This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Asynchronous Massive Access and Neighbor Discovery Using OFDMA

Xu Chen, Lina Liu, Dongning Guo, *Fellow, IEEE,* Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—The fundamental communication problem in the wireless Internet-of-Things (IoT) is to discover a massive number of devices and to provide them with reliable access to shared channels. Oftentimes these devices transmit short messages randomly and sporadically. This paper proposes a novel signaling scheme for grant-free massive access, where each device encodes its identity and/or information in a sparse set of tones. Such transmissions are implemented in the form of orthogonal frequency-division multiple access (OFDMA). Under some mild conditions and assuming device delays to be bounded unknown multiples of sampling intervals, sparse OFDMA is proved to enable arbitrarily reliable asynchronous device identification and message decoding with a codelength that is $O(K(\log K+\log S+\log N))$, where N denotes the device population, K denotes the actual number of active devices, and $\log S$ is essentially equal to the number of information bits each device can send. The computational complexity for discovery and decoding can be made to be $O(K(\log K)(\log K+\log S+\log N)+K^2\log K)$. As a proof of concept, a specific design is proposed to identify up to 200 active devices out of $N=2^{96}$ possible devices with up to 20 samples of delay, moderate signal-to-noise ratios, and fading. If the device population is $N=2^{48}$ instead, each active device can also transmit 48 bits to the access point at the same time. The codelength compares much more favorably with those of standard slotted ALOHA and carrier-sensing multiple access (CSMA) schemes.**

***Index Terms*—Asynchrony, Internet-of-Things, multiaccess, mutiple-access, OFDM, sparse bipartite graph.**

I. Introduction

By some estimate [1], there will be well over one hundred billion connected devices world-wide in the Internet of Things (IoT) by year 2030. There can be over a million low-cost, battery-powered IoT devices within 500 meter range in a densely populated area. The general term of *massive access* describes the setting where a large number of devices need to access a shared medium in the uplink to send messages to some access points. Wireless IoT devices typically transmit short messages randomly and sporadically. If an access point only needs to decode the messages but not the corresponding devices' identities, the setting is referred to as *unsourced random access* [2], [3]. At the other extreme where the sole purpose is to detect and identify active devices within range, the setting is often referred to as *neighbor discovery* or *device identification*. Whereas neighbor discovery is a special case of massive access, the latter can also be regarded as the former if the transmitted data are regarded as all or a segment of the device identities.

Most IoT access solutions in practice either orthogonalize transmissions or suffer from collisions over the air. In particular, using a naive time division multiple access (TDMA) scheme to schedule densely deployed devices would incur large latency. So would a random access mechanism based on classical ALOHA. Since the number of devices far exceeds the frame length, it is also impossible to assign nearly orthogonal sequences to all the devices to support code-division multiple access (CDMA), especially due to device asynchrony. Moreover, massive access using conventional multiuser detection approaches generally involves a polynomial complexity in the number of devices [4], which is also impractical.

Lower frequencies (under 2 GHz) are often used to provide wide coverage and combat blockage for IoT applications. At lower frequencies, the number of antennas that can be deployed is limited by the wavelength. In this paper, we assume all devices and access points are equipped with a single antenna for simplicity.

If a common timing reference such as a beacon signal is available, the relative delays between devices can be made to be quite small. It is, however, hard to eliminate the relative delays, in part because of their different distances to the same access point. For example, a difference of 300 meters implies a free space propagation delay of one microsecond, which spans 20 samples at 20 million samples per second.

A successful massive access solution for the IoT should be *ultra-scalable* to support a massive number of devices, incur low latencies, and allow for some asynchrony between devices.

A. Related Work

To detect and/or decode messages from many transmitters sharing a common medium is a problem in the well established area of multiuser detection. A small-scale neighbor discovery problem is studied as a multiuser detection problem in [4]. In a more challenging case where the device population is several orders of magnitude larger than the number of active devices, schemes inspired by the compressed sensing literature have been proposed in [5]–[13]. In fact, decoding the uncoordinated massive access is closely related to the

X. Chen is with Waymo, Mountain View, CA, 94043 (Email: xuchen2011@u.northwestern.edu).

L. Liu and D. Guo are with Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, 60208 (Email: linaliu2020@u.northwestern.edu; dguo@northwestern.edu).

G. W. Wornell is with Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 (Email: gww@mit.edu).

This work was supported, in part, by NSF under Grant Nos. 1910168 and 2132700, by ARFL through the DAF-MIT AI Accelerator under Cooperative Agreement No. FA8750-19-2-1000, and by ONR under Grant No. N00014-19-1-2665.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

problem of support recovery, which is a special case of compressed sensing. In concept, identifying K active devices out of a total of N devices can be regarded as finding a K-sparse support of an unknown vector of length N. Results in compressed sensing indicate that the codelength L needs to be at least $O(K\log(N/K))$ to allow reliable identification of an arbitrary subset of K active devices among N devices [14]. However, a standard compressed sensing solution is computationally intractable for a very large N. Moreover, while such schemes can reduce the codelength substantially when compared with 802.11 type protocols, they mostly require synchronous transmissions.

Using many antennas at the access point may overcome the preceding fundamental limits on synchronous massive access. In particular, [9] shows that reliable activity detection can be achieved in the asymptotic regime where $K, N, L \to \infty$, as long as L and K are linear in N, and the number of antennas is sufficiently large. It is thus possible to have perfect activity detection with $K \geq L$. Furthermore, [15] proves that the accurate activity detection for $K = O(L^2/\log^2(N/K))$ can be obtained when the number of antennas scales faster than K. This is also the best possible identifiability upper bound. However, the complexities of the proposed algorithms in [9] and [15] are at least linear in device population. A convex optimization based algorithm is proposed to detect active devices using asynchronous CDMA random access [16], but the polynomial complexity scales poorly for massive access.

Compressed sensing algorithms with sublinear complexity for finite discrete alphabets are proposed in [17]–[22], and the algorithms are further extended to handle continuous alphabets in [23]. In [23], it is shown that a codelength of $L = O(K\log(N/K)\log\log(N/K))$ and a computational complexity of $O\left(K\log^{1+r}(N/K)\right)$, where r is an arbitrarily small positive constant, is needed for synchronous active device identification. The proposed scheme and analysis in [23] work for identifying an arbitrarily large fraction of active devices, while this paper aims to correctly identify the *whole* set of devices. Moreover, this paper considers the *asynchronous* multiaccess setting and develops a specific code to accommodate a massive number of devices.

Non-orthogonal multiple access (NOMA) allows multiple devices to share the time and frequency resources via power domain or code domain multiplexing [24], [25], where successive interference cancellation is used to cancel multiuser interference at the receiver. NOMA's decoding algorithms (e.g., message passing [26]) have in general polynomial complexities in the device population. However, whether NOMA is resilient to imperfect channel state information in massive access is not well understood.

The information-theoretic limits of uplink and downlink massive access are studied in [27] and [28], respectively, where the number of devices scales with the codelength in general. It is shown therein that separate device identification and message decoding achieves the capacity. The degree of freedom of massive access fading channels is analyzed in [29]. The capacity of a particular asynchronous set-up is studied in [30]. Successive interference cancellation does not achieve the boundary of the capacity region in the case of asynchronous NOMA [31]. Low-complexity schemes have been proposed for unsourced multiple access [32]–[35]. Specifically, [32] proposes T-fold ALOHA, [33] uses compressed sensing and a tree-code to recover and then stitch sub-blocks, a sparse regression code is used in [34] as an inner code with approximate message passing decoding, and a binary chirp coding scheme is studied in [35]. Unsourced random access has also been extended to the massive MIMO setting [36]. These schemes assume fully synchronous transmissions and the effectiveness of the schemes on asynchronous transmissions is yet to be investigated. There have been various studies on asynchronous transmissions [37]–[41]. Nonetheless, the algorithm proposed in [37] has a linear computational complexity in the device population. Pilot sequence designs were discussed in [38]–[40] with no theoretical guarantees on detection and data transmission performance. The scheme proposed in [41] fits in the context of unsourced random access without activity detection.

The idea of codes on graph has also been applied in random access [42]–[44]. One notable scheme is coded slotted ALOHA, in which transmitters repeat their packets several times in random slots and the access point decodes them using successive cancellation. The scheme is studied in the asynchronous setting in [45]–[47] under the assumption of perfect cancellation. Rateless codes have been proposed for multiple access in machine-to-machine communications [48], where the channel gains are assumed to be known. The preceding schemes suffer from the accumulation of channel estimation errors during successive cancellation, so they do not directly scale to a massive number of users.

B. Contributions

The goal of this work is to develop a suitable signaling scheme for grant-free transmissions by a large number of devices with delay uncertainties. The main contributions of this paper are summarized as follows:

1) We propose a novel signaling scheme, referred to as sparse OFDMA, where a device encodes its information and/or identity onto a sparse set of orthogonal tones. We exploit the fact that the tones' frequencies are invariant to delays. Sparse OFDMA is highly effective in an asynchronous setting, which is particularly appealing in the IoT.
2) With a codelength that is essentially sublinear in the device population, all active devices are correctly detected with high probability in the asymptotic regime where the number of active devices and the maximum delay are sublinear relative to the device population. With a bounded maximum delay, under a mild condition on the number of active devices, the computational complexity is also sublinear in the device population (due to a sparse Fourier transform).
3) The proposed signaling and decoding method are demonstrated to be effective under some practical scenarios using simulations. In one case with moderate signal-to-noise ratios and fading, where the device delays are within 20 samples, we can identify up to 200

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

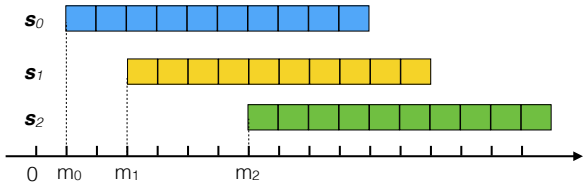

Fig. 1. Illustration of signals from three sample-synchronous, frame-asynchronous users.

active devices out of $N = 2^{96}$ possible devices with 99% accuracy. In this case, if the device population is $N = 2^{48}$ instead, each active device can also transmit 48 bits to the access point at the same time.

C. Paper Organization and Notations

The rest of the paper is organized as follows. Section II presents the system model and main results. Section III describes the signalling scheme of sparse OFDMA. Section IV presents the asynchronous massive access algorithm. Sections V and VI prove theoretical performance guarantees for synchronous and asynchronous transmissions, respectively. Some technical details are relegated to appendices. Section VII presents some numerical results on a practical design. Section VIII concludes the paper.

For ease of notation, the index of a vector or each dimension of a matrix starts from 0 throughout the paper. The elements of a $B \times C$ matrix are denoted as y_b^c, where $c = 0, \cdots, C-1$ and $b = 0, \cdots, B-1$. We write the b-th row vector as $\boldsymbol{y}_b = \left(y_b^0, \cdots, y_b^{C-1}\right)$ and the c-th column vector as $\boldsymbol{y}^c = \left(y_0^c, \cdots, y_{B-1}^c\right)^T$. We denote the real and imaginary parts and the amplitude of a complex-valued variable X as X_R, X_I, and $|X|$, respectively. For a set $\mathcal{K}$, we use $|\mathcal{K}|$ to denote its cardinality. By default, all logarithms are base 2.

II. System Model and Main Scaling Results

We focus on the uplink access problem in a system consisting of one access point and N potential devices in total. Let $\{0, \cdots, N-1\}$ denote the set of device indices. An arbitrary subset of devices are within the range of the access point and active, whose indices form the set $\mathcal{K} \subseteq \{0, \cdots, N-1\}$. Let $K = |\mathcal{K}|$ denote the number of active devices. Each device has a message set with no more than S messages. One of the messages is transmitted when the device is active. For tractability, we assume sample synchrony without frame synchrony, i.e., the delay of each device's signal at the access point is an integer multiple of sampling intervals. We further assume the delay of any device relative to a reference at the access point be no more than M sampling intervals. For a small M, this can be accomplished by using a common beacon to trigger transmissions. Fig. 1 illustrates a small example with three devices with different delays.

As a convention, we refer to the time-domain symbols at a transmitter as chips and refer to the time-domain symbols at an access point as samples. We introduce frequency-domain OFDM symbols in Section III. Let $\boldsymbol{s}_k = (s_{k,0}, \ldots, s_{k,L-1})$ denote the L-chip codeword transmitted by device k. In the absence of frequency selectivity, the received signal at time i is given by

$$x_i = \sum_{k \in \mathcal{K}} a_k s_{k,i-m_k} + w_i \tag{1}$$

for every integer i, where $a_k \in \mathbb{C}$ is the channel coefficient, m_k is the transmission delay of device k, and $w_i \sim \mathcal{CN}(0, 2\sigma^2)$ are independently and identically distributed (i.i.d.) circularly-symmetric complex Gaussian random variables. The discovery scheme is based on a single frame of received signal, so we assume $s_{k,i} = 0$ if $i < 0$ or $i \geq L$.

Theorem 1 (Synchronous massive access): Suppose each device's message set contains no more than S messages. Suppose, out of N devices, an unknown subset of devices transmit. Suppose the maximum delay $M = 0$, so the signals from all transmitting devices are perfectly aligned at the access point. Suppose also the noise variance is fixed and the received signal amplitude of every active device is at least a. Then for every $a, \epsilon > 0$, there exist $\alpha_0, \alpha_1, K_0 > 0$ such that for every N and K satisfying $N \geq K \geq K_0$, there exists a code of length

$$L \leq \alpha_0 K(\log N + \log S + \log K) \tag{2}$$

such that as long as no more than K devices transmit, all their identities and messages will be decoded correctly with probability no less than $1 - \epsilon$. Moreover, this can be accomplished using fewer than $\alpha_1 K(\log K)(\log N + \log S + \log K)$ arithmetic operations.

Theorem 2 (Asynchronous massive access): Suppose each device's message set contains no more than S messages. Suppose, out of N devices, an unknown subset of devices transmit. Suppose the delay of every active device is an integer number of sampling intervals no greater than $M \geq 0$. Suppose also that the noise variance is fixed and the received signal amplitude of every active device is bounded between a and $\bar{a}$. Then for every $a, \bar{a}, \epsilon > 0$ with $a \leq \bar{a}$, there exist $\alpha_0, \alpha_1, K_0 > 0$ such that for every N and K satisfying $N \geq K \geq K_0$, there exists a code of length

$$L \leq \alpha_0 \left((K+M)(\log N + \log S + \log K) + K \log(M+1)\right) \tag{3}$$

such that as long as no more than K devices transmit, all their identities and messages will be decoded correctly with probability no less than $1 - \epsilon$. Moreover, this can be accomplished using fewer than $\alpha_1(K(\log K)(\log N + \log S + \log K) + KM^2 + K^2 M \log(KM+1))$ arithmetic operations.

It is practical to let the decoder assume the amplitudes of the fading coefficients to be within the interval $(a, \bar{a})$. Violations of the bound can be treated as outage. The outage probability can be made arbitrarily small by setting the interval accordingly.

Theorems 1 and 2 provide a scaling law for the complexity and codelength in terms of the device population, the number of active devices, and the delay bound. Synchronous massive access, i.e., $M = 0$, requires a smaller codelength and fewer arithmetic operations than the asynchronous case. Theorem 2 reduces to Theorem 1 in the special case of $M = 0$. It requires a codelength of $K(\log N + \log S)$ for the K devices

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

to transmit their identities and messages. Thus, Theorem 1 has an information-theoretically optimal scaling.

With fixed maximum delay M, when the codebook size $S = O(N)$ and the number of active users K satisfies $K \log K = o(N)$, the codelengths for both the synchronous and asynchronous schemes are sublinear in the device population N according to Theorems 1 and 2. Furthermore, with fixed M, if $K^2 \log K = o(N)$, the number of arithmetic operations involved in the asynchronous scheme is also sublinear in N.

III. Sparse OFDMA Signaling

In this section, we construct a concrete signaling scheme for asynchronous massive access in several incremental steps. Let the spectrum be divided into B orthogonal subcarriers, where B is much smaller than the device population N. It is not possible to design mutually orthogonal OFDM symbols for all N devices. We let each active device transmit several OFDM symbols on a sparse subset of the subcarriers. The scheme is thus referred to as sparse OFDMA. We shall show that sparse OFDMA of moderate symbol duration can accommodate a large number of devices with bursty transmissions.

In Section III-A, we design signals for noiseless device identification with fewer devices than the number of subcarriers, i.e., $N \leq B$. In Section III-B, we augment the signals to enable noiseless device identification where $N > B$ and a single device is active. In Section III-C, we extend to noisy device identification, where $N > B$ and $K \ll N$ devices are active. At last, we fully describe sparse OFDMA signaling for simultaneous device identification and message decoding in Section III-D.

A. Noiseless Device Identification with $N \leq B$

The key idea for addressing arbitrary delays is to use the fact that the frequency of a sinusoidal signal is invariant to the delay, which causes merely a phase shift. Since the delay is bounded by M, we include M chips in the OFDM symbol as a cyclic prefix. Hence, each OFDM symbol contains $B + M$ chips. Since $N \leq B$, device k can be assigned the unique subcarrier k. The transmitted discrete-time signal structure is given by

$$s_{k,i} = g_k \exp\left(\frac{\iota 2\pi k i}{B}\right), \quad i = 0, \cdots, B+M-1, \tag{4}$$

where $g_k \in \mathbb{R}$ is a known design parameter of unit amplitude and $\iota^2 = -1$.

At the receiver side, the signals from all the neighbors arrive after a reference frame start point. The receiver discards the first M samples of each sparse OFDMA symbol and collect the next B samples as $\boldsymbol{y} = (y_0, \cdots, y_{B-1})$, where $y_i = x_{i+M}$, $i = 0, \cdots, B-1$. If each device is assigned a unique subcarrier, performing B-point discrete Fourier transform (DFT) on $\boldsymbol{y}$ yields nonzero at the k-th subcarrier if and only if device k is active. The delay m_k only affects the phase of the DFT value. Therefore, the signaling scheme (4) is sufficient to detect the active devices in a noiseless case with computational complexity of $O(B \log B)$ needed by the Fast Fourier Transform (FFT) algorithm.

B. Noiseless Single Device Identification with $N > B$

Let each device use a single subcarrier. Let $b_k \in \{0, \cdots, B-1\}$ denote the index of the subcarrier used by device k. With $N > B$, it is impossible to assign a distinct subcarrier to each device. So $b_k = b_{k'}$ for some $k \neq k'$. If the signaling scheme given by (4) is applied, devices k and k' are not distinguishable based on the active subcarrier. We resolve this ambiguity by including several OFDM symbols in a frame and embedding the device's identity through coefficient g_k in (4) across the OFDM symbols for transmission.

Let $J = \lceil \log N \rceil$ and let $(k)_2 = (k_1, \cdots, k_J)$ denote the binary representation of device index k. We simply adopt the design of $\left(g_k^0, g_k^1, \cdots, g_k^J\right) = \left(1, (-1)^{k_1}, \cdots, (-1)^{k_J}\right)$. We let device k transmit $\left(\boldsymbol{s}_k^0, \cdots, \boldsymbol{s}_k^J\right)$, where $\boldsymbol{s}_k^j = (s_{k,0}^j, \cdots, s_{k,B+M-1}^j)$ and

$$s_{k,i}^j = g_k^j \exp\left(\frac{\iota 2\pi b_k i}{B}\right), \tag{5}$$

for $i = 0, \cdots, B+M-1$ and $j = 0, \cdots, J$. The code length is thus $(J+1)(B+M)$ chips. For the j-th OFDM symbol, we discard the first M samples and use the next B samples to form a vector $\boldsymbol{y}^j = (y_0^j, \cdots, y_{B-1}^j)$, where

$$y_i^j = x_{i+j(B+M)+M} \tag{6}$$

$$= a_k s_{k,i+M-m_k}^j, \tag{7}$$

$i = 0, \cdots, B-1$. Performing B-point DFT on $\boldsymbol{y}^j$ yields

$$Y_b^j = \frac{1}{B} \sum_{i=0}^{B-1} \exp\left(-\frac{\iota 2\pi b i}{B}\right) a_k s_{k,i+M-m_k}^j \tag{8}$$

$$= \begin{cases} 0, & \text{if } b \neq b_k \\ A_{k,b} g_k^j, & \text{if } b = b_k, \end{cases} \tag{9}$$

where

$$A_{k,b} = a_k \exp\left(\frac{\iota 2\pi b(M - m_k)}{B}\right). \tag{10}$$

As in the $B \geq N$ case, the delay m_k only affects the phase of the received signal from device k.

Subcarrier b is associated with a length-$(J+1)$ vector $\boldsymbol{Y}_b = (Y_b^0, \cdots, Y_b^J)$. It can be seen that g_k^0 serves as a reference symbol capturing the channel coefficients. In our setting, $Y_b^0 = A_{k,b}$. Therefore, the j-th bit of the binary representation of k can be estimated as $k_j = 0$ if $Y_b^j / Y_b^0 = 1$ and $k_j = 1$ if $Y_b^j / Y_b^0 = -1$.

Together, b_k (frequency) and $g_k^0, \ldots, g_k^J$ (gains) carry the device index.

When there is a single active device in the noiseless setting, the signaling of sparse OFDMA consists of $\lceil \log N \rceil$ OFDM symbols. Therefore, the active device can be identified with code length of $O((B+M) \log N)$ chips and computational complexity of $O(B(\log B)(\log N))$.

C. Identification of Multiple Active Devices With and Without Noise

When multiple devices are active, the active devices may use colliding subcarriers, so that the device information cannot always be directly recovered from $Y_{b_k}^j / Y_{b_k}^0$. We propose to let

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

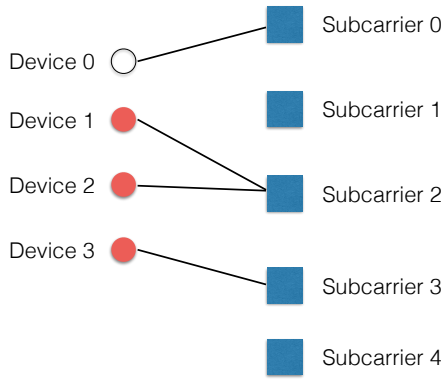

Fig. 2. Bipartite graph representation of sparse OFDMA. Left nodes represent devices and right nodes represent subcarriers. The active devices are marked in red. Subcarriers 0, 1 and 4 are zerotons, subcarrier 3 is a singleton, and subcarrier 2 is a multiton.

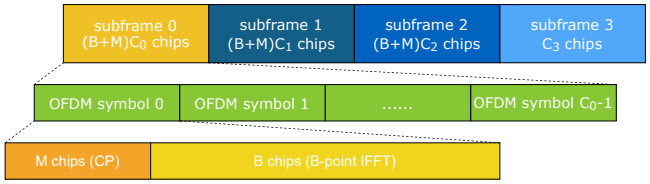

Fig. 3. Frame structure of sparse OFDMA. A frame consists of four subframes. Subframes 0, 1 and 2 have C_0, C_1 and C_2 concatenated OFDM symbols, respectively. After taking B-point IFFT of each OFDM symbol and adding M chips as cyclic prefix, each OFDM symbol will correspond to $B+M$ chips in time domain. In the synchronous case, $M=0$.

each device transmit on *multiple* subcarriers. As in the case of a single active device, we first identify active devices from the singletons. The identified device information is then used to bootstrap the detection of other devices.[1]

The correspondence between the device indices and the subcarriers can be represented by a bipartite graph with N left nodes and B right nodes. The n-th left node is connected with the b-th right node if device n transmits on the b-th subcarrier. We now introduce the following notion:

Definition 1: [Bipartite graph induced by active devices in the sparse OFDMA] The bipartite graph induced by the active devices in sparse OFDMA consists of K left nodes, corresponding to K active devices, and B right nodes, corresponding to B subcarriers, where the k-th left node is connected to the b-th right node if device k transmits on subcarrier b.

We call a subcarrier a *zeroton*, *singleton*, or *multiton*, if no device, a single device, or multiple devices transmit on the subcarrier, respectively. Fig. 2 illustrates an example of bipartite graph with $B=5$ subcarriers and $N=4$ devices, $K=3$ of which are active. In the example, subcarriers 0, 1, and 4 are zerotons, subcarrier 3 is a singleton, and subcarrier 2 is a multiton.

The presence of noise raises additional questions: 1) How can we reliably estimate the channel coefficients? 2) How can we robustly estimate the device information in the noisy setting? 3) How can we determine whether a subcarrier is a zeroton, singleton, or multiton? In the following, we further enhance the signaling scheme to address those three challenges. Specifically, a frame consisting of four subframes is described in Fig. 3, where subframes 0, 1, and 2 are used for device identification and message decoding, and subframe 3 is used for delay estimation.

We first introduce the signaling of the first three subframes, which consist of C_0, C_1, and C_2 OFDM symbols, respectively. Let $C=C_0+C_1+C_2$. Device k is assigned a fixed set of subcarriers, denoted as $\mathcal{B}_k \subseteq \{0,\cdots,B-1\}$. Specifically, we let device k transmit $\left(\boldsymbol{s}_k^0,\cdots,\boldsymbol{s}_k^{C-1}\right)$, where $\boldsymbol{s}_k^c = (s_{k,0}^c,\cdots,s_{k,B+M-1}^c)$ and

$$s_{k,i}^c = g_k^c \sum_{b\in\mathcal{B}_k} \exp\left(\frac{\iota 2\pi b i}{B}\right), \tag{11}$$

for $i=0,\cdots,B+M-1$ and $c=0,\cdots,C-1$.

For the c-th OFDM symbol, as in the previous cases, we discard the first M samples and obtain the remaining B samples as $\boldsymbol{y}^c$. Under the noisy setting, performing B-point DFT on $\boldsymbol{y}^c$ yields

$$Y_b^c = \sum_{k\in\mathcal{K}:b\in\mathcal{B}_k} A_{k,b} g_k^c + W_b^c, \quad b=0,\cdots,B-1, \tag{12}$$

where $A_{k,b}$ is given by (10), and W_b^c are i.i.d. complex Gaussian variables with distribution $\mathcal{CN}(0,2\sigma^2/B)$. The factor B in the noise variance is due to the integration of B samples in the DFT operation.

Let the design vector for device k be

$$\boldsymbol{g}_k = \begin{pmatrix} \mathbf{1} \\ \tilde{\boldsymbol{g}}_k \\ \dot{\boldsymbol{g}}_k \end{pmatrix} \tag{13}$$

where the all-one vector $\mathbf{1}$ of length C_0, $\tilde{\boldsymbol{g}}_k \in \mathbb{R}^{C_1}$, and $\dot{\boldsymbol{g}}_k \in \mathbb{R}^{C_2}$ are the design vectors for the first C_0, next C_1, and the remaining C_2 OFDM symbols, respectively. The values of C_0, C_1, and C_2 will be specified in Section V-A. The all-one segment is used for robust estimation of the channel coefficients. The $\tilde{\boldsymbol{g}}_k$ segment is used to encode the device index information. We let the entries of $\dot{\boldsymbol{g}}_k$ be generated according to i.i.d. ±1 BPSK symbols, with $\mathsf{P}\{\dot{g}_k^c=\pm1\} = 1/2$, $c=0,\cdots,C_2-1$. The $\dot{\boldsymbol{g}}_k$ segment is used to mitigate possible false alarms.

In the absence of noise, we let $\tilde{\boldsymbol{g}}_k = \left(1,(-1)^{k_1},\cdots,(-1)^{k_{\lceil\log N\rceil}}\right)$, which carries the device information. In the noisy setting, the received symbols corresponding to $\tilde{\boldsymbol{g}}_k$ are corrupted in general. In order to robustly estimate the information bits, which is the binary representation of the device's index k, we apply a linear error-control code of length $C_1 = \lceil\lceil\log N\rceil/R\rceil$ symbols, where $R<1$ is the code rate. Let $\boldsymbol{G} \in \mathbb{F}_2^{\lceil\log N\rceil\times C_1}$ be the generator matrix of the error-control code. Let $(r_{k,0},\cdots,r_{k,C_1-1}) = \left(k_1,\cdots,k_{\lceil\log N\rceil}\right)\boldsymbol{G}$, where the operation is over the binary field. We construct C_1 OFDM symbols with $\tilde{g}_k^c = (-1)^{r_{k,c}}$, $c=0,\cdots,C_1-1$. A good set of codes are the low-complexity capacity approaching codes introduced in [50] (more on this in Section V-E). Other

[1]A related, simpler setting for device activity detection is group testing, e.g., in [49], devices who do not violate the energy levels of all subcarriers are declared to be active.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

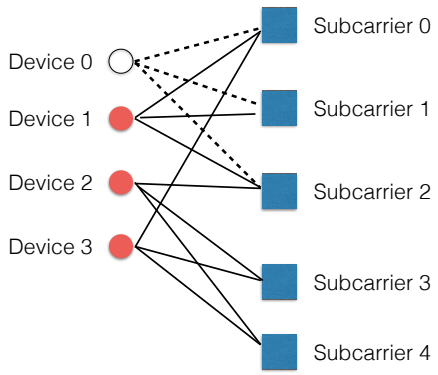

Fig. 4. Bipartite graph representation of sparse OFDMA with $T = 3$.

capacity-achieving codes can be considered, like polar codes [51], but they will yield the same scaling result as described in Theorems 1 and 2.

For device k, we let $|\mathcal{B}_k| = T$ and the set of subcarriers is equally likely to be any T-element subset of $\{0, \cdots, B-1\}$. In other words, every device transmits on exactly T out of B subcarriers. Fig. 4 illustrates an example for the case of $T = 3$. Because an individual device employs a small constant number of subcarriers, the relatively low peak-to-average power ratio of the OFDMA scheme presents much less challenge than standard OFDM to the power amplifier.

The last subframe consists of C_3 chips with low auto-correlation, which are used to estimate device delays. Delays may be estimated by performing the correlation between the received signal and the pilot samples. In particular, the pilot samples are i.i.d. BPSK symbols with length greater than M. Each device employs a random pilot sequence known to the receiver.

Note that a constant number (T) of subcarriers are employed by an individual device, i.e., each device transmits a superposition of several sinusoids. Consequently, the average energy per chip in (11) is equal to T, so the received SNR of each device at the chip level is $O(1)$ and not dependent on N, K, and B. This ensures that every device's received SNR remains practical even as we send N, K, and B to infinity.

D. Massive Access with Device Identification and Message Decoding

Similar to embedding the device index information through $\tilde{\boldsymbol{g}}_k$, each device can encode both a $\lceil \log N \rceil$-bit device index information and a $\lceil \log S \rceil$-bit message information through $\tilde{\boldsymbol{g}}_k$. With a code rate of R, we have $C_1 = \lceil (\lceil \log N \rceil + \lceil \log S \rceil)/R \rceil$ OFDM symbols to carry both the device index and the message. The overall design vector of device k is given by (13). The DFT values at the b-th subcarrier $\boldsymbol{Y}_b$ is a vector of length C and can be written as

$$\boldsymbol{Y}_b = \begin{pmatrix} \bar{\boldsymbol{Y}}_b \\ \tilde{\boldsymbol{Y}}_b \\ \dot{\boldsymbol{Y}}_b \end{pmatrix} \tag{14}$$

$$= \sum_{k \in \mathcal{K}: b \in \mathcal{B}_k} A_{k,b} \begin{pmatrix} \mathbf{1} \\ \tilde{\boldsymbol{g}}_k \\ \dot{\boldsymbol{g}}_k \end{pmatrix} + \begin{pmatrix} \bar{\boldsymbol{W}}_b \\ \tilde{\boldsymbol{W}}_b \\ \dot{\boldsymbol{W}}_b \end{pmatrix} \tag{15}$$

where the dimensions of signals $\bar{\boldsymbol{Y}}_b$, $\tilde{\boldsymbol{Y}}_b$, and $\dot{\boldsymbol{Y}}_b$ are C_0, C_1, and C_2, respectively, so are the dimensions of the noise vectors $\bar{\boldsymbol{W}}_b$, $\tilde{\boldsymbol{W}}_b$, and $\dot{\boldsymbol{W}}_b$.

This work differs from the unsourced random access paradigm of [2] in that each device here employs a unique pilot sequence and the code design depends explicitly on the device population N as well as the maximum number of active devices K. In contrast, in unsourced random access all devices use the same codebook, where what matters is the number of active devices in lieu of the device population. Sparse OFDMA has similarities to coded slotted ALOHA, in which a slot may contain zero, one, or multiple transmissions. While coded slotted ALOHA typically supports only a few users, sparse OFDMA is proposed here to support a massive number of devices without full synchronization.

Some frequently used notation is listed in Table I.

TABLE I
NOTATION

Symbol	Description
K	Number of active devices
N	Total number of devices
M	Maximum delay in samples
S	Number of messages in a device's message set
B	Total number of subcarriers
T	Number of subcarriers used by each device
C_0	Number of OFDM symbols used for channel coefficient estimations
C_1	Number of OFDM symbols used for encoding a message
C_2	Number of OFDM symbols used for singleton verification
C_3	Number of time-domain chips used for delay estimation
C	Sum of C_0, C_1, C_2, i.e., $C = C_0 + C_1 + C_2$
$\tilde{\boldsymbol{g}}_k$	Vector of C_1 bits encoding the transmitted message of device k
$\dot{\boldsymbol{g}}_k$	Vector of C_2 bits for mitigating false alarm of device k
η	Fixed energy threshold

IV. Device Identification and Message Decoding

In this section, we first describe a robust subcarrier detection scheme that can: 1) determine whether a subcarrier is a zeroton, a singleton, or a multiton (as defined in Section III-B);

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

2) detect the device index reliably for singleton subcarriers. We then describe the overall identification and decoding scheme.

A. Robust Subcarrier Detection

In this subsection, we focus on a certain device k that is hashed to subcarrier b. The corresponding frequency-domain values are decomposed into three segments as in (14). We propose the subcarrier detection scheme illustrated as Algorithm 1:

Algorithm 1 Robust-Subcarrier-Detect ($\boldsymbol{Y}$)

Input: Subcarrier values $\boldsymbol{Y} = \left(\bar{\boldsymbol{Y}}^\dagger, \tilde{\boldsymbol{Y}}^\dagger, \dot{\boldsymbol{Y}}^\dagger\right)^\dagger$, where $\bar{\boldsymbol{Y}} \in \mathbb{C}^{C_0}$, $\tilde{\boldsymbol{Y}} \in \mathbb{C}^{C_1}$ and $\dot{\boldsymbol{Y}} \in \mathbb{C}^{C_2}$.
Output: Declaration of zeroton/singleton/multiton, and estimates of index and data if applicable.
if $\|\dot{\boldsymbol{Y}}\|^2 < \eta$ **then**
  declare zeroton and return.
end if
$\hat{\theta} \leftarrow \text{phase}\left(\mathbf{1}^T \bar{\boldsymbol{Y}}/C_0\right)$.
$\tilde{\boldsymbol{Z}} \leftarrow \text{Re}\{\tilde{\boldsymbol{Y}} e^{-\iota\hat{\theta}}\}$.
$(\hat{k}, \hat{\tilde{g}}_{\hat{k}}) \leftarrow \text{Decode}(\tilde{\boldsymbol{Z}})$.
$\dot{A}_{\hat{k}} \leftarrow \dot{\boldsymbol{g}}_{\hat{k}}^\dagger \dot{\boldsymbol{Y}}/C_2$.
if $\|\dot{\boldsymbol{Y}} - \dot{A}_{\hat{k}} \dot{\boldsymbol{g}}_{\hat{k}}\|^2 \le \eta$ **then**
  declare singleton and return $\left(\hat{k}, \hat{\tilde{g}}_{\hat{k}}\right)$.
else
  declare multiton and return.
end if

Algorithm 1 includes the following steps:

1) Zeroton detection: Let $\|\cdot\|$ denote the ℓ^2-norm, which represents the energy in a signal. We declare subcarrier b to be a zeroton if $\|\dot{\boldsymbol{Y}}_b\|^2 < \eta$, where η is some constant threshold.

2) Channel phase estimation: If subcarrier b is not declared to be a zeroton, we estimate the phase of $A_{k,b}$ as

$$\hat{\theta}_b = \angle\left(\frac{1}{C_0}\sum_{c=0}^{C_0-1} \bar{Y}_b^c\right). \tag{16}$$

Suppose device k transmits on a singleton subcarrier and C_0 is large enough, we can obtain sufficiently accurate estimate of the channel phase.

3) Device identification and message decoding: With the phase estimation $\hat{\theta}_b$, we can compensate for the phase of $A_{k,b}$ and then try to decode the device index information, assuming it is a singleton (which we shall verify later in Algorithm 1). We perform (hard) binary decision on $\text{Re}\left\{\tilde{Y}_b^c e^{-\iota\hat{\theta}_b}\right\}$ for all C_1 symbols, and then decode the index and message. The singleton assumption allows us to apply the well-studied point-to-point capacity approaching codes. While more sophisticated multiuser decoding methods may apply to multitons to improve the performance, we show that single-user decoding is sufficient to establish the desired scaling laws in Section V.

4) Singleton verification: Suppose $\hat{k}$ is the decoded index. We estimate the nonzero signal as

$$\dot{A}_{\hat{k},b} = \frac{1}{C_2} \dot{\boldsymbol{g}}_{\hat{k}}^\dagger \dot{\boldsymbol{Y}}_b, \tag{17}$$

where $\dot{g}_k^c$, $c = 0, \cdots, C_2 - 1$, are as described in Section III-C. Then we declare that subcarrier b is a singleton if and only if it passes the energy threshold test, i.e.,

$$\|\dot{\boldsymbol{Y}}_b - \dot{A}_{\hat{k},b} \dot{\boldsymbol{g}}_{\hat{k}}\|^2 \le \eta. \tag{18}$$

The preceding verification scheme is similar to that used for sparse DFT and sparse Walsh-Hadamard transform (WHT), where the singleton verification approach has been proved to be correct with high probability for signal amplitudes lying in a known discrete alphabet [52], [53]. In this paper, we further show that it can effectively identify the singletons for *arbitrary* analog amplitudes that are bounded away from zero, as specified in Theorems 1 and 2.

B. Overall Identification and Decoding Scheme

Algorithm 2 Asynchronous Massive Access via Sparse OFDMA

Input: Subcarrier values $\boldsymbol{Y}_b$, $b = 0, \cdots, B-1$.
Output: Detected active device set $\hat{\mathcal{K}}$ and their messages.
Initialize: Set $\mathcal{B}$ to be the set of all subcarriers. Set $\hat{\mathcal{K}}$ and $\mathcal{L}$ to be the empty set.
for every subcarrier b **do**
  if Robust-Subcarrier-Detect $(\boldsymbol{Y}_b)$ declares a singleton $\left(\hat{k}, \hat{\tilde{\boldsymbol{g}}}_{\hat{k}}\right)$ **then**
    Add $\hat{k}$ to $\mathcal{L}$.
  end if
end for
while fewer than K iterations and $\mathcal{L}$ is not empty **do**
  Pick arbitrary $k \in \mathcal{L}$, remove k from $\mathcal{L}$, and add k to $\hat{\mathcal{K}}$.
  Estimate $\hat{m}_k$ and $\hat{a}_k$ according to (22) and (23).
  Set $\mathcal{S}$ to be the set of subcarriers in $\mathcal{B}$ that are connected with k.
  for every subcarrier $b' \in \mathcal{S}$ **do**
    Cancel the signal of device k from subcarrier b' using (24).
    if Robust-Subcarrier-Detect $(\boldsymbol{Y}_{b'})$ declares a zeroton or a singleton $\left(\hat{k}, \hat{\tilde{\boldsymbol{g}}}_{\hat{k}}\right)$ **then**
      Remove b' from $\mathcal{B}$.
      Add $\hat{k}$ to $\mathcal{L}$ (if a singleton is declared).
    end if
  end for
end while

Once a device index is estimated based on a singleton, its contributions to all its connected subcarriers are canceled out, which may result in new singleton subcarriers. For example, in Fig. 4, device 1 is first detected from the singleton subcarrier 0 and its values are subtracted from subcarrier 2. Then subcarrier 2 becomes a singleton subcarrier and device 2 can be detected from it. There is, however, one challenge. The DFT values

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

from each device at a subcarrier depends on its delay due to (10). We estimate the delay using subframe 3 of the received signal in order to perform successive cancellation.

Subframe 3 consists of C_3 chips. Let $\boldsymbol{s}'_k = (s'_{k,0}, \cdots, s'_{k,C_3-1})$ denote the transmitted chips of device k corresponding to subframe 3. The receiver discards the first M samples of subframe 3 and collects the remaining samples as $\boldsymbol{y}' = (y'_M, \cdots, y'_{C_3-1})$. Let

$$\mathcal{I} = \{M, \cdots, C_3 - 1\} \tag{19}$$

denote the time-domain indices corresponding to subframe 3 with the first M samples skipped. For $i \in \mathcal{I}$,

$$y'_i = x_{(B+M)C+i}. \tag{20}$$

Define the decision statistic based on $\mathcal{I}$ as:

$$\mathcal{T}_k(m) = \sum_{i \in \mathcal{I}} y'_{i+m} (s'_{k,i})^*. \tag{21}$$

Discarding the first M samples in the last subframe guarantees that the cross-correlation of the pilot sequences between different users is performed entirely on the BPSK symbols. We estimate the delay of device k as

$$\hat{m}_k \in \arg \max_{m=0,\cdots,M} |\mathcal{T}_k(m)|. \tag{22}$$

The channel coefficient a_k, which is a deterministic parameter, is then estimated as

$$\hat{a}_k = \frac{1}{C} \boldsymbol{g}_k^\dagger \boldsymbol{Y}_b \exp\left(-\frac{\iota 2\pi b(M - \hat{m}_k)}{B}\right). \tag{23}$$

Note that $\boldsymbol{g}_k^\dagger \boldsymbol{g}_k = C$ due to (13). The DFT values of a connected unprocessed subcarrier b' are then updated according to

$$\boldsymbol{Y}_{b'} \leftarrow \boldsymbol{Y}_{b'} - \hat{a}_k \exp\left(\frac{\iota 2\pi b'(M - \hat{m}_k)}{B}\right) \boldsymbol{g}_k. \tag{24}$$

The overall massive access scheme is described in Algorithm 2. Throughout the algorithm, we maintain three lists: $\hat{\mathcal{K}}$ is a list storing the estimated device indices, $\mathcal{L}$ is a list of detected devices for cancellation, and $\mathcal{B}$ is a list of surviving subcarriers. We first detect all singleton subcarriers and identify their corresponding devices using Algorithm 1. Then we successively cancel each identified device, potentially exposing more singletons along the way to allow more devices to be identified. This process continues until no subcarrier declared as singleton is left.

We next prove that the preceding sparse OFDMA signaling and detection scheme can efficiently identify the active devices and decode their messages. We treat the synchronous case ($M = 0$) in Section V and then the asynchronous case in Section VI.

V. Proof of Theorem 1 (the Synchronous Case)

A. Key Parameters and Propositions

For codeword construction, we choose integer constants

$$T \geq 3 \tag{25}$$

and

$$\beta_0 \geq T(T-1) + 1, \tag{26}$$

e.g., by letting $T = 3$ and $\beta_0 = 7$. We also set the following parameters as

$$B = \beta_0 K \tag{27}$$

$$C_0 = \lceil \log N \rceil \tag{28}$$

$$C_1 = \lceil (\lceil \log N \rceil + \lceil \log S \rceil)/R \rceil \tag{29}$$

$$C_2 = \lceil \beta_1 \log K \rceil \tag{30}$$

where $\beta_1 > 0$ and $R < 1$ are constants to be specified later in the development. Specifically, R is the constant rate of a low-complexity capacity-achieving code for the binary-symmetric channel (BSC) to be further explained in Appendix C. The total number of OFDM symbols is

$$\begin{aligned} C \leq\ & (1 + \lceil 1/R \rceil)\lceil \log N \rceil \\ & + \lceil 1/R \rceil \lceil \log S \rceil + \lceil \beta_1 \log K \rceil. \end{aligned} \tag{31}$$

The codelength $L = BC = \beta_0 KC$ thus satisfies (2) as long as β_0, β_1, and R are positive constants, and by letting

$$\alpha_0 = 2\beta_0(1 + \lceil 1/R \rceil + \lceil \beta_1 \rceil). \tag{32}$$

In this proof, we shall invoke several results about hypergraphs. A hypergraph is a generalization of a graph in which an edge can join any number of vertices. An edge in a hypergraph is also called a hyperedge. A hypergraph is a hypertree if the graph is a tree, i.e., no cycle exists. A hypergraph is a unicyclic component if the graph contains only one cycle. A T-uniform hypergraph is a hypergraph where all the hyperedges have degree-T. Here we let the device nodes be hyperedges and the subcarriers be hypergraph vertices to form a hypergraph. A hyperedge is incident on a vertex in the hypergraph if the corresponding device node is connected to the corresponding subcarrier.

Let G denote the bipartite graph induced by the active devices (see Definition 1). We now characterize the probability that G is such that no two active devices share the same set of subcarriers, so that it is convertible to an equivalent hypergraph (denoted as $G \in \hat{\mathcal{G}}$). Let $d = \binom{B}{T}$. Then (25) and (27) imply that $d > \frac{(\beta_0 K)^T}{2T!}$ for sufficiently large K. Since there are d distinct subsets of T subcarriers, we have

$$\mathsf{P}\{G \in \hat{\mathcal{G}}\} = d(d-1)\cdots(d-K+1)/d^K \tag{33}$$

$$\geq 1 - \zeta/K \tag{34}$$

with some constant ζ [54, The birthday problem] for sufficiently large K.

By the construction of the sparse OFDMA in Section III-C, every device is associated with exactly T subcarriers, so the hypergraph induced by the active devices is a (random) T-uniform hypergraph.

Definition 2: [An ensemble of hypergraphs $\mathcal{G}$] Let $\mathcal{G}$ denote the ensemble of T-uniform hypergraphs with K hyperedges and B vertices that consist of components each of which is either a hypertree or a unicyclic component, where no component has more than $\alpha T \log(\beta_0 K)/(T-1)$ hyperedges

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

with some constant $\alpha > 0$.

In the following, we will first show that there exists α such that G is convertible to a hypergraph in $\mathcal{G}$ (denoted as $G \in \mathcal{G}$) with high probability (Proposition 1). We then characterize the error propagation effects in the case of $G \in \mathcal{G}$ (Proposition 3) and show that Algorithm 1 makes the correct decision with high probability (Proposition 4). It follows then that synchronous massive access via Algorithm 2 succeeds with high probability (Proposition 2). The following propositions will be proved in Sections V-B to V-E.

Proposition 1: Under (25)-(30), there exist constants $K_0' > 0$ and $\nu > 1$, such that for every $K \geq K_0'$, we have

$$\mathsf{P}\left\{G \in \mathcal{G} | G \in \hat{\mathcal{G}}\right\} \geq 1 - \nu/K. \tag{35}$$

Proposition 2: If $G \in \mathcal{G}$, then Algorithm 2 will detect all active devices and decode their messages correctly as long as during its execution Algorithm 1 always makes the correct decision.

Let $\mathcal{S}(t-1)$ be the set of recovered devices in the previous $t-1$ iterations during the execution of Algorithm 2. For each device $\ell \in \mathcal{S}(t-1)$ decoded from a singleton bin b_ℓ, define

$$e_\ell = C^{-1}\boldsymbol{g}_\ell^\dagger \boldsymbol{W}_{b_\ell}. \tag{36}$$

Evidently, $e_\ell \sim \mathcal{CN}(0, 2\sigma^2/(BC))$. In iteration t, the DFT value at subcarrier b can be expressed as

$$\boldsymbol{Y}_b = \sum_{k \in \mathcal{K} \setminus \mathcal{S}(t-1): b \in \mathcal{B}_k} A_{k,b}\boldsymbol{g}_k + \boldsymbol{W}_b + \boldsymbol{V}_b, \tag{37}$$

where the sum is over the set of active devices that are hashed to subcarrier b and not yet recovered, and $\boldsymbol{V}_b$ is due to the residual channel estimation errors from the recovered devices.

Proposition 3: Suppose (25)-(30) hold and $G \in \mathcal{G}$. Suppose also Algorithm 1 always makes the correct decision and for every $\ell \in \mathcal{S}(t-1)$, the amplitude of the error is upper bounded by

$$|e_\ell| \leq \tau(\log K)^{-2}, \tag{38}$$

where

$$\tau = \frac{(T-1)\sqrt{\eta}}{8\alpha\beta_1(1+\log\beta_0)T}. \tag{39}$$

Then, every entry of $\boldsymbol{V}_b$ is bounded by

$$|V_b^c| \leq \frac{\sqrt{\eta}}{4\beta_1 \log K}. \tag{40}$$

Proposition 4: Suppose (25)-(30) hold and $G \in \mathcal{G}$. For large enough K, there exists a constant η such that for every $t = 1, 2, \cdots$, conditioned on that Algorithm 1 makes correct decisions in the first $t-1$ iterations during the execution of Algorithm 2, and e_ℓ is upper bounded by (38) for each decoded device ℓ in the first $t-1$ iterations, Algorithm 1 makes a wrong decision with probability no greater than $7K^{-2}$ in the t-th iteration. Moreover, if device k is decoded from a singleton bin b_k in the t-th iteration, then $|e_k| \geq \tau(\log K)^{-2}$ with probability smaller than K^{-2}.

Assuming Propositions 1-4 hold, we upper bound the massive access error probability P_s as follows. Let $\mathcal{E}$ denote the event that Algorithm 1 makes at least one wrong decision during the execution of Algorithm 2. By Proposition 2, massive access succeeds if $G \in \mathcal{G}$ and that $\mathcal{E}$ does not occur. Evidently,

$$\begin{aligned} P_s &\leq \mathsf{P}\{\mathcal{E} \cup (G \notin \mathcal{G})\} && (41)\\ &\leq \mathsf{P}\{\mathcal{E} \cap (G \in \mathcal{G})\} + \mathsf{P}\{G \notin \mathcal{G}\} && (42)\\ &\leq \mathsf{P}\{\mathcal{E}|G \in \mathcal{G}\} + \mathsf{P}\{G \notin \mathcal{G}|G \in \hat{\mathcal{G}}\} + \mathsf{P}\{G \notin \hat{\mathcal{G}}\}. && (43)\end{aligned}$$

Every time a device is recovered, Algorithm 1 is performed on its connected subcarriers. Since there are K active devices and each of them is connected to T subcarriers, Algorithm 1 runs for at most KT times throughout the detection process. By the union bound and the result of Proposition 4,

$$\begin{aligned} \mathsf{P}\{\mathcal{E}|G \in \mathcal{G}\} &\leq KT\left(\frac{7}{K^2} + \frac{1}{K^2}\right) && (44)\\ &= 8T/K. && (45)\end{aligned}$$

Plugging (34), (35), and (45) into (43), we have that massive access fails with probability

$$P_s \leq (8T + \nu + \zeta)/K. \tag{46}$$

Therefore, given the choice of T, B, and C, massive access fails with probability less than ϵ as long as $K \geq \max\{(8T + \nu + \zeta)/\epsilon, K_0'\}$.

B. Proof of Proposition 1

Throughout the proof, we assume G is a hypergraph, i.e., $G \in \hat{\mathcal{G}}$. For ease of notation, we omit conditioning on $G \in \hat{\mathcal{G}}$.

Let $\mathcal{G}_0$ denote the ensemble of T-uniform hypergraphs with K hyperedges and B vertices consisting of only hypertrees and unicyclic components. For a given constant α, let $\mathcal{G}_1$ denote the ensemble of hypergraphs with the largest component containing fewer than $\alpha T\log(\beta_0 K)/(T-1)$ hyperedges, and $\mathcal{G}_2$ denote the ensemble of hypergraphs with the largest component containing fewer than $\alpha T \log B$ vertices. Evidently, $\mathcal{G} = \mathcal{G}_0 \cap \mathcal{G}_1$.

Due to our choice of K and B, the T-uniform hypergraph is in the so-called subcritical phase, which guarantees some simple structural properties. In particular, [55, Theorem 4] establishes that the hypergraph is entirely composed of hypertrees and unicyclic components with high probability. To be precise, the proof in [55] implies that

$$\mathsf{P}\{G \in \mathcal{G}_0\} \geq 1 - \frac{\nu - 1}{K} \tag{47}$$

for some constant $\nu > 1$ dependent on β_0.

We next show that there exists a constant α, such that $G \in \mathcal{G}_2$ with high probability. The size of a hypergraph is defined as the number of hyperedges K in the graph. The number of vertices in the hypergraph is B. The average vertex degree of a vertex v is defined as the expected number of pairs of (v_i, e_i), where v_i and e_i are some vertex and hyperedge in the graph, such that (v, v_i) are connected via hyperedge e_i.

Lemma 1: G has an average vertex degree of $KT(T-1)/B$.

Proof: Consider a subcarrier b (a vertex). Let $X_k = 1$ if device k uses subcarrier b, and $X_k = 0$, otherwise. Device k has $\binom{B}{T}$ equally likely choices for its subcarriers, where

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

$\binom{B-1}{T-1}$ of those choices include subcarrier b. The average vertex degree is thus calculated as

$$\mathsf{E}\left\{\sum_{k\in\mathcal{K}}(T-1)X_k\right\} = K(T-1)\mathsf{E}\{X_k\} \tag{48}$$

$$= K(T-1)\frac{\binom{B-1}{T-1}}{\binom{B}{T}} \tag{49}$$

$$= \frac{KT(T-1)}{B}. \tag{50}$$

Hence the lemma is proved.

Consider a random hypergraph with B vertices and range T.[2] Since $B > KT(T-1)$ due to (26) and (27), the average *vertex degree* is less than 1 by Lemma 1. Then it has been shown in [56, Theorem 3.6] that, since G's average vertex degree is less than 1, for $K \geq K_0'$ with a large enough K_0', there exists a constant α, such that

$$\mathsf{P}\{G \in \mathcal{G}_2\} \geq 1 - \frac{1}{K}. \tag{51}$$

Let r and s be the number of vertices and hyperedges of a connected component, respectively. If the component is a hypertree, $r = (T-1)s + 1$. If the component is a unicyclic component, $r = (T-1)s$. Therefore, when $G \in \mathcal{G}_0$ and $G \in \mathcal{G}_2$, the number of hyperedges in the largest component is upper bounded by

$$s \leq r/(T-1) \tag{52}$$

$$\leq \alpha T \log B/(T-1) \tag{53}$$

$$= \alpha T \log(\beta_0 K)/(T-1). \tag{54}$$

It implies

$$(\mathcal{G}_0 \cap \mathcal{G}_2) \subset \mathcal{G}_1. \tag{55}$$

Therefore, we have

$$\mathsf{P}\{G \in \mathcal{G}\} = \mathsf{P}\{G \in (\mathcal{G}_0 \cap \mathcal{G}_1)\} \tag{56}$$

$$\geq \mathsf{P}\{G \in (\mathcal{G}_0 \cap \mathcal{G}_2)\} \tag{57}$$

$$\geq 1 - \frac{\nu}{K}, \tag{58}$$

where (56) follows from Definition 2 and definitions of $\mathcal{G}_0$ and $\mathcal{G}_1$, (57) is due to (55), and (58) is due to (47) and (51).

Hence the proof of Proposition 1.

C. Proof of Proposition 2

Every graph in $\mathcal{G}$ consists of only hypertrees and unicyclic components. By [57, Lemma 3.4], every hypertree has a hyperedge (device node) incident on at least $T-1$ singleton subcarriers. Every unicyclic component has a hyperedge that is incident on either $T-1$ or $T-2$ singleton subcarriers. When $T \geq 3$, every hypergraph in $\mathcal{G}$ always has singleton subcarriers. Assuming Algorithm 1 correctly detects the singleton subcarriers and estimates the devices' indices, the detected devices will be removed from the hypergraph. Since removing any device node or hyperedge in $G \in \mathcal{G}$ still yields a hypergraph in $\mathcal{G}$, Algorithm 2 will continue until all active devices are correctly detected.

[2]The size of the largest edge is called range of the hypergraph. A T-uniform hypergraph has a range of T.

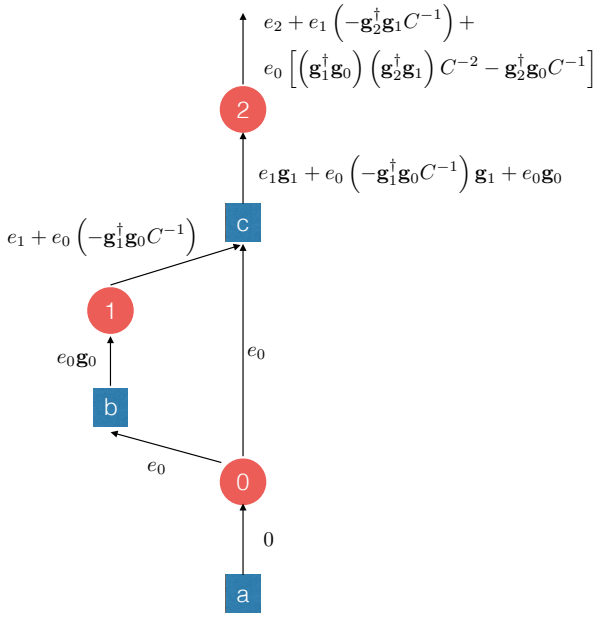

Fig. 5. Error propagation graph for device 2. Device 0 is detected from subcarrier-a in iteration $t = 1$, device 1 is detected from subcarrier-b in iteration $t = 2$, and device 2 is detected from subcarrier-c in iteration $t = 3$.

D. Proof of Proposition 3

We make use of the error propagation graph proposed in [20] to characterize the residual channel estimation errors. The error propagation graph is a directed subgraph induced by the device identification Algorithm 2. It begins from some singleton subcarriers. Every singleton subcarrier points to the device node that should be detected based on it. Every device node in the graph points to all subcarriers it is then cancelled from. Fig. 5 illustrates the error propagation subgraph concerning device 2. In the error propagation graph, device 0 is estimated from singleton subcarrier-a and its values are cancelled from the connected subcarrier-b and subcarrier-c. In a subsequent iteration, subcarrier-b becomes a singleton. Device 1 can be detected and its values are cancelled from subcarrier-c. In a third iteration, subcarrier-c becomes a singleton and device 2 is detected.

Let b_k be the singleton subcarrier used to recover the device k. Let A_k be the channel estimation error of device k defined as

$$A_k = \hat{A}_{k,b_k} - A_{k,b_k}, \tag{59}$$

where A_{k,b_k} is given by (10) and

$$\hat{A}_{k,b_k} = \boldsymbol{g}_k^\dagger \boldsymbol{Y}_{b_k}/C \tag{60}$$

is the estimate according to (23) with no delay in this case.

We will keep track of the errors using the graph in Fig. 5. The error A_k causes cancellation error $A_k\boldsymbol{g}_k$ to every downstream subcarrier node. The errors accumulated by a singleton subcarrier adds to the estimation errors of its downstream device nodes.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Consider the first iteration and a singleton subcarrier b_k due to device k. The DFT value of subcarrier-b_k is given by

$$\boldsymbol{Y}_{b_k} = A_{k,b_k}\boldsymbol{g}_k + \boldsymbol{W}_{b_k}, \tag{61}$$

so the residual estimation error is given by

$$A_k = C^{-1}\boldsymbol{g}_k^\dagger \boldsymbol{W}_{b_k}. \tag{62}$$

In iteration $t \geq 2$, with successive cancellation, the updated frequency value at subcarrier b is

$$\begin{aligned}\boldsymbol{Y}_b &= \sum_{k\in\mathcal{K}\backslash\mathcal{S}(t-1):b\in\mathcal{B}_k} A_{k,b}\boldsymbol{g}_k + \boldsymbol{W}_b \\ &\quad + \sum_{\ell\in\mathcal{S}(t-1):b\in\mathcal{B}_\ell} \left(A_{\ell,b_\ell} - \hat{A}_{\ell,b_\ell}\right)\boldsymbol{g}_\ell \qquad (63)\\ &= \sum_{k\in\mathcal{K}\backslash\mathcal{S}(t-1):b\in\mathcal{B}_k} A_{k,b}\boldsymbol{g}_k + \boldsymbol{W}_b - \sum_{\ell\in\mathcal{S}(t-1):b\in\mathcal{B}_\ell} A_\ell \boldsymbol{g}_\ell. \qquad (64)\end{aligned}$$

Suppose now some subcarrier b_k is a singleton due to device k in some iteration and is used to recover the index, then the channel estimation error is calculated as

$$\begin{aligned}A_k &= C^{-1}\boldsymbol{g}_k^\dagger \boldsymbol{Y}_{b_k} - A_{k,b_k} \qquad (65)\\ &= C^{-1}\boldsymbol{g}_k^\dagger\left(A_{k,b_k}\boldsymbol{g}_k + \boldsymbol{W}_{b_k} - \sum_{\ell\in\mathcal{S}(t-1):b_k\in\mathcal{B}_\ell} A_\ell\boldsymbol{g}_\ell\right) \\ &\quad - A_{k,b_k} \qquad (66)\\ &= C^{-1}\boldsymbol{g}_k^\dagger \boldsymbol{W}_{b_k} - \sum_{\ell\in\mathcal{S}(t-1):b_k\in\mathcal{B}_\ell} C^{-1}A_\ell \boldsymbol{g}_k^\dagger\boldsymbol{g}_\ell \qquad (67)\end{aligned}$$

where we use the fact that $\boldsymbol{g}_k^\dagger\boldsymbol{g}_k = C$ to obtain (67).

Using recursion, the estimation error of $A_{k,b}$ for $k \in \mathcal{S}(t)\backslash\mathcal{S}(t-1)$ is calculated as

$$\begin{aligned}A_k = e_k + &\sum_{(\ell_0,\cdots,\ell_i)\in\mathcal{P}(k),\ell_0\in\mathcal{S}(t-1)} e_{\ell_0}(-\boldsymbol{g}_k^\dagger\boldsymbol{g}_{\ell_i}/C) \\ &\cdot(-\boldsymbol{g}_{\ell_i}^\dagger\boldsymbol{g}_{\ell_{i-1}}/C)\cdots(-\boldsymbol{g}_{\ell_1}^\dagger\boldsymbol{g}_{\ell_0}/C),\end{aligned} \tag{68}$$

where $\mathcal{P}(k) = \{(\ell_0,\cdots,\ell_i) : \ell_0,\cdots,\ell_i, k \text{ is a path of devices in the error propagation graph}\}$, and e_{ℓ_0} is given by (36).

In Fig. 5, for instance, when calculating the estimation error A_2 for device node 2, we take into account device nodes 0 and 1 that have been recovered. In particular, $\mathcal{P}(2) = \{(0,2),(1,2),(0,1,2)\}$. For path $(0,2)$, the product term in the summation is $-e_0\boldsymbol{g}_2^\dagger\boldsymbol{g}_0/C$. For path $(1,2)$, the product term in the summation is $-e_1\boldsymbol{g}_2^\dagger\boldsymbol{g}_1/C$. For path $(0,1,2)$, the product term in the summation is $e_0\boldsymbol{g}_2^\dagger\boldsymbol{g}_1\boldsymbol{g}_1^\dagger\boldsymbol{g}_0/C^2$. Consequently, we have $A_2 = e_2 + e_0(-\boldsymbol{g}_2^\dagger\boldsymbol{g}_0/C + \boldsymbol{g}_2^\dagger\boldsymbol{g}_1\boldsymbol{g}_1^\dagger\boldsymbol{g}_0/C^2) - e_1\boldsymbol{g}_2^\dagger\boldsymbol{g}_1/C$.

Therefore, the DFT value at subcarrier b is calculated as $\boldsymbol{Y}_b$ given by (37), where

$$\begin{aligned}\mathbf{V}_b = -&\sum_{(\ell_0,\cdots,\ell_i)\in\mathcal{P}'(b),\ell_0\in\mathcal{S}(t-1)} e_{\ell_0} \\ &\cdot(-\boldsymbol{g}_{\ell_i}^\dagger\boldsymbol{g}_{\ell_{i-1}}/C)\cdots(-\boldsymbol{g}_{\ell_1}^\dagger\boldsymbol{g}_{\ell_0}/C)\boldsymbol{g}_{\ell_i}\end{aligned} \tag{69}$$

where $\mathcal{P}'(b) = \{(\ell_0,\cdots,\ell_i) : \ell_0,\cdots,\ell_i \text{ is a path of devices leading to subcarrier } b \text{ in the error propagation graph}\}$.

Suppose $G \in \mathcal{G}$, then $|\mathcal{P}'(b)| \leq 2$. Moreover, by Proposition 1, the number of left nodes in each component is less than $\alpha T\log(\beta_0 K)/(T-1)$, which indicates that $|\mathcal{S}(t-1)| \leq \alpha T\log(\beta_0 K)/(T-1) \leq \alpha T\log K/(T-1) + 2\alpha\log\beta_0$. Since the entries of the design parameter $\boldsymbol{g}_\ell$ are i.i.d. BPSK symbols and $\boldsymbol{g}_\ell \in \mathbb{R}^C$, $|-\boldsymbol{g}_{\ell_i}^\dagger\boldsymbol{g}_{\ell_{i-1}}/C| \leq 1$. Thus, for each entry of $\boldsymbol{V}_b$, e_{ℓ_0} has the coefficient satisfying $|(-\boldsymbol{g}_{\ell_i}^\dagger\boldsymbol{g}_{\ell_{i-1}}/C)\cdots(-\boldsymbol{g}_{\ell_1}^\dagger\boldsymbol{g}_{\ell_0}/C)g_{\ell_i}^c| \leq 1$. Combining with the assumption that e_ℓ is upper bounded by (38), each entry of $\boldsymbol{V}_b$ is upper bounded by

$$\begin{aligned}|V_b^c| &\leq |\mathcal{P}'(b)||\mathcal{S}(t-1)||e_\ell| \qquad (70)\\ &\leq \frac{2\alpha T\log(\beta_0 K)}{T-1}\frac{\tau}{(\log K)^2}. \qquad (71)\end{aligned}$$

Plugging (39) into (71) and using the fact $\log(\beta_0 K) \leq (1+\log\beta_0)\log K$ yield (40).

Hence the proof of Proposition 3.

E. Proof of Proposition 4

As described in the proof of Proposition 3, the frequency domain signal in subcarrier-b can be written as (37). The detection error depends on $\boldsymbol{V}_b$ and hence on the number of devices that cause interference. We denote the interference plus noise as

$$\boldsymbol{Z}_b = \boldsymbol{W}_b + \boldsymbol{V}_b. \tag{72}$$

The analysis here is similar to that in [23].

1) Zeroton error detection: Suppose subcarrier b is a zeroton. The DFT values in $\boldsymbol{Y}_b$ are composed of purely noise and interference, i.e., $\boldsymbol{Y}_b = \boldsymbol{Z}_b$. Let $\dot{\boldsymbol{Z}}_b$ and $\dot{\boldsymbol{V}}_b$ be C_2-vectors defined in the same manner as $\dot{\boldsymbol{Y}}_b$ and $\dot{\boldsymbol{W}}_b$ in (14) and (15). The zeroton error E_0 occurs only if $\|\dot{\boldsymbol{Y}}_b\|$ is greater than the threshold η. Conditioned on each element of $\boldsymbol{V}_b$ is bounded by (40), we have

$$\begin{aligned}\|\dot{\boldsymbol{V}}_b\| &\leq \sqrt{C_2}\frac{\sqrt{\eta}}{4\beta_1\log K} \qquad (73)\\ &\leq \frac{\sqrt{\eta}}{2}. \qquad (74)\end{aligned}$$

By the triangle inequality $\|\dot{\boldsymbol{Z}}_b\| \leq \|\dot{\boldsymbol{W}}_b\| + \|\dot{\boldsymbol{V}}_b\|$, the probability of error satisfies

$$\begin{aligned}\mathsf{P}\{E_0\} &= \mathsf{P}\{\|\dot{\boldsymbol{Z}}_b\|^2 \geq \eta\} \qquad (75)\\ &\leq \mathsf{P}\{\|\dot{\boldsymbol{W}}_b\|^2 \geq \eta/4\} \qquad (76)\\ &= \mathsf{P}\left\{Q \geq \frac{B\eta}{4\sigma^2}\right\} \qquad (77)\end{aligned}$$

where $Q \sim \chi^2(2C_2)$ is a standard χ^2 random variable with $2C_2$ degrees of freedom. We then utilize the following lemma proved in [58] to bound the tail probability of Q.

Lemma 2: Let Q be a χ^2 random variable with D degrees of freedom. For any positive x,

$$\mathsf{P}(Q > D + 2\sqrt{Dx} + 2x) \leq \exp(-x). \tag{78}$$

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Hence, choosing $x = \frac{B\eta}{16\sigma^2 C_2}$ and letting $D = 2C_2$ yield

$$\mathsf{P}\{E_0\} \le \mathsf{P}\left\{Q \ge 2C_2 + \sqrt{\frac{B\eta}{2\sigma^2}} + \frac{B\eta}{8\sigma^2 C_2}\right\} \tag{79}$$

$$\le \exp\left(-\frac{B\eta}{16\sigma^2 C_2}\right) \tag{80}$$

where (79) is because $2C_2 + \sqrt{\frac{B\eta}{2\sigma^2}} + \frac{B\eta}{8\sigma^2 C_2} \le \frac{B\eta}{4\sigma^2}$ for sufficiently large K with B and C_2 chosen according to (27) and (30), respectively, and with η being a constant. Moreover, if we pick η as a constant that satisfies

$$\eta \ge \frac{32\sigma^2 \lceil \beta_1 \log K \rceil \log_e(K)}{\beta_0 K}, \tag{81}$$

then the error probability satisfies

$$\mathsf{P}\{E_0\} \le K^{-2}. \tag{82}$$

Note that the right hand side of (81) vanishes as $K \to \infty$. Hence, there exists such a constant η.

2) Singleton error detection: Suppose subcarrier b is a singleton due to device k. Let E_1 denote the subcarrier detection error. A singleton detection error occurs due to one or more of three events: (1) $E_{1,0} = \left\{\|\dot{\boldsymbol{Y}}_b\|^2 < \eta\right\}$; (2) Not $E_{1,0}$ and $E_{1,1} = \left\{\hat{\tilde{\boldsymbol{g}}}_k \ne \tilde{\boldsymbol{g}}_k\right\}$; (3) $E_{1,2} = \left\{\|\dot{\boldsymbol{Y}}_b - \dot{A}_{k,b}\dot{\boldsymbol{g}}_k\|^2 > \eta\right\}$. We prove

$$\mathsf{P}\{E_{1,0}\} \le 2K^{-2}, \tag{83}$$

$$\mathsf{P}\{E_{1,1}\} \le K^{-2}, \tag{84}$$

$$\mathsf{P}\{E_{1,2}\} \le K^{-2}. \tag{85}$$

in Appendices B, C, and D, respectively, using the tools in Appendix A. We thus conclude

$$\mathsf{P}\{E_1\} \le \mathsf{P}\{E_{1,0}\} + \mathsf{P}\{E_{1,1}\} + \mathsf{P}\{E_{1,2}\} \tag{86}$$

$$\le 4K^{-2}. \tag{87}$$

3) Multiton error detection: Let E_2 denote the multiton subcarrier detection error. It is proved in Appendix E that

$$\mathsf{P}\{E_2\} \le 2K^{-2}. \tag{88}$$

Combining (82), (86), and (88), we conclude that the robust subcarrier detection makes the correct decision with probability higher than $1 - 7K^{-2}$.

Note that all devices are decoded from distinct singleton bins subject to independent noise. Consider the decoding of device k from a singleton bin b_k in iteration t. Conditioned on that every device $\ell \in \mathcal{S}(t-1)$ is decoded correctly and the residual error due to e_ℓ is upper bounded by (38), the bin error e_k is a Gaussian variable, i.e., $e_k \sim \mathcal{CN}(0, 2\sigma^2/(BC))$. We have

$$\mathsf{P}\Big\{|e_k| \ge \tau/\log^2 K \Big| \bigcap\nolimits_{\ell\in\mathcal{S}(t-1)} \Big\{(|e_\ell| \le \tau/\log^2 K) \cap (\text{device } \ell \text{ is correctly decode})\Big\}\Big\} \tag{89}$$

$$\le \exp\left(-\frac{\tau^2 BC}{2\sigma^2 \log^4 K}\right)$$

where (89) is due to the fact that $|e_k|^2$ is independent of the decoding of other devices $\ell \in \mathcal{S}(t-1)$ and follows the exponential distribution with mean $\frac{2\sigma^2}{BC}$. With the choice of B given by (27) and C given by (31), the probability is smaller than K^{-2} for large enough K.

We have thus established Propositions 1-4.

F. Complexity

We perform B-point DFT for C symbols. The computational complexity is $O(K(\log K)(\log N + \log S + \log K))$ using FFT operations. In robust subcarrier detection implemented after initialization of Algorithm 2, for each subcarrier, the phase estimation involves $O(\log N)$ operations, the device identification and message decoding involves $O(\log N + \log S)$ operations, and the singleton detection and verification involves $O(\log K)$ operations. The *while* loop of Algorithm 2 will be implemented no more than K times. In each *while* loop, delay estimation is neglected when $M = 0$. For every subcarrier $b' \in \mathcal{S}$, where $|\mathcal{S}| \le T$, the cancellation of signal of the detected device involves $O(\log N + \log S + \log K)$ operations. The robust subcarrier detection for subcarrier b' involves $O(\log N + \log S + \log K)$ operations. Since T is a constant, the *while* loop involves no more than $O(K(\log N + \log S + \log K))$ operations. As a result, the complexity of DFT dominates, leading to the total computational complexity of $O\left(K(\log K)(\log N + \log S + \log K)\right)$. Constant α_1 is introduced as the leading constant, which depends on the $O(K \log K)$ FFT operations.

Hence the proof of Theorem 1.

VI. Proof of Theorem 2 (the Asynchronous Case)

In the asynchronous massive access case, we choose the parameters according to (25)–(30). The last subframe in Fig. 3 consists of

$$C_3 = M + \lceil \beta_2 K \log(KM + 1) \rceil \tag{90}$$

chips, where $\beta_2 \ge 256\bar{a}^2/a^2$. The total codelength in transmitted chips is thus

$$L = (B + M)C + C_3 \tag{91}$$

$$\le (\beta_0 K + M)\big((1 + \lceil 1/R \rceil)\lceil \log N \rceil + \lceil 1/R \rceil \lceil \log S \rceil + \beta_1 \lceil \log K \rceil\big) + M + \lceil \beta_2 K \log(KM + 1) \rceil \tag{92}$$

where (92) is due to (31) and (90). Therefore, the codelength satisfies (3) if we choose

$$\alpha_0 = 2\big(\beta_0(1 + \lceil 1/R \rceil + \beta_1) + \beta_2\big). \tag{93}$$

The FFT operation and channel estimation involve the same number of operations as the synchronous case. In addition, each device needs to estimate its delay once by taking M correlations. The complexity of delay estimation is $O(M(M + K\log(KM + 1))$. Up to K devices need to estimate their delays. The total computational complexity is thus $O\left(K(\log K)(\log N + \log S + \log K)\right) + O\left(KM^2 + K^2 M \log(KM + 1)\right)$.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Lemma 3: Suppose the conditions specified in Theorem 2 hold. Suppose the bipartite graph $G \in \mathcal{G}$ and the parameters are chosen according to (25)–(30). Suppose $\beta_2 \geq 256\bar{a}^2/a^2$ and is chosen to be an integer, and $C_3 = M + \lceil \beta_2 K \log(KM+1) \rceil$ chips are used for delay estimation, the delay of a device estimated according to (22) is correct with probability no less than $1 - 16/K^2$.

The proof of Lemma 3 is relegated to Appendix F.

Denote by $\mathcal{V}$ the event that the delay estimation of all devices based on the signal in subframe 3 is correct. Then we have

$$\mathsf{P}\{\bar{\mathcal{V}}|G \in \mathcal{G}\} \leq 16/K \tag{94}$$

by the union bound and Lemma 3. Conditioned on that the device delays are correctly detected, the residual channel estimation errors can be characterized in the same manner as in the synchronous case. Recall from (43) that the error probability can be upper bounded by

$$P_s \leq \mathsf{P}\{\mathcal{E}|G \in \mathcal{G}\} + \mathsf{P}\{G \notin \mathcal{G}|G \in \hat{\mathcal{G}}\} + \mathsf{P}\{G \notin \hat{\mathcal{G}}\}, \tag{95}$$

where

$$\mathsf{P}\{\mathcal{E}|G \in \mathcal{G}\} \leq \mathsf{P}\{\bar{\mathcal{V}}|G \in \mathcal{G}\} + \mathsf{P}\{\mathcal{E}|\mathcal{V}, G \in \mathcal{G}\}\mathsf{P}\{\mathcal{V}|G \in \mathcal{G}\} \tag{96}$$

$$\leq 16/K + \mathsf{P}\{\mathcal{E}|\mathcal{V}, G \in \mathcal{G}\} \tag{97}$$

$$\leq 16/K + 8T/K, \tag{98}$$

where (94) is used to obtain (97), and (98) is due to (45). We thus have

$$P_s \leq (16 + 8T + \nu + \zeta)/K \tag{99}$$

according to (34), (58), (95), and (98).

Hence Theorem 2 is proved with $K_0 = \max\{(16 + 8T + \nu + \zeta)/\epsilon, K_0'\}$ and α_1 being some constant due to the FFT operation.

VII. Simulation Results

As discussed in Section I, massive access can be regarded as device identification if the transmitted data are regarded as a segment of the identities. Without loss of generality, we focus on the performance of device identification via sparse OFDMA throughout all simulations. The performance of sparse OFDMA will be compared with two random access schemes, namely slotted ALOHA and CSMA. Throughout the simulation, the channel coefficient amplitude is uniformly randomly generated from $[1, 2]$ and the phase is uniformly randomly generated from $[0, 2\pi]$. The received SNR is defined as $\text{SNR} = 10\log(1/(2\sigma^2))$, where σ^2 is the noise variance per dimension.

A. Synchronous device identification

We first investigate the error probability of synchronous device identification via sparse OFDMA. Let the total number of devices be $N = 2^{38}$, which is over 274 billion. (Alternatively, one can have one million devices where each active device can transmit $\log(274000) \approx 18$ bits). We choose the parameters as $B = \lceil 2.5K \rceil$, $C_0 = C_2 = 4$. The code rate $R = 0.5$ is used to encode the device index, and thus $C_1 = 2\lceil \log N \rceil$. The number of subcarriers assigned to each device is $T = 3$.

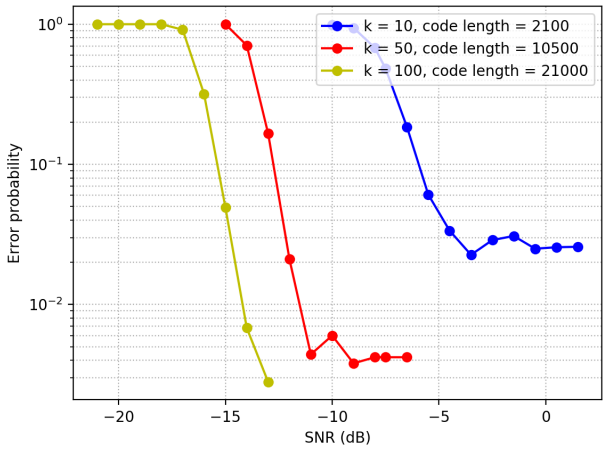

Fig. 6. Error probability of device identification in the case of synchronous transmission, where $N = 2^{38}$.

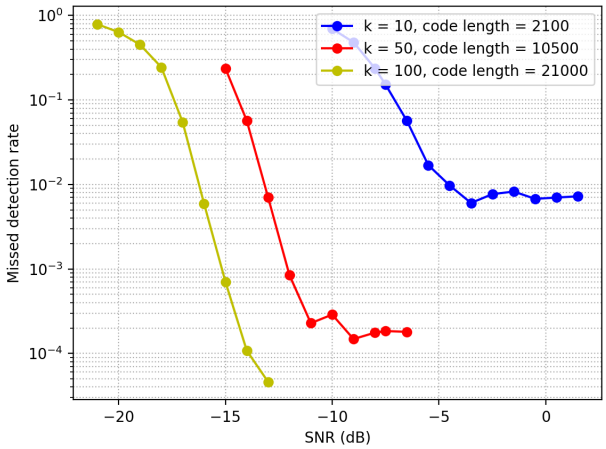

Fig. 7. Rate of missed detection in the case of synchronous transmission, where is $N = 2^{38}$.

Fig. 6 shows the error probability of synchronous device identification. In each simulation, if there exists missed detection or false alarm, an error is registered. Throughout the simulation, the false positive rates is smaller than 10^{-5} for the SNRs in the region of interest, and thus we do not plot the results here.

Fig. 7 shows the miss rate, defined as the average number of misses normalized by the number of active devices K. Simulation shows that with a code length of 10500, sparse OFDM can achieve a miss detection rate lower than 10^{-3} at SNR=-10 dB. In the case of a 20 MHz channel bandwidth, the transmission time is approximately 0.5 ms.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

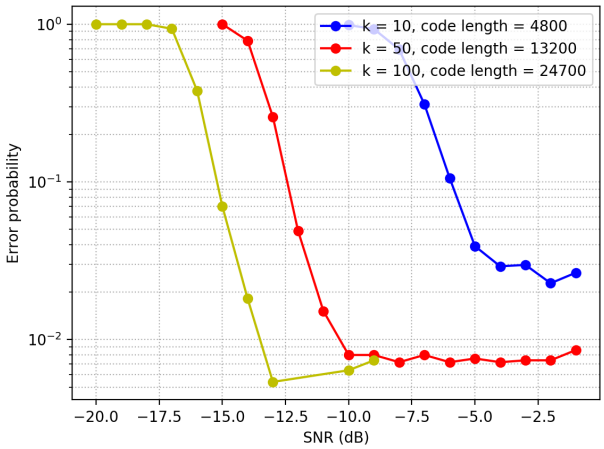

Fig. 8. Error probability of device identification in the case of discrete delay. The device population is $N = 2^{38}$ and the maximum delay is $M = 20$.

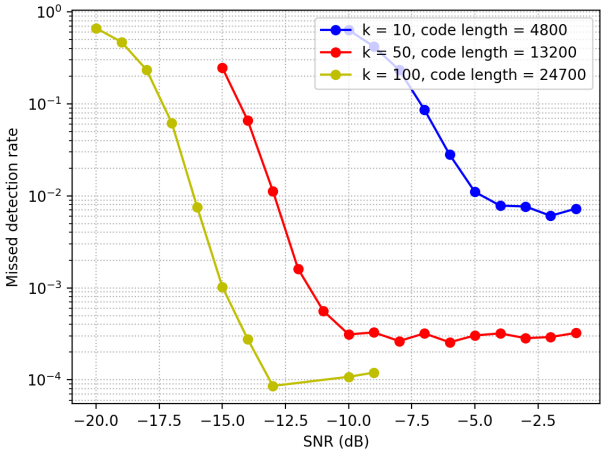

Fig. 9. Rate of missed detection in the case of discrete delay. The device population is $N = 2^{38}$ and the maximum delay is $M = 20$.

B. Asynchronous device identification

We next investigate the error probability of asynchronous device identification. We choose the parameters as $B = \lceil 2.5K \rceil$, $C_0 = C_2 = 4$. The code rate $R = 0.5$ is used to encode the device index, and thus $C_1 = 2\lceil \log N \rceil$. The number of time-domain samples used for delay estimation is $C_3 = 1000$ for both cases of $K = 10$ and $K = 50$, and $C_3 = 2000$ for $K = 100$. The device population is $N = 2^{38}$ and the maximum delay in terms of transmit samples is $M = 20$. Fig. 8 shows the error probability of asynchronous device identification. Fig. 9 shows the missed detection rate (per active user), respectively. As in the synchronous setting, the error probability is low under moderate SNR. In Figs. 6 and 8, we observe error floors which are due to accumulated estimation errors. It would be interesting to investigate how to choose system parameters and the component error-control code to improve the error floor.

We further plot in Fig. 10 the required codelength to achieve

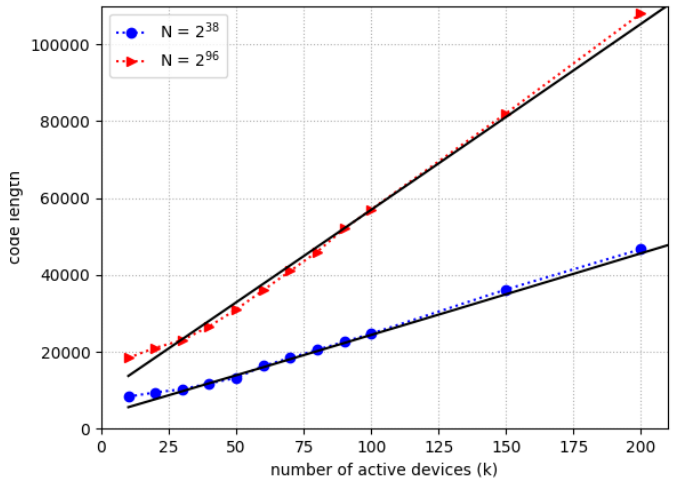

Fig. 10. Required code length to achieve error probability of 10^{-2} at SNR $= -10$ dB. The black solid curve plots $\alpha((K+M)(\log N + \log K) + K\log(M+1))$, where α is some constant to fit the curve on the required codelength.

a fixed error probability at a fixed SNR for different number of active users. In particular, we change the number of subcarriers B and the length of the synchronization subframe C_3 such that the signaling scheme is sufficient to achieve at least an error probability of 10^{-2} at -10 dB SNR. It can be seen that the proposed signaling scheme can even handle a device population of $N = 2^{96}$ with up to $K = 200$ active devices, with a maximum delay of 20 samples. Note that the 96 bits recovered per active device can be split, e.g., into a 48-bit identity and a 48-bit message. We also plot a curve scaling as $(K+M)(\log N + \log K) + K\log(M+1)$, which is the scaling result as in Theorem 2 for asynchronous massive access. It can be seen that as K increases, e.g., $K \geq 50$, the required codelength is aligned with the scaling as shown in the theorem.

C. Comparison with random access

1) Slotted ALOHA: First, consider slotted ALOHA, where every device transmits a frame with probability p independently in each slot over an N_s-slot period. The probability of one given neighbor being missed is equal to the probability that the device is unsuccessful in all N_s slots:

$$\mathsf{P}_{\text{miss,aloha}} = \left(1 - (1-p)^{K-1}p\right)^{N_s}. \tag{100}$$

Setting $p = 1/K$ minimizes $\mathsf{P}_{\text{miss,aloha}}$.

2) CSMA: It is challenging, if not impossible, to implement CSMA-based wireless access. Due to the power asymmetry between devices and access points, a device may not be able to sense another device's transmission in the same cell. Suppose, nonetheless, devices can sense each other and CSMA is used. When the channel is idle, the devices start their timers. The device whose timer expires the first transmits. When the channel becomes busy, the devices stop their timers. Device i has a chance to transmit if its timer is the minimum in some slot. The probability that a given device never gets a chance

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

to transmit is

$$\mathsf{P}_{\text{miss,csma}} = \left(1 - \mathsf{P}\left\{T_1 < \min\{T_2, \cdots, T_K\}\right\}\right)^{N_s}. \quad (101)$$

In order to reliably transmit the device index $\log N$ bits, the number of chips required in each frame is at least $\lceil \log N \rceil / \log(1 + \text{SNR})$. Therefore, the total number of chips required is $N_s \lceil \log N \rceil / \log(1 + \text{SNR})$, where N_s depends on the target miss rate. Under SNR = -10 dB, it can be seen from Fig. 9 that sparse OFDMA can achieve missed detection rate low than 10^{-3} for $K = 10, 50$ or 100. We can calculate the number of chips required by slotted ALOHA and CSMA to achieve a comparable miss detection rate of 10^{-3} at SNR = -10 dB. When $K = 50$, the code length of sparse OFDMA is around 10,500, while slotted ALOHA and CSMA requires more than 90,000 chips to achieve a missed detection rate of 10^{-3}. Sparse OFDMA can effectively reduce the code length by over 80%. Moreover, the code length reduction is even greater for larger K and a lower error probability requirement. When $K = 100$, the code length of sparse OFDMA is around 21,000, while slotted ALOHA and CSMA requires more than 180,000 chips to achieve a missed detection rate of 10^{-3}. Sparse OFDMA can effectively reduce the code length by over 85%. The significant reduction of the code length required by sparse OFDM is due to the code design, which utilizes the sparsity of the active users among the total number of users.

VIII. Conclusion

We have proposed a low-complexity asynchronous neighbor discovery and massive access scheme for very large networks with billions of nodes. The scheme may be suitable for applications in the Internet of Things. The scheme, referred to as sparse OFDMA, applies the recently developed sparse Fourier transform to compressed device identification. Compared with random access schemes, sparse OFDMA requires much shorter code length by exploiting the multiaccess nature of the channel and the multiuser detection gain. Sparse OFDMA is a divide-and-conquer approach of low complexity, where only point-to-point capacity approaching codes are adopted. It provides practical physical layer capability for multipacket reception.

This paper makes the widely adopted assumption that the system is sample-synchronous but not frame-synchronous. It would be interesting to extend the signalling scheme to the fully asynchronous setting. Another interesting direction is to consider multiple antennas or push to a massive number of antennas at the access point. Finally, extending this work to multi-path setting is also important for practical applications in wireless communication systems.

Appendix A
Auxiliary Results on Sub-Gaussian Variables

We introduce the definition of sub-Gaussian variables, which will be used in the proof of the main theorems.

Definition 3: X is σ-subGaussian if there exists $\sigma > 0$ such that

$$\mathsf{E}\left\{\exp(tX)\right\} \le \exp(\sigma^2 t^2/2), \quad \forall t \ge 0. \quad (102)$$

Definition 4 (subGaussian norm): The subGaussian norm of the random variable X is defined as

$$\|X\|_{\phi_2} = \sup_{p \ge 1} p^{-1/2} \left(\mathsf{E}|X|^p\right)^{1/p}. \quad (103)$$

The following three lemmas, which are established in [59], will be used in the proof.

Lemma 4: Suppose X is σ-subGaussian, then aX is $|a|\sigma$-subGaussian.

Lemma 5: Suppose X_1 is σ_1-subGaussian, X_2 is σ_2-subGaussian. Moreover, they are independent. Then $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subGaussian.

Lemma 6: If a random variable X is σ-subGaussian with zero mean, then for any $t > 0$, the following holds

$$\mathsf{P}\{|X| > t\} \le 2e^{-\frac{t^2}{2\sigma^2}}. \quad (104)$$

The following theorem characterizes the properties of subGaussian variables [59].

Theorem 3 (Characterization of subGaussian variables): Let $\mathsf{E}X = 0$. The following are equivalent:

1) $\mathsf{E}(e^{tX}) \le e^{\frac{t^2}{4}}$.
2) $\forall p \ge 1$, $\left(\mathsf{E}|X|^p\right)^{1/p} \le \sqrt{2p}$.

The following theorem is on the concentration of subGaussian random variables.

Theorem 4 (Hanson-Wright inequality [60]): Let $\boldsymbol{Z} = (Z_1, \cdots, Z_n) \in \mathbb{R}^n$ be a random vector with independent components Z_i which satisfy $\mathsf{E}Z_i = 0$ and the subGaussian norm $\|Z_i\|_{\phi_2} \le K$. Let $\boldsymbol{A}$ be a deterministic $n \times n$ matrix. Then, for every $t \ge 0$,

$$\begin{aligned} &\mathsf{P}\left\{|\boldsymbol{Z}^T \boldsymbol{A} \boldsymbol{Z} - \mathsf{E}\left\{\boldsymbol{Z}^T \boldsymbol{A} \boldsymbol{Z}\right\}| > t\right\} \\ &\le 2\exp\left[-\varsigma \min\left(\frac{t^2}{K^4 \|\boldsymbol{A}\|_F^2}, \frac{t}{K^2 \|\boldsymbol{A}\|_{op}}\right)\right], \end{aligned} \quad (105)$$

where the operator norm of $\boldsymbol{A}$ is

$$\|\boldsymbol{A}\|_{op} = \max_{\boldsymbol{x} \ne 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}, \quad (106)$$

the Frobenius norm of $\boldsymbol{A}$ is $\|\boldsymbol{A}\|_F = (\sum_{i,j} |A_{i,j}|^2)^{1/2}$ and ς is an absolute constant that does not depend on $K, \boldsymbol{A}$ and t.

Appendix B
Proof of Singleton Error $E_{1,0}$ (83)

We first establish the following lemma, which is key to the proof of (83).

Lemma 7: Let $B = \beta_0 K$ be given by (27). Let $c \in [0, C_2)$ be a fixed constant. Let $\boldsymbol{Z} = \boldsymbol{W} + \boldsymbol{V}$ be a vector of length C_2, where $\|\boldsymbol{V}\| \le \sqrt{\eta}/2$ and the entries of $\boldsymbol{W}$ are i.i.d. $\mathcal{CN}(0, 2\sigma^2/B)$ random variables. Let $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\dagger$ be a deterministic matrix, where $\boldsymbol{U}$ is a real-valued orthogonal matrix and $\boldsymbol{\Lambda} = \text{diag}\{0, \cdots, 0, 1, \cdots, 1\}$ is a diagonal matrix with c zeros and $C_2 - c$ ones. Let

$$\boldsymbol{S} = \sum_{k=1}^{k_0} A_{k,b} \dot{\boldsymbol{g}}_k, \quad (107)$$

where $k_0 \ge 1$, the entries of $\dot{\boldsymbol{g}}_k$ are i.i.d. BPSK symbols, and $A_{k,b}$ are deterministic variables with $|A_{k,b}| \ge a$. Moreover,

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

$\boldsymbol{S}, \boldsymbol{W}$, and $\boldsymbol{V}$ are mutually independent variables. Then there exists some β_1 such that, for large enough K,

$$\mathsf{P}\left\{\|\boldsymbol{QS}+\boldsymbol{Z}\|^2 \le \eta\right\} \le \frac{2}{K^2}, \tag{108}$$

where $C_2 = \lceil \beta_1 \log K \rceil$ by (30) and η is the constant energy threshold.

Proof: In the following proof, we use $\sum_k$ as a shorthand for $\sum_{k=1}^{k_0}$. The expectation of $\|\boldsymbol{QS}\|^2$ is easily obtained as $\vartheta = (C_2 - c)\sum_k |A_{k,b}|^2$. We have

$$\begin{aligned}&\mathsf{P}\left\{\|\boldsymbol{QS}+\boldsymbol{Z}\|^2 \le \eta\right\}\\ &\le \mathsf{P}\left\{\|\boldsymbol{QS}+\boldsymbol{Z}\|^2 \le \eta \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\}\\ &\quad + \mathsf{P}\left\{\|\boldsymbol{QS}\|^2 \le \vartheta/2\right\}. \end{aligned}\tag{109}$$

To bound the first term on the right hand side of (109), we use the triangle inequality $\|\boldsymbol{Z}\| + \|\boldsymbol{QS}+\boldsymbol{Z}\| \ge \|\boldsymbol{QS}\|$ to obtain

$$\begin{aligned}&\mathsf{P}\left\{\|\boldsymbol{QS}+\boldsymbol{Z}\|^2 \le \eta \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\}\\ &\le \mathsf{P}\left\{\|\boldsymbol{QS}\| - \|\boldsymbol{Z}\| \le \sqrt{\eta} \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\}.\end{aligned}\tag{110}$$

For large enough K, η can be chosen to be an arbitrarily small constant to satisfy (81). For every fixed c, we have $\vartheta \ge (C_2 - c)a^2$, which is greater than 8η for large enough K. Therefore, for large enough K,

$$\begin{aligned}&\mathsf{P}\left\{\|\boldsymbol{QS}\| - \|\boldsymbol{Z}\| \le \sqrt{\eta} \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\}\\ &\le \mathsf{P}\left\{\|\boldsymbol{Z}\| \ge \sqrt{\eta} \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\} &(111)\\ &\le \mathsf{P}\left\{\|\boldsymbol{W}\| \ge \sqrt{\eta}/2 \middle| \|\boldsymbol{QS}\|^2 \ge \vartheta/2\right\} &(112)\\ &= \mathsf{P}\left\{\|\boldsymbol{W}\| \ge \sqrt{\eta}/2\right\} &(113)\\ &\le 1/K^2, &(114)\end{aligned}$$

where (111) follows because conditioned on $\|\boldsymbol{QS}\|^2 \ge \vartheta/2$, $\|\boldsymbol{QS}\| - \|\boldsymbol{Z}\| \le \sqrt{\eta}$ implies that $\|\boldsymbol{Z}\| \ge \sqrt{\eta}$; (112) follows because conditioned on $\|\boldsymbol{V}\| \le \sqrt{\eta}/2$, $\|\boldsymbol{Z}\| \ge \sqrt{\eta}$ implies $\|\boldsymbol{W}\| \ge \sqrt{\eta}/2$; (113) follows because $\boldsymbol{QS}$ is independent of $\boldsymbol{W}$, and (114) follows from (75) and (82).

The second term on the right hand side of (109) is derived in the following steps using the tools in Appendix A. First, using Lemma 4 and Lemma 5 in Appendix A, we show that the real and imaginary parts of $\boldsymbol{S}$ given by (107) are subGaussian variables. Then, by Definition 3 and Theorem 3 in Appendix A, we obtain the upper bounds of the subGaussian norms of the real components and imaginary components of $\boldsymbol{S}$. Finally, we apply Theorem 4 in Appendix A to show that $\|\boldsymbol{QS}\|_2^2$ are concentrated around $(C_2 - c)\sum_k |A_{k,b}|^2$ with high probability. In order to achieve that, it suffices to show that the operator norm of $\boldsymbol{Q}$ is 1 and the Frobenius norm of $\boldsymbol{Q}$ is $C_2 - c$.

Lemma 8: Let $\boldsymbol{S}_R = (S_{0,R}, \cdots, S_{C_2-1,R})$ and $\boldsymbol{S}_I = (S_{0,I}, \cdots, S_{C_2-1,I})$ be the real and imaginary components of $\boldsymbol{S}$, respectively. Let $u = \sqrt{\sum_k (\mathrm{Re}\{A_{k,b}\})^2}$ and $v = \sqrt{\sum_k (\mathrm{Im}\{A_{k,b}\})^2}$. Then $S_{c,R}$ are i.i.d. u-subGaussian random variables with $\mathsf{E}S_{c,R} = 0$ and the subGaussian norm satisfies $\|S_{c,R}\|_{\phi_2} \le 2u$. Similarly, $S_{c,I}$ are i.i.d. v-subGaussian random variables with $\mathsf{E}S_{c,I} = 0$ and the subGaussian norm satisfies $\|S_{c,I}\|_{\phi_2} \le 2v$.

Proof: Since $\dot{g}_k^c$ is BPSK symbol, it is 1-subGaussian with zero mean. Thus, $\mathsf{E}S_{c,R} = 0$. Moreover, $\dot{g}_k^c$ are independent across k. According to Lemma 4 and Lemma 5, $S_{c,R} = \sum_k A_{k,R}\dot{g}_k^c$ is u-subGaussian. $\{S_{c,R}\}_{c=0}^{C_2-1}$ are independent, because $\dot{g}_k^c$ are independent across $c = 0, \cdots, C_2 - 1$.

By Definition 3, $\mathsf{E}\{\exp(tS_{c,R})\} \le \exp\left(u^2t^2/2\right)$. Let $X = S_{c,R}/\sqrt{2}u$. Then $\mathsf{E}\{\exp(tX)\} \le \exp(t^2/4)$. According to Theorem 3, for all $p \ge 1$, $(\mathsf{E}|X|^p)^{1/p} \le \sqrt{2p}$, which yields

$$(\mathsf{E}|S_{c,R}|^p)^{1/p} \le 2u\sqrt{p}. \tag{115}$$

By (103), the subGaussian norm of $S_{c,R}$ is upper bounded as $\|S_{c,R}\|_{\phi_2} \le 2u$. The statement for the imaginary parts follow similarly. ■

In order to apply Theorem 4 to provide a concentration result on $\|\boldsymbol{QS}\|$, we need to first derive the Frobenius norm and the operator norm of $\boldsymbol{Q}$. The Frobenius norm of $\boldsymbol{Q}$ is calculated as

$$\begin{aligned}\|\boldsymbol{Q}\|_F^2 &= \mathrm{tr}\{\boldsymbol{QQ}^\dagger\} &(116)\\ &= \mathrm{tr}\{\boldsymbol{Q}\} &(117)\\ &= C_2 - c, &(118)\end{aligned}$$

where (117) follows by $\boldsymbol{QQ}^\dagger = \boldsymbol{Q}$, and (118) follows because the sum of the eigenvalues of $\boldsymbol{Q}$ is $C_2 - c$.

Since the largest eigenvalue of $\boldsymbol{Q}$ is 1, the operator norm of $\boldsymbol{Q}$ defined by (106) is calculated as

$$\|\boldsymbol{Q}\|_{op} = 1. \tag{119}$$

Moreover, we have

$$\begin{aligned}\mathsf{E}\left\{\|\boldsymbol{QS}_R\|^2\right\} &= \mathsf{E}\left\{\boldsymbol{S}_R^\dagger \boldsymbol{QQ}^\dagger \boldsymbol{S}_R\right\} &(120)\\ &= \mathsf{E}\left\{\boldsymbol{S}_R^\dagger \boldsymbol{QS}_R\right\} &(121)\\ &= \mathsf{E}\{S_{c,R}^2\}\,\mathrm{tr}\{\boldsymbol{Q}\} &(122)\\ &= (C_2 - c)u^2, &(123)\end{aligned}$$

where (122) follows because $\{S_{c,R}\}_{c=0}^{C_2-1}$ are i.i.d. distributed. Similarly, $\mathsf{E}\left\{\|\boldsymbol{QS}_I\|^2\right\} = (C_2 - c)v^2$.

Applying (119), (123) and Theorem 4 with $\boldsymbol{Z} = \boldsymbol{S}_R$, $\boldsymbol{A} = \boldsymbol{Q}$ with $\|\boldsymbol{Q}\|_{op} = 1$, $\|\boldsymbol{Q}\|_F^2 = C_2 - c$ and $K = 2u$ yields

$$\begin{aligned}&\mathsf{P}\left\{\left|\|\boldsymbol{QS}_R\|^2 - (C_2 - c)u^2\right| > t\right\}\\ &\le 2\exp\left(-\varsigma \min\left(\frac{t^2}{16(C_2 - c)u^4}, \frac{t}{4u^2}\right)\right),\end{aligned}\tag{124}$$

where ς is a constant introduced in Theorem 4 in Appendix A.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Letting $t=(C_2-c)u^2/2$, we have

$$\mathsf{P}\left\{\|\boldsymbol{QS}_R\|^2 \le \frac{(C_2-c)u^2}{2}\right\} = \mathsf{P}\left\{\|\boldsymbol{QS}_R\|^2 \le t\right\} \tag{125}$$

$$\le \mathsf{P}\left\{\left|\|\boldsymbol{QS}_R\|^2 - 2t\right| > t\right\} \tag{126}$$

$$\le 2\exp\left(-\frac{\varsigma}{64}(C_2-c)\right), \tag{127}$$

where (127) is due to (124). Similarly, we have

$$\mathsf{P}\left\{\|\boldsymbol{QS}_I\|^2 \le \frac{(C_2-c)v^2}{2}\right\} \le 2\exp\left(-\frac{\varsigma}{64}(C_2-c)\right). \tag{128}$$

Since $\boldsymbol{Q}$ is a real-valued matrix, $\|\boldsymbol{QS}\|^2 = \|\boldsymbol{QS}_R\|^2 + \|\boldsymbol{QS}_I\|^2$. Moreover, $\sum_k |A_{k,b}|^2 = u^2+v^2$. Combining (127) and (128), we have

$$\mathsf{P}\left\{\|\boldsymbol{QS}\|^2 \le \frac{(C_2-c)\sum_k |A_{k,b}|^2}{2}\right\}$$

$$\le \mathsf{P}\left\{\|\boldsymbol{QS}_R\|^2 \le \frac{(C_2-c)u^2}{2}\right\} + \mathsf{P}\left\{\|\boldsymbol{QS}_I\|^2 \le \frac{(C_2-c)v^2}{2}\right\} \tag{129}$$

$$\le 4\exp\left(-\frac{\varsigma}{64}(C_2-c)\right). \tag{130}$$

Combining (109), (114) and (130), there exists some large enough β_1 such that for $C_2 = \lceil \beta_1 \log K \rceil$,

$$\mathsf{P}\left\{\|\boldsymbol{QS}+\boldsymbol{Z}\|^2 \le \eta\right\} \le \frac{2}{K^2}. \tag{131}$$

■

The probability that a singleton is declared to be a zeroton is calculated as

$$\mathsf{P}\left\{E_{1,0}\right\} = \mathsf{P}\left\{\|A_{k,b}\dot{\boldsymbol{g}}_k + \dot{\boldsymbol{Z}}_b\|^2 \le \eta\right\}. \tag{132}$$

Therefore, we can apply Lemma 7 with $\boldsymbol{S} = A_{k,b}\dot{\boldsymbol{g}}_k$, $\boldsymbol{Q}=\boldsymbol{I}$, $\boldsymbol{Z} = \dot{\boldsymbol{Z}}_b$, $k_0 = 1$, and $c=0$ to obtain (83). In this case, β_1 can be chosen to satisfy $\beta_1 \ge 192/\varsigma$ such that $4\exp\left(-\frac{\varsigma}{64}\lceil\beta_1\log K\rceil\right) \le K^{-2}$. ■

Appendix C
Proof of Singleton Error $E_{1,1}$ (84)

We first show that the phase compensation is accurate with high probability. Second, we show that the interference only causes a slight degradation of signal strength with high probability. Third, the device index recovery can be regarded as transmission over a BSC channel and a large enough C_1 can help recover the index information.

We estimate the phase of $\theta = \angle A_{k,b}$ according to (16). The estimate can be expressed as

$$\hat{\theta} = \angle\left(A_{k,b} + \bar{Z}\right), \tag{133}$$

where $\bar{Z} = \sum_{c=0}^{C_0-1}\left(\bar{W}_b^c + \bar{V}_b^c\right)/C_0$. From the geometric interpretation, the maximum phase offsets occurs when the noise is orthogonal to the measurement. Suppose each entry of V_b^c is upper bounded by $\bar{v}$ given by (40). We choose a small θ_0 such that $\theta_0 < \frac{\pi}{3}$ and $\sin\theta_0 > \theta_0/2$, then

$$\mathsf{P}\left\{|\hat{\theta}-\theta| > \theta_0\right\}$$

$$\le \mathsf{P}\left\{\arcsin\frac{|\bar{Z}|}{|A_{k,b}|} > \theta_0\right\} \tag{134}$$

$$\le \mathsf{P}\left\{|\bar{Z}| > a\sin\theta_0\right\} \tag{135}$$

$$\le \mathsf{P}\left\{|\bar{Z}| > \frac{a\theta_0}{2}\right\} \tag{136}$$

$$\le \mathsf{P}\left\{\left|\frac{1}{C_0}\sum_{c=0}^{C_0-1}\bar{W}_b^c\right| > \frac{a\theta_0}{2} - \bar{v}\right\} \tag{137}$$

$$\le 2\mathsf{P}\left\{\left|\frac{1}{C_0}\sum_{c=0}^{C_0-1}\mathrm{Re}(\bar{W}_b^c)\right| > \frac{a\theta_0}{4} - \frac{\bar{v}}{2}\right\} \tag{138}$$

$$\le 4\exp\left(-\frac{(a\theta_0/4-\bar{v}/2)^2 BC_0}{2\sigma^2}\right), \tag{139}$$

where (139) is because $\frac{1}{C_0}\sum_{c=0}^{C_0-1}\mathrm{Re}(\bar{W}_b^c) \sim \mathcal{N}(0,\sigma^2/(BC_0))$.

For $|\hat{\theta}-\theta| \le \theta_0$, considering that $|V_b^c| \le \bar{v}$, we have

$$\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \ge a/2 - \bar{v}, \tag{140}$$

$$\mathrm{Re}\left\{-A_{k,b}e^{-\iota\hat{\theta}} + \tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \le -a/2 + \bar{v}. \tag{141}$$

Thus,

$$\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \ge a/2 - \bar{v}. \tag{142}$$

Hence,

$$\mathsf{P}\left\{\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \le a/2-\bar{v}\right\}$$

$$\le \mathsf{P}\left\{|\hat{\theta}-\theta| > \theta_0\right\} \tag{143}$$

$$\le 4\exp\left(-\frac{(a\theta_0/4-\bar{v}/2)^2 BC_0}{2\sigma^2}\right). \tag{144}$$

To predict $\tilde{g}_k^c$, we make hard binary decision on

$$\mathrm{Re}\left\{\tilde{Y}_b^c e^{-\iota\hat{\theta}}\right\}$$

$$= \mathrm{Re}\left\{A_{k,b}\tilde{g}_k^c e^{-\iota\hat{\theta}} + \tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} + \mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\} \tag{145}$$

$$= \tilde{g}_k^c\,\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} + \mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\}. \tag{146}$$

The prediction error occurs when $\tilde{g}_k^c\,\mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\} < -\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\}$. The flip rate can be upper bounded by

$$\mathsf{P}\left\{\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \le a/2-\bar{v}\right\} +$$

$$\mathsf{P}\left\{\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\} \ge a/2-\bar{v},\right.$$

$$\left.\tilde{g}_k^c\,\mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\} < -\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}} + \tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\}\right\}. \tag{147}$$

Since $\bar{v}$ given by (40) vanishes as K goes to infinity, $a/2-\bar{v}$ can be lower bounded by a constant arbitrarily close to $a/2$.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

The first term is upper bounded by (143), and the second term

$$\begin{aligned}
&\mathsf{P}\left\{\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}}+\tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\}\geq a/2-\bar{v},\right.\\
&\quad\left.\tilde{g}_k^c\,\mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\}<-\mathrm{Re}\left\{A_{k,b}e^{-\iota\hat{\theta}}+\tilde{g}_k^c\tilde{V}_b^c e^{-\iota\hat{\theta}}\right\}\right\}\\
&\leq \mathsf{P}\left\{\tilde{g}_k^c\,\mathrm{Re}\left\{\tilde{W}_b^c e^{-\iota\hat{\theta}}\right\}<-(a/2-\bar{v})\right\} &(148)\\
&\leq \mathsf{P}\left\{-|\tilde{W}_b^c|<-(a/2-\bar{v})\right\} &(149)\\
&\leq \exp\left(-\frac{(a/2-\bar{v})^2B}{2\sigma^2}\right). &(150)
\end{aligned}$$

The flip rate is upper bounded by $4\exp\left(-\frac{(a\theta_0/4-\bar{v}/2)^2BC_0}{2\sigma^2}\right)+\exp\left(-\frac{(a/2-\bar{v})^2B}{2\sigma^2}\right)$. Given $B=\beta_0K$ and $C_0=\lceil\log N\rceil$, the flip rate tends to 0 as K and N goes to infinity. By [50, Theorem 6.1], there exists a code rate that depends only on the given flipping rate of the binary symmetric channel, to encode the $(\lceil\log N\rceil+\lceil\log S\rceil)$-bit information, such that the index can be recovered correctly with probability at least $1-1/N^2$. For large enough K, the flipping rate $e^{-O(K)}$ vanishes. Thus, incorrect device identification and message decoding occurs with probability

$$\mathsf{P}\{E_{1,1}\}\leq N^{-2}\leq K^{-2} \tag{151}$$

with a certain fixed rate R. ∎

Appendix D
Proof of Singleton Error $E_{1,2}$ (85)

Let $\dot{A}_{k,b}=\dot{\boldsymbol{g}}_k^\dagger\dot{\boldsymbol{Y}}_b/C_2$. We have

$$\dot{\boldsymbol{Y}}_b-\dot{A}_{k,b}\dot{\boldsymbol{g}}_k=\left(\boldsymbol{I}-\frac{1}{C_2}\dot{\boldsymbol{g}}_k\dot{\boldsymbol{g}}_k^\dagger\right)\dot{\boldsymbol{Z}}_b. \tag{152}$$

Let $\boldsymbol{Q}=\boldsymbol{I}-\frac{1}{C_2}\dot{\boldsymbol{g}}_k\dot{\boldsymbol{g}}_k^\dagger$. Since $\dot{\boldsymbol{g}}_k^\dagger\dot{\boldsymbol{g}}_k=C_2$, $\dot{\boldsymbol{g}}_k$ is the eigenvector of $\dot{\boldsymbol{g}}_k\dot{\boldsymbol{g}}_k^\dagger/C_2$. Since $\dot{\boldsymbol{g}}_k\dot{\boldsymbol{g}}_k^\dagger/C_2$ is a rank-1 matrix, it has only a single nonzero eigenvalue which is 1. Therefore, the eigenvalue decomposition of $\boldsymbol{Q}$ can be written as

$$\begin{aligned}
\boldsymbol{Q}&=\boldsymbol{U}\Lambda\boldsymbol{U}^\dagger &(153)\\
&=\boldsymbol{U}\mathrm{diag}\{0,1,\cdots,1\}\boldsymbol{U}^\dagger &(154)
\end{aligned}$$

where $\boldsymbol{U}$ is an orthogonal matrix. Then we have

$$\begin{aligned}
\|\dot{\boldsymbol{Y}}_b-\dot{A}_{k,b}\dot{\boldsymbol{g}}_k\|_2^2&=\|\boldsymbol{Q}\dot{\boldsymbol{Z}}_b\|^2 &(155)\\
&=\|\Lambda\boldsymbol{U}^\dagger\dot{\boldsymbol{Z}}_b\|^2 &(156)\\
&=\sum_{c=1}^{C_2-1}|Z_c'|^2, &(157)
\end{aligned}$$

where $\boldsymbol{Z}'=\boldsymbol{U}^\dagger\dot{\boldsymbol{Z}}_b$ and (157) is due to $\Lambda=\mathrm{diag}(0,1,\cdots,1)$. Since $\|\dot{\boldsymbol{Z}}_b\|^2=\|\boldsymbol{Z}'\|^2$, we have

$$\begin{aligned}
\mathsf{P}\left\{\|\dot{\boldsymbol{Y}}_b-\dot{A}_{k,b}\dot{\boldsymbol{g}}_k\|_2^2\geq\eta\right\}&=\mathsf{P}\left\{\sum_{c=1}^{C_2-1}|Z_c'|^2\geq\eta\right\} &(158)\\
&\leq\mathsf{P}\left\{\|\dot{\boldsymbol{Z}}_b\|^2\geq\eta\right\} &(159)\\
&\leq K^{-2}, &(160)
\end{aligned}$$

where (160) follows from (75) and (82).

Appendix E
Proof of Multiton Error (88)

We rewrite the subcarrier values as follows,

$$\dot{\boldsymbol{Y}}_b=\sum_{k\in\mathcal{K}:b\in\mathcal{B}_k}A_{k,b}\dot{\boldsymbol{g}}_k+\dot{\boldsymbol{Z}}_b, \tag{161}$$

where $\dot{\boldsymbol{Z}}_b=\dot{\boldsymbol{W}}_b+\dot{\boldsymbol{V}}_b$.

Suppose the incorrect estimate index from subcarrier-b is j. We have

$$\dot{A}_j=\sum_{k\in\mathcal{K}:b\in\mathcal{B}_k}\frac{1}{C_2}A_{k,b}\dot{\boldsymbol{g}}_j^\dagger\dot{\boldsymbol{g}}_k+\frac{1}{C_2}\dot{\boldsymbol{g}}_j^\dagger\dot{\boldsymbol{Z}}_b. \tag{162}$$

Thus,

$$\begin{aligned}
\dot{\boldsymbol{Y}}_b-\dot{A}_j\dot{\boldsymbol{g}}_j=&\sum_{k\in\mathcal{K}:b\in\mathcal{B}_k}A_{k,b}\left(\boldsymbol{I}-\frac{\dot{\boldsymbol{g}}_j\dot{\boldsymbol{g}}_j^\dagger}{C_2}\right)\dot{\boldsymbol{g}}_k\\
&+\left(\boldsymbol{I}-\frac{\dot{\boldsymbol{g}}_j\dot{\boldsymbol{g}}_j^\dagger}{C_2}\right)\dot{\boldsymbol{Z}}_b. &(163)
\end{aligned}$$

Let

$$\boldsymbol{S}=\sum_{k\in\mathcal{K}:b\in\mathcal{B}_k}A_{k,b}\dot{\boldsymbol{g}}_k \tag{164}$$

and $\boldsymbol{Q}=\boldsymbol{I}-\frac{1}{C_2}\dot{\boldsymbol{g}}_j\dot{\boldsymbol{g}}_j^\dagger$, then the first term in (163) can be written as

$$\sum_{k\in\mathcal{K}:b\in\mathcal{B}_k}A_{k,b}\left(\boldsymbol{I}-\frac{\dot{\boldsymbol{g}}_j\dot{\boldsymbol{g}}_j^\dagger}{C_2}\right)\dot{\boldsymbol{g}}_k=\boldsymbol{Q}\boldsymbol{S}. \tag{165}$$

The multiton subcarrier cannot be detected when $\|\boldsymbol{Y}_b-\dot{A}_{j,b}\dot{\boldsymbol{g}}_j\|^2\leq\eta$. In the following, we upper bound the error probability assuming $b\notin\mathcal{B}_j$. A similar analysis can be carried out for the case of $b\in\mathcal{B}_j$. It is obvious that $\boldsymbol{Q}$ depends on $\boldsymbol{g}_j$ which is independent from $\boldsymbol{S}$. Moreover, $\boldsymbol{V}$ are based on the design parameters from the devices that have been recovered, which are independent of $\boldsymbol{Q}$ and $\boldsymbol{S}$. Conditioned on $\dot{\boldsymbol{g}}_j$, $\boldsymbol{Q}$ is a deterministic matrix, we can apply Lemma 7 with $c=1$, $k_0=|k\in\mathcal{K}:b\in\mathcal{B}_k|$ and $\boldsymbol{Z}=\boldsymbol{Q}\dot{\boldsymbol{Z}}_b$. The condition of the lemma still holds because

$$\|\boldsymbol{Q}\dot{\boldsymbol{V}}_b\|^2=\|\dot{\boldsymbol{V}}_b\|^2-\frac{1}{C_2}\dot{\boldsymbol{V}}_b^\dagger\dot{\boldsymbol{g}}_j\dot{\boldsymbol{g}}_j^\dagger\dot{\boldsymbol{V}}_b\leq\|\dot{\boldsymbol{V}}_b\|^2. \tag{166}$$

The error probability can be upper bounded as

$$\begin{aligned}
\mathsf{P}\left\{\|\boldsymbol{Y}_b-\dot{A}_{j,b}\dot{\boldsymbol{g}}_j\|^2\leq\eta|\dot{\boldsymbol{g}}_j\right\}&=\mathsf{P}\left\{\|\boldsymbol{Q}\boldsymbol{S}+\boldsymbol{Q}\dot{\boldsymbol{Z}}_b\|^2\leq\eta|\dot{\boldsymbol{g}}_j\right\} &(167)\\
&\leq 2K^{-2}, &(168)
\end{aligned}$$

where (168) directly follows from (108). Since (167) holds for every $\dot{\boldsymbol{g}}_j$, we have

$$\mathsf{P}\left\{\|\boldsymbol{Y}_b-\dot{A}_{j,b}\dot{\boldsymbol{g}}_j\|^2\leq\eta\right\}\leq 2K^{-2}. \tag{169}$$

∎

Appendix F
Proof of Lemma 3

The lemma holds trivially in the degenerate case of $M=0$, so we assume M is a natural number in this proof. We focus on

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

the delay estimation for device k. Without loss of generality, we assume the delay is $m_k = 0$. The device experiences the interference from the other $K-1$ devices and noise. The received synchronization pilots in subframe 3 can be written as

$$x_{(B+M)C+i} = a_k s'_{k,i} + \sum_{p\in\mathcal{K}\setminus k} a_p s'_{p,i-m_p} + w_i, \tag{170}$$

where the time-domain samples of the pilots $s'_{k,i}$ are uniform random from $\{+1,-1\}$ and the noise $w_i \sim \mathcal{CN}(0, 2\sigma^2)$.

Let $\mathcal{I}$ be defined as in (19). The number of samples contained in $\mathcal{I}$ is

$$|\mathcal{I}| = \lceil \beta_2 K \log(KM+1) \rceil, \tag{171}$$

where $\beta_2 \geq 256\bar{a}^2/a^2$. Since the random sequence has a length greater than M and the delay is no greater than M, without noise, correlating the second segment with the received sequence will yield an auto-correlation of a sequence with its cyclic shift version. The test metric in (21) is calculated as

$$\mathcal{T}_k(m) = \begin{cases} |\mathcal{I}|a_k + \sum_{i\in\mathcal{I}} w_i s'_{k,i} \\ \quad + \sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p s'_{p,i-m_p} s'_{k,i} \\ \qquad\qquad \text{if } m = 0 \\ \sum_{i\in\mathcal{I}} a_k s'_{k,i+m} s'_{k,i} + \sum_{i\in\mathcal{I}} w_{i+m} s'_{k,i} \\ \quad + \sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p s'_{p,i+m-m_p} s'_{k,i} \\ \qquad\qquad \text{if } m \neq 0. \end{cases} \tag{172}$$

By the choice of the random sequence, when $p \neq k$, $s'_{p,i+m-m_p} s'_{k,i}$ are i.i.d. BPSK symbols, for all p and $i \in \mathcal{I}$. Moreover, when $m \neq 0$, $s'_{k,i+m} s'_{k,i}$ are i.i.d. BPSK symbols for all $i \in \mathcal{I}$. We will show that with a large enough of β_2, the error probability can be lower than $O(1/K^2)$.

Define a threshold

$$\bar{\mathcal{T}} = a|\mathcal{I}|/2. \tag{173}$$

If $|\mathcal{T}_k(0)| > \bar{\mathcal{T}}$ and $|\mathcal{T}_k(m)| < \bar{\mathcal{T}}$ for all $m = 1, \cdots, M$, the delay can be correctly estimated. Therefore, the error probability can be upper bounded as

$$P_e \leq \mathsf{P}\{|\mathcal{T}_k(0)| \leq \bar{\mathcal{T}}\} + \sum_{m=1}^{M} \mathsf{P}\{|\mathcal{T}_k(m)| \geq \bar{\mathcal{T}}\}. \tag{174}$$

For $m = 1, \cdots, M$, we have

$$\begin{aligned} &\mathsf{P}\left\{|\mathcal{T}_k(m)| \geq \bar{\mathcal{T}}\right\} \\ &\leq \mathsf{P}\left\{\left|\operatorname{Re}\{\mathcal{T}_k(m)\}\right| + \left|\operatorname{Im}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}\right\} \qquad (175) \\ &\leq \mathsf{P}\left\{\left|\operatorname{Re}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}/2\right\} \\ &\quad + \mathsf{P}\left\{\left|\operatorname{Im}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}/2\right\}. \qquad (176) \end{aligned}$$

We first derive the upper bound for the first term in (176). Let $r_{k,i}$ be random i.i.d. BPSK symbols for all $k \in \mathcal{K}$ and $i \in \mathcal{I}$. Let $z_i = \operatorname{Re}\{w_i\}$ be the real part of the random noise w_i. We have

$$\begin{aligned} &\mathsf{P}\left\{\left|\operatorname{Re}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}/2\right\} \\ &\leq \mathsf{P}\left\{\left|\sum_{p\in\mathcal{K}}\sum_{i\in\mathcal{I}} \operatorname{Re}\{a_p\} r_{p,i}\right| \geq \bar{\mathcal{T}}/4\right\} \\ &\quad + \mathsf{P}\left\{\left|\sum_{i\in\mathcal{I}} z_{i+m} s'_{k,i}\right| \geq \bar{\mathcal{T}}/4\right\}. \qquad (177) \end{aligned}$$

For the first term on the right hand side of (177), recall that a BPSK symbol is a 1-subGaussian with zero mean. According to Lemma 4 and Lemma 5 in Appendix A, $\sum_p \sum_{i\in\mathcal{I}} \operatorname{Re}\{a_p\} r_{p,i}$ is $\sqrt{\sum_p |\mathcal{I}|(\operatorname{Re}\{a_p\})^2}$-subGaussian variables with zero mean. By Lemma 6 in Appendix A, we have

$$\begin{aligned} &\mathsf{P}\left\{\left|\sum_{p\in\mathcal{K}}\sum_{i\in\mathcal{I}} \operatorname{Re}\{a_p\} r_{p,i}\right| \geq \bar{\mathcal{T}}/4\right\} \\ &\leq 2\exp\left(-\frac{\bar{\mathcal{T}}^2}{32|\mathcal{I}|\sum_p (\operatorname{Re}\{a_p\})^2}\right) \qquad (178) \\ &\leq 2\exp\left(-\frac{a^2|\mathcal{I}|}{128K\bar{a}^2}\right) \qquad (179) \\ &\leq 2\exp\left(-2\log(MK)\right) \qquad (180) \\ &\leq \frac{2}{MK^2}, \qquad (181) \end{aligned}$$

where (180) is due to (171) with $\beta_2 \geq 256\bar{a}^2/a^2$.

For the second term on the right hand side of (177), conditioned on $s'_{k,i}, k \in \mathcal{K}$ and $i \in \mathcal{I}$, the variables $z_{i+m} s'_{k,i}$ are i.i.d. Gaussian variables with zero mean and variance equal to σ^2. Therefore,

$$\begin{aligned} \mathsf{P}\left\{\left|\sum_{i\in\mathcal{I}} z_{i+m} s'_{k,i}\right| \geq \bar{\mathcal{T}}/4\right\} &= 2Q\left(\frac{\bar{\mathcal{T}}}{4\sqrt{|\mathcal{I}|\sigma^2}}\right) \qquad (182) \\ &\leq 2\exp\left(-\frac{\bar{\mathcal{T}}^2}{32|\mathcal{I}|\sigma^2}\right) \qquad (183) \\ &= 2\exp\left(-\frac{a^2|\mathcal{I}|}{128\sigma^2}\right) \qquad (184) \\ &\leq \frac{2}{MK^2}, \qquad (185) \end{aligned}$$

where (183) is due to $Q(x) \leq \exp(-x^2/2)$, and (185) follows from (171).

Thus, by (177), (181) and (185), we have

$$\mathsf{P}\left\{\left|\operatorname{Re}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}/2\right\} \leq \frac{4}{MK^2}. \tag{186}$$

Following the similar derivations, we can obtain $\mathsf{P}\left\{\left|\operatorname{Im}\{\mathcal{T}_k(m)\}\right| \geq \bar{\mathcal{T}}/2\right\} \leq 4/(MK^2)$. It follows that

$$\sum_{m=1}^{M} \mathsf{P}\left\{|\mathcal{T}_k(m)| \geq \bar{\mathcal{T}}\right\} \leq \frac{8}{K^2}. \tag{187}$$

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

For $m=0$, given (172) on $\mathcal{T}_k(m)$, we have

$$\begin{aligned}
&\mathsf{P}\{|\mathcal{T}_k(0)| \le \bar{\mathcal{T}}\} \\
&= \mathsf{P}\left\{\left||\mathcal{I}|a_k + \sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right| \le \bar{\mathcal{T}}\right\} && (188)\\
&\le \mathsf{P}\left\{|\mathcal{I}||a_k| - \left|\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right| \le \bar{\mathcal{T}}\right\} && (189)\\
&\le \mathsf{P}\left\{\left|\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right| \ge \bar{\mathcal{T}}\right\} && (190)\\
&\le \mathsf{P}\left\{\left|\mathrm{Re}\left\{\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right\}\right| \ge \frac{\bar{\mathcal{T}}}{2}\right\} \\
&\quad + \mathsf{P}\left\{\left|\mathrm{Im}\left\{\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right\}\right| \ge \frac{\bar{\mathcal{T}}}{2}\right\}, && (191)
\end{aligned}$$

where (190) is due to the assumption that $|\mathcal{I}||a_k| > 2\bar{\mathcal{T}}$.

Following the similar derivations for $\mathsf{P}\left\{\left|\mathrm{Re}\{\mathcal{T}_k(m)\}\right| \ge \bar{\mathcal{T}}/2\right\}, m=1,\cdots,M$, we have

$$\begin{aligned}
&\mathsf{P}\left\{\left|\mathrm{Re}\left\{\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right\}\right| \ge \frac{\bar{\mathcal{T}}}{2}\right\} \\
&\le \mathsf{P}\left\{\left|\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} \mathrm{Re}\{a_p\} r_{p,i}\right| \ge \frac{\bar{\mathcal{T}}}{4}\right\} \\
&\quad + \mathsf{P}\left\{\left|\sum_{i\in\mathcal{I}} z_i s'_{k,i}\right| \ge \frac{\bar{\mathcal{T}}}{4}\right\} && (192)\\
&\le \frac{4}{K^2}. && (193)
\end{aligned}$$

It can also be verified that $\mathsf{P}\left\{\left|\mathrm{Im}\left\{\sum_{p\in\mathcal{K}\setminus k}\sum_{i\in\mathcal{I}} a_p r_{p,i} + \sum_{i\in\mathcal{I}} w_i s'_{k,i}\right\}\right| \ge \frac{\bar{\mathcal{T}}}{2}\right\} \le 4/K^2$. We thus have

$$\mathsf{P}\{|\mathcal{T}_k(0)| \le \bar{\mathcal{T}}\} \le \frac{8}{K^2}. \quad (194)$$

Hence the proof of Lemma 3 by (174), (187) and (194). ■

APPENDIX G
ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their thorough and helpful reviews.

REFERENCES

[1] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the Internet of Things (IoT) forensics: challenges, approaches, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1191–1221, 2020.

[2] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Information Theory*, 2017, pp. 2523–2527.

[3] G. Liva and Y. Polyanskiy, "On coding techniques for unsourced multiple-access," in *Proc. 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1507–1514.

[4] D. Angelosante, E. Biglieri, and M. Lops, "Neighbor discovery in wireless networks: a multiuser-detection approach," *Physical Communication*, vol. 3, no. 1, pp. 28–36, 2010.

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

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

[7] L. Zhang, J. Luo, and D. Guo, "Neighbor discovery for wireless networks via compressed sensing," *Performance Evaluation*, vol. 70, no. 7, pp. 457–471, 2013.

[8] L. Zhang and D. Guo, "Virtual full duplex wireless broadcasting via compressed sensing," *IEEE/ACM Trans. Networking*, vol. 22, no. 5, pp. 1659–1671, 2014.

[9] L. Liu and W. Yu, "Massive connectivity with massive MIMO part I: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.

[10] A. Thompson and R. Calderbank, "Compressed neighbour discovery using sparse Kerdock matrices," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2286–2290.

[11] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.

[12] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, 2018.

[13] V. K. Amalladinne, K. R. Narayanan, J.-F. Chamberland, and D. Guo, "Asynchronous neighbor discovery using coupled compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4569–4573.

[14] K. D. Ba, P. Indyk, E. Price, and D. P. Woodruff, "Lower bounds for sparse recovery," in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2010, pp. 1190–1197.

[15] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2925–2951, 2021.

[16] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Physical Communication*, vol. 5, no. 2, pp. 129–147, 2012.

[17] X. Li, S. Pawar, and K. Ramchandran, "Sub-linear time compressed sensing using sparse-graph codes," in *Proc. IEEE Int. Symp. Information Theory*, Hong Kong, 2015, pp. 1645–1649.

[18] M. Bakshi, S. Jaggi, S. Cai, and M. Chen, "SHO-FA: Robust compressive sensing with order-optimal complexity, measurements, and bits," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7419–7444, 2016.

[19] W. Zeng, H. Wang, X. Wu, and H. Tian, "Sparse-graph codes and peeling decoder for compressed sensing," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 99, no. 9, pp. 1712–1716, 2016.

[20] X. Chen and D. Guo, "A generalized LDPC framework for sublinear compressive sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, March 2016.

[21] A. Vem, N. T. Janakiraman, and K. Narayanan, "Sub-linear time compressed sensing for support recovery using left and right regular sparse-graph codes," in *2016 IEEE Information Theory Workshop (ITW)*, 2016, pp. 429–433.

[22] S. Pawar and K. Ramchandran, "A hybrid DFT-LDPC framework for fast, efficient and robust compressive sensing," in *Proc. Annual Allerton Conference on Commun., Control, and Computing*, Monticello, IL, 2012, pp. 1943–1950.

[23] X. Li, D. Yin, S. Pawar, R. Pedarsani, and K. Ramchandran, "Sub-linear time support recovery for compressed sensing using sparse-graph codes," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6580–6619, 2019.

[24] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. Vehicular Technology Conference*, 2013, pp. 1–5.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

[25] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[26] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. Vehicular Technology Conference*, 2014, pp. 1–5.

[27] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels." *IEEE Trans. Information Theory*, vol. 63, no. 6, pp. 3516–3539, 2017.

[28] T.-Y. Chen, X. Chen, and D. Guo, "Many-broadcast channels: Definition and capacity in the degraded case," in *Proc. IEEE Int. Symp. Information Theory*, Honolulu, HI, June 2014, pp. 2569–2573.

[29] W. Yu, "On the fundamental limits of massive connectivity," in *Proc. Information Theory and Applications Workshop*, 2017, pp. 1–6.

[30] S. Shahi, D. Tuninetti, and N. Devroye, "The strongly asynchronous massive access channel," *arXiv preprint arXiv:1807.09934*, 2018.

[31] M. Ganji, X. Zou, and H. Jafarkhani, "Asynchronous transmission for multiple access channels: Rate-region analysis and system design for uplink NOMA," *IEEE Transactions on Wireless Communications*, 2021.

[32] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Information Theory*, 2017, pp. 2528–2532.

[33] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6509–6533, 2020.

[34] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Transactions on Information Theory*, vol. 67, no. 10, pp. 6894–6915, 2021.

[35] R. Calderbank and A. Thompson, "CHIRRUP: a practical algorithm for unsourced multiple access," *Information and Inference: A Journal of the IMA*, vol. 9, no. 4, pp. 875–897, 2020.

[36] A. Fengler, O. Musa, P. Jung, and G. Caire, "Pilot-based unsourced random access with a massive MIMO receiver, interference cancellation, and power control," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1522–1534, 2022.

[37] W. Zhu, M. Tao, X. Yuan, and Y. Guan, "Deep-learned approximate message passing for asynchronous massive connectivity," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5434–5448, 2021.

[38] S. Kim, H. Kim, H. Noh, Y. Kim, and D. Hong, "Novel transceiver architecture for an asynchronous grant-free IDMA system," *IEEE Transactions on Wireless Communications*, vol. 18, no. 9, pp. 4491–4504, 2019.

[39] S. Kim, J. Kim, and D. Hong, "A new non-orthogonal transceiver for asynchronous grant-free transmission systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1889–1902, 2020.

[40] W. Zhang, J. Li, X. Zhang, and S. Zhou, "A joint user activity detection and channel estimation scheme for packet-asynchronous grant-free access," *IEEE Wireless Communications Letters*, vol. 11, no. 2, pp. 338–342, 2021.

[41] K. Andreev, S. S. Kowshik, A. Frolov, and Y. Polyanskiy, "Low complexity energy efficient random access scheme for the asynchronous fading MAC," in *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–5.

[42] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, 2015.

[43] E. Paolini, C. Stefanovic, G. Liva, and P. Popovski, "Coded random access: Applying codes on graphs to design random access protocols," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 144–150, 2015.

[44] A. Taghavi, A. Vem, J.-F. Chamberland, and K. Narayanan, "On the design of universal schemes for massive uncoordinated multiple access," in *Proc. IEEE Int. Symp. Information Theory*, Barcelona, Spain, July 2016.

[45] R. De Gaudenzi, O. del Rio Herrero, G. Acar, and E. G. Barrabés, "Asynchronous contention resolution diversity ALOHA: Making CRDSA truly asynchronous," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 11, pp. 6193–6206, 2014.

[46] F. Clazzer, F. Lázaro, G. Liva, and M. Marchese, "Detection and combining techniques for asynchronous random access with time diversity," in *11th International ITG Conference on Systems, Communications and Coding*, 2017, pp. 1–6.

[47] E. Sandgren, A. G. i Amat, and F. Brännström, "On frame asynchronous coded slotted ALOHA: Asymptotic, finite length, and delay analysis," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 691–704, 2016.

[48] M. Shirvanimoghaddam, Y. Li, M. Dohler, B. Vucetic, and S. Feng, "Probabilistic rateless multiple access for machine-to-machine communication," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 12, pp. 6815–6826, 2015.

[49] J. Luo and D. Guo, "Neighbor discovery in wireless ad hoc networks based on group testing," in *Proc. Annual Allerton Conference on Commun., Control, and Computing*, Monticello, IL, 2008, pp. 791–797.

[50] A. Barg and G. Zémor, "Error exponents of expander codes under linear-complexity decoding," *SIAM J. on Discrete Math.*, vol. 17, no. 3, pp. 426–445, 2004.

[51] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[52] S. Pawar and K. Ramchandran, "A robust R-FFAST framework for computing a k-sparse n-length DFT in $\mathrm{O}(k \log n)$ sample complexity using sparse-graph codes," in *Proc. IEEE Int. Symp. Inform. Theory*, Honolulu, HI, June 2014, pp. 1852–1856.

[53] X. Li, J. K. Bradley, S. Pawar, and K. Ramchandran, "The SPRIGHT algorithm for robust sparse Hadamard transforms," in *Proc. IEEE Int. Symp. Inform. Theory*, Honolulu, HI, 2014, pp. 1857–1861.

[54] W. Feller, *An introduction to probability theory and its applications*. New York: John Wiley & Sons, Inc., 1957, vol. 1.

[55] M. Karoński and T. Łuczak, "The phase transition in a random hypergraph," *Journal of Computational and Applied Mathematics*, vol. 142, no. 1, pp. 125–135, 2002.

[56] J. Schmidt-Pruzan and E. Shamir, "Component structure in the evolution of random hypergraphs," *Combinatorica*, vol. 5, no. 1, pp. 81–94, 1985.

[57] E. Price, "Efficient sketches for the set query problem," in *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, California, 2011, pp. 41–56.

[58] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *The Annals of Statistics*, vol. 28, no. 5, pp. 1302–1338, 2000.

[59] O. Rivasplata, "Subgaussian random variables: an expository note," *Internet publication, PDF*, 2012.

[60] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-Gaussian concentration," *Electronic Communications in Probability*, vol. 18, pp. 1–9, 2013.

Xu Chen received the B.E. degree from Sun Yat-sen University, Guangzhou, China, the M.S. degree from Purdue University, West Lafayette, IN, and the Ph.D. degree from Northwestern University, Evanston, IL. From 2009 to 2011, he was a research associate in the Hong Kong Polytechnic University. From 2015 to 2016, he was working as co-founder in Calterah Semiconductor, Shanghai, China. He was a wireless system engineer at Apple in 2016-2018, and a staff perception engineer at Nio in 2018-2019. He received the best paper award in the 2011 International Conference on Advanced Technologies for Communications, the outstanding paper award from the 15th International Conference on Advanced Communication Technology in 2013, and a Best Paper Award at the 2017 IEEE Wireless Communications and Networking Conference. He is currently a staff software engineer at Waymo, Mountain View, CA.

Lina Liu received the B.E. degree in telecommunications engineering from Nanjing University, Nanjing, China, in June 2017 and the M.Phil. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, 2020. She is pursuing her Ph.D. degree at Northwestern University.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3224951

Dongning Guo is a Professor in the Department of Electrical and Computer Engineering at Northwestern University, Evanston, Illinois. He received the B.Eng. degree from the University of Science & Technology of China, Hefei, China, the M.Eng. degree from the National University of Singapore, Singapore, and the M.A. and Ph.D. degrees from Princeton University, Princeton, New Jersey. He was an R&D Engineer at the Center for Wireless Communications, Singapore, from 1998 to 1999. He has held visiting positions at the Norwegian University of Science and Technology in summer 2006, at the Chinese University of Hong Kong in 2010–2011, at Qualcomm New Jersey Research Center in 2011, and at MIT in 2014–2015. Dr. Guo has served on the editorial boards of the IEEE Transactions on Information Theory, the IEEE Transactions on Wireless Communications, and Foundations and Trends in Communications and Information Theory. He has also been a Guest Editor of the IEEE Journal on Selected Areas in Communications.

Dr. Guo is a recipient of the National Science Foundation Faculty Early Career Development (CAREER) Award in 2007. He is also a co-recipient of the IEEE Marconi Prize Paper Award in Wireless Communications in 2010 and the IEEE Wireless Communication and Networking Conference Best Paper Award in 2017. He was elected Fellow of IEEE in 2020.

Gregory W. Wornell (S'83-M'91-SM'00-F'04) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991 he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the Department of Electrical Engineering and Computer Science. At MIT he leads the Signals, Information, and Algorithms Laboratory within the Research Laboratory of Electronics. He is also chair of Graduate Area I (information and system science, electronic and photonic systems, physical science and nanotechnology, and bioelectrical science and engineering) within the EECS department's doctoral program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, the University of California, Berkeley, CA, and Hewlett-Packard Laboratories, Palo Alto, CA.

His research interests and publications span the areas of signal processing, information theory, digital communication, statistical inference, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the Information Theory and Signal Processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching, including the 2019 IEEE Leon K. Kirchmayer Graduate Teaching Award.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: MIT Libraries. Downloaded on December 07,2022 at 16:02:07 UTC from IEEE Xplore. Restrictions apply.