

MIT Open Access Articles

Massive Uncoordinated Multiple Access for Beyond 5G

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Mohammadkarimi, Mostafa, Dobre, Octavia A and Win, Moe Z. 2022. "Massive Uncoordinated Multiple Access for Beyond 5G." IEEE Transactions on Wireless Communications, 21 (5).

As Published: 10.1109/TWC.2021.3117256

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: <https://hdl.handle.net/1721.1/145557>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Massive Uncoordinated Multiple Access for Beyond 5G

Mostafa Mohammadkarimi, *Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE*,
and Moe Z. Win, *Fellow, IEEE*

Abstract—Existing wireless communication systems have been mainly designed to provide substantial gain in terms of data rates. However, 5G and Beyond will depart from this scheme, with the objective not only to provide services with higher data rates. One of the main goals is to support massive machine-type communications (mMTC) in the Internet-of-Things (IoT) applications. Supporting massive uplink communications for devices with sporadic traffic pattern and short-packet size, as it is in many mMTC use cases, is a challenging task, particularly when the control signaling is not negligible in size compared to the payload. In addition, channel estimation becomes challenging for sporadic and short-packet transmission due to the limited number of employed pilots. In this paper, a new uplink multiple access (MA) scheme is proposed for mMTC, which can support a large number of uncoordinated IoT devices with short-packet and sporadic traffic. The proposed uplink MA scheme removes the overheads associated with the device identifier as well as pilots and preambles related to channel estimation. An alternative mechanism for device identification is proposed, where a unique spreading code is dedicated to each IoT device as identifier. This unique code is simultaneously used for the spreading purpose and device identification. Different IoT device identification algorithms which employ sparse signal reconstruction methods are proposed to determine the active IoT devices prior to data detection. Specifically, the Bayesian information criterion model order selection method is employed to develop an IoT device identification algorithm for unknown and time-varying probability of device activity. Our proposed MA scheme benefits from a new non-coherent nonlinear multiuser detection algorithm designed on the basis of unsupervised machine learning techniques to enable data detection without *a priori* knowledge on channel state information. The effectiveness of the proposed MA scheme for known and unknown probability of activity and high overloading factor is supported by simulation results.

Index Terms—Internet-of-Things (IoT), massive machine-type communications (mMTC), Beyond 5G, uplink multiple access, simultaneous sparse signal reconstruction, nonlinear multiuser detection, sporadic transmission, machine learning.

I. INTRODUCTION

MASSIVE UPLINK connectivity is the key factor in the realization of the Internet-of-Things (IoT), as part of 5G and Beyond wireless communication systems [1]. In many IoT applications, massive machine-type communications (mMTC) services are required, where a large number of devices transmit very short packets sporadically. Typically, the number of IoT devices assigned to each base station (BS) in mMTC is in orders of magnitude above what current communication networks are capable to support. Moreover, IoT

devices do not transmit continuously, rather updates are infrequently transmitted to the BS, whenever a measured value changes. Hence, small packets are expected to carry critical payload in mMTC [2]. For example, Sigfox as one of the most adopted solutions for IoT, has been designed to support a packet payload of 12 bytes.

The design of the current wireless communication systems relies on the assumption that the control signaling related to physical (PHY) and media access control (MAC) layers is of negligible size compared to the payload. Thus, heuristic design of control signaling is acceptable and does not affect the overall system performance. However, in mMTC with short-packet transmission, the control signaling can be similar in size with the payload; thus, inefficient design of control signaling leads to highly suboptimal transmission schemes. Excessive control signaling, e.g., the overheads, preambles, and pilots associated with device identifier, exploited for channel estimation, and used for random access procedure, hinders massive connectivity [3]. Thus, efficient multiple access (MA) schemes with highly limited (or non-existent) control signaling are required.

Moreover, channel estimation is another challenge for sporadic and short-packet transmission, especially for a massive number of non-orthogonal transmissions. Existing channel estimation approaches are often based on the assumption that devices are active over long periods so that channel estimation through pilots and preambles is feasible. However, if an IoT device only transmits occasionally, such an assumption cannot longer be valid. Instead, channel estimation has to rely on a single transmission that may be very short, which constrains the number of pilots available to keep the overhead low [4]. Channel estimation becomes more challenging in the grant-free (GF) uplink MA scheme, where resources are randomly selected by devices.

The conventional request-grant and GF uplink MA schemes cannot provide massive connectivity since the maximum number of supported devices is limited by the number and scheduling granularity of orthogonal resources. Hence, uplink MA schemes based on non-orthogonal transmission are required to boost up the uplink connectivity in mMTC. In addition, the handshaking mechanism in the conventional MA schemes require excessive control signaling in the random access procedure, which makes them inefficient for mMTC [5].

Motivated by these facts, a new uplink MA scheme for short-packet and sporadic traffic in mMTC is proposed in this paper. The main idea behind the proposed MA scheme is to reduce the control signaling while simultaneously supporting a massive number of uncoordinated IoT devices with a single BS. The proposed MA scheme is designed based on asynchronous direct-sequence spread spectrum (DS-SS) with non-orthogonal spreading codebook, and is capable of supporting

M. Mohammadkarimi is with the Institute for Digital Communications, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany (e-mail: mostafa.mohammadkarimi@fau.de).

O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, Canada, (e-mail: odobre@mun.ca).

M. Z. Win is with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Boston, MA, USA (e-mail: moewin@mit.edu).

undetermined DS-SS systems in static networks, where the BS and IoT devices are immobile.

To remove the control signaling associated with the IoT device identifier, a unique spreading code is dedicated to each IoT device which is simultaneously used for the spreading purpose and device identification. In a nutshell, instead of allocating a fragment of the IoT packet to the signaling associated with the MAC address (device identifier), the unique spreading code is used as IoT device identifier. Moreover, the MA scheme relies on an unsupervised machine learning technique to enable non-coherent data detection, thus removing the need of preambles and pilots used for channel estimation. The lack of preambles and pilots further reduces the control signaling.

Our proposed approach for removing the device identifier relies on sparsity-aware IoT device identification at the BS to determine the active IoT devices before data detection. Based on the sporadic traffic pattern of the IoT devices as well as lack of knowledge about the channel state information (CSI) of the IoT devices, the squared ℓ_2 -norm sparse signal reconstruction (SSR) and Bayesian information criterion (BIC) $\ell_1 - \ell_2$ mixed-norm simultaneous sparse signal reconstruction (SSSR) IoT device identification algorithms are developed. In the former algorithm, the IoT identification problem is formulated as an SSR using the generalized cross-validation (GCV) approach followed by parallel hypothesis testing. The latter algorithm formulates the IoT device identification problem as a BIC model order selection SSSR problem. Unlike existing sparsity-aware activity detection algorithms, the BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification algorithm can determine active devices, when the probability of activity is unknown and time-varying.

The proposed uplink MA scheme is also equipped with a new non-coherent nonlinear multiuser detection (MUD) algorithm to detect data of the active IoT devices, applied after the IoT device identification algorithms. We propose the non-coherent 2-mean clustering (2-MC)-MUD algorithm based on 2-MC unsupervised machine learning and differential coding to detect data without channel estimation at the BS.

A. Related Works

The problem of jointly detecting both activity and data in code division multiple access (CDMA), when the probability of activity is known at the BS has been studied in [6]–[8]. However, the developed algorithms either assume perfect CSI or estimate CSI at the BS. Also, they have considered coordinated CDMA with perfectly synchronized transmissions and solve the problem through compressive sensing recovery algorithms based on either the convex relaxation or the greedy optimization algorithms. Besides, these algorithms do not propose any solution how to set the tuning parameter in the optimization problem in the signal recovery procedure. The degree of sparsity, and thus, the false alarm and correct identification rates depend on the value of the tuning parameter.

The authors in [9] studied the problem of joint channel estimation and detection of activity and data for sporadic communications. However, they considered coordinated transmission with *a priori* known probability of activity for devices. Also, the accuracy of the joint activity and data detection depends

on the length of pilot employed for channel estimation. In [6], the authors develop sparsity-aware maximum a posteriori probability (MAP) detector for the coordinated CDMA and obtain a close-form expression for the tuning parameter as a function of the probability of user activity. However, this algorithm assumes that the channel matrix is perfectly known at the BS. The problem of sparsity-aware successive interference cancellation (SIC) for mMTC has been investigated in [10], [11]. The authors in [12] have employed compressive sensing MUD in slotted ALOHA. Approximate message passing (AMP)-based user activity detection algorithm which exploits the statistics of the wireless channel for the uplink has been investigated in [13], [14]. Recently, device activity detection and channel estimation for massive connectivity in massive multiple-input and multiple-output has been studied in [15].

The main contributions of this work are as follows:

- A new uplink MA scheme is proposed for mMTC. The proposed MA scheme exhibits the following advantages:
 - It is capable to support thousands of uncoordinated IoT devices;
 - It supports sporadic traffic pattern and short-packet;
 - It significantly reduces packet time on-air since it is designed for underdetermined DS-SS (number of devices is larger than the spreading factor);
 - It removes the control signaling associated with the device identifier to reduce uplink packet overhead;
 - It removes preambles and pilot employed for channel estimation to reduce uplink overhead;
 - It exhibits high scalability in terms of adding new IoT devices (high overloading factor) without negatively affecting the system performance.
- A new mechanism for the IoT device identification at the BS is developed instead of using device identifier. Since the active IoT devices in the network do not use a device identifier in order to identify themselves to the BS, the squared ℓ_2 -norm SSR and the BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT identification algorithms are proposed to detect active IoT devices. The proposed algorithms exhibit the following advantages:
 - They can detect active IoT devices without knowledge of the CSI;
 - They remove the need for matched-filter (MF) implementation for all spreading codes; thus reducing the complexity of the receiver;
 - The BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification can identify active IoT devices when the probability of activity is unknown and time-varying.
- The statistical performance analysis of the squared ℓ_2 -norm SSR IoT device identification algorithm is presented, and closed-form expressions for the correct identification and false alarm rates are derived.
- A new non-coherent nonlinear MUD algorithm, i.e., 2-MC-MUD in combination with differential coding is designed for short packet transmission. The proposed 2-MC-MUD algorithm exhibits the following advantages:
 - It supports both coordinated and uncoordinated DS-SS transmission irrespective of the traffic pattern;

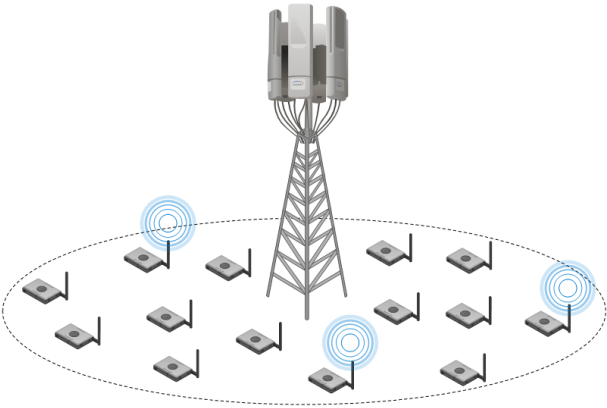


Fig. 1: Single-hop IoT network with sporadic traffic pattern.

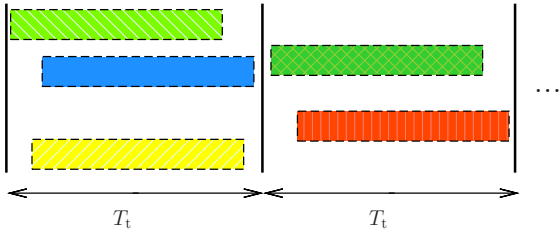


Fig. 2: Received packets at the BS.

- It does not require knowledge of the CSI at the BS.

B. Notations

Random variables are displayed in sans serif upright fonts, while their realizations are shown in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x ; a random vector and its realization are denoted by \mathbf{x} and \mathbf{x} ; a random matrix and its realization are denoted by \mathbf{X} and \mathbf{X} , respectively. Sets and random sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by \mathcal{X} and \mathcal{X} , respectively.

The identity matrix and zero vector are shown by \mathbf{I} and $\mathbf{0}$, and the indicator function is defined as $\mathbb{I}\{x\} = 1$ if x is true; otherwise, $\mathbb{I}\{x\} = 0$. The cardinality of a set, which measures the number of elements of the set, is denoted by $\text{card}(\cdot)$. The ℓ_0 quasi-norm of vector $\mathbf{a}_j = [a_{0,j} \ a_{1,j} \ \dots \ a_{m-1,j}]^\dagger$ and the $\ell_0 - \ell_0$ quasi-norm of matrix $\mathbf{A} \triangleq [\mathbf{a}_0 \ \mathbf{a}_1 \ \dots \ \mathbf{a}_{n-1}]$ are respectively defined as $\|\mathbf{a}_j\|_0 \triangleq \text{card}(\{i \in \mathcal{I} | a_{i,j} \neq 0\})$, and $\|\mathbf{A}\|_0 \triangleq \text{card}(\{i \in \mathcal{I} | \exists j, j = 0, 1, \dots, n-1, a_{i,j} \neq 0\})$, where $\mathcal{I} \triangleq \{0, 1, \dots, m\}$. We use $\text{tr}(\mathbf{B})$, \mathbf{B}^{-1} and $\det(\mathbf{B})$ to show the trace, inverse, and determinant of a square matrix \mathbf{B} . We also employ $\text{diag}(\mathbf{B})$ to represent the diagonal elements of \mathbf{B} in vector form. Throughout the paper, $(\cdot)^*$, $(\cdot)^\dagger$, and $(\cdot)^H$ show the complex conjugate, transpose, and Hermitian transpose, respectively. Also, $|\cdot|$, $\lfloor \cdot \rfloor$, and \otimes represent the absolute value operator, floor function, and Kronecker product, respectively. $\mathbb{E}\{\cdot\}$ is the statistical expectation, \hat{x} is an estimate of x , \cup is the logical OR operator, and \cap is the logical AND operator. The complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The remaining of the paper is organized as follows. Section II introduces the system model. Section III describes the proposed IoT device identification algorithms and presents their analytical performance evaluation. In Section IV, the data detection problem is discussed, and the nonlinear 2-MC-MUD algorithm is proposed. Simulation results are provided in Section V, and conclusions are drawn in Section VI.

II. SYSTEM MODEL

Consider K_u IoT devices communicating with a single IoT BS in a single-hop communication, as shown in Fig. 1. It is considered that the IoT devices transmit data in short packets over independent doubly block fading channel, where the fading channel is block fading in time and in frequency. The probability of packet transmission for each IoT device is assumed to be P_a .¹ The IoT devices transmit their packet after receiving a beacon signal transmitted by the IoT BS. This signal is periodically transmitted with period $T_t = N_s T_s + \tau_{\max}$, where N_s is the number of symbols per IoT packet, T_s is the symbol duration, and τ_{\max} is the known maximum delay of the single-hop IoT network. It is assumed that T_t equals the coherence time of the fading channel. The coherence time of the wireless channel in suburban locations is approximately 0.5s. The coherence bandwidth of wireless channel is typically in the range of 100 kHz to 1MHz. Thus, the coherence block is in the range of 5×10^4 to 5×10^5 chips.

We denote $\mathcal{X}_u \triangleq \{0, 1, \dots, K_u - 1\}$ and \mathcal{X}_a the total and active IoT devices in the network, respectively. The round-trip delay of the k th IoT device is shown by $\tau_k \triangleq 2\ell_k/c$, $\tau_k \in [0, \tau_{\max}]$, where c is the speed of light, and ℓ_k is the distance between the k -th IoT device and the BS. We consider that $\tau_k, k, \in \mathcal{X}_u$, is known at the receiver. Fig. 2 illustrates the received IoT packets at the BS.

As illustrated in Fig. 3, for each IoT device, the payload bits \mathbf{d}_k , $k \in \mathcal{X}_a$, are encoded by the channel encoder to increase the reliability of packet transmission. Then, the encoded data is passed through the differential encoding block. Differential encoding is employed to remove the need of channel estimation in the MUD at the BS to enable non-coherent detection. After differential encoding, the data is multiplied by a unique spreading waveform. It is considered that the spreading waveforms of the IoT devices do not change over time. Finally, the DS-SS signal is binary phase-shift-keying (BPSK) modulated and then transmitted.

The impulse response of the doubly block fading channel for the k th IoT device is given as $h_k(t) \approx \check{g}_k \delta(t - \tau_k)$, where \check{g}_k is the fading coefficient of the k th IoT device, which is constant during a packet but changes to an independent value for the next packet. The received baseband signal over doubly block fading channel in each transmission period with respect to the timing reference of the BS is modeled as

$$\begin{aligned} r(t) &= \sum_{k=0}^{K_u-1} \sum_{n=0}^{N_s-1} \check{g}_k \sqrt{\eta_k P_k} e^{j\varphi_k} \mathbf{b}_{k,n} s_k(t - nT_s - \tau_k) + w(t) \\ &= \sum_{k=0}^{K_u-1} \sum_{n=0}^{N_s-1} \mathbf{g}_k \mathbf{b}_{k,n} s_k(t - nT_s - \tau_k) + w(t), \end{aligned} \quad (1)$$

¹Both known and unknown probability of activity are studied in this paper.

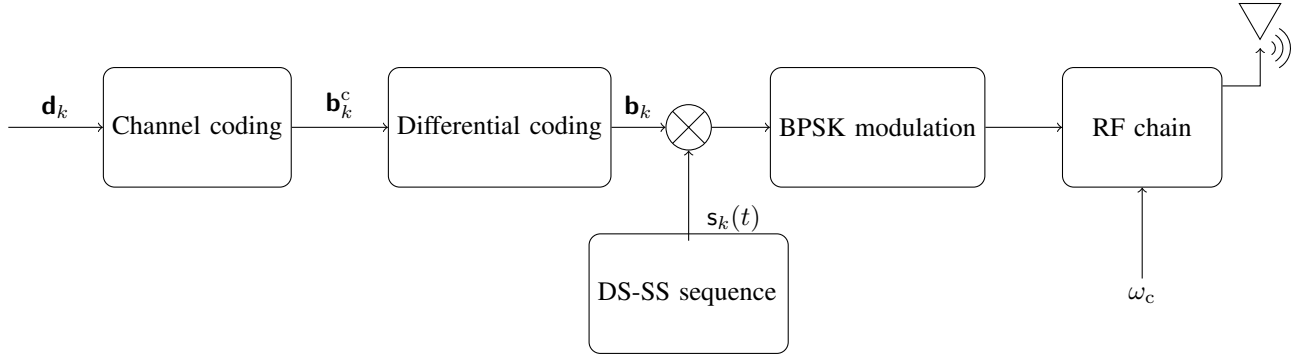


Fig. 3: Block diagram that illustrates packet transmission at IoT devices.

where $t \in [0, T_t]$, $\mathbf{g}_k \triangleq \check{\mathbf{g}}_k \sqrt{\eta_k p_k} e^{j\varphi_k}$, and $\check{\mathbf{g}}_k$, φ_k , and $\{\mathbf{b}_{k,n}, n = 0, 1, \dots, N_s - 1\}$ respectively denote the fading channel coefficient, carrier phase (CP), and symbol stream of the k th IoT device, which are unknown at the BS. Also, $\eta_k = (\frac{\lambda_c}{4\pi\ell_k})^2$ and p_k denote the pathloss and transmit power of the k th IoT device, respectively, where λ_c is the wavelength of the carrier signal. It is considered that $\check{\mathbf{g}}_k \sim \mathcal{CN}(\mu_k, \sigma_k^2)$, and the envelope of the CSI, i.e., $|\check{\mathbf{g}}_k|$ has a Rician distribution with K -factor $|\mu_k|^2/(\sigma_k^2)$. The symbol stream for the inactive IoT devices is modeled as transmitting zeros during the packet, i.e., $\mathbf{b}_{k,n} = 0$, $n = 0, 1, \dots, N_s - 1$, while active IoT devices employ BPSK modulation with $\mathbb{E}\{|\mathbf{b}_{k,n}|^2\} = 1$. The DS-SS signaling waveform of the k th IoT device, $s_k(t)$, is given by

$$s_k(t) = \sum_{m=0}^{N_c-1} c_k^{(m)} \psi(t - mT_c), \quad t \in [0, T_s], \quad (2)$$

where T_c is the chip duration, $\mathbf{c}_k = [c_k^{(0)} \ c_k^{(1)} \ \dots \ c_k^{(N_c-1)}]^\dagger$ is the spreading sequence of $\{+1, -1\}$ assigned to the k th IoT device, N_c is the spreading factor, and $\psi(t)$ is the chip waveform with unit power. It is assumed that $\psi(t)$ is a rectangular pulse confined within $[0, T_c]$. To support massive connectivity, $K_u > N_c$, which leads to non-orthogonal transmission. The baseband additive complex Gaussian noise at the output of the receive filter with bandwidth $1/T_c$ is denoted by $w(t)$.

Fig. 4 shows the block diagram of the proposed receiver at the IoT BS. As seen, the received baseband signal is passed through the chip MF and sampled at the chip rate. The output of the sampled chip MF for the i th chip at the j th observation symbol is obtained as

$$\begin{aligned} \mathbf{r}_j^{(i)} &\triangleq \int_{jT_s+iT_c}^{jT_s+(i+1)T_c} \mathbf{r}(t) \psi(t - jT_s - iT_c) dt \\ &= \sum_{k=0}^{K_u-1} \mathbf{g}_k \mathbf{u}_{k,j}^{(i)} + \mathbf{w}_j^{(i)} \quad i = 0, 1, \dots, N_c - 1, \end{aligned} \quad (3)$$

where

$$\mathbf{w}_j^{(i)} \triangleq \int_{jT_s+iT_c}^{jT_s+(i+1)T_c} \mathbf{w}(t) \psi(t - jT_s - iT_c) dt, \quad (4)$$

and

$$\mathbf{u}_{k,j}^{(i)} \triangleq \int_{jT_s+iT_c}^{jT_s+(i+1)T_c} \sum_{n=0}^{N_s-1} \mathbf{b}_{k,n} s_k(t - nT_s - \tau_k) \psi(t - jT_s - iT_c) dt. \quad (5)$$

By employing (4), one can show that the joint probability density function (PDF) of the corresponding noise vector associated with the j th observation vector, i.e., $\mathbf{w}_j \triangleq [\mathbf{w}_j^{(0)} \ \mathbf{w}_j^{(1)} \ \dots \ \mathbf{w}_j^{(N_c-1)}]^\dagger$ is characterized by $\mathbf{w}_j \sim \mathcal{CN}(\mathbf{0}_{N_c}, \sigma_w^2 \mathbf{I})$ with $\sigma_w^2 \triangleq N_0/T_c$, where $N_0/2$ is the power spectral density of the white noise. The integral in (5) represents the area under the received signal waveform of the k th IoT device during the i th chip-matched filtering duration at the j th observation symbol.

Let us write the delay of the k th IoT as

$$\tau_k \triangleq \alpha_k T_s + \beta_k T_c + \xi_k, \quad (6)$$

with $\alpha_k \triangleq \lfloor \tau_k/T_s \rfloor$, $\beta_k \triangleq \lfloor \tau_k/T_c \rfloor - \alpha_k N_c$, and $\xi_k \in [0, T_c]$.

Based on the values of α_k , β_k , and ξ_k , $\mathbf{u}_{k,j}^{(i)}$ in (5) is expressed as a function of $\mathbf{b}_{k,j-\alpha_k}$ and $\mathbf{b}_{k,j-\alpha_k-1}$ as [16]

$$\begin{aligned} \mathbf{u}_{k,j}^{(i)} &\triangleq \sum_{n=0}^{N_s-1} \sum_{m=0}^{N_c-1} c_k^{(m)} \mathbf{b}_{k,n} \\ &\times \int_{jT_s+iT_c}^{jT_s+(i+1)T_c} \psi(t - nT_s - mT_c - \tau_k) \psi(t - jT_s - iT_c) dt \\ &= \mathbf{b}_{k,j-\alpha_k-1} x_k^{(i)} (1 - \xi_k) + \mathbf{b}_{k,j-\alpha_k} x_k^{(i)} (\xi_k), \end{aligned} \quad (7)$$

where

$$x_k^{(i)}(\nu) \triangleq \sum_{m=0}^{N_c-1} c_k^{(m)} \int_{iT_c}^{(i+1)T_c} \psi(t - mT_c - \nu T_c) \psi(t - iT_c) dt, \quad (8)$$

with $\nu \in [0, 1)$. We can write (7) in vector form as follows

$$\mathbf{u}_{k,j} = \mathbf{b}_{k,j-\alpha_k-1} \mathbf{x}_{k,0} + \mathbf{b}_{k,j-\alpha_k} \mathbf{x}_{k,1} \quad (9)$$

where $\mathbf{b}_{k,j} = 0$ when $j \notin [0, N_s - 1]$, and

$$\mathbf{u}_{k,j} \triangleq [\mathbf{u}_{k,j}^{(0)} \ \mathbf{u}_{k,j}^{(1)} \ \dots \ \mathbf{u}_{k,j}^{(N_c-1)}]^\dagger \quad (10a)$$

$$\mathbf{x}_{k,1} \triangleq [x_k^{(0)}(\xi_k) \ x_k^{(1)}(\xi_k) \ \dots \ x_k^{(N_c-1)}(\xi_k)]^\dagger \quad (10b)$$

$$\mathbf{x}_{k,0} \triangleq [x_k^{(0)}(1 - \xi_k) \ x_k^{(1)}(1 - \xi_k) \ \dots \ x_k^{(N_c-1)}(1 - \xi_k)]^\dagger.$$

For the rectangular chip waveform pulse-shaping $\psi(t)$, one can easily obtain

$$\begin{bmatrix} \mathbf{x}_{k,1} \\ \mathbf{x}_{k,0} \end{bmatrix} = (1 - \xi_k) \begin{bmatrix} \mathbf{0}_{\beta_k} \\ \mathbf{c}_k \\ \mathbf{0}_{N_c - \beta_k} \end{bmatrix} + \xi_k \begin{bmatrix} \mathbf{0}_{\beta_k+1} \\ \mathbf{c}_k \\ \mathbf{0}_{N_c - \beta_k - 1} \end{bmatrix}. \quad (11)$$

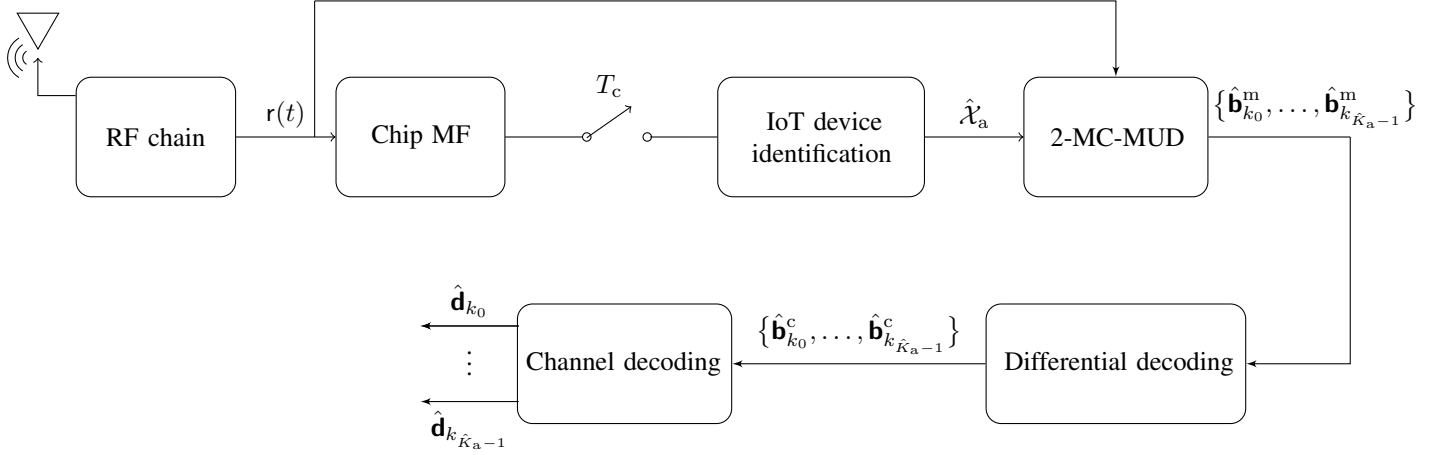


Fig. 4: Block diagram of the proposed receiver at the BS.

Let us define $\mathbf{X}_k \triangleq [\mathbf{x}_{k,0} \ \mathbf{x}_{k,1}]$. By employing (3) and (9), the j th observation vector, i.e., $\mathbf{r}_j \triangleq [r_j^{(0)} \ r_j^{(1)} \ \dots \ r_j^{(N_c-1)}]^\dagger$, is written as follows

$$\mathbf{r}_j = \mathbf{X}\mathbf{G}\mathbf{b}_j + \mathbf{w}_j = \mathbf{X}\mathbf{h}_j + \mathbf{w}_j, \quad (12)$$

where

$$\mathbf{X} \triangleq [\mathbf{X}_0 \ \mathbf{X}_1 \ \dots \ \mathbf{X}_{K_u-1}], \quad (13)$$

$$\mathbf{G} \triangleq \begin{bmatrix} \mathbf{g}_0 & & 0 \\ & \mathbf{g}_1 & \\ 0 & & \ddots \\ & & & \mathbf{g}_{K_u-1} \end{bmatrix} \otimes \mathbf{I}_2, \quad (14)$$

$$\mathbf{b}_j \triangleq \begin{bmatrix} \mathbf{b}_{0,j-\alpha_0-1} & \mathbf{b}_{0,j-\alpha_0} & \mathbf{b}_{1,j-\alpha_1-1} & \mathbf{b}_{1,j-\alpha_1} & \dots \\ & & & & & \mathbf{b}_{K_u-1,j-\alpha_{(K_u-1)}-1} & \mathbf{b}_{K_u-1,j-\alpha_{(K_u-1)}} \end{bmatrix}^\dagger, \quad (15)$$

and

$$\mathbf{h}_j \triangleq [\mathbf{h}_{0,j,0} \ \mathbf{h}_{0,j,1} \ \mathbf{h}_{1,j,0} \ \mathbf{h}_{1,j,1} \ \dots \ \mathbf{h}_{K_u-1,j,0} \ \mathbf{h}_{K_u-1,j,1}]^\dagger \quad (16)$$

with

$$\mathbf{h}_{k,j,f} \triangleq \mathbf{g}_k \mathbf{b}_{k,j-\alpha_k-1+f}, \quad f \in \{0, 1\}. \quad (17)$$

Finally, by stacking the N_t observation vectors, the observation matrix is written as follows

$$\mathbf{R}_T = \mathbf{X}\mathbf{G}\mathbf{B}_T + \mathbf{W}_T = \mathbf{X}\mathbf{H}_T + \mathbf{W}_T, \quad (18)$$

where $\mathbf{R}_T \triangleq [\mathbf{r}_0 \ \mathbf{r}_1 \ \dots \ \mathbf{r}_{N_t-1}]$, $\mathbf{B}_T \triangleq [\mathbf{b}_0 \ \mathbf{b}_1 \ \dots \ \mathbf{b}_{N_t-1}]$, $\mathbf{W}_T \triangleq [\mathbf{w}_0 \ \mathbf{w}_1 \ \dots \ \mathbf{w}_{N_t-1}]$, and $\mathbf{H}_T \triangleq [\mathbf{h}_0 \ \mathbf{h}_1 \ \dots \ \mathbf{h}_{N_t-1}]$. In (18), \mathbf{X} is referred to as dictionary.

As seen in Fig. 4, after chip-matched filtering and sampling, the IoT device identification algorithm is applied to the measurement matrix \mathbf{R} to detect the active IoT devices. The outcome of the IoT device identification algorithm is a set of IoT devices $\hat{\mathcal{X}}_a$. Then, the MUD algorithm is applied to detect data of the IoT devices in $\hat{\mathcal{X}}_a$.

III. IOT DEVICE IDENTIFICATION

Device identification is the first step in uplink MA schemes that devices do not use control signaling in order to identify themselves to the BS. In this case, the BS needs to determine the active devices before data detection. In this section, different IoT device identification algorithms are developed.

A. IoT Device Identification: Problem Formulation

For the sake of decreasing the complexity, a portion of the observation window can be employed for IoT device identification. Let us consider a truncated observation window of length L as follows

$$\mathbf{R} = \mathbf{X}\mathbf{G}\mathbf{B} + \mathbf{W} = \mathbf{X}\mathbf{H} + \mathbf{W}, \quad (19)$$

where $\mathbf{R} \triangleq [\mathbf{r}_{\bar{\alpha}} \ \mathbf{r}_{\bar{\alpha}+1} \ \dots \ \mathbf{r}_{\bar{\alpha}+L-1}]$, $\mathbf{B} \triangleq [\mathbf{b}_{\bar{\alpha}} \ \mathbf{b}_{\bar{\alpha}+1} \ \dots \ \mathbf{b}_{\bar{\alpha}+L-1}]$, $\mathbf{W} \triangleq [\mathbf{w}_{\bar{\alpha}} \ \mathbf{w}_{\bar{\alpha}+1} \ \dots \ \mathbf{w}_{\bar{\alpha}+L-1}]$, and $\mathbf{H} \triangleq [\mathbf{h}_{\bar{\alpha}} \ \mathbf{h}_{\bar{\alpha}+1} \ \dots \ \mathbf{h}_{\bar{\alpha}+L-1}]$ with $1 \leq L \leq N_s + \alpha_{\min} - \bar{\alpha}$, where $\bar{\alpha}$ is an arbitrary positive integer, $\bar{\alpha} > \alpha_{\max} \triangleq \max\{\alpha_0, \alpha_1, \dots, \alpha_{K_u-1}\}$, and $\alpha_{\min} \triangleq \min\{\alpha_0, \alpha_1, \dots, \alpha_{K_u-1}\}$.² Fig. 5 shows the underdetermined system of linear equations in (18), and Fig. 6 illustrates truncated observation windows for IoT device identification in (19).

The activity of an IoT device is defined for an entire packet, i.e., the rows of \mathbf{H} corresponding to the active and inactive IoT devices are non-zero and zero, respectively. Thus, the problem of IoT device identification for the k th IoT device, $k \in \mathcal{X}_u$, can be expressed as the following binary hypothesis testing:

$$H_{1k}: \quad \mathbf{h}_{k,\bar{\alpha},L} \neq \mathbf{0} \quad (20)$$

$$H_{0k}: \quad \mathbf{h}_{k,\bar{\alpha},L} = \mathbf{0},$$

where

$$\mathbf{h}_{k,\bar{\alpha},L} \triangleq [\mathbf{h}_{k,\bar{\alpha}}^\dagger \ \mathbf{h}_{k,\bar{\alpha}+1}^\dagger \ \dots \ \mathbf{h}_{k,\bar{\alpha}+L-1}^\dagger]^\dagger, \quad (21a)$$

$$\mathbf{h}_{k,j} \triangleq [\mathbf{h}_{k,j,0} \ \mathbf{h}_{k,j,1}]^\dagger, \quad (21b)$$

and H_{0k} and H_{1k} are the null and alternative hypotheses denoting that the k th IoT device is active and inactive, respectively. As seen in (20), the IoT device identification problem is formulated as K_u parallel binary hypothesis testing problems.

The first step in IoT device identification is to reconstruct $\mathbf{h}_{k,\bar{\alpha},L}$, $k \in \mathcal{X}_u$, from the truncated observation matrix in (19). However, (19) represents an underdetermined system of linear equations since $N_c < K_u$. Hence, it is not uniquely solvable.

Let us denote the number of active IoT devices by the random variable $k_a = \text{card}(\mathcal{X}_a)$. For $P_a \ll 1$, $\mathbb{P}\{k_a \ll K_u\} = 1$, and thus, \mathbf{B} and \mathbf{H} in (19) are sparse matrices. Moreover,

²The number of non-zero elements of the observation window for the active IoT devices is considered the same to facilitate theoretical analysis.

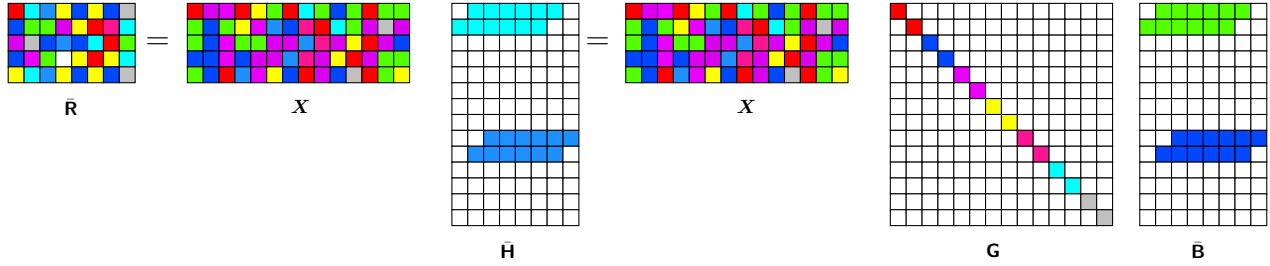


Fig. 5: Underdetermined system of linear equations for $K_u = 7$, $K_a = 2$, $N_t = 8$, $N_c = 5$, $\alpha_{\max} = 1$, and $N_s = 6$. Due to the asynchronicity among the IoT devices, the matrix of the transmitted symbols \mathbf{B} includes two rows for each IoT device.

the columns of $\mathbf{H}(\mathbf{B})$ share the same sparsity profiles. This sparse structure is referred to as block-sparse. The block-sparse structure of \mathbf{H} can be observed in Fig. 5.

The sparse structure of \mathbf{H} can be employed to reconstruct the columns of \mathbf{H} from the underdetermined linear observation model in (19). When each column of \mathbf{H} is individually reconstructed from its corresponding column in \mathbf{R} , it is referred to as SSR. The SSR for the columns of \mathbf{H} , i.e., \mathbf{h}_j , $\bar{\alpha} \leq j \leq \bar{\alpha} + L - 1$, is formulated as follows

$$\hat{\mathbf{h}}_j = \arg \min_{\mathbf{h}_j} \frac{1}{2} \|\mathbf{r}_j - \mathbf{X}\mathbf{h}_j\|_F^2 + \lambda_{\ell_0} \|\mathbf{h}_j\|_0, \quad (22)$$

where λ_{ℓ_0} is the tuning parameter which balances both approximation error and sparsity level of the solution.

The ℓ_0 -minimization in (22) is both numerically unstable and NP-hard since the ℓ_0 quasi-norm is a discrete-value function. One approach to the SSR is to replace the ℓ_0 quasi-norm by a convex function with common sparsity profile that leads to a solution very close to the one of the original problem. This method is called convex relaxation and converts the combinatorial problem in (22) into a convex optimization problem which can be solved in polynomial time. Different convex functions can be employed to relax $\|\mathbf{h}_j\|_0$ in (22). A common family of convex functions is the ℓ_q norm, given as

$$\|\mathbf{h}_j\|_q = \left(\sum_{k=0}^{K_u-1} \sum_{f=0}^1 |h_{k,j,f}|^q \right)^{\frac{1}{q}}. \quad (23)$$

The recovered vectors by the ℓ_q norm minimization can be employed to infer the active IoT set \mathcal{X}_a .

On the other hand, the block-sparse structure of \mathbf{H} can be employed to improve the reconstruction of \mathbf{H} in (22). This method of signal reconstruction is referred to as SSSR. Opposite to SSR, the SSSR simultaneously exploits the column sparsity along with the block-sparse structure in the optimization problem in order to reconstruct the matrix \mathbf{H} . The SSSR of \mathbf{H} , given \mathbf{R} and the dictionary \mathbf{X} , is expressed as

$$\hat{\mathbf{H}} = \arg \min_{\mathbf{H}} \frac{1}{2} \|\mathbf{R} - \mathbf{X}\mathbf{H}\|_F^2 + \lambda_{\ell_0}^{\ell_0} \|\mathbf{H}\|_0, \quad (24)$$

where $\lambda_{\ell_0}^{\ell_0}$ is the tuning parameter which balances both approximation error and sparsity level of the solution. Similar to the ℓ_0 -minimization in (22), the $\ell_0 - \ell_0$ -minimization in (24) is unstable and NP-hard. Therefore, the quasi-norm $\|\mathbf{H}\|_0$ is replaced with the $\ell_p - \ell_q$ ($p, q \geq 1$) mixed-norm as

$$J_{p,q}(\mathbf{H}) = \sum_{k=0}^{K_u-1} \|\mathbf{h}_{k,\bar{\alpha},L}\|_q^p \quad (25)$$

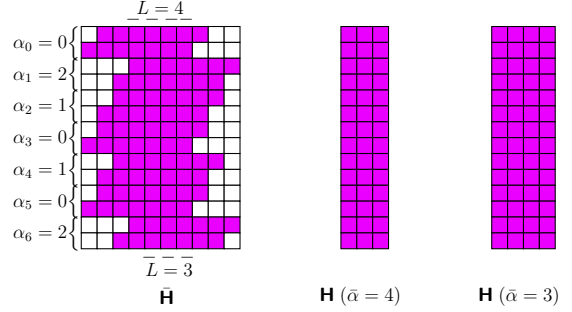


Fig. 6: Different observation windows for IoT device identification ($K_u = 7$, $N_s = 7$, $\alpha_{\max} = 2$, $\alpha_{\min} = 0$, $1 \leq L \leq 4$). The purple color is employed to show the packet of the IoT devices, which is zero for inactive and non-zero for active IoT devices.

to convert the combinatorial problem in (24) into a convex optimization problem. The recovered matrix by the relaxed SSSR can also be employed to infer the active IoT set.

B. IoT Device Identification for Known Probability of Activity

Here, we propose an algorithm for IoT device identification when the probability of activity P_a is known at the BS. Convex relation through squared ℓ_2 -norm followed by a threshold setting mechanism is employed for IoT device identification.

1) *Squared ℓ_2 -Norm SSR IoT Device Identification*: The squared ℓ_2 -norm convex relaxation form of (22) is given by

$$\hat{\mathbf{h}}_j = \arg \min_{\mathbf{h}_j} \frac{1}{2} \|\mathbf{r}_j - \mathbf{X}\mathbf{h}_j\|_F^2 + \lambda \|\mathbf{h}_j\|_2^2, \quad (26)$$

where \mathbf{h}_j is given in (16). The squared ℓ_2 -norm SSR algorithm formulates the IoT identification problem as a ridge regression (RD) estimation problem as in (26) followed by K_u parallel binary hypothesis testing problems. This is because the RD does not set the coefficients of $\hat{\mathbf{h}}_j$ to zero. The optimal solution of (26) is obtained as [17]

$$\hat{\mathbf{h}}_j = \left(\mathbf{X}^\dagger \mathbf{X} + 2\lambda \mathbf{I} \right)^{-1} \mathbf{X}^\dagger \mathbf{r}_j, \quad (27)$$

which is a simple linear estimator of \mathbf{r}_j that shrinks ordinary least-squares (LS) estimates towards zero. The tuning parameter λ for SSR can be obtained through cross-validation and generalized cross-validation (GCV) [18], [19]. The latter is a method of model selection that is widely employed; in this case, λ is obtained as follows [18]

$$\lambda_{\text{cv}} = \arg \min_{\lambda} \frac{\|(\mathbf{I} - \mathbf{Q})\mathbf{r}_j\|_2^2}{[\text{tr}(\mathbf{I} - \mathbf{Q})]^2}, \quad (28)$$

where $\mathbf{Q} \triangleq \mathbf{X}(\mathbf{X}^\dagger \mathbf{X} + 2\lambda \mathbf{I})^{-1} \mathbf{X}^\dagger$. In [20], it has been shown that the optimal tuning parameter of the RD estimator for $\mathbf{r}_j =$

$\mathbf{X}\mathbf{h}_j + \mathbf{w}_j$ in terms of minimum mean squared error can be approximated as follows

$$\lambda_j^{\text{op}} \approx \frac{\sigma_w^2 \text{tr}[\bar{\Sigma}_{\mathbf{X}}^{-1}]}{\mathbf{h}_j^H \bar{\Sigma}_{\mathbf{X}}^{-1} \mathbf{h}_j + 3 \text{tr}[\bar{\Sigma}_{\mathbf{X}}^{-2}]}, \quad (29)$$

where $\bar{\Sigma}_{\mathbf{X}} \triangleq \mathbf{X}^\dagger \mathbf{X}$. As observed, λ_j^{op} depends on \mathbf{h}_j which is unknown and needs to be estimated by the RD estimator. In this case, for moderate and high signal-to-noise ratio (SNR) range, an approximation of (29) can be obtained by replacing $\mathbf{h}_j^H \bar{\Sigma}_{\mathbf{X}}^{-1} \mathbf{h}_j$ with its expected value [21]. Since the elements of \mathbf{h}_j are uncorrelated, by employing $\mathbb{E}\{|h_{k,j,0}|^2\} = \mathbb{E}\{|h_{k,j,1}|^2\} = P_a \eta_k p_k (\sigma_k^2 + |\mu_k|^2)$, $k \in \mathcal{X}_u$, we can show that $\mathbb{E}\{\mathbf{h}_j^H \bar{\Sigma}_{\mathbf{X}}^{-1} \mathbf{h}_j\} = P_a (\mathbf{\Gamma}^\dagger \otimes \mathbf{1}^\dagger) \bar{\mathbf{\Lambda}}_{\mathbf{X}}$, where $\mathbf{\Gamma} \triangleq [\gamma_0 \ \gamma_1 \ \dots \ \gamma_{K_u-1}]^\dagger$, $\gamma_k \triangleq \eta_k p_k (\sigma_k^2 + |\mu_k|^2)$, $\mathbf{1}_2 = [1 \ 1]^\dagger$, and $\bar{\mathbf{\Lambda}}_{\mathbf{X}} \triangleq \text{diag}(\bar{\Sigma}_{\mathbf{X}}^{-1})$. Substituting $\mathbb{E}\{\mathbf{h}_j^H \bar{\Sigma}_{\mathbf{X}}^{-1} \mathbf{h}_j\} = P_a (\mathbf{\Gamma}^\dagger \otimes \mathbf{1}^\dagger) \bar{\mathbf{\Lambda}}_{\mathbf{X}}$ into (29), results in

$$\lambda^{\text{opt}} \approx \frac{\sigma_w^2 \text{tr}[\bar{\Sigma}_{\mathbf{X}}^{-1}]}{P_a (\mathbf{\Gamma}^\dagger \otimes \mathbf{1}^\dagger) \bar{\mathbf{\Lambda}}_{\mathbf{X}} + 3 \text{tr}[\bar{\Sigma}_{\mathbf{X}}^{-2}]}. \quad (30)$$

As seen in (30), λ^{opt} is inversely proportional to P_a .

By substituting $\mathbf{r}_j = \mathbf{X}\mathbf{h}_j + \mathbf{w}_j$ in (12) into (27), $\hat{\mathbf{h}}_j$ can be written as a linear function of \mathbf{h}_j as

$$\hat{\mathbf{h}}_j = \mathbf{\Omega} \mathbf{h}_j + \mathbf{w}'_j, \quad (31)$$

where

$$\mathbf{\Omega} \triangleq \begin{bmatrix} \Omega_{0,0} & \Omega_{0,1} & \dots & \Omega_{0,2K_u-1} \\ \Omega_{1,0} & \Omega_{1,1} & \dots & \Omega_{1,2K_u-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{2K_u-1,0} & \Omega_{2K_u-1,1} & \dots & \Omega_{2K_u-1,2K_u-1} \end{bmatrix} \quad (32)$$

$$= \mathbf{I} - 2\lambda^{\text{opt}} (\bar{\Sigma}_{\mathbf{X}} + 2\lambda^{\text{opt}} \mathbf{I})^{-1},$$

and

$$\mathbf{w}'_j \triangleq \begin{bmatrix} w'_{0,j,0} \\ w'_{0,j,1} \\ \vdots \\ w'_{K_u-1,j,0} \\ w'_{K_u-1,j,1} \end{bmatrix} = (\bar{\Sigma}_{\mathbf{X}} + 2\lambda^{\text{opt}} \mathbf{I})^{-1} \mathbf{X}^\dagger \mathbf{w}_j. \quad (33)$$

In (33), \mathbf{w}'_j is zero-mean complex Gaussian colored noise vector with covariance matrix given by

$$\mathbf{\Sigma}^{\mathbf{w}'_j} \triangleq \begin{bmatrix} \Sigma_{0,0}^{\mathbf{w}'_j} & \Sigma_{0,1}^{\mathbf{w}'_j} & \dots & \Sigma_{0,2K_u-1}^{\mathbf{w}'_j} \\ \Sigma_{1,0}^{\mathbf{w}'_j} & \Sigma_{1,1}^{\mathbf{w}'_j} & \dots & \Sigma_{1,2K_u-1}^{\mathbf{w}'_j} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{2K_u-1,0}^{\mathbf{w}'_j} & \Sigma_{2K_u-1,1}^{\mathbf{w}'_j} & \dots & \Sigma_{2K_u-1,2K_u-1}^{\mathbf{w}'_j} \end{bmatrix}$$

$$= \mathbb{E}\{\mathbf{w}'_j (\mathbf{w}'_j)^H\} = \sigma_w^2 (\bar{\Sigma}_{\mathbf{X}} + 2\lambda^{\text{opt}} \mathbf{I})^{-2} \bar{\Sigma}_{\mathbf{X}}, \quad (34)$$

where $\Sigma_{2k_1+f_1, 2k_2+f_2}^{\mathbf{w}'_j} = \mathbb{E}\{w'_{k_1, j, f_1} (w'_{k_2, j, f_2})^*\}$.

The elements of $\hat{\mathbf{h}}_j$ in (31) associated with the k th IoT device, i.e., $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$ can be written as follows

$$\hat{h}_{k,j,f} = \Omega_{2k+f, 2k+f} \mathbf{h}_{k,j,f} + \Omega_{2k+f, 2k+\bar{f}} \mathbf{h}_{k,j,\bar{f}} \quad (35)$$

$$+ \sum_{n \neq k} \left\{ \Omega_{2k+f, 2n+f} \mathbf{h}_{n,j,f} + \Omega_{2k+f, 2n+\bar{f}} \mathbf{h}_{n,j,\bar{f}} \right\} + \mathbf{w}'_{k,j,f},$$

where $f, \bar{f} \in \{0, 1\}$ and $\bar{f} \triangleq f + (-1)^f$. The second term on the right-hand side of (35) represents the effect of the multiuser interference caused by the active IoT devices in the network.

Due to the central limit theorem (CLT), $\hat{h}_{k,j,f}$, $f \in \{0, 1\}$, in (35) given hypothesis H_{0k} and H_{1k} can be accurately approximated by complex Gaussian random variables for sufficiently small values of K -factor $\kappa_k \triangleq |\mu_k|^2 / \sigma_k^2$ and large enough $P_a K_u$. Simulation results show that for $\kappa_k \triangleq |\mu_k|^2 / \sigma_k^2 < 0.2$, Gaussian assumption is valid. In fact, the lower κ_k , the more reliable the Gaussian assumption is. It should be mentioned that the random variables $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$ are not joint Gaussian random variables as shown in Fig. 7. The mean, variance, and cross-correlation of $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$ are given in Lemma 1.

Lemma 1. *First and second order statistics of the reconstructed signal for the k th IoT device in (35), i.e., $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$, are given as follows*

$$\mathbb{E}\{\hat{h}_{k,j,0} | H_{tk}\} = \mathbb{E}\{\hat{h}_{k,j,1} | H_{tk}\} = 0, \quad t \in \{0, 1\} \quad (36)$$

$$\Sigma_{f,f}^{tk} \triangleq \text{Var}\{\hat{h}_{k,j,f} | H_{tk}\} = \mathbb{E}\{|\hat{h}_{k,j,f}|^2 | H_{tk}\} \quad (37)$$

$$= t\gamma_k \left(\Omega_{2k+f, 2k+f}^2 + \Omega_{2k+f, 2k+\bar{f}}^2 \right)$$

$$+ P_a \sum_{n \neq k} \gamma_n \left(\Omega_{2k+f, 2n+f}^2 + \Omega_{2k+f, 2n+\bar{f}}^2 \right) + \Sigma_{2k+f, 2k+f}^{\mathbf{w}'_j},$$

and

$$\Sigma_{0,1}^{tk} = \text{Cov}\{\hat{h}_{k,j,0}, \hat{h}_{k,j,1} | H_{tk}\} = \mathbb{E}\{\hat{h}_{k,j,0} \hat{h}_{k,j,1}^* | H_{tk}\}$$

$$= t\gamma_k \left(\Omega_{2k, 2k} \Omega_{2k+1, 2k} + \Omega_{2k+1, 2k+1} \Omega_{2k, 2k+1} \right)$$

$$+ P_a \sum_{n \neq k} \gamma_n \left(\Omega_{2k, 2n} \Omega_{2k+1, 2n} + \Omega_{2k+1, 2n+1} \Omega_{2k, 2n+1} \right)$$

$$+ \Sigma_{2k, 2k+1}^{\mathbf{w}'_j}, \quad (38)$$

where $\gamma_k = \eta_k p_k (\sigma_k^2 + |\mu_k|^2)$, $\Sigma_{1,0}^{tk} = \Sigma_{0,1}^{tk}$, $t, f \in \{0, 1\}$, $\bar{f} \triangleq f + (-1)^f$, and $\Sigma_{2k+f, 2k+f}^{\mathbf{w}'_j}$ is given in (34) (Proof in Appendix I) \square

Since the joint PDF of $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$ given H_{tk} , i.e., $p(\hat{h}_{k,j,0}, \hat{h}_{k,j,1} | H_{tk})$, $t \in \{0, 1\}$, cannot be expressed in a tractable mathematical form and since there is high correlation between $\hat{h}_{k,j,0}$ and $\hat{h}_{k,j,1}$, we can either use $\hat{h}_{k,j,0}$ or $\hat{h}_{k,j,1}$ to identify the transmission state of the k th IoT device. Moreover, the in-phase and quadrature components of $\hat{h}_{k,j,f}$ can be accurately approximated by correlated joint Gaussian random variables due to the CLT for sufficiently small values of K -factor $\kappa_k \triangleq |\mu_k|^2 / \sigma_k^2$ and large enough $P_a K_u$. To verify the credibility of Gaussian assumption, we evaluate the kurtosis and skewness for $\text{Re}\{\hat{h}_{k,j,0}\}$ and $\text{Re}\{\hat{h}_{k,j,1}\}$ in Table I.

Similar to the proof of Lemma 1, we can show that the distribution of the reconstructed signal for the k th IoT device is given as follows

$$\begin{bmatrix} \text{Re}\{\hat{h}_{k,j,0}\} \\ \text{Im}\{\hat{h}_{k,j,0}\} \end{bmatrix} \sim \begin{cases} \mathcal{N}(\mathbf{0}, \mathbf{C}_{0,0}^{0k}), & H_{0k} \\ \mathcal{N}(\mathbf{0}, \mathbf{C}_{0,0}^{1k}), & H_{1k} \end{cases}, \quad (39)$$

and

$$\begin{bmatrix} \text{Re}\{\hat{h}_{k,j,1}\} \\ \text{Im}\{\hat{h}_{k,j,1}\} \end{bmatrix} \sim \begin{cases} \mathcal{CN}(\mathbf{0}, \mathbf{C}_{1,1}^{0k}), & H_{0k} \\ \mathcal{CN}(\mathbf{0}, \mathbf{C}_{1,1}^{1k}), & H_{1k} \end{cases}, \quad (40)$$

Table I: Credibility of Gaussian assumption for $\hat{\mathbf{h}}_{k,j,0}$ and $\hat{\mathbf{h}}_{k,j,1}$.

Variable	Kurtosis	Skewness	Variance
Gaussian (theory)	3	0	0.002 {0.2657}
$\text{Re}\{\hat{\mathbf{h}}_{k,j,0}\}$	3.105	0.0224	0.001
$\text{Re}\{\hat{\mathbf{h}}_{k,j,1}\}$	3.014	-0.0233	{0.2643}

where

$$\mathbf{C}_{f,f}^{tk} = \begin{bmatrix} \bar{\Sigma}_{f,f}^{tk} & \rho_{f,f}^{tk} \\ \rho_{f,f}^{tk} & \bar{\Sigma}_{f,f}^{tk} \end{bmatrix}, \quad (41)$$

$$\begin{aligned} \rho_{f,f}^{tk} &= \mathbb{E}\left\{\text{Re}\{\hat{\mathbf{h}}_{k,j,f}\} \text{Im}\{\hat{\mathbf{h}}_{k,j,f}\} | H_{tk}\right\} \\ &= t\bar{\mu}_k \tilde{\mu}_k \eta_k p_k \left(\Omega_{2k+f,2k+f}^2 + \Omega_{2k+f,2k+\bar{f}}^2 \right) \\ &\quad + P_a \sum_{n \neq k} \bar{\mu}_n \tilde{\mu}_n \eta_n p_n \left(\Omega_{2k+f,2n+f}^2 + \Omega_{2k+f,2n+\bar{f}}^2 \right), \end{aligned} \quad (42)$$

$$\begin{aligned} \bar{\Sigma}_{f,f}^{tk} &\triangleq \text{Var}\left\{\text{Re}\{\hat{\mathbf{h}}_{k,j,f}\} | H_{tk}\right\} = \mathbb{E}\left\{\left(\text{Re}\{\hat{\mathbf{h}}_{k,j,f}\}\right)^2 | H_{tk}\right\} \\ &= t(\sigma_k^2/2 + |\bar{\mu}_k|^2) \eta_k p_k \left(\Omega_{2k+f,2k+f}^2 + \Omega_{2k+f,2k+\bar{f}}^2 \right) \\ &\quad + P_a \sum_{n \neq k} (\sigma_n^2/2 + |\bar{\mu}_n|^2) \eta_n p_n \left(\Omega_{2k+f,2n+f}^2 + \Omega_{2k+f,2n+\bar{f}}^2 \right) \\ &\quad + \Sigma_{2k+f,2k+f}^w/2, \end{aligned} \quad (43)$$

and

$$\begin{aligned} \tilde{\Sigma}_{f,f}^{tk} &\triangleq \text{Var}\left\{\text{Im}\{\hat{\mathbf{h}}_{k,j,f}\} | H_{tk}\right\} = \mathbb{E}\left\{\left(\text{Im}\{\hat{\mathbf{h}}_{k,j,f}\}\right)^2 | H_{tk}\right\} \\ &= t(\sigma_k^2/2 + |\tilde{\mu}_k|^2) \eta_k p_k \left(\Omega_{2k+f,2k+f}^2 + \Omega_{2k+f,2k+\bar{f}}^2 \right) \\ &\quad + P_a \sum_{n \neq k} (\sigma_n^2/2 + |\tilde{\mu}_n|^2) \eta_n p_n \left(\Omega_{2k+f,2n+f}^2 + \Omega_{2k+f,2n+\bar{f}}^2 \right) \\ &\quad + \Sigma_{2k+f,2k+f}^w/2, \end{aligned} \quad (44)$$

with $\bar{\mu}_k \triangleq \text{Re}\{\mu_k\}$ and $\tilde{\mu}_k \triangleq \text{Im}\{\mu_k\}$.

The larger the ratio of the variances in (39) and (40), i.e., $\Sigma_{0,0}^{1k}/\Sigma_{0,0}^{0k}$, and $\Sigma_{1,1}^{1k}/\Sigma_{1,1}^{0k}$, the better identification performance. Accordingly, we use the reconstructed signal in (35) for the identification of the k th IoT device as follows

$$\check{\mathbf{h}}_{k,j} \triangleq \begin{bmatrix} \text{Re}\{\hat{\mathbf{h}}_{k,j,1}\} \\ \text{Im}\{\hat{\mathbf{h}}_{k,j,1}\} \end{bmatrix} \mathbb{I}\{\varrho_k < 0\} + \begin{bmatrix} \text{Re}\{\hat{\mathbf{h}}_{k,j,0}\} \\ \text{Im}\{\hat{\mathbf{h}}_{k,j,0}\} \end{bmatrix} \mathbb{I}\{\varrho_k \geq 0\}, \quad (45)$$

where $\check{\mathbf{h}}_{k,j} = [\check{\mathbf{h}}_{k,j,0}, \check{\mathbf{h}}_{k,j,1}]^\dagger$, $k \in \mathcal{X}_u$, and

$$\varrho_k = \left(\frac{\Sigma_{0,0}^{1k}}{\Sigma_{0,0}^{0k}} - \frac{\Sigma_{1,1}^{1k}}{\Sigma_{1,1}^{0k}} \right). \quad (46)$$

Using the reconstructed signal in the form of (45) enables us to derive closed-form expressions for the correct identification and false alarm rates. In order to identify the transmission state of the k th IoT device, $k \in \mathcal{X}_u$, based on $\check{\mathbf{h}}_{k,j}$, the maximum likelihood ratio (MLR) test can be used [22].

Lemma 2. *The optimal MLR decision rule for IoT device identification based on the reconstructed signal $\check{\mathbf{h}}_{k,j}$, $k \in \mathcal{X}_u$, in (45) is given by*

$$\mathbf{d}_k = \begin{cases} H_{1k}, & \phi(\check{\mathbf{h}}_{k,j}) \geq \theta_k \\ H_{0k}, & \phi(\check{\mathbf{h}}_{k,j}) < \theta_k \end{cases}, \quad (47)$$

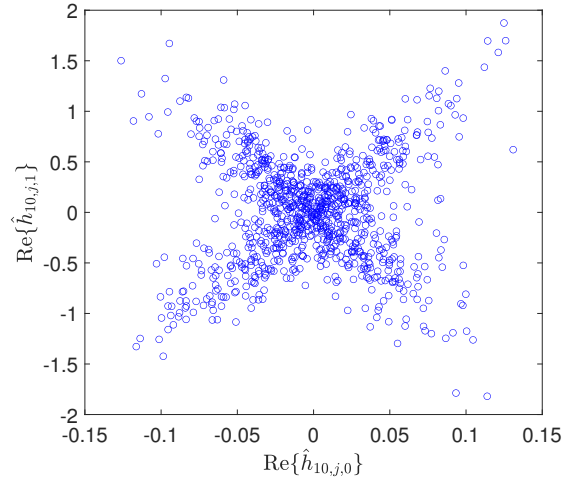


Fig. 7: Scatter plot of $\text{Re}\{\hat{\mathbf{h}}_{10,j,1}\}$ and $\text{Re}\{\hat{\mathbf{h}}_{10,j,0}\}$ when device $k = 10$ is active, $K_u = 1536$, $N_c = 512$, $P_a = 0.01$, and $\text{SNR} = 16$ dB.

where

$$\phi(\check{\mathbf{h}}_{k,j}) = \sum_{n=0}^1 \chi_{f,f}[n] z_{k,j}^2[n], \quad (48)$$

$$\chi_{f,f}[n] \triangleq \frac{\lambda_{f,f}[n]}{\lambda_{f,f}[n] + 1}. \quad (49)$$

In (49), $\lambda_{f,f}[0]$ and $\lambda_{f,f}[1]$ are the eigenvalues of the symmetric matrix $\mathbf{B}_{f,f}^{1k} \triangleq (\mathbf{A}_{f,f}^{0k})^\dagger \mathbf{C}_{f,f}^{1k} \mathbf{A}_{f,f}^{0k}$, and

$$[z_{k,j}[0], z_{k,j}[1]]^\dagger \triangleq (\mathbf{V}_{f,f}^{1k})^\dagger (\mathbf{A}_{f,f}^{0k})^\dagger \check{\mathbf{h}}_{k,j}, \quad (50)$$

where $\mathbf{V}_{f,f}^{1k}$ is the modal matrix of $\mathbf{B}_{f,f}^{1k}$, $\mathbf{A}_{f,f}^{0k} \triangleq \mathbf{V}_{f,f}^{0k} (\mathbf{A}_{f,f}^{0k})^{-\frac{1}{2}}$, $(\mathbf{V}_{f,f}^{0k})^\dagger \mathbf{C}_{f,f}^{0k} \mathbf{V}_{f,f}^{0k} = \mathbf{A}_{f,f}^{0k}$, and $f = 1$ for $\varrho_k < 0$, and $f = 0$ for $\varrho_k \geq 0$. The threshold value for the k th IoT device, i.e., θ_k , is set for a desirable probability of false alarm $P_k^{(f)} \triangleq \mathbb{P}\{\mathbf{d}_k = H_{1k} | H_{0k}\}$ as follows

$$P_k^{(f)} = \frac{1}{2\pi} \int_{\theta_k}^{+\infty} \int_{-\infty}^{+\infty} \prod_{n=0}^1 \frac{\exp(-j\omega x)}{\sqrt{1 - 2j\chi_{f,f}[n]\omega}} d\omega dx. \quad (51)$$

By using (48), we can obtain the correct identification rate for the k th IoT device as follows

$$P_k^{(c)} = \frac{1}{2\pi} \int_{\theta_k}^{+\infty} \int_{-\infty}^{+\infty} \prod_{n=0}^1 \frac{\exp(-j\omega x)}{\sqrt{1 - 2j\lambda_{f,f}[n]\omega}} d\omega dx. \quad (52)$$

(Proof in Appendix II) \square

The decisions corresponding to the L measurements for the k th IoT device, i.e., $d_{k,\ell}$, $\ell = 1, 2, \dots, L$, can be fused together as

$$\mathbf{D}_k = \begin{cases} H_{1k}, & \sum_{\ell=1}^L d_{k,\ell} \geq n_k \\ H_{0k}, & \sum_{\ell=1}^L d_{k,\ell} < n_k \end{cases}, \quad (53)$$

where n_k is an integer value [23]. A formal description of the proposed squared ℓ_2 -norm SSR IoT device identification algorithm is given in Algorithm 1.

C. IoT Device Identification for Unknown Probability of Activity

Here, we develop an IoT device identification algorithm for the case of unknown probability of activity P_a at the BS. The

Algorithm 1 Squared ℓ_2 -norm SSR IoT device identification

Input: \mathbf{X} , \mathbf{R} , $P_k^{(t)}$, n_k , $k \in \mathcal{X}_u$

Output: Active IoT set $\hat{\mathcal{X}}_a$

Initialization: $\hat{\mathcal{X}}_a = \emptyset$

- 1: **for** $k = 0, 1, \dots, K_u - 1$ **do**
 - 2: Obtain θ_k by using (51)
 - 3: Obtain $\mathbf{h}_{k,j}$ by employing (27) and (45)
 - 4: Compute $\phi(\mathbf{h}_{k,j})$ for $j = 0, 1, \dots, L - 1$ using (48)
 - 5: Identify the transmission state of the k th IoT device by employing (47) and then (53)
 - 6: **if** $D_k = H_{1k}$ **then**
 - 7: $\hat{\mathcal{X}}_a \leftarrow \{\hat{\mathcal{X}}_a, k\}$
 - 8: **end if**
 - 9: **end for**
-

convex relaxation through $\ell_1 - \ell_2$ mixed-norm is employed for signal reconstruction, which can directly identify the active IoT devices through the non-zero elements of the reconstructed signal. This is attributed to the $\ell_1 - \ell_2$ mixed-norm ability to provide sparse estimates. Note that the proposed $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification algorithm does not require knowledge of transmit power by IoT devices.

1) *BIC $\ell_1 - \ell_2$ Mixed-Norm SSSR IoT Device Identification Algorithm:* Let us consider $P_a \in [0, P_{\max}]$, where P_a is unknown at the BS, but the maximum probability of activity P_{\max} is *a priori* known at the BS.

Since in-phase and quadrature components tend to be either zero or non-zero simultaneously, this provides additional grouping in SSSR. By stacking the in-phase and quadrature components (\mathbf{X} is a real-valued matrix), we can write (19) as

$$\mathbf{Y} = \mathbf{X}\mathbf{U} + \mathbf{V}, \quad (54)$$

where

$$\mathbf{Y} \triangleq [\text{Re}\{\mathbf{R}\} \quad \text{Im}\{\mathbf{R}\}] \quad (55a)$$

$$\mathbf{U} \triangleq [\text{Re}\{\mathbf{H}\} \quad \text{Im}\{\mathbf{H}\}] \quad (55b)$$

$$\mathbf{V} \triangleq [\text{Re}\{\mathbf{W}\} \quad \text{Im}\{\mathbf{W}\}]. \quad (55c)$$

For block-sparse matrix \mathbf{U} in (54), the $\ell_1 - \ell_2$ mixed-norm SSSR is given as follows

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\mathbf{U}\|_F^2 + N_d \lambda_g \sum_{k=0}^{K_u-1} \|\mathbf{u}_{G_k}\|_2, \quad (56)$$

where $N_d \triangleq 2LN_c$

$$\mathbf{u}_{G_k} \triangleq [\mathbf{U}_{2k}, \mathbf{U}_{2k+1}, \cdot], \quad (57)$$

and $\mathbf{U}_{k,\cdot}$ is the k th row of \mathbf{U} . In (56), λ_g represents the tuning parameter which is unknown and time-varying. The degrees of sparsity depends on λ_g ; the larger λ_g is, the sparser the estimate is. For unknown λ_g , model order selection methods can be employed to identify active IoT devices. By extending the BIC model order selection method in [24] to multiple measurement vectors, the reconstructed matrix $\hat{\mathbf{U}}$ is given by

$$\hat{\mathbf{U}} = \hat{\mathbf{U}}^{(\hat{\lambda})}, \quad (58)$$

where

$$\hat{\lambda} = \arg \min_{\lambda \in [\lambda_L, \lambda_U]} C_{\text{BIC}}(\lambda), \quad (59)$$

$$C_{\text{BIC}}(\lambda) \triangleq \log\left(\frac{1}{N_d} \|\mathbf{Y} - \mathbf{X}\hat{\mathbf{U}}^{(\lambda)}\|_F^2\right) + \log(N_d) \frac{\text{df}}{N_d}, \quad (60)$$

$$\hat{\mathbf{U}}^{(\lambda)} = \arg \min_{\mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\mathbf{U}\|_F^2 + N_d \lambda \sum_{k=0}^{K_u-1} \|\mathbf{u}_{G_k}\|_2, \quad (61)$$

and df is the degree of freedom which is given as follows

$$\text{df} = \sum_{k=0}^{K_u-1} \mathbb{I}\{\|\hat{\mathbf{u}}_{G_k}\|_2 > 0\} + (2L - 1) \sum_{k=0}^{K_u-1} \frac{\|\hat{\mathbf{u}}_{G_k}\|_2}{\|\hat{\mathbf{u}}_{G_k}^{\text{LS}}\|_2}, \quad (62)$$

where $\hat{\mathbf{u}}_{G_k}^{\text{LS}}$ is the LS estimate for the k th IoT device signal.

The Karush–Kuhn–Tucker (KKT) optimality conditions of the optimization problem in (61) is given as

$$-\boldsymbol{\Psi}_k + N_d \lambda \frac{\mathbf{u}_{G_k}}{\|\mathbf{u}_{G_k}\|_2} = \mathbf{0} \quad \text{if } \mathbf{u}_{G_k} \neq \mathbf{0}^\dagger \quad (63a)$$

$$\|\boldsymbol{\Psi}_k\|_2 \leq N_d \lambda \quad \text{if } \mathbf{u}_{G_k} = \mathbf{0}^\dagger, \quad (63b)$$

where

$$\begin{aligned} \boldsymbol{\Psi}_k &\triangleq \nabla_{\mathbf{u}_{G_k}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\mathbf{U}\|_F^2 \\ &= [\mathbf{X}_{:,2k}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U}) \quad \mathbf{X}_{:,2k+1}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U})], \end{aligned} \quad (64)$$

and $\mathbf{X}_{:,k}$ is the k th column of \mathbf{X} . Let us write $\boldsymbol{\Psi}_k$ as

$$\boldsymbol{\Psi}_k = \boldsymbol{\varphi}_k - \mathbf{u}_{G_k} \boldsymbol{\Lambda}_k \quad (65)$$

where

$$\boldsymbol{\varphi}_k = [\mathbf{X}_{:,2k}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U}_{-\{2k,\cdot\}}) \quad \mathbf{X}_{:,2k+1}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U}_{-\{2k+1,\cdot\}})], \quad (66)$$

and

$$\begin{aligned} \boldsymbol{\Lambda}_k &\triangleq \text{diag}\{\mathbf{X}_{:,2k}^\dagger \mathbf{X}_{:,2k}, \dots, \mathbf{X}_{:,2k}^\dagger \mathbf{X}_{:,2k}, \\ &\quad \mathbf{X}_{:,2k+1}^\dagger \mathbf{X}_{:,2k+1}, \dots, \mathbf{X}_{:,2k+1}^\dagger \mathbf{X}_{:,2k+1}\} \end{aligned} \quad (67)$$

with $\mathbf{U}_{-\{i,\cdot\}}$ as the matrix \mathbf{U} with the i th row being set to $\mathbf{0}^\dagger$. The dimension of the diagonal matrix $\boldsymbol{\Lambda}_k$ is $4L \times 4L$.

From (63a) and (65), we have $\mathbf{u}_{G_k} (N_d \lambda \mathbf{I} / \|\mathbf{u}_{G_k}\|_2 + \boldsymbol{\Lambda}_k) = \boldsymbol{\varphi}_k$ when the k th IoT device is active. In contrast, when the k th IoT device is inactive, $\boldsymbol{\Psi}_k = \boldsymbol{\varphi}_k$. Hence, we can write

$$\mathbf{u}_{G_k} = \mathbb{I}\{\|\boldsymbol{\varphi}_k\|_2 > N_d \lambda\} \boldsymbol{\varphi}_k \left(\frac{N_d \lambda}{\|\mathbf{u}_{G_k}\|_2} \mathbf{I} + \boldsymbol{\Lambda}_k \right)^{-1}. \quad (68)$$

To solve the optimization (61), we can use block-coordinate descent algorithm, where consists of solving each \mathbf{u}_{G_k} in (57) at a time. By starting from a sparse solution like, $\hat{\mathbf{U}} = \mathbf{0}$, at each iteration, we check for a given k whether \mathbf{u}_{G_k} is optimal or not based on the conditions in (63). If $\|\boldsymbol{\varphi}_k\|_2 \leq N_d \lambda$, $\hat{\mathbf{u}}_{G_k} = \mathbf{0}$; otherwise, \mathbf{u}_{G_k} at the t th iteration is iteratively updated as

$$\hat{\mathbf{u}}_{G_k}^{[t]} = \mathbb{I}\{\|\boldsymbol{\varphi}_k^{[t]}\|_2 > N_d \lambda\} \boldsymbol{\varphi}_k^{[t]} \left(\frac{N_d \lambda}{\|\hat{\mathbf{u}}_{G_k}^{[t-1]}\|_2} \mathbf{I} + \boldsymbol{\Lambda}_k \right)^{-1}, \quad (69)$$

Algorithm 2 BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification algorithm

Input: \mathbf{Y} , \mathbf{X} , \mathbf{A}_k , $k = 0, 1, \dots, K_u - 1$, $\lambda \in [\lambda_L, \lambda_U]$, M_C , and M_G
Output: Active IoT set $\hat{\mathcal{X}}_a$
Initialization: $\hat{\mathcal{X}}_a = \emptyset$, $i = 1$, $Golden = 1$, $\lambda_1 = \lambda_L$

- 1: **while** $Golden = 1$ **do**
 $\hat{\mathbf{U}}^{[0]} = \hat{\mathbf{U}}^{RD}$, $t = 1$, $SSSR = 1$
- 2: **while** $SSSR = 1$ **do**
- 3: **for** $k = 0, 1, \dots, K_u - 1$ **do**
 Obtain $\boldsymbol{\varphi}_k^{[t]}$ by employing (70)
- 4: **if** $\|\boldsymbol{\varphi}_k^{[t]}\|_2 \leq N_c \lambda_i$ **then**
 $\hat{\mathbf{u}}_{G_k}^{[t]} = \mathbf{0}$
 else
 Update $\mathbf{u}_{G_k}^{[t]}$ as in (69)
 end if
- 5: **end for**
- 6: **if** $(\|\hat{\mathbf{U}}^{[t]} - \hat{\mathbf{U}}^{[t-1]}\| \geq \epsilon_c) \cap (t < M_C)$ **then**
- 7: $t \leftarrow t + 1$
- 8: **else**
 $\hat{\mathbf{U}}^{(\lambda_i)} = \hat{\mathbf{U}}^{[t]}$, $SSSR = 0$
 end if
- 9: **end while**
 $i \leftarrow i + 1$
- 10: **if** $i = 2$ **then**
 $\lambda_i = \lambda_U$, $Golden = 1$
- 11: **else if** $(|\lambda_{i-1} - \lambda_{i-2}| \geq \epsilon_g) \cap (i < M_G + 1)$ **then**
 Find λ_i using the Golden selection search
 $Golden = 1$
- 12: **else**
 $\hat{\lambda} = \arg \min_{\lambda \in \{\lambda_{i-1}, \lambda_{i-2}\}} C_{BIC}(\lambda)$,
 $\hat{\mathbf{U}} = \hat{\mathbf{U}}^{(\hat{\lambda})}$, $Golden = 0$,
- 13: **end if**
- 14: **end while**
- 15: $\hat{\mathcal{X}}_a = \{\forall k \in \{0, \dots, K_u - 1\} \mid \|\hat{\mathbf{u}}_{G_k}\|_2 \neq 0\}$.

where

$$\boldsymbol{\varphi}_k^{[t]} = [\mathbf{X}_{:,2k}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U}_{-\{2k\}}^{[t-1]}) \quad \mathbf{X}_{:,2k+1}^\dagger (\mathbf{Y} - \mathbf{X}\mathbf{U}_{-\{2k+1\}}^{[t-1]})] \quad (70)$$

This procedure continues until the absolute difference of successive iterations becomes smaller than the tolerance value ϵ_c .

2) *Efficient One-dimensional Search:* Efficient one dimensional iterative search algorithms can be used to solve the optimization problem in (59). In an iterative search method, the interval $[\lambda_L, \lambda_U]$ is repeatedly reduced on the basis of function evaluations until a reduced bracket $[\lambda_L, \lambda_U]$ is achieved which is sufficiently small. These methods can be applied to any function and differentiability of the function is not essential. An iterative search method in which iterations can be performed until the desired accuracy in either the minimizer or the minimum value of the objective function is achieved is the golden-section search method [25]. For a strictly unimodal function with an extremum inside the interval, the Golden-section search method finds that extremum, while for an in-

terval containing multiple extrema (possibly including the interval boundaries), it converges to one of them.

Convergence of the Optimization Problem in (61): It has been shown that for an optimization problem whose objective function is the sum of a smooth and convex function and a non-smooth but block-separable convex function, block-coordinate descent optimization converges towards the global minimum of the problem [26]. In (61), $\|\mathbf{Y} - \mathbf{X}\mathbf{U}\|_F^2$ is a smooth and differentiable convex function and $\sum_{k=0}^{K_u-1} \|\mathbf{u}_{G_k}\|_2$ is a separable penalty function, where $\|\mathbf{u}_{G_k}\|_2$ is a continuous and convex function with respect to \mathbf{u}_{G_k} . Thus, block-coordinate descent converges to the global minimum.

A formal description of the $\ell_1 - \ell_2$ mixed-norm SSSR IoT identification algorithm is summarized in Algorithm 2. In Algorithm 2, M_G and M_C , denote the maximum number of iterations for the Golden selection search and the block-coordinate descent optimization, respectively.

IV. DATA DETECTION

The next step after IoT device identification is to detect the data of devices identified as active. Since channel state information (CSI) is unknown, the existing MUD algorithms, such as SIC cannot be employed. In this section, we propose a new nonlinear MUD algorithm which does not require channel estimation for data detection.

A. 2-MC-MUD Algorithm

Fig. 8 shows the block diagram of the proposed 2-MC-MUD algorithm. The output of the IoT device identification algorithm is a set of IoT devices $\hat{\mathcal{X}}_a$. Since the delay of the IoT devices are known, we can apply sequence matched filtering to the small set of active IoT devices. Without loss of generality, we assume that $\hat{\mathcal{X}}_a \triangleq \{k_0, k_1, \dots, k_{\hat{K}_a-1}\}$ and $\tau_{k_0} \leq \tau_{k_1} \leq \dots \leq \tau_{k_{\hat{K}_a-1}}$, where $\hat{K}_a \triangleq \text{card}(\hat{\mathcal{X}}_a)$.

We consider a bank of \hat{K}_a single-user matched filter (MF) for the identified active IoT devices in $\hat{\mathcal{X}}_a$, as shown in Fig. 8. The output of the MF after synchronized sampling and normalization by N_c for the k_n th IoT device is expressed as [27]

$$\begin{aligned} y_{k_n,i} &\triangleq \frac{1}{N_c} \int_{\tau_{k_n+iT_s}}^{\tau_{k_n}+(i+1)T_s} r(t) s_{k_n}(t - iT_s - \tau_{k_n}) dt \quad (71) \\ &= \mathbf{g}_{k_n} \mathbf{b}_{k_n,i} + \sum_{k_j < k_n} \mathbf{g}_{k_j} \mathbf{b}_{k_j,i+1} \rho_{k_n k_j} + \sum_{k_j < k_n} \mathbf{g}_{k_j} \mathbf{b}_{k_j,i} \rho_{k_j k_n} \\ &\quad + \sum_{k_j > k_n} \mathbf{g}_{k_j} \mathbf{b}_{k_j,i} \rho_{k_n k_j} + \sum_{k_j > k_n} \mathbf{g}_{k_j} \mathbf{b}_{k_j,i-1} \rho_{k_j k_n} + \mathbf{w}_{k_n,i}, \end{aligned}$$

where $\mathbf{w}_{k_n,i} \triangleq \sigma_w \int_{\tau_{k_n+iT_s}}^{\tau_{k_n}+(i+1)T_s} w(t) s_{k_n}(t - iT_s - \tau_{k_n}) dt$, $\rho_{k_n k_j} \triangleq \frac{1}{N_c} \int_{\tau_{k_j}}^{\tau_{k_n}} s_{k_n}(t) s_{k_j}(t - \tau_{k_j}) dt$, and $\rho_{k_j k_n} \triangleq \frac{1}{N_c} \int_0^{\tau_{k_j}} s_{k_n}(t) s_{k_j}(t + T_s - \tau_{k_j}) dt$.

The output of the single-user MF in (71) for the k_n th IoT device can be written as

$$y_{k_n,i} = \mathbf{g}_{k_n} \mathbf{b}_{k_n,i} + \mathbf{v}_{k_n,i}, \quad i = 0, 1, \dots, N_s - 1, \quad (72)$$

where $\mathbf{v}_{k_n,i}$ represents the effect of noise and multiuser interference on the k_n th IoT device.

For data detection without any sign ambiguity, the phase of \mathbf{g}_{k_n} , $k_n \in \hat{\mathcal{X}}_a$, is leastwise required to be known at the BS.

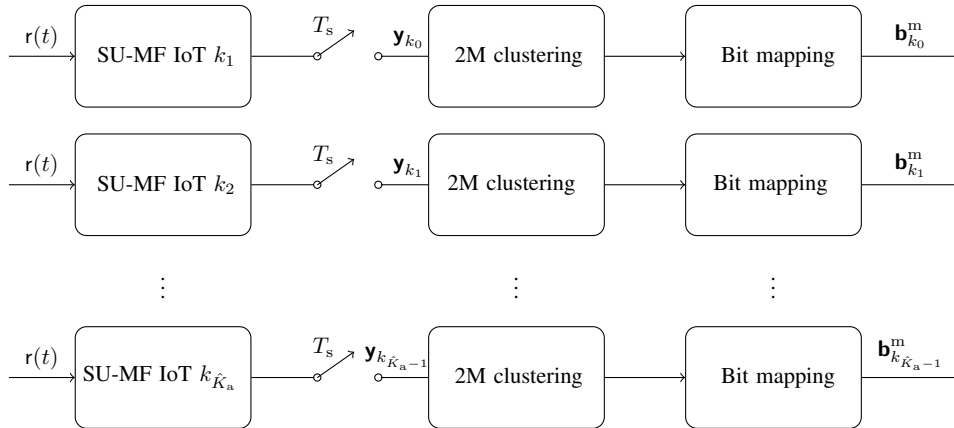


Fig. 8: Block diagram of the proposed 2-MC-MUD algorithm.

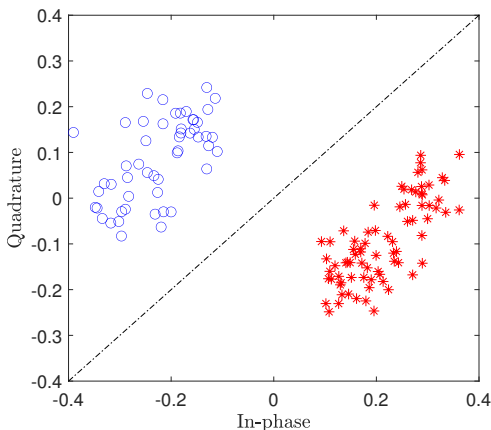


Fig. 9: Separated symbols by the proposed 2-MC-MUD algorithm for an active IoT device at 10 dB SNR, $K_u = 768$, and $N_c = 512$.

However, by employing differential coding at IoT devices, a MUD algorithm can be developed which removes the need for such *a priori* knowledge. Differential coding is a coding technique used for non-coherent data detection. Instead of encoding a bit sequence directly, it encodes the difference between the bit sequence as $\mathbf{b}_{k_n,i} = \mathbf{b}_{k_n,i-1} \oplus \mathbf{b}_{k_n,i}^c$, $k_n \in \mathcal{X}_a$, [21], where \oplus is the modulo-2 addition and $\mathbf{b}_{k_n,i}^c$ is the i th bit at the output of the channel encoder of the k_n th IoT device.

Since \mathbf{g}_{k_n} , $k_n \in \mathcal{X}_a$, remains unchanged during the short packet, the received symbols of the active IoT device k_n in (72) form two clusters corresponding to the transmitted bits 1 and 0. The main idea behind the proposed MUD algorithm is to extract these two clusters regardless of which cluster is labeled 1 or 0. By extracting the two clusters, the data stream of the active IoT device k_n can be detected without any prior knowledge about the CSI and carrier phase (CP) due to differential coding.

To extract these two clusters for each active IoT device, the 2-MC algorithm can be employed. By applying the 2-MC algorithm to $y_{k_n,i}$, $i = 0, 1, \dots, N_s - 1$, in (72), the two clusters are separated based on the nearest mean criterion regardless to the label. The 2-MC minimizes the within-cluster sum of squares (WCSS), i.e., the sum of the squared Euclidean distance [28]. Let us define $\mathcal{U} \triangleq \{0, 1, \dots, N_s - 1\}$. The 2-MC algorithm partitions \mathcal{U} into two sets $\mathcal{U}_{k_n,0}$ and $\mathcal{U}_{k_n,1}$

by minimizing the WCSS as follows

$$\begin{aligned} \arg \min_{\mathcal{U}} \quad & \sum_{i \in \mathcal{U}_{k_n,0}} |y_{k_n,i} - \mu_{k_n,0}|^2 + \sum_{i \in \mathcal{U}_{k_n,1}} |y_{k_n,i} - \mu_{k_n,1}|^2, \\ \text{subject to} \quad & \mu_{k_n,0} = \frac{1}{\text{card}(\mathcal{U}_{k_n,0})} \sum_{i \in \mathcal{U}_{k_n,0}} y_{k_n,i}, \\ & \mu_{k_n,1} = \frac{1}{\text{card}(\mathcal{U}_{k_n,1})} \sum_{i \in \mathcal{U}_{k_n,1}} y_{k_n,i}. \end{aligned} \quad (73)$$

The minimization problem in (73) can be solved by different methods. One of the most common algorithm is the Lloyd's algorithm which uses an iterative refinement technique [29]. Given initial mean values $\mu_{k_n,0}^{[0]}$ and $\mu_{k_n,1}^{[1]}$ for $\mu_{k_n,0}$ and $\mu_{k_n,1}$ in (73), the Lloyd's algorithm proceeds by alternating between the assignment and updating steps as follows:

Assignment Step: The element of \mathcal{U} at iteration t , i.e., $\mathcal{U}^{[t]}$ is assigned to $\mathcal{U}_{k_n,0}^{[t]}$ when

$$\mathcal{U}_{k_n,0}^{[t]} = \left\{ i : |y_{k_n,i} - \mu_{k_n,0}^{[t]}| \leq |y_{k_n,i} - \mu_{k_n,1}^{[t]}| \right\}. \quad (74)$$

Otherwise, it is assigned to $\mathcal{U}_{k_n,1}^{[t]}$.

Updating Step: The mean of the the clusters $\mathcal{U}_{k_n,0}^{[t]}$ and $\mathcal{U}_{k_n,1}^{[t]}$ are updated as

$$\mu_{k_n,1}^{[t+1]} = \frac{1}{\text{card}(\mathcal{U}_{k_n,1}^{[t]})} \sum_{i \in \mathcal{U}_{k_n,1}^{[t]}} y_{k_n,i}, \quad (75a)$$

$$\mu_{k_n,0}^{[t+1]} = \frac{1}{\text{card}(\mathcal{U}_{k_n,0}^{[t]})} \sum_{i \in \mathcal{U}_{k_n,0}^{[t]}} y_{k_n,i}. \quad (75b)$$

The 2-MC algorithm converges when the assignment step does not change. Fig. 9 shows the output of the 2-MC algorithm for an active IoT device. As seen, the sequence at the output of the MF is portioned into two clusters regardless of the label.

After partitioning \mathcal{U} into two clusters $\mathcal{U}_{k_n,0}$ and $\mathcal{U}_{k_n,1}$, $y_{k_n,i}$, $i = 0, 1, \dots, N_s - 1$, is mapped into a binary sequence $\mathbf{b}_{k_n}^m \triangleq [\mathbf{b}_{k_n,0}^m \ \mathbf{b}_{k_n,1}^m \ \dots \ \mathbf{b}_{k_n,N_s-1}^m]^\dagger$ with elements as

$$\mathbf{b}_{k_n,i}^m = \mathbb{I}\{i \in \mathcal{U}_{k_n,1}\}. \quad (76)$$

Then, by applying differential decoding to the mapped binary sequence $\mathbf{b}_{k_n}^m$, the channel coded data stream for the active IoT device k_n is obtained as follows

Algorithm 3 : 2-MC-MUD algorithm

Input: $r(t)$, $\hat{\mathcal{X}}_a$, $\hat{K}_a = \text{card}(\hat{\mathcal{X}}_a)$
Output: $\hat{\mathbf{b}}_{k_n}$, $k_n \in \hat{\mathcal{X}}_a$
 1: **for** $n = 0, 1, \dots, \hat{K}_a - 1$ **do**
 2: Set initial value for $\mathcal{U}_{k_n,1}^{[0]}$ and $\mathcal{U}_{k_n,0}^{[0]}$
 3: Obtain $y_{k_n,i}$, $i = 0, 1, \dots, N_s - 1$, by employing (71)
 4: **while** $\mathcal{U}_{k_n,1}^{[t+1]} \neq \mathcal{U}_{k_n,1}^{[t]}$ **do**
 5: obtain $\mathcal{U}_{k_n,1}^{[t]}$ and $\mathcal{U}_{k_n,0}^{[t]}$ by employing (74)
 6: $\mu_{k_n,1}^{[t+1]} \leftarrow \mathcal{U}_{k_n,1}^{[t]}$ by employing (75a)
 7: $\mu_{k_n,0}^{[t+1]} \leftarrow \mathcal{U}_{k_n,0}^{[t]}$ by employing (75b)
 8: **end while**
 9: Obtain the binary mapped sequence $\mathbf{b}_{k_n}^m$ through (76)
 10: Apply differential decoding in (77) to $\mathbf{b}_{k_n}^m$ to obtain $\hat{\mathbf{b}}_{k_n}^c$
 11: Apply channel decoding to $\hat{\mathbf{b}}_{k_n}^c$ to obtain \mathbf{d}_{k_n}
 12: **end for**

$$\hat{\mathbf{b}}_{k_n,i}^c = \mathbf{b}_{k_n,i}^m \oplus \mathbf{b}_{k_n,i-1}^m. \quad (77)$$

Finally, $\hat{\mathbf{b}}_{k_n}^c \triangleq [\hat{\mathbf{b}}_{k_n,0}^c \ \hat{\mathbf{b}}_{k_n,1}^c \ \dots \ \hat{\mathbf{b}}_{k_n,N_s-2}^c]^\dagger$ is decoded by the channel decoder, and the data stream of the active IoT device k_n is obtained. The proposed 2-MC-MUD algorithm is summarized in Algorithm 3.

B. Complexity Analysis

The complexity of the proposed squared ℓ_2 -norm SSR IoT device identification algorithm is $\mathcal{O}(K_u L N_c^2 + L N_c^3)$. The complexity of the proposed BIC ℓ_1 - ℓ_2 mixed-norm SSSR IoT device identification algorithm per each iteration is $\mathcal{O}(N_c L K_u^2)$, where the maximum number of iterations is $M_G M_C$ (M_G and M_C are the maximum number of iterations for the Golden selection search and the block-coordinate descent optimization, respectively). The complexity of the proposed 2-MC-MUD for single user matched filtering and clustering is $\mathcal{O}(N_c^2 N_s k_a)$ and $\mathcal{O}(K_u N_s k_a)$, respectively, where $k_a = \text{card}(\mathcal{X}_a)$.

V. SIMULATION RESULTS

In this section, we examine the performance of the proposed IoT device identification algorithms and the 2-MC-MUD algorithm through several simulation experiments.

A. Simulation Setup

We considered an IoT network with $K_u = 1024$ IoT devices. It is assumed that the spreading sequences of the IoT devices are random binary codes with spreading factor $N_c = 512$. Unless otherwise specified, the overloading factor is $OF \triangleq K_u/N_c = 2$. Each IoT packet is 128 bits with payload length of 40 bits. A polar code with list size 8 and code rate 40/128 was adopted for channel coding. The delay of the IoT devices was generated as uniform distributions $\alpha_k \sim \mathcal{U}_d[0, 5]$, $\beta_k \sim \mathcal{U}_d[0, 511]$, and $\xi_k \sim \mathcal{U}_c[0, 1)$. The effect of the unknown CSI and CP for each IoT device was modeled as independent complex Gaussian random variables with mean $\mu_k = \sqrt{0.1} + j\sqrt{0.1}$ and variance $\sigma_k^2 = 1$, $k \in \mathcal{X}_u$, i.e., Rician fading with K -factor 0.2 was considered. The average system SNR was defined as $\vartheta \triangleq \bar{P}_a \sum (|\mu_k|^2 + \sigma_k^2) p_k \eta_k / \sigma_w^2$, where $p_k = \zeta / \eta_k$ (ζ changes according to ϑ), $\bar{P}_a = P_a$ for Algorithm 1, $\bar{P}_a = P_{\max}/2$ for Algorithm 2 (for the case of time-varying P_a),

and $\sigma_w^2 = 1$ is the variance of the additive noise. The range of tuning parameter for the BIC minimization in (59) was set as $\lambda = [0 \ 500]$, $\epsilon_g = 2$, and $M_G = 20$. The performance of the proposed IoT device identification algorithms were evaluated in terms of system correct identification and system false alarm rates, which are calculated as follows

$$P_C = \frac{1}{N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \frac{\text{card}(\hat{\mathcal{X}}_a[n] \cap \mathcal{X}_a[n])}{\text{card}(\mathcal{X}_a[n])},$$

and

$$P_F = \frac{1}{N_{\text{MC}}} \sum_{n=1}^{N_{\text{MC}}} \frac{\text{card}(\hat{\mathcal{X}}_a[n]) - \text{card}(\hat{\mathcal{X}}_a[n] \cap \mathcal{X}_a[n])}{K_u - \text{card}(\mathcal{X}_a[n])},$$

where $N_{\text{MC}} = 10^4$ is the number of Monte Carlo trials, and $\mathcal{X}_a[n]$ and $\hat{\mathcal{X}}_a[n]$ are the true and estimated active IoT device set at the n th Monte Carlo trial, respectively. Also, the performance of the proposed 2-MC-MUD algorithm was evaluated in terms of average packet error rate (PER) in the presence IoT device identification error.

B. Simulation Results

Fig. 10 depicts P_C and P_F of the proposed squared ℓ_2 -norm SSR IoT device identification algorithm (Algorithm 1) versus SNR for different values of P_a and $P_f^{(k)}$, $K_u = 1024$, and $L = 1$. The threshold values θ_k , $k \in \mathcal{X}_u$, are set by using (51). As seen, Algorithm 1 can offer high correct identification error rate even for a single observation vector. Also, there is an insignificant gap between P_C obtained in the simulation experiment and the theoretical result in (52). Similarly, P_F matches the preset false alarm rate, i.e., $P_f^{(k)} \in \{0.03, 0.04, 0.05\}$, $k \in \mathcal{X}_u$. We note from Fig. 10 that the theoretical results more accurately match the simulation results at higher P_a and at lower SNRs since the CLT is more reliable and the deviation from the Gaussian distribution is lower.

In Fig. 11, we compare the performance of Algorithm 1 and Algorithm 2 with the Alternating Direction Method of Multipliers (ADMM) using GCV [30] for $P_a = 0.02$, $K_u = 1024$, and $L = 21$. For Algorithm 1, the threshold values θ_k , $k \in \mathcal{X}_u$, are set for $P_k^f = 0.05$ in (51), and $n_k = 4$ and 5 are used for hard decision combining in (53). As seen, while Algorithm 1 outperforms Algorithm 2 and the ADMM-GCV algorithm at lower values of SNR, Algorithm 2 exhibits superior performance at higher SNRs. We notice that the performance improvement of Algorithm 1 for $L = 21$ is not very high compared to $L = 1$ in Fig. 10 since $p(\hat{h}_{k,j_1,m}, \hat{h}_{k,j_2,m} | H_{tk}) \neq p(\hat{h}_{k,j_1,m} | H_{tk}) p(\hat{h}_{k,j_2,m} | H_{tk})$, $j_1 \neq j_2$, $t \in \{0, 1\}$.

Fig. 12 illustrates P_C and P_F of the proposed BIC ℓ_1 - ℓ_2 mixed-norm SSSR IoT device identification algorithm versus SNR for different overloading factors when the probability of activity varies uniformly in the range $P_a \in [0, 0.06]$. As seen, the proposed algorithm exhibits high correct identification rate for high overloading factors, such as $OF = 2$ (1024 devices), when P_a is unknown and time-varying. Also, the false alarm rate of the proposed algorithm is significantly low for the SNR values lower than 15 dB due to the capability of the BIC ℓ_1 - ℓ_2 mixed-norm SSSR algorithm to exploit block sparsity.

Fig. 13 compares the performance of the developed MA scheme when the proposed 2-MC-MUD and differentially co-

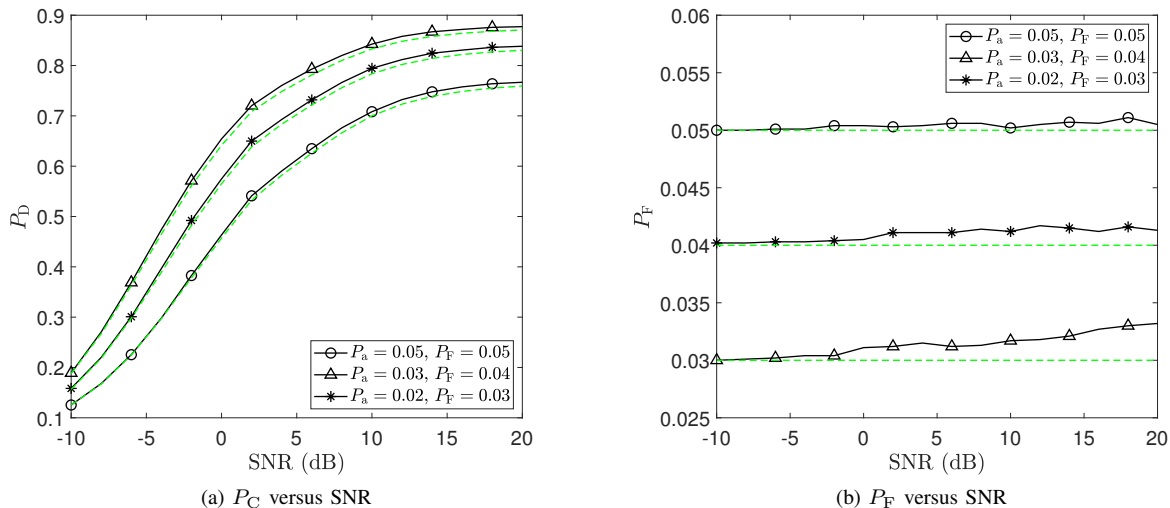


Fig. 10: The system correct identification rate, P_C , and false alarm rate, P_F , of the proposed squared ℓ_2 -norm SSR IoT device identification algorithm (Algorithm 1) versus overloading factor for different values of $P_a, P_F^{(k)} \in \{0.03, 0.04, 0.05\}$, $K_u = 1024$, and $L = 1$.

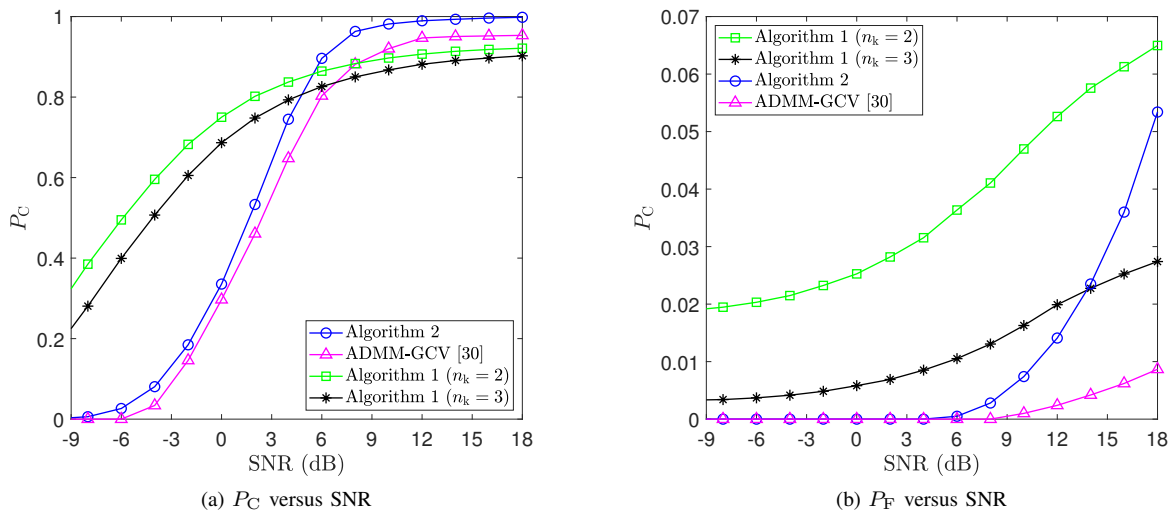


Fig. 11: Performance comparison of the proposed IoT device identification algorithms with the ADMM-GCV algorithm [30] for $P_a = 0.02$, $K_u = 1024$, and $L = 21$.

herent decorrelation (DCD)-MUD [27] algorithms are employed for non-coherent data detection. As seen, the proposed MUD algorithm outperforms the DCD-MUD [27]. This superiority in performance is related to the capability of the 2-MC algorithm to accurately separate the two clusters of data.

VI. CONCLUSION

A new uncoordinated uplink MA for mMTC with short-packet and sporadic traffic was proposed in this paper. The proposed MA scheme reduces the control signaling associated with the MAC and PHY layers. Instead of transmitting the device identifier using a portion of bits in a packet, the squared ℓ_2 -norm SSR and BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification algorithms were proposed to identify active IoT devices through the assigned unique non-orthogonal spreading code to each IoT device. To further reduce the overhead, we removed the preambles and pilots used for channel estimation by developing the non-coherent 2-MC-MUD algorithm based on unsupervised machine learning.

APPENDIX I

By applying the statistical expectation to (35) and employing $\mathbb{E}\{h_{k,j,1}|H_{tk}\} = 0$ and $\mathbb{E}\{h_{k,j,0}|H_{tk}\} = 0$, we can write $\mathbb{E}\{\hat{h}_{k,j,f}|H_{tk}\} = 0$, $t, f \in \{0, 1\}$. To obtain the variance of $\hat{h}_{k,j,f}$ given H_{tk} , i.e., $\Sigma_{f,f}^{tk}$, we use the variance sum law as follows

$$\text{Var}\left\{\sum_i a_i z_i\right\} = \sum_i \left(|a_i|^2 \text{Var}\{z_i\} + \sum_{j \neq i} a_i a_j^* \text{Cov}\{z_i, z_j\}\right). \quad (78)$$

Since $h_{k_1,j,f}$, $h_{k_2,j,\bar{f}}$, and $w'_{k,j,f}$, $k, k_1, k_2 \in \mathcal{X}_u$, in (35), are zero-mean and uncorrelated random variables, by applying (78) to (35), we obtain (79) at the top of next page.

By using (17), $\text{Var}\{h_{k,j,f}|H_{1k}\}$, $f \in \{0, 1\}$, in (79) can be written as

$$\begin{aligned} \text{Var}\{h_{k,j,f}|H_{tk}\} &= \text{Var}\{g_k \mathbf{b}_{k,j-\alpha_k-1+f}|H_{tk}\} \\ &= \mathbb{E}\left\{\text{Var}\{g_k \mathbf{b}_{k,j-\alpha_k-1+f}|g_k, H_{tk}\}\right\} \\ &\quad + \text{Var}\left\{\mathbb{E}\{g_k \mathbf{b}_{k,j-\alpha_k-1+f}|g_k, H_{tk}\}\right\}. \end{aligned} \quad (80)$$

$$\begin{aligned} \Sigma_{f,f}^{tk} &= \text{Var}\{\hat{\mathbf{h}}_{k,j,f}|H_{tk}\} = \mathbb{E}\{|\hat{\mathbf{h}}_{k,j,f}|^2|H_{tk}\} = t\Omega_{2k+f,2k+f}^2 \text{Var}\{\mathbf{h}_{k,j,f}|H_{tk}\} + t\Omega_{2k+f,2k+f}^2 \text{Var}\{\mathbf{h}_{k,j,\bar{f}}|H_{tk}\} \\ &+ \sum_{n \neq k} \Omega_{2k+f,2n+f}^2 \text{Var}\{\mathbf{h}_{n,j,f}|H_{tk}\} + \sum_{n \neq n} \Omega_{2k+f,2n+\bar{f}}^2 \text{Var}\{\mathbf{h}_{n,j,\bar{f}}|H_{tk}\} + \text{Var}\{\mathbf{w}'_{k,j,f}\}. \end{aligned} \quad (79)$$

$$\begin{aligned} \Sigma_{0,1}^{1k} &= \text{Cov}\{\hat{\mathbf{h}}_{k,j,0}, \hat{\mathbf{h}}_{k,j,1}|H_{tk}\} = \mathbb{E}\{\hat{\mathbf{h}}_{k,j,0} \hat{\mathbf{h}}_{k,j,1}^*|H_{tk}\} = t(\Omega_{2k,2k} \Omega_{2k+1,2k}) \mathbb{E}\{|\mathbf{h}_{k,j,0}|^2|H_{tk}\} \\ &+ t(\Omega_{2k+1,2k+1} \Omega_{2k,2k+1}) \mathbb{E}\{|\mathbf{h}_{k,j,1}|^2|H_{tk}\} + \sum_{n \neq k} \Omega_{2k,2n} \Omega_{2k+1,2n} \mathbb{E}\{|\mathbf{h}_{n,j,0}|^2|H_{tk}\} \\ &+ \sum_{n \neq k} \Omega_{2k+1,2n+1} \Omega_{2k,2n+1} \mathbb{E}\{|\mathbf{h}_{n,j,1}|^2|H_{tk}\} + \mathbb{E}\{\mathbf{w}'_{k,j,0} (\mathbf{w}'_{k,j,1})^*|H_{tk}\}. \end{aligned} \quad (86)$$

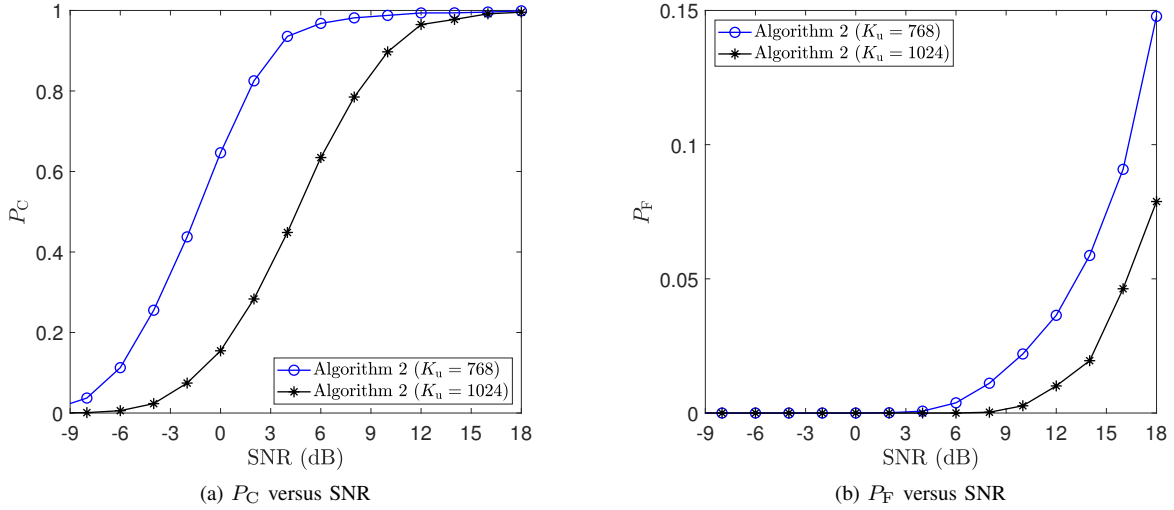


Fig. 12: The system correct identification rate, P_C , and false alarm rate, P_F , of the proposed BIC $\ell_1 - \ell_2$ mixed-norm SSSR IoT device identification algorithm (Algorithm 2) versus SNR and overloading factor for random probability of activity $P_a \in [0, 0.06]$ and $L = 21$.

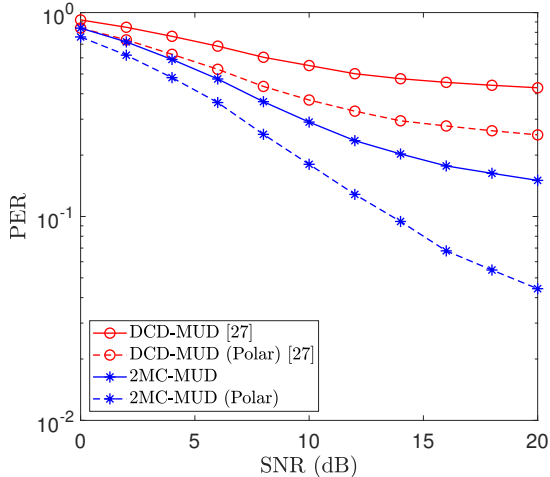


Fig. 13: PER of the proposed MA scheme when 2-MC-MUD and DCD-MUD [27] algorithms are employed for non-coherent data detection at the BS ($K_u = 768$, $OF = 1.5$, and $P_a \in [0, 0.06]$).

Since $\text{Var}\{\mathbf{b}_{k,j-\alpha_k-1+f}|H_{tk}\} = t$, t , $f \in \{0, 1\}$, we can write

$$\mathbb{E}\left\{\text{Var}\{\mathbf{g}_k \mathbf{b}_{k,j-\alpha_k-1+f} | \mathbf{g}_k, H_{tk}\}\right\} = \mathbb{E}\{|\mathbf{g}_k|^2\} \text{Var}\{\mathbf{b}_{k,j-\alpha_k-1+f}|H_{tk}\} = t(\sigma_k^2 + |\mu_k|^2) \eta_k p_k. \quad (81)$$

By substituting (81) and $\mathbb{E}\{\mathbf{g}_k \mathbf{b}_{k,j-\alpha_k-1+f} | \mathbf{g}_k, H_{tk}\} = 0$, $f \in \{0, 1\}$, into (80), we obtain

$$\text{Var}\{\mathbf{h}_{k,j,f}|H_{tk}\} = t(\sigma_k^2 + |\mu_k|^2) \eta_k p_k. \quad (82)$$

Similar to (80), for $n \neq k$ and $f \in \{0, 1\}$, we can write

$$\begin{aligned} \text{Var}\{\mathbf{h}_{n,j,f}|H_{tk}\} &= \text{Var}\{\mathbf{g}_n \mathbf{b}_{n,j-\alpha_n-1+f}|H_{tk}\} \\ &= \mathbb{E}\left\{\text{Var}\{\mathbf{g}_n \mathbf{b}_{n,j-\alpha_n-1+f} | \mathbf{g}_n, H_{tk}\}\right\} \\ &+ \text{Var}\left\{\mathbb{E}\{\mathbf{g}_n \mathbf{b}_{n,j-\alpha_n-1+f} | \mathbf{g}_n, H_{tk}\}\right\}. \end{aligned} \quad (83)$$

Because $\text{Var}\{\mathbf{b}_{n,j-\alpha_n-1+f}|H_{tk}\} = \text{Var}\{\mathbf{b}_{n,j-\alpha_n-1+f}\} = P_a$, $n \neq k$, $f \in \{0, 1\}$, we can write

$$\mathbb{E}\left\{\text{Var}\{\mathbf{g}_n \mathbf{b}_{n,j-\alpha_n-1+f} | \mathbf{g}_n, H_{tk}\}\right\} = \mathbb{E}\{|\mathbf{g}_n|^2\} \text{Var}\{\mathbf{b}_{n,j-\alpha_n-1+f}|H_{tk}\} = P_a(\sigma_n^2 + |\mu_n|^2) \eta_n p_n. \quad (84)$$

By substituting (84) and $\mathbb{E}\{\mathbf{g}_n \mathbf{b}_{n,j-\alpha_n-1+f} | \mathbf{g}_n, H_{tk}\} = 0$, $f \in \{0, 1\}$, into (83), we obtain

$$\text{Var}\{\mathbf{h}_{n,j,f}|H_{tk}\} = P_a(\sigma_n^2 + |\mu_n|^2) \eta_n p_n. \quad (85)$$

Finally, by substituting (82), (85), $\text{Var}\{\mathbf{w}'_{k,j,f}\} = \Sigma_{2k+f,2k+f}^{w'}$ into (79), (37) is derived. For the cross-correlation of $\hat{\mathbf{h}}_{k,j,0}$ and $\hat{\mathbf{h}}_{k,j,1}$, we obtain (86), where by substituting (82) and (85) into (86), and then by using $\mathbb{E}\{\mathbf{w}'_{k,j,0} (\mathbf{w}'_{k,j,1})^* | H_{tk}\} = \Sigma_{2k,2k+1}^{w'}$, results in (38).

APPENDIX II

By employing the MLR test, the transmission state of the k th IoT device is identified as active, i.e., $\mathbf{d}_k = H_{1k}$, if

$$\frac{p(\check{\mathbf{h}}_{k,j}|H_{1k})}{p(\check{\mathbf{h}}_{k,j}|H_{0k})} = \frac{2\pi |C_{f,f}^{0k}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \check{\mathbf{h}}_{k,j}^\dagger (C_{f,f}^{1k})^{-1} \check{\mathbf{h}}_{k,j}\right)}{2\pi |C_{f,f}^{1k}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \check{\mathbf{h}}_{k,j}^\dagger (C_{f,f}^{0k})^{-1} \check{\mathbf{h}}_{k,j}\right)} > \lambda,$$

where $\lambda = (1 - P_a)/P_a$. A canonical form of the above detector is given by [22]

$$\mathbf{h}_{k,j}^\dagger (\mathbf{C}_{f,f}^{0k})^{-1} (\mathbf{C}_{f,f}^{1k}) (\mathbf{C}_{f,f}^{1k} + \mathbf{C}_{f,f}^{0k})^{-1} \check{\mathbf{h}}_{k,j} > \theta_k, \quad (87)$$

where θ_k is determined based on desirable false alarm rate for the k th IoT device. Let us write $\mathbf{C}_{f,f}^{0k} = \mathbf{V}_{f,f}^{0k} \mathbf{\Lambda}_{f,f}^{0k} (\mathbf{V}_{f,f}^{0k})^{-1}$, where $\mathbf{V}_{f,f}^{0k}$ is a square matrix whose columns are eigenvectors of $\mathbf{C}_{f,f}^{0k}$, and $\mathbf{\Lambda}_{f,f}^{0k}$ is a diagonal matrix where its i th diagonal element is the eigenvalue associated with the i th column of $\mathbf{C}_{f,f}^{0k}$. We define $\mathbf{A}_{f,f}^{0k} = \mathbf{V}_{f,f}^{0k} (\mathbf{\Lambda}_{f,f}^{0k})^{-\frac{1}{2}}$ and $\mathbf{B}_{f,f}^{1k} \triangleq (\mathbf{A}_{f,f}^{0k})^\dagger \mathbf{C}_{f,f}^{1k} \mathbf{A}_{f,f}^{0k}$. Taking into account $(\mathbf{V}_{f,f}^{0k})^\dagger \mathbf{V}_{f,f}^{0k} = \mathbf{I}$, we can show that $(\mathbf{A}_{f,f}^{0k})^\dagger \mathbf{C}_{f,f}^{0k} \mathbf{A}_{f,f}^{0k} = \mathbf{I}$. Then, using this result, the canonical detector in (87) can be written as follows

$$\mathbf{h}_{k,j}^\dagger \mathbf{A}_{f,f}^{0k} \mathbf{B}_{f,f}^{1k} (\mathbf{B}_{f,f}^{1k} + \mathbf{I})^{-1} (\mathbf{A}_{f,f}^{0k})^\dagger \check{\mathbf{h}}_{k,j} > \theta_k. \quad (88)$$

Based on eigenvalue decomposition of $\mathbf{B}_{f,f}^{0k}$, we have

$$\mathbf{B}_{f,f}^{1k} \triangleq (\mathbf{A}_{f,f}^{0k})^\dagger \mathbf{C}_{f,f}^{1k} \mathbf{A}_{f,f}^{0k} = \mathbf{V}_{f,f}^{1k} \mathbf{\Lambda}_{f,f}^{1k} (\mathbf{V}_{f,f}^{1k})^{-1}, \quad (89)$$

where $\mathbf{V}_{f,f}^{1k}$ and $\mathbf{\Lambda}_{f,f}^{1k}$ are the eigenvector and eigenvalue matrices of $\mathbf{B}_{f,f}^{1k}$, respectively. Since $\mathbf{B}_{f,f}^{1k}$ is a symmetric matrix, we have $(\mathbf{V}_{f,f}^{1k})^\dagger \mathbf{V}_{f,f}^{1k} = \mathbf{I}$. By letting $\mathbf{z}_{k,j} \triangleq [z_{k,j}[0], z_{k,j}[1]]^\dagger \triangleq (\mathbf{V}_{f,f}^{1k})^\dagger (\mathbf{A}_{f,f}^{0k})^\dagger \check{\mathbf{h}}_{k,j}$ in (88), we obtain

$$\mathbf{z}_{k,j}^\dagger \mathbf{\Lambda}_{f,f}^{1k} (\mathbf{\Lambda}_{f,f}^{1k} + \mathbf{I})^{-1} \mathbf{z}_{k,j} > \theta_k, \quad (90)$$

which is equivalent to the test statistics in (48) and (49). Note that the matrix $\mathbf{A}_{f,f}^{0k} \mathbf{V}_{f,f}^{1k}$ diagonalizes both $\mathbf{C}_{f,f}^{0k}$ and $\mathbf{C}_{f,f}^{1k}$.

By employing (48) and (49), the false alarm rate for the k th IoT device is derived as follows

$$\begin{aligned} P_k^{(f)} &= \mathbb{P}\{\mathbf{d}_k = H_{1k} | H_{0k}\} \\ &= \mathbb{P}\left\{ \sum_{n=0}^1 \chi_{f,f}[n] z_{k,j}^2[n] \geq \theta_k | H_{0k} \right\}. \end{aligned} \quad (91)$$

To obtain the PDF of $\mathbf{U} \triangleq \sum_{n=0}^1 \chi_{f,f}[n] z_{k,j}^2[n]$ in (91), we need to derive its characteristic function (CF) and then express the PDF as the inverse Fourier transform. Since $z_{k,j}^2[n]$ in (48) under hypothesis H_{0k} follows the central Chi-squared (χ^2) distribution with 1 degrees of freedom and the fact that $z_{k,j}^2[0]$ and $z_{k,j}^2[1]$ are independent random variables, we obtain the CF of \mathbf{U} as follows

$$\phi_{\mathbf{U}}(\omega) \triangleq \mathbb{E}\{\exp(j\omega \mathbf{U})\} = \prod_{n=0}^1 \frac{1}{\sqrt{1 - 2j\chi_{f,f}[n]\omega}}, \quad (92)$$

where $\chi_{f,f}[n]$ is given (49). Taking the inverse Fourier transform of $\phi_{\mathbf{U}}(\omega)$, we have

$$p_{\mathbf{U}}(u | H_{0k}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \prod_{n=0}^1 \frac{\exp(-j\omega u)}{\sqrt{1 - 2j\chi_{f,f}[n]\omega}} d\omega. \quad (93)$$

Using (93), the false alarm rate for the k th IoT device is obtained as in (51). Following the same procedure, the correct identification rate for the k th IoT device in (52) is derived.

REFERENCES

[1] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sept. 2016.

[2] M. Zhong, Y. Yang, H. Yao, X. Fu, O. A. Dobre, and O. Postolache, "5G and IoT: Towards a new era of communications and measurements," *IEEE Instrum. Meas. Mag.*, vol. 22, no. 6, pp. 18–26, Dec. 2019.

[3] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[4] A. Zanella, M. Zorzi, A. F. dos Santos, P. Popovski, N. Pratas, C. Stefanovic, A. Dekorsy, C. Bockelmann, B. Busropan, and T. A. Norp, "M2M massive wireless access: challenges, research issues, and ways forward," in *Proc. IEEE Globecom*, 2013, pp. 151–156.

[5] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Veh. Technol. Mag.*, vol. 13, no. 4, pp. 40–50, Oct. 2018.

[6] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, Feb. 2011.

[7] H. F. Schepker and A. Dekorsy, "Compressive sensing multi-user detection with block-wise orthogonal least squares," in *Proc. IEEE VTC*, Yokohama, Japan, May 2012, pp. 1–5.

[8] X. Zhang, F. Labeau, Y.-C. Liang, and J. Fang, "Compressive sensing based multiuser detection via iterative reweighted approach in M2M communications," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 764–767, Oct. 2018.

[9] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *Proc. ISWCS*, 2013, pp. 1–5.

[10] B. Knoop, F. Monsees, C. Bockelmann, D. Wübben, S. Paul, and A. Dekorsy, "Sparsity-aware successive interference cancellation with practical constraints," in *Proc. VDE WSA*, 2013, pp. 1–8.

[11] J. Ahn, B. Shim, and K. B. Lee, "Sparsity-aware ordered successive interference cancellation for massive machine-type communications," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 134–137, Feb. 2018.

[12] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, Č. Stefanović *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE access*, vol. 6, pp. 28 969–28 992, June 2018.

[13] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[15] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 66, no. 11, pp. 2933–2946, June 2018.

[16] D. S. Pham, A. M. Zoubir, R. F. Brcic, and Y. H. Leung, "A nonlinear m -estimation approach to robust asynchronous multiuser detection in Non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1624–1633, May 2007.

[17] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[18] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.

[19] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, Dec. 2009.

[20] P. S. Boonstra, B. Mukherjee, and J. M. Taylor, "A small-sample choice of the tuning parameter in ridge regression," *Statistica Sinica*, vol. 25, no. 3, p. 1185, July 2015.

[21] A. Bhattacharya, *Digital Communication*. Tata McGraw-Hill, 2005.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory*, 1998.

[23] M. A. Al-Jarrah, M. A. Yaseen, A. Al-Dweik, O. A. Dobre, and E. Alsusa, "Decision fusion for IoT-based wireless sensor networks," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1313–1326, Feb. 2019.

[24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society*, vol. 68, no. 1, pp. 49–67, Aug. 2006.

[25] A. Antoniou and W.-S. Lu, *Practical optimization: algorithms and engineering applications*. Springer Science & Business Media, 2007.

[26] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[27] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[28] K. Teknomo, "K-means clustering tutorial," *Medicine*, vol. 100, no. 4, p. 3, 2006.

- [29] C. Tang and C. Monteleoni, “On Lloyd’s algorithm: New theoretical insights for clustering in practice,” in *Proc. Artificial Intelligence and Statistics*, 2016, pp. 1280–1289.
- [30] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.