_v^2 \sigma_x^2 + 1}{2(1-\lambda)} \right)} \quad (45)
\end{aligned}$$

symbols chosen to be verified in a total of  $\mathcal{G}$  possible vectors. In this way, the diversity order achieved by the AA-VGL-DF is given by (47).

## VI. NUMERICAL RESULTS

In this section, we evaluate the performance of the AA-VGL-DF and other relevant mMTC detection schemes. We consider an underdetermined mMTC system with  $N = 128$  devices and a single base-station equipped with  $M = 64$  antennas. The evaluated schemes experience an independent and identically-distributed (i.i.d.) random flat-fading channel model and the values  $a_{m,n}$  of (7) are taken from complex Gaussian distribution of  $\mathcal{CN}(0, 1)$ . The active devices radiate QPSK symbols with power values drawn uniformly at random in  $[0.1, 0.3]$  and the activity probabilities are given by a beta-binomial distribution, as described in Section II. Each transmission slot has 128 symbols, split into 60 metadata and 68 data. This balance between pilots and data is suggested in [36]. For systems that need explicit channel estimation, we considered the scheme described in Section V-B.

Initially, we verify the Symbol Error Rate (SER) performance and the Spectral Efficiency of the AA-VGL-DF for the six sparsity scenarios shown in Fig. 1. Fig. 6 shows that as lower is the activity probability of devices, better is the SER performance of AA-VGL-DF. The result of Fig. 7 illustrates the achievable spectral efficiency of the system with the AA-VGL-DF detection scheme and shows that, as the sparsity increases, the spectral efficiency also increases. This is due to the reduced number of block collisions and better detection performance, thus reducing the interference. Both plots consider the average SNR as  $10 \log(N \sigma_x^2 / \sigma_v^2)$ .

Given the SER results of those different scenarios, we choose the beta-distribution with  $\alpha = 4$  and  $\beta = 8$  as it provides an intermediary sparsity, to compare the SER and Bit Error Rate (BER) performances of the AA-VGL-DF and other relevant mMTC detection schemes.

The numerical results of both uncoded and coded systems are averaged over  $10^5$  runs. The performance of AA-VGL-DF is compared with other relevant schemes, as the linear mean squared error (LMMSE), unsorted SA-SIC [11], SA-SIC with A-SQRD [13], AA-RLS Linear, AA-RLS-DF [19] and a version of AA-RLS-DF with the internal list of this work. Besides that, we analyze a version with AA-VGL-DF with perfect activity user detection (AUD) and, as a lower bound,

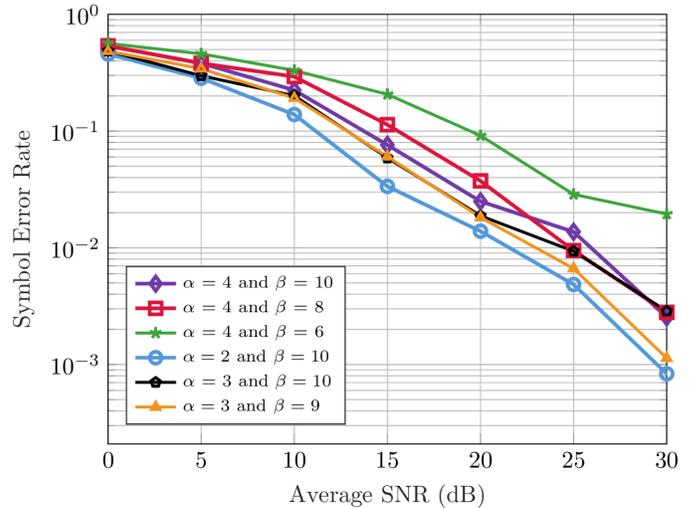


Fig. 6: Symbol Error Rate vs. Average SNR of the AA-VGL-DF algorithm in different sparsity scenarios. The activity pattern of devices is determined by a random variable with beta distribution, as shown in Fig. 1. In the legend, are shown the  $\alpha$  and  $\beta$  parameters of each considered distribution, for  $N = 128$  and  $M = 64$ .

the Oracle LMMSE detector, which has the knowledge of the index of nonzero entries, is considered.

Fig. 8 shows the symbol error rate performance of the considered algorithms. LMMSE has a poor performance as the system is underdetermined. Due to error propagation, the unsorted SA-SIC does not perform well. SA-SIC with A-SQRD is effective since it considers the activity probabilities, but under imperfect CSI conditions, its performance is not so good. In contrast, as AA-RLS-DF does not need explicit channel estimation, it is more efficient. The decision-feedback scheme provides a SER gain due to the interference cancellation, which also happens by including the internal list. The proposed schemes with lists of candidates obtain results that outperform the other relevant schemes, approaching the lower bound. The AA-VGL-DF with perfect AUD surpasses the lower bound for high SNRs, where the filter weights are better adjusted and the list schemes are able to correct more errors.

For the coded systems with IDD, Fig. 9 shows the BER of the already considered algorithms under the scheme proposed in Section IV. The LDPC matrix has 256 columns and 128

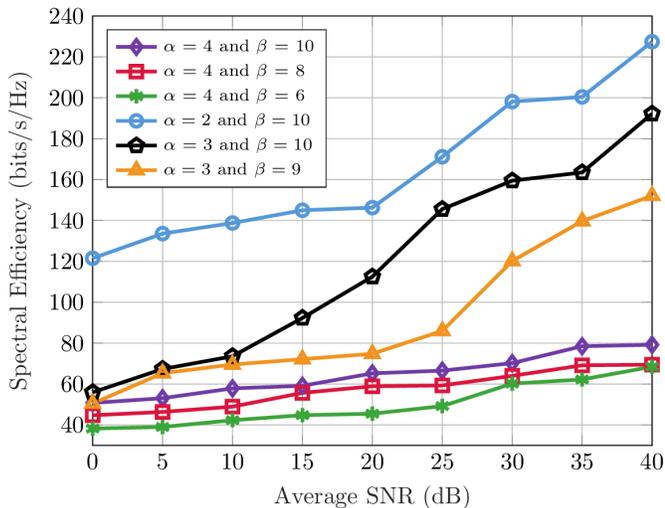


Fig. 7: Spectral Efficiency vs. Average SNR of the AA-VGL-DF algorithm with imperfect CSI in different sparsity scenarios. In the legend, are shown the  $\alpha$  and  $\beta$  parameters of each considered distribution, for  $N = 128$  and  $M = 64$ .

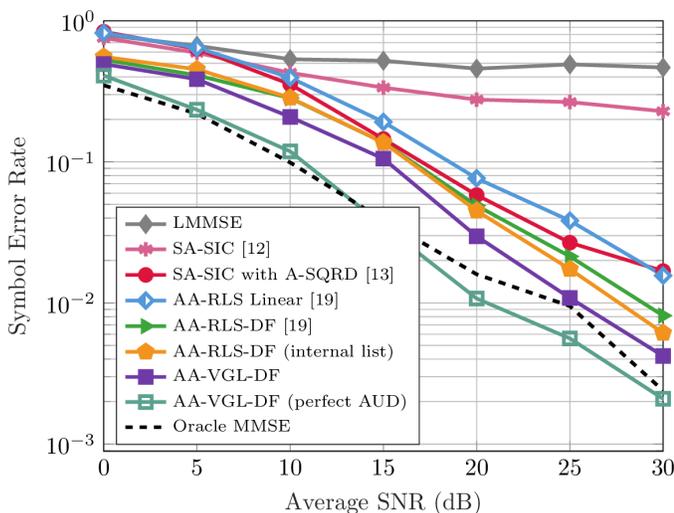


Fig. 8: Symbol Error Rate values vs. Average SNR. Parameters of proposals are  $\lambda = 0.92$ ,  $\gamma = 0.001$  and  $\xi = 10$ . The pattern activity of the  $N = 128$  devices is modelled with a beta-binomial distribution with  $\alpha = 4$  and  $\beta = 8$ . We consider imperfect CSI in the approaches which depends of the channel estimation.

rows, avoiding length-4 cycles and with 6 ones per column. The Sum-Product Algorithm (SPA) decoder is used and the average SNR is  $10 \log(NR\sigma_x^2/\sigma_v^2)$ , where  $R = 1/2$  is the rate of the LDPC code. The sparsity of the mMTC approach degrades the expected efficiency of LMMSE-PIC, obtaining little variation in relation to LMMSE and LMMSE-SIC. The hierarchy of performance of the other considered algorithms is the same as the uncoded case but with better error rate values. The iterative scheme matches the results for low bit error rate values. Fig. 10 exhibits the spectral efficiency of the considered algorithms. The filter refinement promoted by the internal and external lists provokes a better spectral efficiency than the other detection schemes. The oracle LMMSE is the

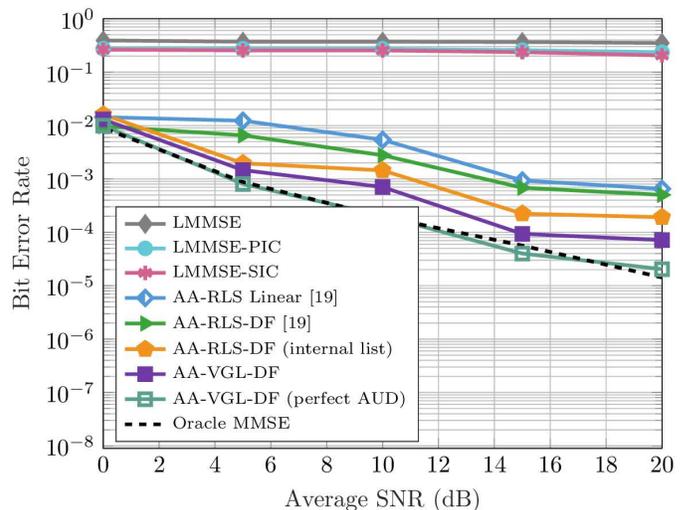


Fig. 9: Bit Error Rate values vs. Average SNR. LDPC with block length of 128, symbol rate  $R = 0.5$ , refined by 2 decoding iterations for the same scenario of Fig. 8.

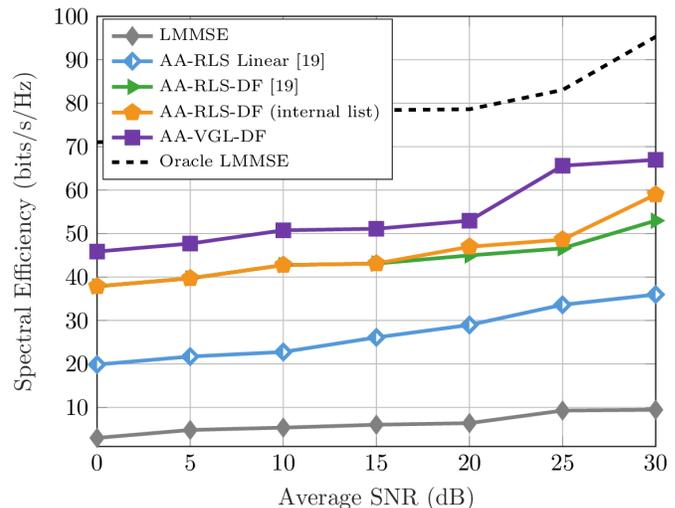


Fig. 10: Spectral Efficiency vs. Average SNR. Parameters of proposals are  $\lambda = 0.92$ ,  $\gamma = 0.001$  and  $\xi = 10$ . The pattern activity of the  $N = 128$  devices is modelled with a beta-binomial distribution with  $\alpha = 4$  and  $\beta = 8$  and imperfect CSI.

upper bound of the system.

## VII. CONCLUDING REMARKS

In this paper, we have proposed and investigated the AA-VGL-DF detection scheme, for mMTC. Considering different sparsity scenarios, we have presented a list-based DF detector along with an  $l_0$ -norm regularized RLS algorithm. In order to mitigate error propagation, we employ two lists schemes, based on constellation points that generate candidates for detection. Simulations have shown that AA-VGL-DF significantly outperforms existing approaches with a competitive computational complexity. AA-VGL-DF is also compared and analysed in terms of spectral efficiency and diversity. We have also incorporated into AA-VGL-DF an IDD scheme based on LDPC codes modified to the mMTC scenario.

## REFERENCES

- [1] P. Popovski *et al.*, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *IEEE Access*, vol. 6, pp. 55765-55779, 2018.
- [2] T. Salam *et al.*, "Data Aggregation in Massive Machine Type Communication: Challenges and Solutions," *IEEE Access*, vol. 7, pp. 41921-41946, 2019.
- [3] E. de Carvalho *et al.*, "Random Pilot and Data Access in Massive MIMO for Machine-Type Communications," *IEEE Trans. on Wir. Comm.*, vol. 16, no. 12, pp. 7703-7717, Dec. 2017.
- [4] D. Hu, L. He and X. Wang, "Semi-Blind Pilot Decontamination for Massive MIMO Systems," *IEEE Trans. on Wir. Comm.*, vol. 15, no. 1, pp. 525-536, Jan. 2016.
- [5] H. Yin *et al.*, "Robust Pilot Decontamination Based on Joint Angle and Power Domain Discrimination," *IEEE Trans. on Signal Proc.*, vol. 64, no. 11, pp. 2990-3003, June 1, 2016.
- [6] A. Adhikary *et al.*, "Joint Spatial Division and Multiplexing—The Large-Scale Array Regime," *IEEE Trans. on Inf. Theory*, vol. 59, no. 10, pp. 6441-6463, Oct. 2013.
- [7] E. Björnson *et al.*, "Massive MIMO Has Unlimited Capacity," *IEEE Trans. on Wir. Comm.*, vol. 17, no. 1, pp. 574-590, Jan. 2018.
- [8] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26-35, May 2014.
- [9] L. Liu *et al.*, "Massive Connectivity With Massive MIMO—Part I: Device Activity Detection and Channel Estimation," *IEEE Trans. on Signal Proc.*, vol. 66, no. 11, pp. 2933-2946, June 1, 2018.
- [10] H. Zhu *et al.*, "Exploiting Sparse User Activity in Multiuser Detection," *IEEE Trans. on Comm.*, vol. 59, no. 2, pp. 454-465, Feb. 2011.
- [11] B. Knoop *et al.*, "Compressed sensing K-best detection for sparse multi-user communications," *22nd EUSIPCO*, Lisbon, 2014, pp. 1726-1730.
- [12] B. Knoop *et al.*, "Sparsity-Aware Successive Interference Cancellation with Practical Constraints," *17th ITG WSA*, Germany, 2013, pp. 1-8.
- [13] J. Ahn *et al.*, "Sparsity-Aware Ordered Successive Interference Cancellation for Massive Machine-Type Communications," in *IEEE Wir. Comm. Letters*, vol. 7, no. 1, pp. 134-137, Feb. 2018.
- [14] X. Zhang *et al.*, "Compressive Sensing Based Multiuser Detection via Iterative Reweighted Approach in M2M Communications," *IEEE Wir. Comm. Letters*, Early Access, 2018.
- [15] R. B. Di Renna and R. C. de Lamare, "Activity-Aware Multiple Feedback SIC for Massive Machine-Type Communications," *SCC 2019, 12th Int. ITG Conf. on Sys. Commu. and Coding*, Germany, Feb. 2019.
- [16] J. G. A. Skellam, "A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials," *Journal of Royal Statistical Society, Series B*, 1948. Available: <https://www.jstor.org/stable/pdf/2983779.pdf>.
- [17] 3GPP TR 37.868, "Study on RAN Improvements for Machine-type Communications; (Release 10)," *Technical Specification Group Radio Access Network*, V0.8.1 (2011-08), Feb. 26, 2014. Accessed on: Dec. 13, 2019. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2630>.
- [18] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Upper Saddle River, N.J.: Prentice Hall PTR, 2002.
- [19] R. B. Di Renna and R. C. de Lamare, "Adaptive Activity-Aware Iterative Detection for Massive Machine-Type Communications," in *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1631-1634, Dec. 2019.
- [20] R. C. De Lamare and R. Sampaio-Neto, "Minimum Mean-Squared Error Iterative Successive Parallel Arbitrated Decision Feedback Detectors for DS-CDMA Systems," *IEEE Transactions on Communications*, vol. 56, no. 5, pp. 778-789, May 2008.
- [21] P. Li, R. C. de Lamare and R. Fa, "Multiple Feedback Successive Interference Cancellation Detection for Multiuser MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2434-2439, August 2011.
- [22] R. C. de Lamare and R. Sampaio-Neto, "Reduced-Rank Adaptive Filtering Based on Joint Iterative Optimization of Adaptive Filters," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 980-983, Dec. 2007.
- [23] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Reduced-Rank Processing Based on Joint and Iterative Interpolation, Decimation, and Filtering," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2503-2514, July 2009.
- [24] R. C. de Lamare and R. Sampaio-Neto, "Reduced-Rank Space-Time Adaptive Interference Suppression With Joint Iterative Least Squares Algorithms for Spread-Spectrum Systems," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1217-1228, March 2010.
- [25] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Reduced-Rank Equalization Algorithms Based on Alternating Optimization Design Techniques for MIMO Systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2482-2494, July 2011.
- [26] P. S. Bradley *et al.*, "Feature selection via concave minimization and support vector machines," in *Proc. 13th ICML*, 1998, pp. 82-90.
- [27] A. G. D. Uchoa, C. T. Healy and R. C. de Lamare, "Iterative Detection and Decoding Algorithms for MIMO Systems in Block-Fading Channels Using LDPC Codes," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2735-2741, April 2016.
- [28] Xiaodong Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," in *IEEE Trans. on Comm.*, vol. 47, no. 7, pp. 1046-1061, July 1999.
- [29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, Cambridge, 1st edition, 2005.
- [30] S. Loyka and F. Gagnon, "Performance analysis of the V-BLAST algorithm: an analytical approach," *IEEE Trans. on Wir. Comm.*, vol. 3, no. 4, pp. 1326-1337, July 2004.
- [31] R. C. de Lamare, "Adaptive and Iterative Multi-Branch MMSE Decision Feedback Detection Algorithms for Multi-Antenna Systems," *IEEE Trans. on Wir. Comm.*, vol. 12, no. 10, pp. 5294-5308, October 2013.
- [32] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," in *IEEE Trans. Inf. Theory*, vol. 49, no. 5, May 2003.
- [33] H. Zhang, H. Dai and B. L. Hughes, "Analysis of the diversity-multiplexing tradeoff for ordered MIMO SIC receivers," *IEEE Trans. Commun.*, vol. 57, no. 1, Jan. 2009.
- [34] M. K. Varanasi, "Decision feedback multiuser detection: a systematic approach," *IEEE Trans. on Inf. Theory*, vol. 45, no. 1, pp. 219-240, Jan. 1999.
- [35] A. A. Rontogiannis *et al.*, "A square-root adaptive V-BLAST algorithm for fast time-varying MIMO channels," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 265-268, May 2006.
- [36] L. Liu *et al.*, "Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things," in *IEEE Sig. Proc. Mag.*, vol. 35, no. 5, pp. 88-99, Sept. 2018.