NORTHWESTERN UNIVERSITY

The Internet of Things: Fundamental Limits and Practical Algorithms

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Electrical Engineering and Computer Science

By

Xu Chen

EVANSTON, ILLINOIS

December 2016

ABSTRACT

The Internet of Things: Fundamental Limits and Practical Algorithms

Xu Chen

By 2020, there will be more than 200 billion sensor enabled objects world-wide in the Internet of Things (IoT). The biggest challenge of future IoT is to provide ultra-scalable wireless access for a massive number of devices. The goal of this thesis is to build up a model for systems with massive access, study the fundamental limits, and design practical signaling schemes and signal processing algorithms. The thesis mainly consists of two parts, which are presented in Chapter 2 and Chapter 3, respectively. Chapter 2 is devoted to the modeling of the IoT and the study of the fundamental limits from the perspective of information theory. Chapter 3 is devoted to the design of low-complexity practical signal processing algorithms for neighbor discovery.

Classical multiuser information theory studies the fundamental limits of models with a fixed (often small) number of users as the coding blocklength goes to infinity. In Chapter 2, we introduce a new many-user paradigm, where the number of users and the blocklength simultaneously tend to infinity. This paradigm is motivated by emerging systems whose

massive number of users is comparable or far exceeds the blocklength, such as in machine-to-machine communication systems and the IoT. The focus of the thesis is the Gaussian many-access channel, which is used to model the uplink transmission of the IoT. The many-access channel consists of a single receiver and many transmitters with fixed power, where all or a subset of users may transmit in a given block and need to be identified. The conventional notion of capacity in bits per channel use is ill-suited for the task, as Cover and Thomas recognized that the rate per sender vanishes. A new notion of capacity is introduced and characterized for the Gaussian many-access channel. The capacity can be achieved by first detecting the set of active users and then decoding their messages.

To achieve the capacity of the many-access channels, an essential step is device identification, also known as neighbor discovery. In wireless neighbor discovery, an access point needs to identify all the active devices in its surrounding areas. In Chapter 3, a novel low-complexity wireless neighbor discovery scheme, referred to as sparse orthogonal frequency division multiplexing (sparse-OFDM) is proposed. In the IoT, the number of devices is very large while every device accesses the network with a small probability, so the number of active devices in a frame is much smaller than the total local device population. Sparse OFDM is a one-shot transmission scheme with low complexity, which exploits both the parallel channel access offered by OFDM and the bursty nature of transmissions. The scheme is inspired by the sublinear algorithms for computing sparse Fourier transform and compressive sensing. When the transmission delay of each device is an integer number of symbol intervals, analysis and simulation show that sparse OFDM enables successful asynchronous neighbor discovery using a much smaller transmission length than the random access schemes.

Acknowledgements

At the very beginning, I would like to express my sincere gratitude to my advisor, Professor Dongning Guo, for his inspiring discussions, invaluable advice and continuous support during the course of my Ph.D. study. His enthusiasm for research and endless dedication to every piece of our research work has taught me how to become a truly scholar. I also greatly appreciate that Professor Guo allows me to work on my startup in the last year of my PhD, which turns out to be invaluable experience in my life.

I would like to thank Professor Randall Berry and Professor Michael Honig, for serving in my thesis committee and giving me insightful comments. I have gained a lot by taking their courses at Northwestern University.

I am indebted to many colleagues in the Communications and Networking Laboratory at Northwestern University: Binnan Zhuang, Ka Hung Hui, Suvarup Saha, Fei Teng, Khalid Zeineddine, Tsung-yi Chen, Cheng Chen, Chang Liu, Xu Wang, Zhiyi Zhou and Hao Ge. I am also grateful to meet Yin Xia, Jiang Wang, Huanyu Cheng, Dajun Yue, Yang Yang and Yu Cheng during my Ph.D. study. Our precious friendship will continue after school. My Ph.D. life will not be so much fun without you all.

Most importantly, I want to express my deep gratitude to my parents and my wife, Wei Zhang. Their selfless love is always my best support. To them I dedicate this dissertation.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Background

The Internet of Things (IoT) is becoming a major growth area that touches all consumers and all sectors of the economy. It is believed that over 200 billion devices will be connected by year 2020 and the number will continue to increase in the foreseeable future.

The IoT communication will be a key enabler of a wide range of new applications in healthcare, transportation systems, smart grid, smart home, smart city, and public safety, to name a few. In recent years, the wireless industry have become increasingly interested in the IoT services as the next major source of revenue as the technology and market for smart phones mature.

The IoT has unique features that is very different from that of cellular network [**1**]. First, there are a massive number of devices in the system. A majority of the IoT devices are expected to be wireless and their density will be a few orders of magnitude higher than smart phones. Second, each device has a random on-off activity and the transmitted message length is usually short. Third, there is usually much more uplink traffic than downlink. These unique features pose new challenges of modeling the system and providing ultra-scalable wireless access for the IoT. This motivates the study of this thesis on two important questions:

(1) How should we model the IoT to incorporate its unique features and what is the fundamental limit of the system?

(2) How do we design practical signal processing algorithms to approach the fundamental limit? In particular, the algorithms should be of low-complexity and is scalable for the massive number of devices.

1.2. Modeling of the IoT

Classical information theory characterizes the fundamental limits of communication systems by studying the asymptotic regime of infinite coding blocklength. The prevailing models in multiuser information theory assume a fixed (usually small) number of users, where fundamental limits as the coding blocklength goes to infinity are studied. Even in the large-system analysis of multiuser systems [**2**–**4**], the blocklength is sent to infinity before the number of users is sent to infinity. In some sensor networks and emerging machine-to-machine communication systems, a massive and ever-increasing number of wireless devices may need to share the spectrum in a given area. This motivates us to rethink the assumption of fixed user population. Here we propose a new *many-user paradigm*, where the number of users is allowed to increase without bound with the blocklength.

In general, the theory that assumes a fixed number of users does not apply to systems where the number of users is comparable or even larger than the blocklength. For example, dense sensor networks, the IoT or machine-to-machine communication systems with many thousands of devices in a given cell. A key reason is that for many functions $f(k, n)$, letting $k \to \infty$ after $n \to \infty$ may yield a different result than letting n and $k = k_n$ (as a function

of n) simultaneously tend to infinity,[1] i.e.,

$$\lim_{k\to\infty}\lim_{n\to\infty} f(k,n) \neq \lim_{n\to\infty} f(k_n,n)\,. \tag{1.1}$$

This new paradigm in multiuser information theory models where k_n can grow arbitrarily large with n is referred to as the many-user regime.

One motivating example is the design of ultra-scalable IoT communication systems where the number of users k is comparable or even larger than the blocklength n, and the message transmitted to each user could be very short. The many-user regime therefore becomes a better performance indicator in the context of IoT where $k_n = O(n)$ and the number of bits to be transmitted for each user may be sublinear in n. We are interested in the fundamental limits in this regime.

In this thesis, we propose a Gaussian many-access channel to model the uplink transmission in the IoT [**5**]. The many-access channel consists of a single receiver and many transmitters with per user power constraint, where all or a subset of users may transmit in a given block and need to be identified. The rate for each user measured in bit per channel use vanishes as k_n grows, indicating that the traditional notion of capacity becomes ill-suited for the task. A new notion of capacity is introduced and characterized for the Gaussian many-access channel. The capacity can be achieved by first detecting the set of active users and then decoding their messages.

A many-broadcast channel has also been proposed to model the downlink transmission in the IoT. The many-broadcast channel consists of a single transmitter with fixed power,

[1]Take the function $f(n,k) = \log(1+k/n)$ as an example. Taking the limits separately gives 0 or ∞ while taking the limit simultaneously with $k_n = n$ yields $\lim_{n\to\infty} f(n,n) = \log 2$.

and the number of receivers grows as the blocklength [**6**]. It has been shown that typicality set decoding can only achieve the reliable communication if the number of receivers grows sublinearly with the blocklength. The maximum-likelihood decoding is critical in the many-user regime to achieve reliable communication when the number of users grows linearly with the blocklength. The result was published in other papers and is not reported in the thesis.

1.3. Practical Neighbor Discovery Algorithms for the IoT

Before any data communication takes place in a wireless network, an access point needs to first identify all active devices in its surrounding areas (aka its neighborhood). The identity of a node is usually equivalent to its network interface address (NIA). The goal of wireless device identification, also known as *neighbor discovery*, is to obtain the NIAs of all active neighbors. For example, if the NIA is represented by 20 bits, and every NIA is valid, then there are potentially $2^{20} \approx 1$ million identities. If there are 4 billion (distinct) identities, then the NIA is represented by at least 32 bits (like the IP address).

The study of the fundamental limit of Gaussian many-access channel shows that reliable neighbor discovery is an essential step to achieve the capacity. In the IoT, many devices are of low power and low cost. A practical neighbor discovery scheme should involve low complexity on the device side. In this thesis, we aim to propose a low-complexity neighbor discovery algorithm by exploiting the bursty transmission nature.

Assuming the usual radio frequency processing and (discrete-time) sampling, the baseband model for device identification can be described as follows. Suppose there are N possible identities in total, which can also be viewed as the cardinality of the entire space

of valid NIAs. (N is typically an extremely large number. The actual number of devices is usually smaller.) Device j is identified by its unique signature consisting of L symbols, denoted as $\boldsymbol{s}_j = [s_{j,0}, \cdots, s_{j,L-1}]$, where $s_{j,i} = 0$ for $i < 0$ or $i \geq L$. The device transmits the signature during the discovery period in order to be identified.[2] Assume that device j transmits the signature $\boldsymbol{s}_j$ with delay of m_j symbols relative to the beginning of the period.

Consider the neighborhood of a receiver (e.g., an access point). Suppose K devices are in the neighborhood. Let a_j denote the channel coefficient between device j and the receiver. If device j is not in the neighborhood (or does not exist), then $a_j = 0$, whereas if device j is in the neighborhood, $a_j \neq 0$ represents the channel coefficient, which includes the effects of large-scale fading (aka path loss) as well as small-scale block fading. The received signal at time i can be expressed as

$$y_i = \sum_{j=1}^{N} a_j s_{j,i-m_j} + z_i \tag{1.2}$$

where z_i denotes additive white Gaussian noise. The task of the receiver is to identify which (K) coefficients a_j are nonzero out of a very large number (N) of them based on the observation through the linear model (1.2).

Conventionally, device identification is through contention-based random access. Namely, each device repeatedly sends its identity packets enough times with random delays, so that the access point will receive at least one of those packets free of collision. Due to contention, many retransmissions are needed to guarantee reliable identification.

[2]If a device repeats a certain signal (e.g., its identity itself), the entire transmission is regarded as the signature.

Contention-based identification has been studied by many authors (e.g. [**7**,**8**]). Such a scheme is sufficient for cellular networks and local area networks, because identification is carried out infrequently and a device, once identified, typically transmits a large amount of data. On the contrary, in the IoT with a massive number of devices, where each device typically has very small amount of data to transmit, the overhead becomes significant, and scalability of the scheme becomes crucial.

It has also been proposed to accommodate IoT devices in the LTE family standards. However, the number of IoT devices and the aggregate traffic, especially in the uplink, will easily outgrow the amount of resources reserved for contentions in LTE and LTE-A [**9**]. An alternative ultra-scalable wireless access technique is absolutely needed to facilitate future growth of the IoT.

Consider the model described by (1.2). Suppose the number of active devices K is much smaller than the population size N, which is usually the case in the IoT. The neighbor discovery problem boils down to recovering the support of an extremely sparse vector: $\boldsymbol{a} = [a_0, \cdots, a_{N-1}]^T$. If the delays m_j are 0 for all devices, the problem clearly belongs to a class of problems commonly referred to as *compressed sensing* (also known as *sparse recovery*) [**10**]. The wisdom of compressed sensing is that, due to the sparse nature of the desired vector, usually the number of measurements (or symbols y_i here) needed to recover $\boldsymbol{a}$ can be much smaller than the dimensionality (N) of the unknown signal.

Based on the insights from compressed sensing and multiuser detection, we believe the key to ultra-scalability is to allow many nodes to transmit simultaneously to an access point, and be decoded and identified jointly by the access point. In fact, synchronous neighbor discovery schemes inspired by compressed sensing have been proposed [**10**]. It

has been shown to have significant throughput gain over contention-based random access schemes (especially if the overhead of framing and feedback is also accounted for). Asynchronous neighbor discovery has also been studied in [**11**], but it involves a complexity that grows polynomial in the total device population N.

In this thesis, we propose a neighbor discovery scheme that achieves three objectives: 1) small transmission length, 2) low computational complexity, and 3) reliable detection for asynchronous transmssion. A novel low-complexity wireless neighbor discovery scheme, referred to as sparse orthogonal frequency division multiplexing (sparse-OFDM) is proposed. By judiciously designing the codeword structure, sparse OFDM relates the neighbor discovery algorithm to sparse Fourier transform. Thus, the recently proposed sublinear algorithms can be leveraged [**12**]. Compared with the random access schemes, sparse OFDM requires much shorter transmission length. Sparse OFDM adopts well-established point-to-point capacity approaching codes and involves low complexity. It provides practical physical layer capability for multipacket reception.

CHAPTER 2

Gaussian Many-Access Channel

2.1. Introduction

Classical information theory characterizes the fundamental limits of communication systems by studying the asymptotic regime of infinite coding blocklength. The prevailing models in multiuser information theory assume a fixed (usually small) number of users, where fundamental limits as the coding blocklength goes to infinity are studied. Even in the large-system analysis of multiuser systems [**2**–**4**], the blocklength is sent to infinity before the number of users is sent to infinity.[1] In some sensor networks and emerging machine-to-machine communication systems, a massive and ever-increasing number of wireless devices with bursty traffic may need to share the spectrum in a given area. This motivates us to rethink the assumption of fixed population of fully buffered users. Here we propose a new *many-user paradigm*, where the number of users is allowed to increase without bound with the blocklength.[2]

In this chapter, we introduce the many-access channel (MnAC) to model systems consisting of a single receiver and many transmitters, the number of which is comparable to or even larger than the blocklength [**16**, **17**]. We study the asymptotic regime where the number of transmitting devices (k) increases with the blocklength (n). The model

[1]The same can be said of the many-user broadcast coding strategy for the point-to-point channel proposed in [**13**], and the CEO problem [**14**].

[2]The only existing model of this nature is found in [**15**], in which the authors sought for uniquely-decodable codes for a noiseless binary adder channel with the number of users increasing with the blocklength.

also accommodates random access, namely, it allows each transmitter to be active with certain probability in each block. We assume synchronous transmission in the model, while the capacity of strong asynchronous MnAC was studied in [**18**].

In general, the classical theory does not apply to systems where the number of users is comparable or larger than the blocklength, such as in a machine-to-machine communication system with many thousands of devices in a given cell. One key reason is that, for many functions of two variables f, $\lim_{k\to\infty}\lim_{n\to\infty} f(k,n) \neq \lim_{n\to\infty} f(k_n,n)$, i.e., letting $k\to\infty$ after $n\to\infty$ may yield a different result than letting n and $k=k_n$ (as a function of n) simultaneously tend to infinity. Moreover, the traditional notion of rate in bits per channel use is ill-suited for the task in the many-user regime as noted (for the Gaussian multiaccess channel) in [**19**, pp. 546-547] by Cover and Thomas, "when the total number of senders is very large, so that there is a lot of interference, we can still send a total amount of information that is arbitrary large even though the rate per individual sender goes to 0."

Capacity of the conventional multiaccess channel is well understood [**20**–**22**]. The capacity can be established using the fact that joint typicality holds with high probability as the blocklength grows to infinity. This argument, however, does not directly apply to models where the number of users also goes to infinity. Specifically, joint typicality requires the simultaneous convergence of the empirical joint entropy of every subset of the input and output random variables to the corresponding joint entropy. Even though convergence holds for every subset due to the law of large numbers, the asymptotic equipartition property is not guaranteed because the number of those subsets increases exponentially with the number of users [**6**]. Resorting to strong typicality does not resolve this because

the empirical distribution over an increasing alphabet (due to increasing number of users) does not converge.

In general, the received signal of the Gaussian MnAC is a noisy superposition of the codewords chosen by the active users from their respective codebooks. The detection problem boils down to identifying codewords based on their superposition. It is closely related to sparse recovery, also known as compressed sensing, which has been studied in a large body of works [**23**–**32**]. Information-theoretic limits of exact support recovery was considered in [**26**], and stronger necessary and sufficient conditions have been derived subsequently [**28**, **29**, **32**]. Using existing results in the sparse recovery literature, it can be shown that the message length (in bits) that can be transmitted reliably by each user should be in the order of $\Theta(n(\log k_n)/k_n)$.

In this work, we provide a sharp characterization of the capacity of Gaussian many-access channels as well as the user identification cost. As an achievable scheme, each user's transmission consists of a signature that identifies the user, followed by a message-bearing codeword. The decoder first identifies the set of active users based on the superposition of their unique signatures. (This is in fact a compressed sensing problem [**10**, **33**].) It then decodes the messages from the identified active users. The length of the signature matches the capacity penalty due to user activity uncertainty. The proof techniques find their roots in Gallager's error exponent analysis [**34**]. Also studied is a more general setup where groups of users have heterogeneous channel gains and activity patterns. Again, separate identification and decoding is shown to achieve the capacity region.

Unless otherwise noted, we use the following notational conventions: x denotes a scalar, $\boldsymbol{x}$ denotes a column vector, and $\underline{\boldsymbol{x}}$ denotes a matrix. The corresponding uppercase

letters X, $\boldsymbol{X}$, and $\underline{\boldsymbol{X}}$ denote the corresponding random scalar, random vector and random matrix, respectively. Given a set A, let $\boldsymbol{x}_A = (x_i)_{i\in A}$ denote the subset of variables of $\boldsymbol{x}$ whose indices are in A and let $\underline{\boldsymbol{x}}_A = (\boldsymbol{x}_i)_{i\in A}$ be the matrix formed by columns of $\underline{\boldsymbol{x}}$ whose indices are in A. Let $x_n \leq_n y_n$ denote $\limsup_{n\to\infty}(x_n - y_n) \leq 0$. That is, x_n is essentially asymptotically dominated by y_n. All logarithms are natural. The binary entropy function is denoted as $H_2(p) = -p\log p - (1-p)\log(1-p)$.

The rest of the chapter is organized as follows. Section 2.2 presents the system model and main capacity results. Section 2.3 gives the proof of converse for the MnAC capacity. Section 2.4 proves the random user identification cost. Section 2.5 shows that the MnAC capacity is achievable using separate identification and decoding. Section 2.6 discusses the challenges of applying successive decoding in MnAC. Section 2.7 analyzes the capacity of MnAC with heterogeneous channel gains and activity patterns. Concluding remarks are given in Section 3.8.

2.2. System Model and Main Results

Let n denote the number of channel uses, i.e., the blocklength. Let the number of users be a function of n and be explicitly denoted as ℓ_n, so that it is tied to the blocklength. The received symbols in a block form a column vector of length n:

$$\boldsymbol{Y} = \sum_{k=1}^{\ell_n} \boldsymbol{S}_k(w_k) + \boldsymbol{Z} \tag{2.1}$$

where w_k is the message of user k, $\boldsymbol{S}_k(w_k) \in \mathbb{R}^n$ is the corresponding n-symbol codeword, and $\boldsymbol{Z}$ is a Gaussian noise vector with independent standard Gaussian entries. Suppose each user accesses the channel independently with identical probability α_n during any

given block. If user k is inactive, it is thought of as transmitting the all-zero codeword $\boldsymbol{S}_k(0) = \mathbf{0}$.

Definition 1. *Let $\mathcal{S}_k$ and $\mathcal{Y}$ denote the input alphabet of user k and the output alphabet, respectively. An (M, n) symmetric code with power constraint P for the MnAC channel $(\mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_{\ell_n}, p_{Y|S_1,\cdots,S_{\ell_n}}, \mathcal{Y})$ consists of the following mappings:*

(1) *The encoding functions $\mathcal{E}_k : \{0, 1, \ldots, M\} \to \mathcal{S}_k^n$ for every user $k \in \{1, \cdots, \ell_n\}$, which maps any message w to the codeword $\boldsymbol{s}_k(w) = [s_{k1}(w), \cdots, s_{kn}(w)]^T$. In particular, $\boldsymbol{s}_k(0) = \mathbf{0}$, for every k. Every codeword $\boldsymbol{s}_k(w)$ satisfies the power constraint:*

$$\frac{1}{n}\sum_{i=1}^{n} s_{ki}^2(w) \leq P. \tag{2.2}$$

(2) *Decoding function $\mathcal{D} : \mathcal{Y}^n \to \{0, 1, \ldots, M\}^{\ell_n}$, which is a deterministic rule assigning a decision on the messages to each possible received vector.*

The average error probability of the (M, n) code is:

$$\mathsf{P}_e^{(n)} = \mathsf{P}\left\{\mathcal{D}(\boldsymbol{Y}) \neq (W_1, \ldots, W_{\ell_n})\right\}, \tag{2.3}$$

where $W_1, \cdots, W_{\ell_n}$ are independent, and for every $k \in \{1, \cdots, \ell_n\}$,

$$\mathsf{P}\left\{W_k = w\right\} = \begin{cases} 1 - \alpha_n, & w = 0, \\ \frac{\alpha_n}{M}, & w \in \{1, \ldots, M\}. \end{cases} \tag{2.4}$$

The preceding model reduces to the conventional ℓ-user multiaccess channel in the special case where $\ell_n = \ell$ is fixed and $\alpha_n = 1$ as the blocklength n varies.

2.2.1. The Message-Length Capacity

Definition 2 (Asymptotically achievable message length)**.** *We say a positive nondecreasing sequence of message lengths* $\{v(n)\}_{n=1}^{\infty}$*, or simply,* $v(\cdot)$*, is asymptotically achievable for the MnAC if there exists a sequence of* $(\lceil \exp(v(n)) \rceil, n)$ *codes according to Definition 1 such that the average error probability* $\mathsf{P}_e^{(n)}$ *given by* (2.3) *vanishes as* $n \to \infty$*.*

It should be clear that by asymptotically achievable message length we really mean a function of the blocklength. The base of $\exp(\cdot)$ should be consistent with the unit of the message length. If the base of $\exp(\cdot)$ is 2 (resp. e), then the message length is measured in bits (resp. nats).

Definition 3 (Symmetric message-length capacity)**.** *For the MnAC channel described by* (2.1)*, a positive nondecreasing function* $B(n)$ *of the blocklength* n *is said to be the symmetric message-length capacity of the MnAC channel if, for any* $0 < \epsilon < 1$*,* $(1-\epsilon)B(n)$ *is an asymptotically symmetric achievable message length, whereas* $(1+\epsilon)B(n)$ *is not.*

For the special case of a (conventional) multiaccess channel, the symmetric capacity $B(n)$ in Definition 3 is asymptotically linear in n, so that $\lim_{n\to\infty} B(n)/n$ is equal to the symmetric capacity of the multiaccess channel (in, e.g., bits per channel use). From this point on, by "capacity" we mean the message-length capacity in contrast to the conventional capacity. In many-user information theory, $B(n)$ need not grow linearly with the blocklength.

Let $\underline{\boldsymbol{S}}_k = [\boldsymbol{S}_k(1), \cdots, \boldsymbol{S}_k(M)]$ denote the matrix consisting of all but the first all-zero codeword of user k. Let $\underline{\boldsymbol{S}} = [\underline{\boldsymbol{S}}_1, \cdots, \underline{\boldsymbol{S}}_{\ell_n}] \in \mathbb{R}^{n\times(M\ell_n)}$ denote the concatenation of the codebooks of all users. For ease of analysis, we often use the following equivalent model

for the Gaussian MnAC (2.1):

$$\boldsymbol{Y} = \underline{\boldsymbol{S}}\boldsymbol{X} + \boldsymbol{Z}, \tag{2.5}$$

where $\boldsymbol{Z}$ is defined as in (2.1) and $\boldsymbol{X} \in \mathbb{R}^{M\ell_n}$ is a vector indicating the codewords transmitted by the users. Specifically, $\boldsymbol{X} = [\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \cdots, \boldsymbol{X}_{\ell_n}^T]^T$, where $\boldsymbol{X}_k \in \mathbb{R}^M$ indicates the codeword transmitted by user k, $k = 1, \cdots, \ell_n$, i.e.,

$$\boldsymbol{X}_k = \begin{cases} \boldsymbol{0} & \text{with probability } 1 - \alpha_n \\ \boldsymbol{e}_m & \text{with probability } \frac{\alpha_n}{M}, \;\; m = 1, \ldots, M \end{cases} \tag{2.6}$$

where $\boldsymbol{e}_m$ is the binary column M-vector with a single 1 at the m-th entry. Let

$$\mathcal{X}_m^\ell = \left\{ \boldsymbol{x} = \left[\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_\ell^T\right]^T : \boldsymbol{x}_i \in \{\boldsymbol{0}, \boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}, \text{for every } i \in \{1, \cdots, \ell\} \right\}. \tag{2.7}$$

The signal $\boldsymbol{X}$ must take its values in $\mathcal{X}_M^{\ell_n}$.

The following theorem is a main result of the work.

Theorem 1 (Symmetric capacity of the Gaussian many-access channel)**.** *Let n denote the coding blocklength, ℓ_n denote the total number of users, and α_n denote the probability a user is active, independent of other users. Suppose ℓ_n is nondecreasing with n and*

$$\lim_{n\to\infty} \alpha_n = \alpha \in [0, 1]. \tag{2.8}$$

Denote the average number of active users as

$$k_n = \alpha_n \ell_n. \tag{2.9}$$

Then the symmetric message-length capacity $B(n)$ of the Gaussian many-access channel described by (2.1), with each user's SNR being no greater than P, is characterized as follows:

1) Suppose ℓ_n and k_n are both unbounded, $k_n = O(n)$, and

$$\ell_n e^{-\delta k_n} \to 0 \tag{2.10}$$

for every $\delta > 0$. Let θ denote the limit of

$$\theta_n = \frac{2\ell_n H_2(\alpha_n)}{n \log(1 + k_n P)}, \tag{2.11}$$

which may be ∞.

If $\theta < 1$, then

$$B(n) = \frac{n}{2k_n} \log(1 + k_n P) - \frac{H_2(\alpha_n)}{\alpha_n}. \tag{2.12}$$

If $\theta > 1$, then a user cannot send even 1 bit reliably.

If $\theta = 1$, then message length $\frac{\epsilon n}{2k_n} \log(1 + k_n P)$ is not achievable for any $\epsilon > 0$.

2) If ℓ_n is unbounded and k_n is bounded, then message length ϵn is not achievable for any $\epsilon > 0$.

3) If ℓ_n is bounded, i.e., $\ell_n = \ell < \infty$ for sufficiently large n, then

$$B(n) = \begin{cases} \frac{n}{2} \log(1 + P) & \text{if } \alpha = 0, \\ \frac{n}{2\ell} \log(1 + \ell P) & \text{if } \alpha > 0. \end{cases} \tag{2.13}$$

A heuristic understanding of the expression of $B(n)$ in (2.12) is as follows: If a genie-aided receiver revealed the set of active users to the receiver, the total number of bits that can be communicated through the MnAC with k_n users would be approximately $(n/2)\log(1+k_nP)$, so that the symmetric capacity is

$$B_1(n) = \frac{n}{2k_n}\log(1+k_nP). \tag{2.14}$$

The total uncertainty in the activity of all ℓ_n users is $\ell_n H_2(\alpha_n) = k_n H_2(\alpha_n)/\alpha_n$, so the capacity penalty on each of the k_n active users is $H_2(\alpha_n)/\alpha_n$. If every user is always active, i.e., $\alpha_n = 1$, the penalty term is zero and the capacity resembles that of a multiaccess channel.

By the current definition, the symmetric capacity (2.12) can be reduced to

$$B'(n) = \frac{n}{2k_n}\log k_n - \frac{H_2(\alpha_n)}{\alpha_n}, \tag{2.15}$$

because $\log(1+k_nP) = \log k_n + o(\log k_n)$. We prefer the form of (2.12) for its connection to the original capacity formular for the Gaussian multiaccess channel.

Fig. 2.1 illustrates the capacity $B(n)$ given by (2.12) in the special case where $P = 10$ (i.e., the SNR is 10 dB), $k_n = n/4$, with different scalings of user number ℓ_n. The purpose is to show the trend of the capacity as the blocklength increases rather than the capacity at finite length. The message-length capacity $B(n)$ scales sub-linearly in n. Moreover, $B(n)$ depends on the scaling of k_n and ℓ_n, whose effects cannot be captured by the conventional multiaccess channels. In particular, if ℓ_n grows too quickly (e.g., $\ell_n = n^3$), an average user cannot transmit a single bit reliably.

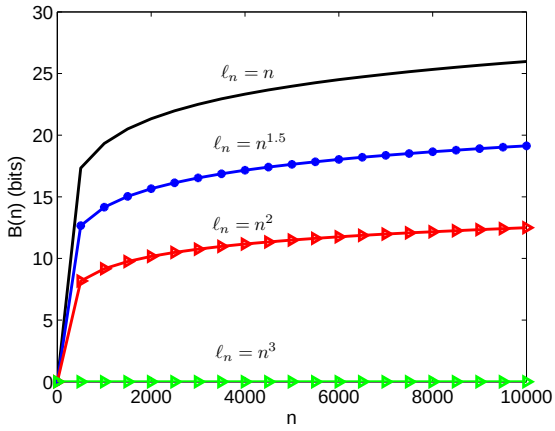

Figure 2.1. Plot of $B(n)$ given by (2.12), where $P = 10$, $k_n = n/4$.

The assumptions in Case 1) of Theorem 1 prohibit two uninteresting cases: i) The average number of active users k_n grows faster than linear in the blocklength n; and ii) the total number of users ℓ_n grows exponentially in n. For example, if $k_n = n(\log n)^2$, an average user will not be able to transmit a single bit reliably as n increases to infinity.

Time sharing with power allocation, which can achieve the capacity of the conventional multiaccess channel [**19**], is inadequate for the MnAC in general. For example, if $k_n = 2n$, not a single channel use can be guaranteed for every active user. Moreover, if $k_n = n$ and each user applies all energy in a single exclusive channel use, the resulting data rate is generally poor.

2.2.2. The User Identification Cost

As a by-product in the proof of Theorem 1, we can derive the fundamental limits of random user identification (without data transmission), where every user is active with certain probability and the receiver aims to detect the set of active users. To quantify the cost of user identification, we denote the total number of users as ℓ and let other parameters

depend on ℓ. (This is in contrast to the setting in Section 2.2.1.) The probability of a user being active is denoted as α_ℓ, and the average number of active users is denoted as $k_\ell = \alpha_\ell \ell$. Suppose n_0 symbols are used for user identification purpose. Let $\boldsymbol{X}^a \in \mathbb{R}^\ell$ be a random vector, which consists of independent and identically distributed (i.i.d.) Bernoulli entries with mean α_ℓ. Then the received signal is

$$\boldsymbol{Y}^a = \underline{\boldsymbol{S}}^a \boldsymbol{X}^a + \boldsymbol{Z}^a, \tag{2.16}$$

where $\boldsymbol{Z}^a$ consists of n_0 i.i.d. standard Gaussian entries, and $\underline{\boldsymbol{S}}^a = [\boldsymbol{S}_1^a \cdots, \boldsymbol{S}_\ell^a]$ with $\boldsymbol{S}_k^a \in \mathbb{R}^{n_0}$ being the signature of user k. Moreover, the realization of the signature must satisfy the following power constraint:

$$\frac{1}{n_0} \sum_{i=1}^{n_0} (\boldsymbol{s}_{ki}^a)^2 \leq P. \tag{2.17}$$

Definition 4 (Minimum user identification cost)**.** *We say the identification is erroneous in case of any miss or false alarm. For the channel described by* (2.16)*, the minimum user identification cost is said to be $n(\ell)$ if $n(\ell) > 0$ and for every $0 < \epsilon < 1$, the probability of erroneous identification vanishes as $\ell \to \infty$ if the signature length $n_0 = (1+\epsilon)n(\ell)$, whereas the error probability is strictly bounded away from zero if $n_0 = (1-\epsilon)n(\ell)$.*

As in the case of capacity, the definition focuses on the asymptotics of $\ell \to \infty$, so the minimum cost function $n(\cdot)$ is not unique. The random user identification problem has been studied in the context of compressed sensing problem [**26**, **35**]. The following theorem gives a sharp characterization of how many channel uses n_0 are needed for reliable identification.

Theorem 2 (Minimum identification cost through the Gaussian many-access channel)**.** *Let the total number of users be ℓ, where each user is active with the same probability. Suppose the average number of active users k_ℓ satisfies*

$$\lim_{\ell\to\infty} \ell e^{-\delta k_\ell} = 0 \tag{2.18}$$

for every $\delta > 0$. Let

$$n(\ell) = \frac{\ell H_2(k_\ell/\ell)}{\frac{1}{2}\log(1+k_\ell P)}. \tag{2.19}$$

Suppose $n(\ell)/k_\ell$ has finite limit or diverges to infinity. The asymptotic identification cost is characterized as follows:

1) If $\lim_{k_\ell\to\infty} n(\ell)/k_\ell > 0$, then the minimum user identification cost is $n(\ell)$.

2) If $\lim_{k_\ell\to\infty} n(\ell)/k_\ell = 0$, then a signature length of $n_0 = \epsilon k_\ell$ yields vanishing error probability for any $\epsilon > 0$; on the other hand, if $n_0 \le (1-\epsilon)n(\ell)$, then the identification error cannot vanish as $\ell \to \infty$.

Note that (2.18) implies $k_\ell \to \infty$ as $\ell \to \infty$. In the special case where $k_\ell = \lceil \ell^{1/d} \rceil$ for some $d > 1$, the minimum user identification cost is $n(\ell) = 2(d-1)k_\ell + o(k_\ell)$, which is linear in the number of active users. The minimum cost function $n(\ell)$ is illustrated in Fig. 2.2.

In the following, we first prove the converse of Theorem 1, which can be particularized to prove the converse of Theorem 2. Then we prove the achievability of Theorem 2, which is an essential step leading to the achievability of Theorem 1 eventually.

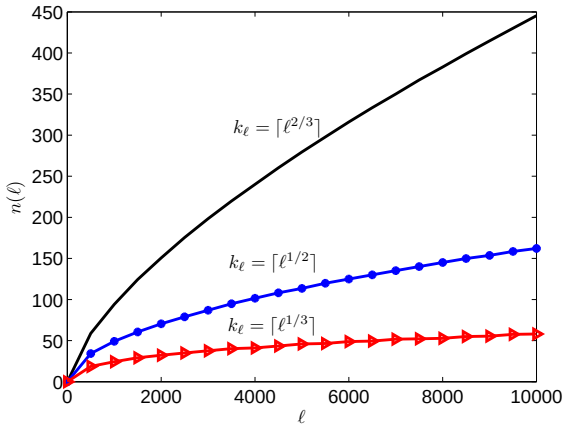

Figure 2.2. Plot of $n(\ell)$ specified in Theorem 2, where $P = 10$, i.e., SNR = 10 dB.

2.3. Proof of the Converse of Theorem 1

We prove the converse for the three cases in Theorem 1, respectively.

2.3.1. Converse for Case 1): unbounded ℓ_n and unbounded k_n

This proof requires more work than a straightforward use of Fano's inequality, because the size of the joint input alphabet may increase rapidly with the blocklength. To overcome this difficulty, define for every given $\delta \in (0, 1)$,

$$\mathcal{B}_m^\ell(\delta, k) = \left\{ \boldsymbol{x} \in \mathcal{X}_m^\ell : 1 \leq ||\boldsymbol{x}||_0 \leq (1+\delta)k \right\}, \tag{2.20}$$

which can be thought of as an ℓ_0 ball but the origin. Since $\boldsymbol{X}$ in (2.5) is a binary vector, whose expected support size is k_n, it is found in $\mathcal{B}_M^{\ell_n}(\delta, k_n)$ with high probability for large n.

Based on the input distribution described in Section 2.2,

$$H(\boldsymbol{X}) = \ell_n H(\boldsymbol{X}_1) = \ell_n (H_2(\alpha_n) + \alpha_n \log M). \tag{2.21}$$

Let $E = 1\{\hat{\boldsymbol{X}} \neq \boldsymbol{X}\}$ indicate whether the receiver makes an error, where $\hat{\boldsymbol{X}}$ is the estimation of $\boldsymbol{X}$. Consider an (M, n) code satisfying the power constraint (2.2) with $\mathsf{P}_e^{(n)} = P\{E = 1\}$. The input entropy $H(\boldsymbol{X})$ can be calculated as

(2.22)

$$H(\boldsymbol{X}) = H(\boldsymbol{X}|\boldsymbol{Y}) + I(\boldsymbol{X}; \boldsymbol{Y})$$

$$= H\left(\boldsymbol{X}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}|\boldsymbol{Y}\right) + I(\boldsymbol{X}; \boldsymbol{Y}) \tag{2.23}$$

$$= H\left(1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}|\boldsymbol{Y}\right) + H\left(\boldsymbol{X}|1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}, \boldsymbol{Y}\right) + I(\boldsymbol{X}; \boldsymbol{Y}), \tag{2.24}$$

where we used the chain rule of the entropy to obtain (2.24). Because the error indicator E is determined by $\boldsymbol{X}$ and $\boldsymbol{Y}$, we can further obtain

(2.25)

$$H(\boldsymbol{X}) = H\left(1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}|\boldsymbol{Y}\right) + H\left(\boldsymbol{X}, E|\boldsymbol{Y}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right) + I(\boldsymbol{X}; \boldsymbol{Y})$$

$$= H\left(1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}|\boldsymbol{Y}\right) + H\left(E|\boldsymbol{Y}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right)$$

$$+ H\left(\boldsymbol{X}|E, \boldsymbol{Y}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right) + I(\boldsymbol{X}; \boldsymbol{Y}) \tag{2.26}$$

$$\leq H_2\left(P\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right) + H_2\left(\mathsf{P}_e^{(n)}\right)$$

$$+ H\left(\boldsymbol{X}|E, \boldsymbol{Y}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right) + I(\boldsymbol{X}; \boldsymbol{Y}) \tag{2.27}$$

$$\leq 2\log 2 + H\left(\boldsymbol{X}|E, \boldsymbol{Y}, 1\left\{\boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n)\right\}\right) + I(\boldsymbol{X}; \boldsymbol{Y}). \tag{2.28}$$

In the following, we will upper bound $I(\boldsymbol{X};\boldsymbol{Y})$ and $H\left(\boldsymbol{X}|E,\boldsymbol{Y},1\left\{\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\right\}\right)$.

Lemma 1. *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ follow the distribution described by* (2.5)*, then*

$$I(\boldsymbol{X};\boldsymbol{Y})\leq\frac{n}{2}\log\left(1+k_nP\right). \tag{2.29}$$

Proof. See Appendix A.1. □

Lemma 2. *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ follow the distribution described by* (2.5)*. If k_n is an unbounded sequence satisfying* (2.10)*, then for large enough n,*

$$H\left(\boldsymbol{X}|E,\boldsymbol{Y},1\left\{\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\right\}\right)\leq 4\mathsf{P}_e^{(n)}\left(k_n\log M+k_n+\ell_nH_2(\alpha_n)\right)+\log M. \tag{2.30}$$

Proof. See Appendix A.2. □

Combining (2.21), (2.28), and Lemmas 1 and 2, we can obtain

$$\begin{aligned}\ell_nH_2(\alpha_n)+k_n\log M\leq{}&2\log 2+4\mathsf{P}_e^{(n)}(k_n\log M+k_n+\ell_nH_2(\alpha_n))+\\&\log M+\frac{n}{2}\log(1+k_nP).\end{aligned} \tag{2.31}$$

Dividing both sides of (2.31) by k_n and rearranging the terms, we have

$$\begin{aligned}&\left(1-4\mathsf{P}_e^{(n)}\right)\log M-\frac{1}{k_n}\log M+\left(1-4\mathsf{P}_e^{(n)}\right)\frac{H_2(\alpha_n)}{\alpha_n}\\&\leq B_1(n)+\frac{2\log 2}{k_n}+4\mathsf{P}_e^{(n)},\end{aligned} \tag{2.32}$$

where $B_1(n)$ is defined as (2.14). Since $k_n \to \infty$, we have for large enough n,

$$\left(1 - 4\mathsf{P}_e^{(n)} - \frac{1}{k_n}\right)\left(\log M + \frac{H_2(\alpha_n)}{\alpha_n}\right) \leq B_1(n) + \delta + 4\mathsf{P}_e^{(n)}. \tag{2.33}$$

Since $\mathsf{P}_e^{(n)}$ vanishes and $k_n \to \infty$ as n increases and δ can be chosen arbitrarily small, according to (2.33), given any $\epsilon > 0$, there exists some δ and for large enough n such that the following holds:

$$\log M \leq (1+\epsilon)B_1(n) - \frac{H_2(\alpha_n)}{\alpha_n} \tag{2.34}$$

$$= (1 + \epsilon - \theta_n)B_1(n), \tag{2.35}$$

where θ_n is defined as (2.11), whose limit is denoted as θ. Since (2.35) holds for arbitrary ϵ, if $\theta > 1$, there exists a small enough ϵ such that $\log M < 0$ for large enough n. It implies $B(n) = 0$, meaning that an average user cannot send a single bit of information reliably. If $\theta = 1$, then (2.35) implies that for large enough n, $\log M < \epsilon B_1(n)$ for any $\epsilon > 0$.

If $\theta < 1$, $B(n)$ given by (2.12) can be written as

$$B(n) = (1 - \theta_n)B_1(n). \tag{2.36}$$

The message length can be further upper bounded as

$$\log M \leq \left(1 + \frac{\epsilon}{1-\theta_n}\right) B(n), \tag{2.37}$$

which implies $\log M \leq (1+\epsilon)B(n)$ for any arbitrarily small ϵ. Thus, the converse for Case 1) is established.

We have the following result on the "overhead factor" θ_n.

Proposition 1. *Let θ_n be defined as in (2.11). Consider the regime $k_n = \Theta(n)$. The following holds as $n \to \infty$:*

(1) *If $\ell_n = \lceil an \rceil$ for some constant $a > 0$, then $\theta_n \to 0$ as $n \to \infty$.*

(2) *If $\ell_n = \lceil an^d \rceil$ for some constant $a > 0$, $d > 1$ and $c = \lim_{n\to\infty} \frac{k_n}{n}$, then $\theta_n \to 2c(d-1)$.*

Proof. The proof is straightforward from (2.11) as $n \to \infty$. □

Proposition 1 demonstrates the overhead of active user identification as a function of the growth rate of ℓ_n. When ℓ_n grows linearly in n, the cost of detecting the set of active users is negligible when amortized over n channel uses. On the other hand, when ℓ_n grows too quickly in n, θ_n could be larger than 1, meaning that an average user cannot even transmit a single bit reliably over a block. For user identification not to use up all channel uses, we need

$$d < 1 + \frac{1}{2} \limsup_{n\to\infty} \frac{n}{k_n}. \tag{2.38}$$

This explains the capacity trends in Fig. 2.1.

2.3.2. Converse for Case 2): unbounded ℓ_n and bounded k_n

The converse claim is basically that no linear growth in message length is achievable. Suppose that, to the contrary, $\limsup_{n\to\infty} B(n)/n = C$ for some $C > 0$. There must exist some $k_0 \geq 1$ such that $\frac{1}{2k_0} \log(1 + k_0 P) < C$. Then C is at least the symmetric capacity of the conventional multiaccess channel with k_0 users. However, as $n \to \infty$,

there is a non-vanishing probability that the number of active users is greater than $2k_0$. Letting each user transmit a message length of $B(n)$ would yield a strictly positive error probability. Hence the converse is proved.

2.3.3. Converse for Case 3): bounded ℓ_n

If $\alpha_n \to 0$, a transmitting user sees no interference with probability $(1-\alpha_n)^{\ell_n - 1} \to 1$. The converse is obvious because $\frac{1}{2}\log(1+P)$ is the conventional capacity for the point-to-point channel. The achievable message length cannot exceed $\frac{n}{2}\log(1+P)$ asymptotically.

If $\alpha_n \to \alpha > 0$, the number of active users is a binomial random variable. (The channel is nonergodic.) The probability that all ℓ users are active is $\alpha^\ell > 0$. Hence the converse follows from the symmetric rate $\frac{1}{2\ell}\log(1+\ell P)$ for the conventional multiaccess channel with ℓ users.

2.4. Proof of Theorem 2

In this section, we prove the converse and achievability of the minimum user identification cost (Theorem 2). It is a crucial step in the proof of the achievability part of Theorem 1.

2.4.1. Converse

In either of the two cases in Theorem 2, it suffices to show that the probability of error cannot vanish if $n_0 = (1-\epsilon)n(\ell)$ for any $0 < \epsilon < 1$. The converse of Theorem 2 follows exactly from that of Theorem 1 by replacing $M = 1$ and letting $n = n_0$. According to (2.34), in order to achieve vanishing error probability for random user identification, for

any $0 < \epsilon < 1$,

$$(1+\epsilon)\frac{n_0}{2k_\ell}\log(1+k_\ell P) \geq \frac{H_2(\alpha_\ell)}{\alpha_\ell}. \tag{2.39}$$

Therefore, the length of the signature must satisfy

$$n_0 > (1-\epsilon)\frac{\ell H_2(\alpha_\ell)}{\frac{1}{2}\log(1+k_\ell P)} \tag{2.40}$$

for sufficiently large ℓ.

2.4.2. Achievability

Let $n(\ell)$ be given by (2.19). Pick an arbitrary fixed $\epsilon \in (0, P)$. In the following, we will show that we can achieve vanishing error probability in identification using signature length

$$n_0 = \begin{cases} (1+\epsilon)\, n(\ell), & \text{if } \lim_{k_\ell \to \infty} n(\ell)/k_\ell > 0 \\ \epsilon k_\ell, & \text{if } \lim_{k_\ell \to \infty} n(\ell)/k_\ell = 0. \end{cases} \tag{2.41}$$

We provide a user identification scheme whose error probability is upper bounded by e^{-ck_ℓ} for some positive constant c dependent on ϵ. Let the signatures of each user $\boldsymbol{S}_k^a$ be generated according to i.i.d. Gaussian distribution with zero mean and variance

$$P' = P - \epsilon. \tag{2.42}$$

The receiver searches the binary activity vector that best explains the received signal. We restrict the search to be among all binary ℓ-vectors whose weight does not exceed the

average k_ℓ by a small fraction, and formulate it as an optimization problem:

$$\begin{aligned} \text{minimize} \quad & \| \boldsymbol{Y}^a - \underline{\boldsymbol{S}}^a \boldsymbol{x} \|_2^2 \\ \text{subject to} \quad & \boldsymbol{x} \in \{0,1\}^\ell \\ & \sum_{i=1}^{\ell} x_i \leq (1+\delta_\ell) k_\ell, \end{aligned} \tag{2.43}$$

where δ_ℓ controls the search region of $\boldsymbol{x}$. We choose δ_ℓ to be some monotone decreasing sequence such that $\delta_\ell^2 k_\ell$ is unboundedly increasing and $\delta_\ell \log k_\ell \to 0$. Specifically, we let

$$\delta_\ell = k_\ell^{-\frac{1}{3}}. \tag{2.44}$$

Denote $\mathcal{E}_d$ as the event of detection error and $\mathcal{F}_j$ as the event that the signature of the j-th user violates the power constraint (2.2), $j = 1, \cdots, \ell$. The probability of error in the stage of activity identification $P_e^{(\ell)}$ is thus calculated as

$$\mathsf{P}_e^{(\ell)} \leq \mathsf{P}\left\{ \mathcal{E}_d \cup \left(\bigcup_{j \in \{1,\cdots,\ell\}} \mathcal{F}_j \right) \right\} \tag{2.45}$$

$$\leq \mathsf{P}\left\{\mathcal{E}_d\right\} + \ell \mathsf{P}\left\{\mathcal{F}_1\right\} \tag{2.46}$$

using the union bound and the fact that all codewords are identically distributed.

Furthermore,

$$\ell \mathsf{P}\left\{\mathcal{F}_1\right\} = \ell \mathsf{P}\left\{ \sum_{i=1}^{n_0} (S_{1i}^a)^2 > n_0 P \right\} \tag{2.47}$$

$$\leq \ell e^{-c n_0}, \tag{2.48}$$

where c is some positive number (which depends on ϵ) due to large deviation theory for the sum of i.i.d. Gaussian random variables [**36**]. In either case of (2.41), $n_0 \geq_n ak_\ell$ for some $a > 0$, so (2.48) implies

$$\ell \mathsf{P}\left\{\mathcal{F}_1\right\} \leq_\ell \ell e^{-\delta k_\ell} \tag{2.49}$$

for some $\delta > 0$, which vanishes as $\ell \to \infty$ by assumption (2.18).

We next derive an upper bound of the probability of detection error $\mathsf{P}\left\{\mathcal{E}_d\right\}$. Clearly,

$$\mathsf{P}\{\mathcal{E}_d\} = \mathsf{E}\left\{\mathsf{P}\{\mathcal{E}_d|\boldsymbol{X}^a\}\right\} \tag{2.50}$$

$$\leq \mathsf{P}\{\boldsymbol{X}^a \notin \mathcal{B}_1^\ell(\delta_\ell, k_\ell)\} + \sum_{\boldsymbol{x} \in \mathcal{B}_1^\ell(\delta_\ell, k_\ell)} \mathsf{P}\{\mathcal{E}_d|\boldsymbol{X}^a = \boldsymbol{x}\}\mathsf{P}\left\{\boldsymbol{X}^a = \boldsymbol{x}\right\}. \tag{2.51}$$

The support size of the transmitted signal $\boldsymbol{X}^a$ given by (2.16) follows the binomial distribution $\mathrm{Bin}(\ell, k_\ell/\ell)$. By the Chernoff bound for binomial distribution [**37**],

$$\mathsf{P}\{\boldsymbol{X}^a \notin \mathcal{B}_1^\ell(\delta_\ell, k_\ell)\} = \mathsf{P}\left\{\sum_{i=1}^{\ell} X_i^a > (1+\delta_\ell)k_\ell\right\} + \mathsf{P}\left\{\sum_{i=1}^{\ell} X_i^a = 0\right\} \tag{2.52}$$

$$\leq \exp\left(-k_\ell \delta_\ell^2/3\right) + (1 - k_\ell/\ell)^\ell, \tag{2.53}$$

which vanishes due to (2.44) and the fact that $(1 - k_\ell/\ell)^\ell$ vanishes for unbounded k_ℓ. In other words, the number of active user is smaller than $(1+\delta_\ell)k_\ell$ with high probability. In order to prove Theorem 2, it suffices to show that the second term on the right-hand side (RHS) of (2.51) vanishes.

Pick arbitrary $\boldsymbol{x}^* \in \mathcal{B}_1^\ell(\delta_\ell, k_\ell)$. Let its support be A^*, which must satisfy $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$. We write $\mathsf{P}\{\mathcal{E}_d|\boldsymbol{X}^a = \boldsymbol{x}^*\}$ interchangeably with $\mathsf{P}\{\mathcal{E}_d|A^*\}$, because there is a

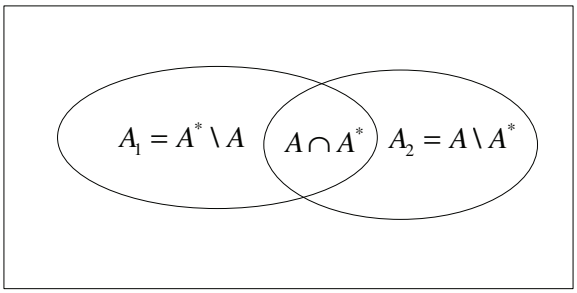

Figure 2.3. The set relationship.

one-to-one mapping between x^* and A^*. In the remainder of this subsection, we analyze the decoding error probability conditioned on a fixed A^* and drop the conditioning on A^* for notational convenience, i.e., $P\{\mathcal{E}_d\}$ implicitly means $\mathsf{P}\{\mathcal{E}_d|A^*\}$. The randomness lies in the signatures $\underline{\boldsymbol{S}}^a$ and the received signal $\boldsymbol{Y}^a$ from $\boldsymbol{x}^*$. Define

$$T_A = \left\| \boldsymbol{Y}^a - \sum_{i \in A} \boldsymbol{S}_i^a \right\|_2^2 - \left\| \boldsymbol{Y}^a - \sum_{i \in A^*} \boldsymbol{S}_i^a \right\|_2^2. \tag{2.54}$$

According to the decoding rule (2.43), a detection error may occur only if there is some $A \subseteq \{1, \cdots, \ell\}$ such that $A \neq A^*$, such that $|A| \leq (1+\delta_\ell)k_\ell$, and $T_A \leq 0$. Hence,

$$\mathcal{E}_d \subseteq \bigcup_{\substack{A \subseteq \{1,\cdots,\ell\}: \\ |A| \leq (1+\delta_\ell)k_\ell, A \neq A^*}} \{T_A \leq 0\}. \tag{2.55}$$

In the following, we divide the exponential number of error events in (2.55) into a relatively small number of classes. We will show that the probability of error of each class vanishes and so does the overall error probability. Specifically, we write the union over A according to the cardinality of the sets $A^* \cap A$ and $A \backslash A^*$. Let $w_1 = |A_1|$ and $w_2 = |A_2|$, where $A_1 = A^* \backslash A$ represents the set of misses and $A_2 = A \backslash A^*$ represents the set of false alarms. Then (w_1, w_2) must satisfy $w_1 \leq |A^*|$, $w_2 \leq |A|$, and $|A^*| + w_2 = |A| + w_1$.

According to the decoding rule (2.43), (w_1, w_2) must be found in the following set:

$$\mathcal{W}^{(\ell)} = \{(w_1, w_2) : w_1 \in \{0, 1, \cdots, |A^*|\}, w_2 \in \{0, 1, \cdots, (1+\delta_\ell)k_\ell\},$$

$$w_1 + w_2 > 0, |A^*| + w_2 \leq (1+\delta_\ell)k_\ell + w_1\}. \tag{2.56}$$

Further define the event $\mathcal{E}_{w_1,w_2}$ as

$$\mathcal{E}_{w_1,w_2} = \bigcup_{\substack{A\subseteq\{1,\cdots,\ell\}:\\ |A^*\setminus A|=w_1, |A\setminus A^*|=w_2}} \{T_A \leq 0\}. \tag{2.57}$$

By (2.55), $\mathcal{E}_d \subseteq \bigcup_{(w_1,w_2)\in\mathcal{W}^{(\ell)}} \mathcal{E}_{w_1,w_2}$. Hence

$$\mathsf{P}\{\mathcal{E}_d\} \leq \sum_{(w_1,w_2)\in\mathcal{W}^{(\ell)}} \mathsf{P}\{\mathcal{E}_{w_1,w_2}\}. \tag{2.58}$$

We will show that when ℓ is large enough, there exists some constant $c_0 > 0$ such that $\mathsf{P}\{\mathcal{E}_{w_1,w_2}\} \leq e^{-k_\ell c_0}$ for all $(w_1, w_2) \in \mathcal{W}^{(\ell)}$.

Define

$$\mathcal{A}_1(w_1) = \{A_1 : A_1 \subseteq A^*, |A_1| = w_1\} \tag{2.59}$$

and

$$\mathcal{A}_2(w_2) = \{A_2 : A_2 \subseteq \{1, \cdots, \ell\}\setminus A^*, |A_2| = w_2\}. \tag{2.60}$$

Then for any A leading to an error event in $\mathcal{E}_{w_1,w_2}$ specified by (2.57), it can be written as $A = A_2 \cup (A^* \backslash A_1)$, for some $A_1 \in \mathcal{A}_1(w_1)$ and $A_2 \in \mathcal{A}_2(w_2)$. Therefore, (2.57) gives

$$\mathcal{E}_{w_1,w_2} = \bigcup_{A_1 \in \mathcal{A}_1(w_1)} \bigcup_{A_2 \in \mathcal{A}_2(w_2)} \{T_A \leq 0\}, \tag{2.61}$$

which implies

$$1\{\mathcal{E}_{w_1,w_2}\} \leq \sum_{A_1 \in \mathcal{A}_1(w_1)} \left(\sum_{A_2 \in \mathcal{A}_2(w_2)} 1\{T_A \leq 0\} \right)^{\rho} \tag{2.62}$$

for every $\rho \in [0,1]$. As a result,

$$\mathsf{P}\{\mathcal{E}_{w_1,w_2}\} = \mathsf{E}\{1\{\mathcal{E}_{w_1,w_2}\}\} \tag{2.63}$$

$$\leq \sum_{A_1 \in \mathcal{A}_1(w_1)} \mathsf{E}\left\{ \left(\sum_{A_2 \in \mathcal{A}_2(w_2)} 1\{T_A \leq 0\} \right)^{\rho} \right\} \tag{2.64}$$

where the expectation is taken over $(\underline{\boldsymbol{S}}^a, \boldsymbol{Y}^a)$. We further calculate the expectation by first conditioning on $(\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a)$ as follows:

$$\mathsf{P}\{\mathcal{E}_{w_1,w_2}\} \leq \sum_{A_1 \in \mathcal{A}_1(w_1)} \mathsf{E}\left\{ \mathsf{E}\left\{ \left(\sum_{A_2 \in \mathcal{A}_2(w_2)} 1\{T_A \leq 0\} \right)^{\rho} \middle| \underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a \right\} \right\} \tag{2.65}$$

$$\leq \sum_{A_1 \in \mathcal{A}_1(w_1)} \mathsf{E}\left\{ \left[\mathsf{E}\left\{ \sum_{A_2 \in \mathcal{A}_2(w_2)} 1\{T_A \leq 0\} \middle| \underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a \right\} \right]^{\rho} \right\}, \tag{2.66}$$

where the expectation is taken first with respect to the probability measure $p_{\underline{\boldsymbol{S}}^a_{\{1,\cdots,\ell\}\backslash A^*} | \underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a}$ and then with respect to the probability measure $p_{\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a}$; and Jensen's inequality is applied in (2.66) to the concave function x^{ρ}, $0 < \rho \leq 1$. Note that $\underline{\boldsymbol{S}}^a_{\{1,\cdots,\ell\}\backslash A^*}$ and $\underline{\boldsymbol{S}}^a_{A^*}$ are independent and $\boldsymbol{Y}^a$ only depends on $\underline{\boldsymbol{S}}^a_{A^*}$, we have $p_{\underline{\boldsymbol{S}}^a_{\{1,\cdots,\ell\}\backslash A^*} | \underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a}(\underline{\boldsymbol{s}}_1 | \underline{\boldsymbol{s}}_2, \boldsymbol{y}) =$

$p_{\underline{\boldsymbol{S}}^a_{\{1,\cdots,\ell\}\setminus A^*}}(\underline{\boldsymbol{s}}_1)$. The inner expectation in (2.66) is taken with respect to the probability measure $p_{\underline{\boldsymbol{S}}^a_{A_2}}$ for each $A_2 \in \mathcal{A}_2(w_2)$. Since the entries of $\underline{\boldsymbol{S}}^a$ are i.i.d., the inner expectation yields identical results for all $A_2 \in \mathcal{A}_2(w_2)$ and the outer expectation yields identical results for all $A_1 \in \mathcal{A}_1(w_1)$.

The number of choices for A_1 is $\binom{|A^*|}{w_1}$, whereas the number of choices for A_2 is no greater than $\binom{\ell}{w_2}$. Therefore, we apply the union bound to obtain

$$\mathsf{P}\left\{\mathcal{E}_{w_1,w_2}\right\} \leq \binom{|A^*|}{w_1}\binom{\ell}{w_2}^{\rho} \mathsf{E}\left\{\left[\mathsf{E}\left\{1\left\{T_A \leq 0\right\}\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\}\right]^{\rho}\right\}, \tag{2.67}$$

where A is now a fixed representative choice with $|A^*\backslash A| = w_1$ and $|A\backslash A^*| = w_2$.

We will obtain an upper bound of the detection error probability by further upper bounding $\mathsf{E}\left\{1\left\{T_A \leq 0\right\}\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\}$. Let

$$p_{Y|\boldsymbol{S}_A}(y_i|\boldsymbol{s}_{A,i}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(y_i - \sum_{k\in A} s_{ki}\right)^2\right). \tag{2.68}$$

The conditional distribution of $\boldsymbol{y}$ given that the codewords $\underline{\boldsymbol{s}}_A$ are transmitted is given by $p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{y}|\underline{\boldsymbol{s}}_A) = \prod_{i=1}^n p_{Y|\boldsymbol{S}_A}(y_i|\boldsymbol{s}_{A,i})$, where n is the dimension of $\boldsymbol{y}$. Then for any $\lambda \geq 0$, the following holds due to (2.54):

$$\mathsf{E}\left\{1\left\{T_A \leq 0\right\}\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\} = \mathsf{E}\left\{1\left\{\frac{p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_A)}{p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_{A^*})} \geq 1\right\}\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\} \tag{2.69}$$

$$\leq \mathsf{E}\left\{\left(\frac{p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_A)}{p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_{A^*})}\right)^{\lambda}\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\} \tag{2.70}$$

$$= p^{-\lambda}_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_{A^*})\mathsf{E}\left\{p^{\lambda}_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{Y}^a|\underline{\boldsymbol{S}}^a_A)\middle|\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y}^a\right\}, \tag{2.71}$$

where (2.71) follows because $(\underline{\boldsymbol{S}}^a_{A^*}, \boldsymbol{Y})$ is independent of $\underline{\boldsymbol{S}}^a_{A_2}$. For every function $g\left(\boldsymbol{S}^a_{A^*}, \boldsymbol{Y}^a\right)$, $\mathsf{E}\left\{g\left(\boldsymbol{S}^a_{A^*}, \boldsymbol{Y}^a\right)\right\} = \int_{\mathbb{R}^{n_0}} \mathsf{E}\left\{g\left(\boldsymbol{S}^a_{A^*}, \boldsymbol{y}\right) p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{y}|\underline{\boldsymbol{S}}^a_{A^*})\right\} d\boldsymbol{y}$. Combining (2.67) and (2.71) yields

(2.72)
$$\mathsf{P}\left\{\mathcal{E}_{w_1,w_2}\right\} \leq \binom{|A^*|}{w_1}\binom{\ell}{w_2}^{\rho} \int_{\mathbb{R}^{n_0}} \mathsf{E}\left\{p^{1-\lambda\rho}_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{y}|\underline{\boldsymbol{S}}^a_{A^*}) \left(\mathsf{E}\left\{p^{\lambda}_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{y}|\underline{\boldsymbol{S}}^a_A)\middle|\underline{\boldsymbol{S}}^a_{A^*}\right\}\right)^{\rho}\right\} d\boldsymbol{y}.$$

Due to the memoryless channel property, i.e., $p_{\boldsymbol{Y}|\underline{\boldsymbol{S}}_A}(\boldsymbol{y}|\underline{\boldsymbol{S}}^a_A) = \prod_{i=1}^{n_0} p_{Y|\boldsymbol{S}_A}(y_i|\boldsymbol{S}^a_{A,i})$, we obtain

$$\mathsf{P}\left\{\mathcal{E}_{w_1,w_2}\right\} \leq \binom{|A^*|}{w_1}\binom{\ell}{w_2}^{\rho} \left(m_{\lambda,\rho}(w_1,w_2)\right)^{n_0} \tag{2.73}$$

where

$$m_{\lambda,\rho}(w_1,w_2) = \int_{\mathbb{R}} \mathsf{E}\left\{p^{1-\lambda\rho}_{Y|\boldsymbol{S}_A}(y|\boldsymbol{S}^a_{A^*}) \left(\mathsf{E}\left\{p^{\lambda}_{Y|\boldsymbol{S}_A}(y|\boldsymbol{S}^a_A)\middle|\boldsymbol{S}^a_{A^*}\right\}\right)^{\rho}\right\} dy. \tag{2.74}$$

The first two terms of the RHS of (2.73) can be upper bounded as [**19**, Page 353]

$$\binom{|A^*|}{w_1}\binom{\ell}{w_2}^{\rho} \leq \exp\left(|A^*| H_2\left(\frac{w_1}{|A^*|}\right) + \rho\ell H_2\left(\frac{w_2}{\ell}\right)\right). \tag{2.75}$$

Moreover, by the Gaussian distribution of the codewords, the last term of the RHS of (2.73) can be explicitly calculated (see Appendix A.3) to obtain

$$m_{\lambda,\rho}(w_1,w_2) = \exp\left(\frac{1-\rho}{2}\log(1+\lambda w_2 P') - \frac{1}{2}\log\left(1+\lambda(1-\lambda\rho)w_2P' + \lambda\rho(1-\lambda\rho)w_1P'\right)\right). \tag{2.76}$$

Therefore, by (2.73)-(2.76),

$$\mathsf{P}\{\mathcal{E}_{w_1,w_2}\} \leq \exp\left(-k_\ell h_{\lambda,\rho}(w_1,w_2)\right), \tag{2.77}$$

where

$$\begin{aligned} h_{\lambda,\rho}(w_1,w_2) &= \frac{n_0}{2k_\ell}\log\left(1+\lambda(1-\lambda\rho)w_2P'+\lambda\rho(1-\lambda\rho)w_1P'\right) \\ &\quad - \frac{(1-\rho)n_0}{2k_\ell}\log\left(1+\lambda w_2P'\right) - \frac{|A^*|}{k_\ell}H_2\left(\frac{w_1}{|A^*|}\right) - \frac{\rho\ell}{k_\ell}H_2\left(\frac{w_2}{\ell}\right). \end{aligned} \tag{2.78}$$

To show the capacity achievability, we next show that by choosing λ and ρ properly, for large enough ℓ, $h_{\lambda,\rho}(w_1,w_2)$ is strictly greater than some positive constant for all $(w_1,w_2) \in \mathcal{W}^{(\ell)}$.

Lemma 3. *Fix $\epsilon \in (0,P)$. Let $P' = P-\epsilon$. Let $n(\ell)$ be given by (2.19) and n_0 be given by (2.41). Suppose $n(\ell)/k_\ell$ has finite limit or diverges to infinity. There exists $\ell^* > 0$ and $c_0 > 0$ such that for every $\ell \geq \ell^*$ the following holds: If the true signal $x^a \in \mathcal{B}_1^\ell(\delta_\ell, k_\ell)$, i.e., $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$, then for every $(w_1,w_2) \in \mathcal{W}^{(\ell)}$ with $\mathcal{W}^{(\ell)}$ defined as in (2.56), there exist $\lambda \in [0,\infty)$ and $\rho \in [0,1]$ such that*

$$h_{\lambda,\rho}(w_1,w_2) \geq c_0. \tag{2.79}$$

Proof. See Appendix A.4. □

Lemma 3 and (2.77) imply

$$\mathsf{P}\{\mathcal{E}_{w_1,w_2}|A^*\} \leq e^{-c_0k_\ell}, \tag{2.80}$$

for all $\ell \geq \ell^*$, $(w_1, w_2) \in \mathcal{W}^{(\ell)}$, and $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$. Then as long as $\ell \geq \ell^*$, for any $\boldsymbol{x} \in \mathcal{B}_1^\ell(\delta_\ell, k_\ell)$,

$$\mathsf{P}\{\mathcal{E}_d|\boldsymbol{X}^a = \boldsymbol{x}\} \leq \sum_{(w_1,w_2)\in\mathcal{W}^{(\ell)}} \mathsf{P}\{\mathcal{E}_{w_1,w_2}|\boldsymbol{X}^a = \boldsymbol{x}\} \tag{2.81}$$

$$\leq \sum_{(w_1,w_2)\in\mathcal{W}^{(\ell)}} e^{-c_0 k_\ell} \tag{2.82}$$

$$\leq 4k_\ell^2 e^{-c_0 k_\ell}, \tag{2.83}$$

where (2.83) is due to $w_1 \leq 2k_\ell$ and $w_2 \leq 2k_\ell$. Therefore, the first term on the RHS of (2.51) vanishes as ℓ increases. So does $\mathsf{P}\{\mathcal{E}_d\}$. Thus we can achieve arbitrarily reliable identitifcation with SNR $P' = P - \epsilon$ and signature length n_0 given by (2.41). Since ϵ can be arbitrarily small, the achievability of Theorem 2 is established.

2.5. Proof of the Achievability of Theorem 1

2.5.1. Achievability for Case 3) with bounded ℓ_n

As ℓ_n is nondecreasing, $\ell_n \to \ell$ for some constant ℓ. If $\alpha_n \to \alpha > 0$, with some positive probability all ℓ users are active. Hence the achievability capacity follows from the result for the conventional multiaccess channel with ℓ users.

If $\alpha_n \to 0$, a transmitting user experiences a single-user channel with probability $(1-\alpha_n)^{\ell_n - 1} \to 1$. Therefore, it can achieve a vanishing error probability with the conventional capacity for the point-to-point channel.

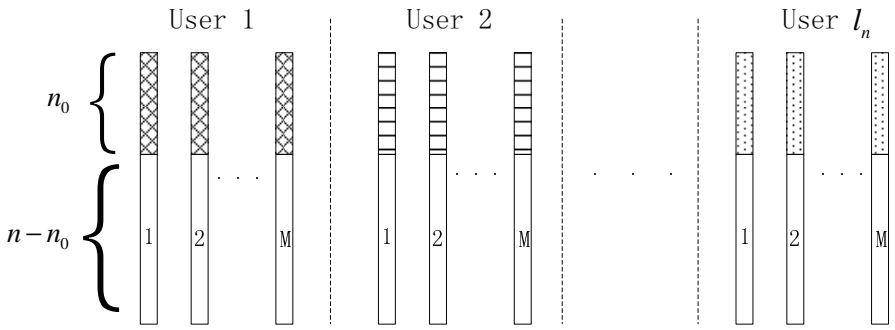

Figure 2.4. Codebook structure. Each user maintains M codewords with each consisting of a message-bearing codeword prepended by a signature.

2.5.2. Achievability for Case 1) and Case 2) with unbounded ℓ_n

We first assume unbounded k_n and establish the achievability result. The case of bounded k_n is then straightforward.

We consider a two-stage approach: In the first stage, the set of active users are identified based on their unique signatures. In the second stage, the messages from the active users are decoded. Let θ_n and its limit θ be defined as in Theorem 1. We consider the cases of $\theta = 0$ and $\theta > 0$ at the same time. Fix $\epsilon \in (0, \min(1, P))$. Specifically, the following scheme is used:

- *Codebook construction:* The codebooks of the ℓ_n users are generated independently. Let

$$n_0 = \begin{cases} \epsilon n, & \text{if } \theta = 0 \\ (1+\epsilon)\,\theta_n n, & \text{otherwise .} \end{cases} \tag{2.84}$$

For user k, codeword $\boldsymbol{s}_k(0) = \mathbf{0}$ represents silence. User k also generates

$$M = \lceil \exp\left[(1-\epsilon)B(n)\right] \rceil \tag{2.85}$$

codewords as follows. First, generate M random sequences of length $n-n_0$, each according to i.i.d. Gaussian distribution with zero mean and variance $P' = P-\epsilon$. Then generate one signature of length n_0 with i.i.d. $\mathcal{N}(0, P')$, denoted by $\boldsymbol{S}_k^a$, and prepend this signature to every codeword to form M codewords of length n. In other words, the w-th codeword of user k takes the shape of $\boldsymbol{S}_k(w) = \begin{pmatrix} \boldsymbol{S}_k^a \\ \boldsymbol{S}_k^b(w) \end{pmatrix}$. The matrix of the concatenated codebooks of all users is illustrated in Fig. 2.4.

- *Transmission:* For user k to be silent, it is equivalent to transmitting $\boldsymbol{s}_k(0)$. Otherwise, to send message $w_k \neq 0$, user k transmits $\boldsymbol{S}_k(w_k)$.
- *Channel:* Each user is active independently with probability α_n. The active users transmit simultaneously. The received signal is $\boldsymbol{Y}$ given by (2.5).
- *Two-stage detection and decoding:* Upon receiving $\boldsymbol{Y}$, the decoder performs the following:

  (1) Active user identification: Let $\boldsymbol{Y}^a$ denote the first n_0 entries of $\boldsymbol{Y}$, corresponding to the superimposed signatures of all active users subject to noise. $\boldsymbol{Y}^a$ is mathematically described by (2.16). The receiver estimates $\boldsymbol{X}^a$ according to (2.43). The output of this stage is a set $A \subseteq \{1, \cdots, \ell_n\}$ that contains the detected active users.

  (2) Message decoding: Let $\boldsymbol{Y}^b$ denote the last $n - n_0$ entries of $\boldsymbol{Y}$, corresponding to the superimposed message-bearing codewords. The receiver solves

the following optimization problem:

$$\text{minimize} \quad ||\boldsymbol{Y}^b - \underline{\boldsymbol{S}}^b \left[\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_{\ell_n}^T\right]^T ||^2 \tag{2.86}$$

$$\text{subject to} \quad \boldsymbol{x}_k \in \mathcal{X}_M^1, k = 1, \cdots, \ell_n \tag{2.87}$$

$$\boldsymbol{x}_k = \boldsymbol{0}, \quad \forall k \notin A \tag{2.88}$$

$$\boldsymbol{x}_k \neq \boldsymbol{0}, \quad \forall k \in A \tag{2.89}$$

Basically the receiver performs the maximum likelihood decoding for the set of users in the purported active user set A. The position of 1 in each recovered nonzero $\boldsymbol{x}_k$ indicates the message from user k.

Theorem 3 (Achievability of the Gaussian many-access channel)**.** *Let θ_n be defined as* (2.11) *and $B(n)$ be defined as* (2.12)*. Suppose $\lim_{n\to\infty} \theta_n < 1$. For the MnAC given by* (2.1)*, for any given constant $\epsilon \in (0, 1)$, the message length of $(1-\epsilon)B(n)$ is asymptotically achievable using the preceding scheme.*

The remainder of this section is devoted to the proof of Theorem 3. In Section 2.5.3, we show that the set of active users can be accurately identified in stage 1. In Section 2.5.4, we show that the users' messages can be accurately decoded in stage 2 assuming knowledge of the active users. The results are combined in Section 2.5.5 to establish the achievability part of Theorem 3.

2.5.3. Optimal User Identification

We shall invoke Theorem 2 (proved in Section 2.4) to quantify the cost of reliable user identification. To adapt to the notation in this section, we apply Theorem 2 with ℓ and k_ℓ

being replaced by ℓ_n and k_n, respectively. With the change of notations, $n(\ell)$ as defined in Theorem 2 can be written as

$$n(\ell) = \frac{\ell_n H_2(k_n/\ell_n)}{\frac{1}{2}\log(1+k_n P)} \tag{2.90}$$

$$= \theta_n n, \tag{2.91}$$

where θ_n is given by (2.11).

According to Theorem 2, choosing the signature length $n_0 = (1+\epsilon)\theta_n n$ and $n_0 = \epsilon k_n$ yields vanishing error probability in user identification for the case of $\lim_{n\to\infty} \theta_n n/k_n > 0$ and $\lim_{n\to\infty} \theta_n n/k_n = 0$, respectively, where $\epsilon \in (0,1)$ is an arbitrary constant. In the following, we make use of this result to prove that choosing n_0 according to (2.84) guarantees reliable user identification.

First, consider $\theta = 0$. By (2.84), the signature length is $n_0 = \epsilon n$ for some ϵ. In the case of $\lim_{n\to\infty} \theta_n n/k_n > 0$, since θ_n vanishes, it must have $n_0 \geq_n (1+\epsilon)\theta_n n$. In the case of $\lim_{n\to\infty} \theta_n n/k_n = 0$, since $k_n = O(n)$, $n_0 = \epsilon n$ implies $n_0 \geq_n \epsilon' k_n$ for some $\epsilon' > 0$. By Theorem 2, the choice of n_0 is sufficient for reliable user identification.

Second, consider $\theta > 0$. By (2.84), the signature length is $n_0 = (1+\epsilon)\theta_n n$. Since $k_n = O(n)$, it must have $\lim_{n\to\infty} \theta_n n/k_n > 0$. Thus, the signature length n_0 obviously achieves reliable user identification by Theorem 2.

2.5.4. Achieving the Capacity of MnAC with Known User Activities

In previous work [**16**], we studied the capacity of the Gaussian MnAC where all users are always active and the number of users is sublinear in the blocklength, i.e., $k_n = o(n)$. In

that case, random coding with Feinstein's suboptimal decoding, which suffices to achieve the capacity of conventional multiaccess channel capacity, can achieve the capacity of the Gaussian MnAC. Proving the capacity achievability for faster scaling of the number of active users is much more challenging, mainly because the exponential number of possible error events prevents one from using the simple union bound. Here, we derive the capacity of the MnAC for the case where the number of users may grow as quickly as linearly with the blocklength by lower bounding the error exponent of the error probability due to maximum-likelihood decoding. The results also complement a related study of many-broadcast models in [**6**].

Theorem 4 (Capacity of the Gaussian many-access channel without random access)**.** *For the MnAC with k_n always-active users, suppose the number of channel uses is n and the number of users k_n grows as $O(n)$, the symmetric capacity is*

$$B_1(n) = \frac{n}{2k_n} \log(1 + k_n P). \tag{2.92}$$

In particular, for any $\epsilon \in (0, 1)$, there exists a sequence of codebooks with message lengths (in nats) $B_1(n)(1 - \epsilon)$ such that the average error probability is arbitrarily small for sufficiently large n.

In the following, we will prove Theorem 4. We can model the MnAC with known user activities using (2.5) with $\alpha_n = 1$, i.e., $k_n = \ell_n$. Upon receiving the length-n vector $\boldsymbol{y}$, we

estimate $\boldsymbol{x} = \left[\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_{k_n}^T\right]^T$ using the maximum likelihood decoding:

$$\text{minimize} \quad ||\boldsymbol{y} - \underline{\boldsymbol{s}}\boldsymbol{x}||^2 \tag{2.93}$$

$$\text{subject to} \quad \boldsymbol{x}_k = \boldsymbol{e}_m, \quad \text{for some } m = 1, \cdots, M. \tag{2.94}$$

Define $\mathcal{F}_j$ as the event that user j's codeword violates the power constraint (2.2), $j = 1, \cdots, k_n$. Define $\mathcal{E}_k$ as the error event that k users are received in error. Suppose $P\{\mathcal{E}_k|A^*\}$ is the probability of $\mathcal{E}_k$ given that the true signal is $\boldsymbol{x}^*$ with support A^*. By symmetry of the codebook construction, the average error probability can be calculated as

$$\mathsf{P}_e^{(n)} \leq \mathsf{P}\left\{\bigcup_{k=1}^{k_n} \mathcal{E}_k \cup \bigcup_{j=1}^{k_n} \mathcal{F}_j\right\} \tag{2.95}$$

$$\leq \frac{1}{M^{k_n}} \sum_{A^*} \sum_{k=1}^{k_n} \mathsf{P}\{\mathcal{E}_k|A^*\} + \sum_{j=1}^{k_n} \mathsf{P}\left\{\mathcal{F}_j\right\}. \tag{2.96}$$

Let A be the support of the estimated $\boldsymbol{x}$ according to the maximum likelihood decoding. Define A_1 and A_2 in the same manner as that in Section 2.5.3, i.e., $A_1 = A^* \backslash A$ and $A_2 = A \backslash A^*$. In this case, $|A| = |A^*| = k_n$ and $|A_2| = |A_1| = k$. Further denote $\gamma = k/k_n$ as the fraction of users subjected to errors. Then we write $\mathsf{P}\{\mathcal{E}_k|A^*\}$ and $\mathsf{P}\{\mathcal{E}_\gamma|A^*\}$ interchangeably. In the following analysis, we consider a fixed A^* and drop the conditioning on A^* for notational convenience. Following similar arguments leading to (2.73), letting

$\lambda = \frac{1}{1+\rho}$ and considering $\binom{k_n}{\gamma k_n}$ possible sets of A_1 and $M^{\gamma k_n}$ possible sets of A_2, we have

(2.97)
$$\mathsf{P}\{\mathcal{E}_\gamma\} \leq \binom{k_n}{\gamma k_n} M^{\gamma k_n \rho} \left(\int_{\mathbb{R}} \mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_{A^*}) \left(\mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_A) \middle| \boldsymbol{S}_{A^*} \right\} \right)^{\rho} \right\} dy \right)^n$$

(2.98)
$$= \binom{k_n}{\gamma k_n} M^{\gamma k_n \rho} \left(\int_{\mathbb{R}} \mathsf{E}\left\{ \mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_{A^*}) \middle| \boldsymbol{S}_{A^*\cap A} \right\} \left(\mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_A) \middle| \boldsymbol{S}_{A^*} \right\} \right)^{\rho} \right\} dy \right)^n.$$

By symmetry, $\mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_A) \middle| \boldsymbol{S}_{A^*} \right\} = \mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\frac{1}{\rho+1}}(y|\boldsymbol{S}_{A^*}) \middle| \boldsymbol{S}_{A^*\cap A} \right\}$, which results in

$$\mathsf{P}\{\mathcal{E}_\gamma\} \leq \binom{k_n}{\gamma k_n} M^{\gamma k_n \rho} \exp(-nE_0(\gamma,\rho)), \tag{2.99}$$

where $E_0(\gamma,\rho)$ is defined by

$$E_0(\gamma,\rho) = -\log\left[\int_{\mathbb{R}} \mathsf{E}\left\{ \left[\mathsf{E}\left\{ \left(p_{Y|\boldsymbol{S}_A}(y|\boldsymbol{S}_A)\right)^{\frac{1}{\rho+1}} \middle| \boldsymbol{S}_{A^*} \right\} \right]^{1+\rho} \right\} dy \right]. \tag{2.100}$$

By the inequality $\binom{k_n}{\gamma k_n} \leq \exp(k_n H_2(\gamma))$, we can further upper bound $\mathsf{P}\{\mathcal{E}_\gamma\}$ as

$$\mathsf{P}\{\mathcal{E}_\gamma\} \leq \exp\left[-nf(\gamma,\rho)\right], \tag{2.101}$$

where

$$f(\gamma,\rho) = E_0(\gamma,\rho) - \gamma\rho\frac{k_n}{n}v(n) - \frac{k_n}{n}H_2(\gamma), \tag{2.102}$$

and $v(n) = \log M$. Intuitively, $E_0(\gamma,\rho)$ in (2.101) is an achievable error exponent for the error probability caused by a particular A being detected in favor of A^* and the terms

$k_n H_2(\gamma) + \gamma\rho k_n v(n)$ correspond to the cardinality of all possible A leading to the error event $\mathcal{E}_\gamma$.

By particularizing (2.76) with $w_1 = w_2 = \gamma k_n$ and $\lambda = \frac{1}{1+\rho}$, we can derive $E_0(\gamma, \rho)$ explicitly as

$$E_0(\gamma,\rho) = -\log m_{\lambda,\rho}(w_1, w_2)|_{w_1=w_2=\gamma k_n, \lambda=\frac{1}{1+\rho}} \tag{2.103}$$

$$= \frac{\rho}{2}\log\left(1 + \frac{\gamma k_n P'}{\rho+1}\right). \tag{2.104}$$

The achievable error exponent for $P(\mathcal{E}_\gamma)$ is determined by the minimum error exponent over the range of γ, i.e.,

$$E_r = \min_{\frac{1}{k_n}\le\gamma\le 1} \max_{0\le\rho\le 1} f(\gamma,\rho). \tag{2.105}$$

The following Lemma is key to establishing Theorem 4.

Lemma 4. *Let M be such that the message length $v(n) = \log M$ is given by*

$$v(n) = (1-\epsilon)\frac{n}{2k_n}\log(1 + k_n P'). \tag{2.106}$$

Suppose $k_n = O(n)$, there exists n^* and $c_0 > 0$ such that for every $n \ge n^*$,

$$P\{\mathcal{E}_k | A^*\} \le e^{-c_0 n} \tag{2.107}$$

holds uniformly for all $1 \le k \le k_n$ and for all $|A^*|$.

Proof. See Appendix A.5. □

Due to Lemma 4, for large enough n,

$$\sum_{k=1}^{k_n} P\{\mathcal{E}_k|A^*\} \leq k_n e^{-c_0 n} \tag{2.108}$$

which vanishes as n increases. Moreover, following the same argument as (2.48), the second term of the RHS of (2.96) vanishes and hence $\mathsf{P}_e^{(n)}$ given by (2.96) can be proved to vanish. As a result, Theorem 4 is established.

2.5.5. Achieving the Capacity of MnAC with On-off Random Access

In this subsection, we combine the results of Section 2.5.3 and Section 2.5.4 to prove the achievability result for Case 1) and Case 2) in Theorem 3. We first prove the case of unbounded k_n, and the case of bounded k_n follows naturally. Let θ denote the limit of θ_n.

Case 1): unbounded ℓ_n and unbounded k_n.

We further divide this case into two sub-cases.

Sub-case a: $0 < \theta < 1$.

We need to show that the message length $(1-\epsilon)B(n)$ is asymptotically achievable for any fixed $\epsilon \in (0,1)$.

The detection errors are caused by activity identification error or message decoding error. It has been shown by (2.53) that with high probability the number of active users is no more than $(1+\delta_n)k_n$. As a result, Theorem 2 and Theorem 4 conclude that the message length

$$\frac{(1-\epsilon')(n-n_0)}{2(1+\delta_n)k_n} \log\left(1+(1+\delta_n)k_n P\right), \tag{2.109}$$

where $n_0 = (1+\epsilon')\theta_n n$, is asymptotically achievable for any $\epsilon' > 0$.

In order to prove the achievability, it suffices to show that there exists ϵ' such that the message length given by (2.109) is asymptotically greater than

$$(1-\epsilon)B(n) = \frac{(1-\epsilon)(1-\theta_n)n}{2k_n}\log\left(1+k_nP\right). \tag{2.110}$$

The intuition of proof is that for sufficiently large n, $(1+\delta_n)k_n$ is approximately k_n, and we can always find a small enough ϵ' such that $(1-\epsilon')(n-n_0)$ is greater than $(1-\epsilon)(1-\theta_n)n$.

We choose some small enough $\epsilon' > 0$ such that

$$(1-\epsilon')^2 - \epsilon'(1-\epsilon')^2\frac{1+\theta}{1-\theta} > 1-\epsilon. \tag{2.111}$$

This is feasible because the left-hand side of (2.111) is equal to 1 if $\epsilon' = 0$.

Since $\log\left(1+(1+\delta_n)k_nP\right)/\log(1+k_nP) \to 1$ and $\delta_n \to 0$ as n increases, we have

$$\frac{\log\left(1+(1+\delta_n)k_nP\right)}{(1+\delta_n)} \geq_n (1-\epsilon')\log(1+k_nP). \tag{2.112}$$

The difference between (2.109) and $(1-\epsilon)B(n)$ is calculated as

$$\begin{aligned}
&\frac{(1-\epsilon')(n-n_0)}{2(1+\delta_n)k_n}\log\left(1+(1+\delta_n)k_nP\right) - (1-\epsilon)B(n) \\
&\geq_n \left[\frac{(1-\epsilon')^2(1-n_0/n)}{1-\theta_n} - (1-\epsilon)\right]B(n) && (2.113)\\
&= \left[(1-\epsilon')^2 - \epsilon'(1-\epsilon')^2\frac{\theta_n}{1-\theta_n} - (1-\epsilon)\right]B(n) && (2.114)\\
&\geq_n \left[(1-\epsilon')^2 - \epsilon'(1-\epsilon')^2\frac{1+\theta}{1-\theta} - (1-\epsilon)\right]B(n) && (2.115)
\end{aligned}$$

where (2.115) is due to $\theta_n \leq_n (1+\theta)/2$. By (2.111), the RHS of (2.115) is greater than zero. It means that for large enough n, the achievabile message length (2.109) is greater than $(1-\epsilon)B(n)$, which establishes the achievability.

Sub-case b: $\theta = 0$.

The proof for the case of vanishing θ_n is analogous. We need to show that message length $(1-\epsilon)B_1(n)$ is asymptotically achievable for any fixed $\epsilon \in (0,1)$.

The number of active users is no more than $(1+\delta_n)k_n$ with high probability. As a result, Theorem 2 and Theorem 4 conclude that the message length

$$\frac{(1-\epsilon')(n-n_0)}{2(1+\delta_n)k_n}\log\left(1+(1+\delta_n)k_nP\right), \tag{2.116}$$

where $n_0 = \epsilon' n$, is asymptotically achievable for any $\epsilon' > 0$.

In order to prove Theorem 3, it suffices to show that there exists ϵ' such that the message length given by (2.116) is asymptotically greater than

$$(1-\epsilon)B_1(n) = (1-\epsilon)\frac{n}{2k_n}\log\left(1+k_nP\right). \tag{2.117}$$

Choose some small enough $\epsilon' > 0$ such that

$$(1-\epsilon')^3 > (1-\epsilon). \tag{2.118}$$

The difference between (2.116) and $(1-\epsilon)B_1(n)$ is calculated as

$$\frac{(1-\epsilon')(n-n_0)}{2(1+\delta_n)k_n}\log\left(1+(1+\delta_n)k_nP\right)-(1-\epsilon)B_1(n)$$

$$\geq_n \left[(1-\epsilon')^2(1-n_0/n)-(1-\epsilon)\right]B_1(n) \tag{2.119}$$

$$=\left[(1-\epsilon')^3-(1-\epsilon)\right]B(n), \tag{2.120}$$

where (2.119) is due to (2.112). By the choice of ϵ' given by (2.118), (2.120) is greater than zero. It concludes that for large enough n, the achievable message length (2.116) is greater than $(1-\epsilon)B_1(n)$, which establishes the achievability.

Case 2): unbounded ℓ_n and bounded k_n

In the case of unbounded ℓ_n and bounded k_n, there is nonvanishing probability that the number of active users is equal to any finite number. The number of active users is no longer fewer than $(1+\delta_n)k_n$ with high probability. Let s_n be any increasing sequence. There is high probability that the number of users is fewer than $(1+\delta_n)s_n$. As a result, by treating s_n as the unbounded k_n as in Case 1), we can apply the established achievable results for Case 1). The achievability result for Case 2) is summarized in the following theorem.

Theorem 5. *Let s_n be any increasing sequence satisfying $s_n = O(n)$, $\ell_n e^{-\delta s_n} \to 0$ for every $\delta > 0$ and*

$$\lim_{n\to\infty}\frac{2\ell_n H_2(s_n/\ell_n)}{n\log(1+s_nP)}<1. \tag{2.121}$$

Then any message length given by

$$(1-\epsilon)\left(\frac{n}{2s_n}\log(1+s_nP)-H_2(s_n/\ell_n)\right) \tag{2.122}$$

is asymptotically achievable

Proof. See Appendix A.6. □

2.6. On Successive Decoding for Many-Access Channels

In conventional multiaccess channels, the sum capacity can be achieved by successive decoding. A natural question is: Can the sum capacity of the MnAC be achieved using successive decoding? We consider the system model where all users have the same power constraints, assuming no random activity and the number of users being $k_n = an$ for some $a > 0$. We provide a negative answer for the case where Gaussian random codes are used and successive decoding is applied. Throughout the discussion in this section, we do not seek to achieve the symmetric capacity, but the sum capacity achieved by successive decoding.

Suppose Gaussian random codes are used, i.e., each user generates its codewords as i.i.d. Gaussian random variables with zero mean and variance P. Thus the codewords of other users look like Gaussian noise to any given user. The first user to be decoded has the largest interference from all the other $k_n - 1$ users and its signal-to-interference-plus-noise ratio (SINR) is $Q = \frac{P}{1+(k_n-1)P}$. Suppose the first user transmits with message length

$$v(n) = (1-\epsilon)nC, \tag{2.123}$$

where $C = \frac{1}{2}\log(1+Q)$. We will show that the error probability is strictly bounded from zero. The intuition is that the error probability usually decays at the rate of $\exp(-\delta nC)$, where δ is some positive constant dependent on ϵ. In the MnAC setting, if the interference due to many users is so large that nC converges to a finite constant, the error exponent is not large enough to drive the error probability to zero as the blocklengh increases.

Lemma 5. *Suppose Gaussian random codes are used and successive decoding is applied. There exist universal constants $A_1 > 0$ and $A_2 > 0$, such that the error probability of the first user is lower bounded as*

$$\mathsf{P}_e^{(n)} \geq \mathsf{Q}(x)e^{-\frac{A_1 T x^3}{S^{3/2}}}\left(1 - \frac{A_2 T x}{S^{3/2}}\right) - e^{-(\lambda-1)(n-1)\epsilon C}, \tag{2.124}$$

where $\mathsf{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-\frac{u^2}{2})du$, $S = 2nQ(2+Q)$,

$$x = \frac{2(\lambda\epsilon n + 1 - \lambda\epsilon)C(1+Q)}{\sqrt{S}}, \tag{2.125}$$

and

$$T = n\mathsf{E}\left\{(-Q(1-Z^2) - 2\sqrt{Q}Z)^3\right\} \tag{2.126}$$

with Z being a standard Gaussian random variable.

Proof. See Appendix A.7. □

Let $k_n = an$ for some constant $a > 0$. Then, as $n \to \infty$,

$$nQ \to \frac{1}{a}, \tag{2.127}$$

$$S \to \frac{4}{a}, \tag{2.128}$$

$$T \to 0, \tag{2.129}$$

$$nC \to \frac{1}{2a}, \tag{2.130}$$

$$x \to \frac{\epsilon\lambda}{2\sqrt{a}}. \tag{2.131}$$

Therefore,

$$\lim_{n\to\infty} \mathsf{P}_e^{(n)} \geq \mathsf{Q}\left(\frac{\epsilon\lambda}{2\sqrt{a}}\right) - e^{-\frac{(\lambda-1)\epsilon}{2a}}. \tag{2.132}$$

Using the lower bound of $\mathsf{Q}(x) \geq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2}$, it can be seen that when the exponential term is dominating, there exist some small enough $\lambda\epsilon$ such that the first term in (2.132) is greater than the second term. In this case, the error probability is strictly bounded away from zero. Fig. 2.5 plots the numerical results of the RHS of (2.132) for different values of a and λ. It can be seen that for the different values of a, there exists some λ that makes the lower bound of error probability (2.132) strictly greater than zero.

2.7. Many-Access Channel with Heterogeneous User Groups

In this section, we will generalize the characterization of capacity region to the case where groups of users have heterogeneous channel gains and activity patterns. Suppose ℓ_n users can be divided into a finite number of J groups, where group j consists of $\beta^{(j)}\ell_n$ users with $\sum_{j=1}^{J}\beta^{(j)} = 1$. Further assume every user in group j has the same

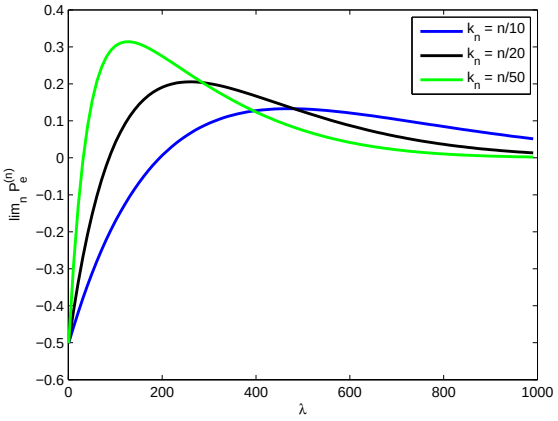

Figure 2.5. Lower bound of error probability given by (2.132) for successive decoding with $\epsilon = 10^{-3}$.

power constraint $P^{(j)}$. Each user in group j transmits with probability $\alpha_n^{(j)}$. We refer to such MnAC with heterogeneous channel gains and activity patterns as the configuration $\left(\{\alpha_n^{(j)}\}, \{\beta^{(j)}\}, \{P^{(j)}\}, \ell_n\right)$. The error probability is defined as the probability that the receiver incorrectly detects the message of any user in the system. The problem is what is the maximum achievable message length for users in each group such that the average error probability vanishes.

Definition 5 (Asymptotically achievable message length tuple)**.** *Consider a MnAC of configuration* $\left(\{\alpha_n^{(j)}\}, \{\beta^{(j)}\}, \{P^{(j)}\}, \ell_n\right)$. *A sequence of* $\left(\lceil\exp(v^{(1)}(n))\rceil, \cdots, \lceil\exp(v^{(J)}(n))\rceil\right.$ $\left.n\right)$ *code for this configuration consists of a* $\left(\lceil\exp(v^{(j)}(n))\rceil, n\right)$ *symmetry code for every user in group* j *according to Definition 1,* $j = 1, \cdots, J$.

We say a message length tuple $\left(v^{(1)}(n), \cdots, v^{(J)}(n)\right)$ *is asymptotically achievable if there exists a sequence of* $\left(\lceil\exp(v^{(1)}(n))\rceil, \cdots, \lceil\exp(v^{(J)}(n))\rceil, n\right)$ *codes such that the average error probability vanishes as* $n \to \infty$.

Definition 6 (Capacity region of the many-access channel)**.** *Consider a MnAC of configuration* $\left(\{\alpha_n^{(j)}\}, \{\beta^{(j)}\}, \{P^{(j)}\}, \ell_n\right)$*. The capacity region is the set of asymptotically achievable message length tuples. In particular, for every* $\left(B^{(1)}(n), \cdots, B^{(J)}(n)\right)$ *in the capacity region, if the users transmit with message length tuple* $\big((1-\epsilon)B^{(1)}(n), \cdots, (1-\epsilon)B^{(J)}(n)\big)$*, the average error probability vanishes as* $n \to \infty$*. If any user transmits with message length outside the capacity region, reliable communication cannot be achieved.*

Theorem 6. *Consider a MnAC of configuration* $\left(\{\alpha_n^{(j)}\}, \{\beta^{(j)}\}, \{P^{(j)}\}, \ell_n\right)$*. Suppose* $\ell_n \to \infty$ *and* $\alpha_n^{(j)} \to \alpha^{(j)} \in [0,1]$*. Let the average number of active users in group* j *be* $k_n^{(j)} = \alpha_n^{(j)}\beta^{(j)}\ell_n = O(n)$*, such that* $\ell_n e^{-\delta k_n^{(j)}} \to 0$ *for every* $\delta > 0$ *and every* $j = 1, \cdots, J$*. Let* $\theta_n^{(j)}$ *be defined as*

$$\theta_n^{(j)} = \frac{2\beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right)}{n \log k_n^{(j)}}. \tag{2.133}$$

and let $\theta^{(j)}$ *denote its limit. Suppose* $\log k_n^{(j_1)} / \log k_n^{(j_2)} \to 1$ *for any* $j_1, j_2 \in \{1, \cdots, J\}$*. If* $\sum_{j=1}^J \theta^{(j)} < 1$*, then the message length capacity region is characterized as*

$$\sum_{j=1}^J k_n^{(j)} B^{(j)}(n) \le \frac{n}{2} \log\left(\sum_{j=1}^J k_n^{(j)}\right) - \sum_{j=1}^J \beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right). \tag{2.134}$$

If $\sum_{j=1}^J \theta^{(j)} > 1$*, then some user cannot transmit a single bit reliably.*

It is interesting to note that as far as the asymptotic message lengths are concerned, the impact of the transmit power is inconsequential. Also, the only limitation on the

message is their weighted average. This is in contrast to the classical multiaccess channel, where the sum rate of each subset of users is subject to a separate upper bound in general.

2.7.1. Converse

The proof of converse follows similarly as in Section 2.3. We only sketch the proof here. Consider the system model described by (2.5). Suppose the message length transmitted by each user in group j is $v^{(j)}(n)$, $j = 1, \cdots, J$. Let $\tilde{\boldsymbol{X}}_j$ denote a vector, which stacks the vectors $\boldsymbol{X}_k$, for all k belonging to group j. Since there are a total of $\beta^{(j)}\ell_n$ users in group j and the distributions of $\boldsymbol{X}_k$ are the same for all k in the same group j, we have

$$H\left(\tilde{\boldsymbol{X}}_j\right) = \beta^{(j)}\ell_n H(\boldsymbol{X}_k) \tag{2.135}$$

$$= \beta^{(j)}\ell_n \left(H_2\left(\alpha_n^{(j)}\right) + \alpha_n^{(j)} v^{(j)}(n)\right). \tag{2.136}$$

Define $\mathcal{J} \subseteq \{1, \cdots, J\}$. Further denote $\tilde{\boldsymbol{X}}_{\mathcal{J}}$ as the vector consisting of $\{\tilde{\boldsymbol{X}}_j : j \in \mathcal{J}\}$. Thus,

$$H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}\right) = \sum_{j\in\mathcal{J}} H\left(\tilde{\boldsymbol{X}}_j\right). \tag{2.137}$$

Applying the chain rule, we have

$$H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}\right) = I\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}; \boldsymbol{Y}\right) + H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\boldsymbol{Y}\right) \tag{2.138}$$

$$= H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\tilde{\boldsymbol{X}}_{\{1,\cdots,J\}\setminus\mathcal{J}}\right) - H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\boldsymbol{Y}\right) + H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\boldsymbol{Y}\right) \tag{2.139}$$

$$\leq I\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}; \boldsymbol{Y}|\tilde{\boldsymbol{X}}_{\{1,\cdots,J\}\setminus\mathcal{J}}\right) + H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\boldsymbol{Y}\right). \tag{2.140}$$

Following the argument in Lemma 1, we have

$$I\left(\tilde{\boldsymbol{X}}_{\mathcal{J}};\boldsymbol{Y}|\tilde{\boldsymbol{X}}_{\{1,\cdots,J\}\setminus\mathcal{J}}\right)\leq\frac{n}{2}\log\left(1+\sum_{j\in\mathcal{J}}k_n^{(j)}P^{(j)}\right). \tag{2.141}$$

In order to achieve vanishing error probability, following the argument in Lemma 2, we have

$$H\left(\tilde{\boldsymbol{X}}_{\mathcal{J}}|\boldsymbol{Y}\right)=o\left(\sum_{j\in\mathcal{J}}k_n^{(j)}v^{(j)}(n)+\beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right)\right). \tag{2.142}$$

Combining (2.136), (2.137), (2.140), (2.141), and (2.142), we have for large enough n,

$$(1-\epsilon)\sum_{j\in\mathcal{J}}\left[k_n^{(j)}v^{(j)}(n)+\beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right)\right]\leq\frac{n}{2}\log\left(1+\sum_{j\in\mathcal{J}}k_n^{(j)}P^{(j)}\right), \tag{2.143}$$

for every $\epsilon>0$.

Since the power in each group is bounded, we have $\log\left(1+\sum_{j\in\mathcal{J}}k_n^{(j)}P^{(j)}\right)/\log\left(\sum_{j\in\mathcal{J}}k_n^{(j)}\right)$ tends to 1 as n increases. Thus, (2.143) implies that for every $\epsilon>0$ and every $\mathcal{J}\subseteq\{1,\cdots,J\}$,

$$\sum_{j\in\mathcal{J}}k_n^{(j)}v^{(j)}(n)\leq(1+\epsilon)\frac{n}{2}\log\left(\sum_{j\in\mathcal{J}}k_n^{(j)}\right)-\sum_{j\in\mathcal{J}}\beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right). \tag{2.144}$$

As in (2.15), we have dropped the power terms in the capacity expression to ease the rest of the proof. By (2.144), we have

$$\sum_{j\in\mathcal{J}}k_n^{(j)}v^{(j)}(n)\leq\left[1+\epsilon-\sum_{j\in\mathcal{J}}\theta_n^{(j)}\xi_n^{(\mathcal{J},j)}\right]\frac{n}{2}\log\left(\sum_{j\in\mathcal{J}}k_n^{(j)}\right), \tag{2.145}$$

where

$$\xi_n^{(\mathcal{J},j)} = \frac{\log\left(k_n^{(j)}\right)}{\log\left(\sum_{j\in\mathcal{J}} k_n^{(j)}\right)}. \tag{2.146}$$

For any $\mathcal{J}_1, \mathcal{J}_2 \subseteq \{1, \cdots, J\}$, we have

$$\frac{\log\left(\min_{j\in\mathcal{J}_1} k_n^{(j)}\right)}{\log\left(\max_{j\in\mathcal{J}_2} k_n^{(j)}\right) + \log J} \leq \frac{\log\left(\sum_{j\in\mathcal{J}_1} k_n^{(j)}\right)}{\log\left(\sum_{j\in\mathcal{J}_2} k_n^{(j)}\right)} \leq \frac{\log\left(\max_{j\in\mathcal{J}_1} k_n^{(j)}\right) + \log J}{\log\left(\min_{j\in\mathcal{J}_2} k_n^{(j)}\right)}. \tag{2.147}$$

Taking the limit of $n \to \infty$ on both sides of (2.147), by the assumption that $\log k_n^{(j_1)} / \log k_n^{(j_2)}$ tends to 1 for any j_1, j_2, we have

$$\frac{\log\left(\sum_{j\in\mathcal{J}_1} k_n^{(j)}\right)}{\log\left(\sum_{j\in\mathcal{J}_2} k_n^{(j)}\right)} \to 1. \tag{2.148}$$

It implies that $\xi_n^{(\mathcal{J},j)} \to 1$ for all $j \in \mathcal{J}$. If $\sum_{j=1}^{J} \theta^{(j)} > 1$, particularizing (2.145) with $\mathcal{J} = \{1, \cdots, J\}$ implies that for large enough n, $v^{(j)}(n) = 0$ for all $j = 1, \cdots, J$.

If $\sum_{j=1}^{J} \theta^{(j)} < 1$, the achievable message length can be further upper bounded as

$$\sum_{j\in\mathcal{J}} k_n^{(j)} v^{(j)}(n) \leq \left(1 + \frac{\epsilon}{1 - \sum_{j\in\mathcal{J}} \theta_n^{(j)} \xi_n^{(\mathcal{J},j)}}\right) B_{\mathcal{J}}(n), \tag{2.149}$$

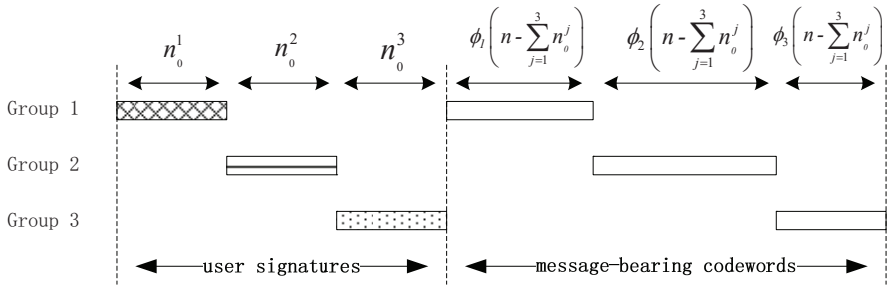

Figure 2.6. Transmission scheme for $J = 3$ groups.

where

$$B_{\mathcal{J}}(n) = \frac{n}{2}\log\left(\sum_{j\in\mathcal{J}} k_n^{(j)}\right) - \sum_{j\in\mathcal{J}} \beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right). \tag{2.150}$$

Applying (2.149) with $\mathcal{J} = \{1, \cdots, J\}$ and $\xi_n^{(\mathcal{J},j)} \to 1$, the achievable message length tuple must satisfy

$$\sum_{j\in\{1,\cdots,J\}} k_n^{(j)} v^{(j)}(n) \le (1+\epsilon) B_{\{1,\cdots,J\}}(n) \tag{2.151}$$

for every $\epsilon > 0$. Thus, the converse part of Theorem 6 is established.

Note that by (2.149), any achievable message length tuple must satisfy

$$\sum_{j\in\mathcal{J}} k_n^{(j)} v^{(j)}(n) \le (1+\epsilon) B_{\mathcal{J}}(n) \tag{2.152}$$

for all $\mathcal{J} \subseteq \{1, \cdots, J\}$. However, in the regime of unbounded k_n, (2.149) implies that these constraints are dominated by the one for $\mathcal{J} = \{1, \cdots, J\}$, because $B_{\mathcal{J}}(n) \ge_n B_{\{1,\cdots,J\}}(n)$ for all $\mathcal{J} \subseteq \{1, \cdots, J\}$.

2.7.2. Achievability

We need to prove that the region of the achievable message length tuple covers the region specified by (2.134). In particular, we will show that the message length tuple satisfying

$$\sum_{j=1}^{J} k_n^{(j)} v^{(j)}(n) \leq (1-\epsilon)\left[\frac{n}{2}\log\left(\sum_{j=1}^{J} k_n^{(j)}\right) - \sum_{j=1}^{J}\beta^{(j)}\ell_n H_2\left(\alpha_n^{(j)}\right)\right] \tag{2.153}$$

is asymptotically achievable for every $\epsilon > 0$.

One achievable scheme is to detect active users in each group and their transmitted messages in a time-division manner. In particular, in the first stage, we let users in group 1 transmit the signatures before group 2, and so on. The signature length transmitted by users in group j is $n_0^{(j)}$, $j = 1, \cdots, J$. In the second stage, we let each group share the remaining time resource $n - \sum_{j=1}^{J} n_0^{(j)}$. Users in group 1 transmit their message-bearing codewords before group 2, and so on. The time resource allocated to group j in the second stage is $\phi_j\left(n - \sum_{j=1}^{J} n_0^{(j)}\right)$, where $\phi_j \geq 0$ and $\sum_{j=1}^{J}\phi_j = 1$. At the receiver side, the receiver performs user identification according to the group order, and then decode the transmitted messages according to the group order. The overall scheme is illustrated in Fig. 2.6.

Let $\theta_n^{(j)}$ be given by (2.133), which can be regarded as the fraction of resource to detect the active users in group j. According to Theorem 2 and Theorem 4, the message length tuple satisfying

$$v^{(j)}(n) = (1-\epsilon')\phi^{(j)}\frac{n - \sum_{j'=1}^{J} n_0^{(j')}}{2k_n^{(j)}}\log\left(k_n^{(j)}\right), \tag{2.154}$$

where

$$n_0^{(j)} = \begin{cases} (1+\epsilon'/2)\theta_n^{(j)} n, & \text{if } \theta^{(j)} > 0 \\ \frac{\epsilon'}{2J} n, & \text{if } \theta^{(j)} = 0 \end{cases}, \tag{2.155}$$

is achievable for every $\epsilon' \in (0,1)$.

If $\theta^{(j')} > 0$, by (2.148),

$$\frac{n_0^{(j')}}{2} \log(k_n^{(j)}) = (1+\epsilon'/2)\beta^{(j')} \ell_n H_2\left(\alpha_n^{(j')}\right) \frac{\log\left(k_n^{(j)}\right)}{\log\left(k_n^{(j')}\right)} \tag{2.156}$$

$$\leq_n (1+\epsilon')\beta^{(j')} \ell_n H_2\left(\alpha_n^{(j')}\right). \tag{2.157}$$

If $\theta^{(j')} = 0$,

$$\frac{n_0^{(j')}}{2} \log(k_n^{(j)}) = \frac{\epsilon'}{2J} \frac{n}{2} \log k_n^{(j)}. \tag{2.158}$$

Therefore,

$$\sum_{j'=1}^{J} \frac{n_0^{(j')}}{2} \log(k_n^{(j)}) \leq_n \frac{\epsilon'}{2} \frac{n}{2} \log k_n^{(j)} + \sum_{j'=1}^{J} (1+\epsilon')\beta^{(j')} \ell_n H_2\left(\alpha_n^{(j')}\right). \tag{2.159}$$

By (2.154), the achievable message length is calculated as

(2.160)

$$k_n^{(j)} v^{(j)}(n) \geq_n (1-\epsilon')\phi^{(j)} \left[(1-\epsilon'/2)\frac{n}{2} \log k_n^{(j)} - (1+\epsilon') \sum_{j=1}^{J} \beta^{(j)} \ell_n H_2\left(\alpha_n^{(j)}\right) \right]$$

$$\geq_n (1-\epsilon')\phi^{(j)} \left[(1-\epsilon')\frac{n}{2} \log \left(\sum_{j=1}^{J} k_n^{(j)} \right) - (1+\epsilon') \sum_{j=1}^{J} \beta^{(j)} \ell_n H_2\left(\alpha_n^{(j)}\right) \right]. \tag{2.161}$$

According to (2.161), there must exist some small enough ϵ' such that for large enough n,

$$k_n^{(j)} v^{(j)}(n) \geq \phi^{(j)}(1-\epsilon) \left[\frac{n}{2} \log \left(\sum_{j=1}^{J} k_n^{(j)} \right) - \sum_{j=1}^{J} \beta^{(j)} \ell_n H_2 \left(\alpha_n^{(j)} \right) \right] \tag{2.162}$$

for all $j = 1, \cdots, J$.

Since (2.162) holds for any $\phi_j > 0$, by varying the convex combination due to $\phi^{(j)}$, $j = 1, \cdots, J$, the region spanned by the achievable message tuple (2.154) covers the region specified by (2.153). The achievability result is thus established.

2.8. Conclusion

In this chapter, we have proposed a model of many-access channel, where the number of users scales with the coding blocklength as a first step towards the study of many-user information theory. New notions of message length and symmetric capacity have been defined. The symmetric capacity of a many-access channel is shown to be a function in the channel uses, consisting of two terms. The first term is the symmetric capacity of many-access channel with knowledge of the set of active users and the second term can be regarded as the cost of user identification in random access channels. Separate identification and decoding has been shown to be capacity achieving. The detection scheme can be extended to achieve the capacity region of a many-access channel with a finite number of groups experiencing different channel gains.

The results presented in this work reveal the capacity growth in the asymptotic regime. The holy grail is a many-user information theory for finite but large number of users and finite but large block length that applies accurately in practice. The challenge of developing such a theory is difficult to overestimate (see, e.g., [**38**, **39**]).

The many-access channel model together with the capacity result and the compressed sensing based identification technique will provide insights for the optimal design in emerging applications with massive sporadic access [**40**–**42**], such as in the Internet of Things and machine-to-machine communication, where the number of devices in a cell may far exceed the blocklength.

CHAPTER 3

Asynchronous Neighbor Discovery

3.1. Introduction

By some estimate [**43**], there will be more than 200 billion sensor enabled objects world-wide in the Internet of Things (IoT) by year 2020. There can be over a million such devices within 500 meter range in a densely populated area. For any wireless device to function in the IoT or an ad hoc network, the first step is to discover access points and/or other communication parties within range and also be discovered by them. This is called neighbor discovery.

Neighbor discovery is an essential step for medium access protocols and routing protocols. There has been a large body of research works on neighbor discovery for general networks [**8**,**44**,**45**]. In conventional networks, the overhead of neighbor discovery is often thought of as amortized over the long data transmission. However, a typical IoT device makes bursty transmissions and the message is usually short in a single transmission. It has been shown that the neighbor discovery overhead may considerably reduce the data throughput in systems involving a massive number of devices [**17**]. It is critical to minimize the neighbor discovery overhead.

Due to its unique features [**1**], the IoT poses additional challenges for designing an ultra-scalable scheme. The total number of IoT devices is extremely large and it is hard

to achieve perfect synchronization. Many IoT devices are of low cost and low power. A scalable scheme should have relatively low computational complexity at the device side.

The current technology and protocols may be inadequate to address the challenges. For example, a naive time division multiple access (TDMA) scheme to schedule all the devices would incur too much latency. Neighbor discovery using conventional multiuser detection approaches, say code division multiple access (CDMA) or orthogonal frequency-division multiple access (OFDMA), generally involves a complexity that scales polynomial in the number of devices [**45**], which is also unaffordable considering the large latency and power consumption. In this chapter, we propose an efficient neighbor discovery scheme that tackles the above-mentioned issues.

3.1.1. Related Work

Network layer approaches: Network layer protocol designs for neighbor discovery can be categorized into randomized and deterministic algorithms. A main objective is to optimize random access probability or the transmit schedule of each device such that the system throughput is maximized [**7**, **8**, **44**, **46**, **47**]. Instead of purely avoiding collision, the FlashLinQ technology developed by Qualcomm assigns channels based on signal-to-interference ratios and achieves superior performance over carrier sense multiple access with collision avoidance (CSMA/CA) systems [**48**, **49**]. The neighbor discovery algorithm proposed in this chapter can be regarded as a physical layer technique, which can be optimized with the network layer protocols to improve the system performance.

Coded random access: Recently, the idea of codes on graph has been applied in random access [**40**, **41**]. One scheme is named coded slotted ALOHA, where the packets are

repeatedly transmitted in different slots and are decoded using successive cancellation. These works assume synchronization transmission and perfect interference cancellation. The asynchrounous model has been studied in [**50**–**52**], where the asynchronicity is modeled to cause interference. However, perfect interference cancellation was still assumed. Rateless codes have been proposed for multiple access in machine-to-machine communications [**53**], where the channel gains are assumed to be known. As the number of users increases, the imperfect channel estimatioon is detrimental to the performance of successive cancellation.

In common with coded slotted ALOHA, our neighbor discovery scheme also applies the design of erasure correcting codes in the successive cancellation framework. Our proposed scheme is different in many aspects. Our scheme is a one-shot transmission. Moreover, we carefully characterize the error propagation effects due to residual channel estimation error.

Multi-user detection approaches: Neighbor discovery can be formulated as multiuser detection from the perspective of physical layer processing [**45**]. Due to the bursty traffic patterns, the number of active devices is typically orders of magnitude smaller than the total local device population. Based on this crucial observation, low-complexity neighbor discovery algorithms inspired by compressed sensing were proposed [**10**, **33**]. These algorithms can reduce the transmission length by over 50% compared with the 802.11 type protocols, but they require synchronous transmissions. The LASSO algorithm was proposed to detect active devices in asynchronous CDMA random access [**11**], but it involves a high complexity when the total number of users is large. Ideally, the complexity of a desirable scheme only scales polynomial with the number of active users.

3.1.2. Our Contributions

In this work, we propose a novel scheme, referred to as sparse orthogonal frequency-division multiplexing (sparse-OFDM), for *asynchronous* neighbor discovery. A key feature that distinguishes the proposed scheme from previous schemes is that sparse OFDM exploits both the parallel channel access enabled by OFDM and the bursty transmission nature in the IoT. Specifically, sparse OFDM judiciously allocates the sparsely separated channels to the devices. The resulting signal structure relates neighbor discovery to the sparse Fourier transform, studied in, e.g., [**12**], which applies to time-domain signals whose Fourier transform domain representation is sparse. The main features of sparse OFDM are as follows:

(1) When the number of active devices and the maximum delay in terms of sample points is sublinear relative to the device population, sparse OFDM can correctly detect the active devices with high probability. It only requires sublinear computational complexity and a sublinear number of transmit symbols in terms of the device population.

(2) Sparse OFDM is a one-shot transmission as opposed to scheduling the devices to transmit many frames in random access protocols. It utilizes the low-complexity point-to-point capacity-approaching codes, while at the same time exploits the multiuser diversity from successive cancellation.

(3) Sparse OFDM is inspired by the recent development of the sparse Fourier transform [**12**] and sparse Hadamard transform [**54**]. The previous works assume the signal amplitudes belong to a known discrete alphabet, and the signal amplitude

can be perfectly recovered [**12**, **55**, **56**]. In this work, we assume arbitrary signal amplitudes, and characterize the effect of error propagation due to imperfect signal estimations leveraging the results on random hypergraphs.

(4) Sparse OFDM provides practical physical-layer capability for multipacket reception. The scheme can be jointly designed with the random access protocol to further optimize the performance [**57**,**58**]. Moreover, sparse OFDM can be easily adapted to the case of peer-to-peer broadcasting, where each device has multiple bits of information to send. Sparse OFDM is particularly appealing in the IoT, where a typical message is short.

3.1.3. Chapter Organization

The rest of the chapter is organized as follows. Section 3.2 presents the system model and main results. Section 3.3 describes the signalling scheme of sparse OFDM. Section 3.4 presents the asynchronous neighbor discovery algorithm. Section 3.5 and Section 3.6 provide proofs of the theoretical performance guarantees for synchronous and asynchronous transmission, respectively. Section 3.7 presents the numerical results. Section 3.8 concludes the chapter.

Throughout the chapter, the index of a vector or each dimension of a matrix starts from 0. 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 $\boldsymbol{y}^c = \left(y_0^c, \cdots, y_{B-1}^c\right)$. We denote the real and imaginary parts of a variable X as X_R and X_I, respectively. All logarithms are base 2.

3.2. System Model and Main Results

Consider a network with N devices in total. Let $\mathcal{K} \subseteq \{0, \cdots, N-1\}$ denote an arbitrary set of active devices, where $K = |\mathcal{K}|$ is the number of active devices. We assume symbol synchronicity without frame synchronicity, i.e., the delay of each device's transmission is an integer number of symbol intervals. Fig. 3.1 shows the three-user model. Moreover, the delay of any device is no greater than M symbol intervals due to propagation delay and clock/timing differences between the transmitting and receiving devices. To yield scalable results, we let both the number of active devices K and the maximum delay M scale up with N as $N \to \infty$ in general. We assume all the devices are aware of a reference frame start point of their neighbors. This can be easily achieved by using a common a beacon signal. Device k transmits an L-symbol codeword, described as $s_{k,0}, \ldots, s_{k,L-1}$. It suffices to consider a single receiver and its discovery problem. In the absence of frequency selectivity, the received signal at every (integer) time i is given by

$$x_i = \sum_{k \in \mathcal{K}} a_k s_{k,i-m_k} + w_i, \tag{3.1}$$

where $a_k \in \mathbb{C}$ is the channel coefficient, m_k is the transmission delay of device k, and w_i are independently and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables with distribution $\mathcal{CN}(0, 2\sigma^2)$. The discovery scheme is based on signals within a single codeword duration. From each receiver's point of view, the signal $s_{k,i} = 0$ for $i < 0$ and $i \geq L$ for all k.

We shall design a transmission and detection scheme with small transmission length and low computational complexity. Each codeword is appropriately designed so that the

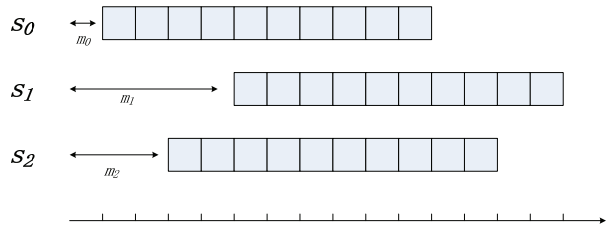

Figure 3.1. Frame-asynchronous symbol-synchronous three-user model.

codelength L is sublinear in N when K and M are sublinear in N. The following two theorems are the key results of this work:

Theorem 7. *Suppose the device transmissions are perfectly synchronized to the receiver's frame, i.e., $M = 0$. Suppose the noise variance is fixed and the channel amplitudes of all active devices is at least $\underline{a}$. For every $\underline{a}, \epsilon > 0$, 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 $L \leq \alpha_0 K \log N$, such that $\mathsf{P}\{\hat{\mathcal{K}} \neq \mathcal{K}\} \leq \epsilon$ for every subset $\mathcal{K}$ of active users of size not larger than K, where $\hat{\mathcal{K}}$ is the estimated set of active devices. In addition, the number of arithmetic operations needed for computing $\hat{\mathcal{K}}$ is no greater than $\alpha_1 K (\log K)(\log N)$.*

Theorem 8. *Suppose the transmission delay of each device is an integer number of symbol intervals upper bounded by M. Suppose the noise variance is fixed and the channel amplitude of every active device lies in the region of $[\underline{a}, \bar{a}]$. For every $\underline{a}, \bar{a}, \epsilon > 0$, 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 $L \leq \alpha_0((K + M) \log N + K \log(K + M))$, such that $\mathsf{P}\{\hat{\mathcal{K}} \neq \mathcal{K}\} \leq \epsilon$ for every subset $\mathcal{K}$ of active users of size not larger than K, where $\hat{\mathcal{K}}$ is the estimated set of active devices. In addition, the number of arithmetic operations needed for computing $\hat{\mathcal{K}}$ is no greater than $\alpha_1 K \left((\log K)(\log N) + KM \log(K + M)\right)$.*

Theorem 7 implies that arbitrarily reliable synchronous neighbor discovery is achieved asymptotically with codelength of $O(K \log N)$ and $O(K(\log K)(\log N))$ arithmetic operations. Theorem 8 implies that arbitrarily reliable asynchronous neighbor discovery is achievable asymptotically with codelength of $O((K + M) \log N + K \log(K + M))$ and $O\left(K(\log K)(\log N) + K^2 M \log(K + M)\right)$ arithmetic operations. Synchronous neighbor discovery, i.e., $M = 0$, requires a smaller codelength and fewer arithmetic operations than the asynchronous case. In both theorems, a lower bound on the signal strengths is needed for successful sparse recovery [**26**]. In a practical system, if the channel gain between two devices is too small, they are not regarded as neighbors of each other.

The maximum relative delay M is usually small in practice. The delay depends on the timing difference and the maximum distance between a device and the receiver. For example, a distance of 300 meters implies free space propagation delay of one microsecond, which spans 20 samples if the sampling frequency is 20 MHz. Suppose the maximum delay M is constant. When the number of active users K is sublinear in terms of the device population N, i.e., $K = o(N)$, the transmission lengths for both the synchronous and asynchronous schemes are sublinear in the number of devices. When $K = o(\sqrt{N})$, the number of arithmetic operations involved in the asynchronous scheme is also sublinear in N.

3.3. Sparse OFDM Signaling

Our scheme inherits the idea of OFDM. OFDM divides the spectrum into B orthogonal subcarriers. The subcarriers are assigned to different devices for transmission. In conventional OFDM, we need $B \geq N$ subcarriers if we need to schedule the transmissions

of all devices at the same time. If the number of devices is large, there will be many narrow subcarriers. The proposed scheme is referred to as sparse OFDM, because we divide the spectrum into $B \ll N$ sparsely spaced subcarriers.

In the following, to facilitate the exposition, we will describe our signaling scheme in three steps. First, we consider noiseless neighbor discovery where the total device population N is smaller than the number of available OFDM frequency bins B. Second, we consider noiseless neighbor discovery where $B < N$ and a single device is active. Third, we consider the general noisy neighbor discovery, where $B < N$ and $K < N$ devices are active.

3.3.1. Device Identification in the Case of $B \geq N$

The key idea for addressing arbitrary delay is to use the fact that the frequency of a sinusoidal signal is invariant to delay, where the delay merely causes a phase shift. Suppose each OFDM symbol contains $B + M$ samples, where M can be regarded as a cyclic prefix length accounting for the unknown delay. Since $N \leq B$, we can choose a discrete frequency $b_k \in \{0, \ldots, B-1\}$ to uniquely identify device k. The discrete-time signal structure is given by

$$s_{k,i} = g_k \exp\left(\frac{\iota 2\pi b_k i}{B}\right), \quad i = 0, \cdots, B+M-1, \tag{3.2}$$

where $g_k \in \mathbb{R}$ is a known design parameter of unit amplitude.

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 OFDM symbol and collect the remaining 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 tone, performing B-point discrete Fourier transform (DFT) on $\boldsymbol{y}$ yields a tone at the b_k-th frequency bin if and only if device k is active. Therefore, the signaling scheme (3.2) is sufficient to detect the active devices in a noiseless B-device case with a computational complexity of $O(B \log B)$ needed by the Fast Fourier Transform (FFT) algorithm. The delay m_k only affects the phase of the DFT value, so we can uniquely identify the user based on its frequency. This important insight will lead to the signaling design for a general case.

3.3.2. Single Device Identification in the Case of $B < N$

Suppose there are $N > B$ devices and only a single device k is transmitting. We still want to apply the signaling scheme (3.2). Suppose performing a B-point DFT on $\boldsymbol{y}$ yields a tone at the b-th frequency bin, we still cannot identify which device is active, because there may be multiple devices assigned to this bin. The way to resolve this problem is to transmit multiple OFDM symbols and embed the device information through coefficient g_k in (3.2).

Let $C = \lceil \log N \rceil$ and $(k)_2 = (k_1, \cdots, k_C)$ denote the binary representation of device index k. We design $\left(g_k^0, \cdots, g_k^C\right) = \left(1, (-1)^{k_1}, \cdots, (-1)^{k_C}\right)$. We apply the signaling scheme (3.2) to $C+1$ OFDM symbols, with the c-th symbol having $g_k = g_k^c$ in (3.2). In particular, we let device k transmit $\left(\boldsymbol{s}_k^0, \cdots, \boldsymbol{s}_k^C\right)$, where $\boldsymbol{s}_k^c = (s_{k,0}^c, \cdots, s_{k,B+M-1}^c)$ and

$$s_{k,i}^c = g_k^c \exp\left(\frac{\iota 2\pi b_k i}{B}\right), \tag{3.3}$$

for $i = 0, \cdots, B+M-1$ and $c = 0, \cdots, C$. The common codelength is thus $(C+1)(B+M)$ symbols. Denote $\boldsymbol{y}^c = (y_0^c, \cdots, y_{B-1}^c)$ with $y_i^c = x_{i+c(B+M)+M}$, $i = 0, \cdots, B-1$, i.e.,

discard the first M samples in each OFDM symbol interval. Performing B-point DFT on the c-th OFDM symbol $\boldsymbol{y}^c$ yields

$$Y_b^c = \frac{1}{B}\sum_{i=0}^{B-1} \exp\left(-\frac{\iota 2\pi bi}{B}\right) y_i^c \tag{3.4}$$

$$= \sum_{k\in\mathcal{K}:b_k=b} A_k g_k^c, \quad b = 0, \cdots, B-1, \tag{3.5}$$

where

$$A_k = a_k \exp\left(\iota 2\pi b_k (M - m_k)/B\right). \tag{3.6}$$

As in the $B \leq N$ case, the delay m_k only affects the phase of the frequency value of the received signal from device k. The frequency binning effectively separates the devices.

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

The two design parameters b_k (frequency) and g_k (gain) play important roles. In particular, g_k is used to carry the device index information. The frequency b_k is designed to separate devices into different frequency bins. By design, the relationship between the device index and its transmit frequency is represented by a bipartite graph. In the bipartite graph, the devices represent left nodes and the B frequency bins represent right nodes. Left node i is connected with right node j if device i transmits at the j-th frequency bin. We call a frequency bin a *zeroton*, *singleton* or *multiton*, if no device, a single device, or more than one device transmit at the frequency tone, respectively. Fig. 3.2(a) illustrates

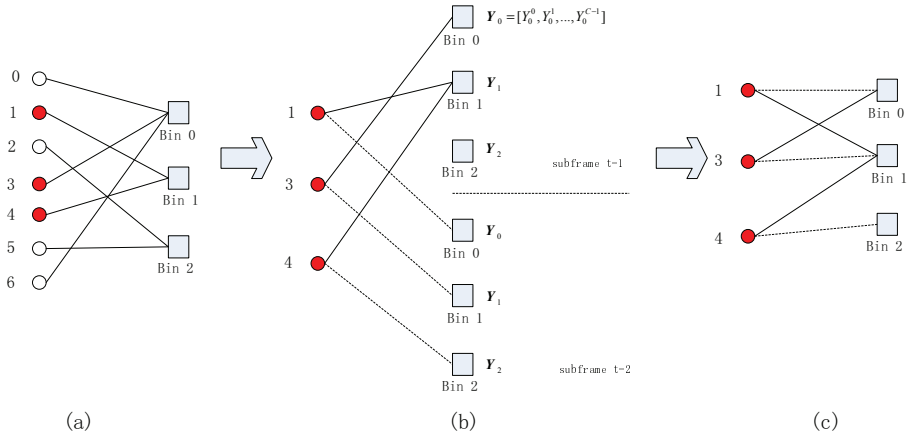

Figure 3.2. Bipartite graph representation of sparse OFDM. Left nodes represent devices and right nodes represent frequency bins. The active devices are marked in red. (a) The bipartite graph of sparse OFDM for a single subframe. (b) The bipartite graph of sparse OFDM for two subframes to resolve collision, where only the active (red) devices are shown. (c) Worst-interference bipartite graph.

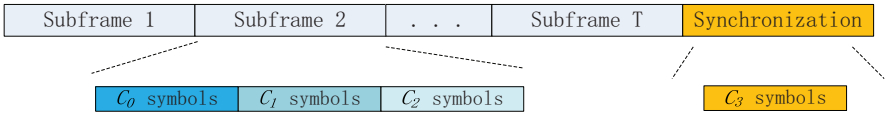

Figure 3.3. Frame structure of sparse OFDM. A frame consists of T subframes, where every subframe contains $C_0 + C_1 + C_2 + C_3$ OFDM symbols.

an example of bipartite graph with a total of $N = 7$ devices, $K = 3$ active devices, and $B = 3$ frequency bins.

When there is a single active device in the noiseless setting, the device can be identified with $O((B + M) \log N)$ samples and computational complexity of $O(B(\log N)(\log B))$.

3.3.3. Identification of Multiple Active Devices With and Without Noise

When multiple devices are active, the devices may use colliding tones, so that the device information cannot always be directly recovered from $Y_{b_k}^{j+1}/Y_{b_k}^0$, $j = 0, \cdots, \lceil \log N \rceil$. The

idea is to let the devices transmit at random frequency bins for multiple subframes. We first identify active devices from the singleton bins and then use the identified device information to bootstrap the detection of other devices.

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 distinguish a frequency bin to be a zeroton, singleton or multiton? In the following, we further enhance the signaling scheme to address these three challenges.

The overall frame structure is described in Fig. 3.3. The frame structure contains T subframes used for device identification and one additional subframe for synchronization pilots.

Signaling for subframes: We first introduce the signaling of the subframes. Each subframe consits of three segments and the i-th segment consists of C_i OFDM symbols, $i = 1, \cdots, 3$. Every OFDM symbol in the first three segments has the same signal structure (3.2), except that g_k in (3.2) is replaced by judiciously designed symbols. For ease of notation, the subframe index is suppressed. Let the length-C design vector for device k be

$$\boldsymbol{g}_k = \begin{bmatrix} \mathbf{1} \\ \tilde{\boldsymbol{g}}_k \\ \dot{\boldsymbol{g}}_k \end{bmatrix} \tag{3.7}$$

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 OFDM symbols, the second C_1 OFDM symbols, and the remaining C_2

OFDM symbols, respectively. The total number of OFDM symbols in each subframe is

$$C = C_0 + C_1 + C_2. \tag{3.8}$$

Under the noisy setting, performing B-point FFT on the c-th OFDM symbol yields

$$Y_b^c = \sum_{k \in \mathcal{K}: b_k = b} A_k g_k^c + W_b^c, \quad b = 0, \cdots, B-1, \tag{3.9}$$

where W_b^c are i.i.d. complex Gaussian variables with distribution $\mathcal{CN}(0, 2\sigma^2/B)$. For each subframe, the frequency values at the b-th bin $\boldsymbol{Y}_b$ is a vector of length C and can be written as

$$\boldsymbol{Y}_b = \begin{bmatrix} \bar{\boldsymbol{Y}}_b \\ \tilde{\boldsymbol{Y}}_b \\ \dot{\boldsymbol{Y}}_b \end{bmatrix} \tag{3.10}$$

$$= \sum_{k \in \mathcal{K}: b_k = b} A_k \begin{bmatrix} \mathbf{1} \\ \tilde{\boldsymbol{g}}_k \\ \dot{\boldsymbol{g}}_k \end{bmatrix} + \begin{bmatrix} \bar{\boldsymbol{W}}_b \\ \tilde{\boldsymbol{W}}_b \\ \dot{\boldsymbol{W}}_b \end{bmatrix}. \tag{3.11}$$

Note that the assigned frequency bin b_k, design vector $\boldsymbol{g}_k$ and received frequency vector $\boldsymbol{Y}_b$ may be different for different subframes.

The design vector is inspired by the generalized low-density parity-check (LDPC) framework for sublinear compressive sensing [**59**]. Specifically, the first all-one segment is used to estimate the channel coefficients. The second segment is used to estimate the device index. In the absence of noise, $\tilde{\boldsymbol{g}}_k = \left(1, (-1)^{k_1}, \cdots, (-1)^{k_C}\right)$ carries the device

information. Under the noisy setting, the values of $\tilde{\boldsymbol{g}}_k$ are corrupted. We thus apply error-control code to encode the device index into vector $\tilde{\boldsymbol{g}}_k$ with $C_1 = \lceil \log N \rceil / R$ OFDM symbols, where R is the code rate. The third segment is used for singleton verification. We let the entries of $\dot{\boldsymbol{g}}_k$ be generated according to i.i.d. Radamacher (± 1) variables. The number of symbols C_0, C_1 and C_2 will be specified later.

For device k, we let b_k be taken independently uniformly at random from $\{0, \cdots, B-1\}$ for each subframe (it may be different across subframes). Performing FFT on the OFDM symbols results in hashing of the devices into B bins uniformly at random in each subframe. Fig. 3.2 (b) illustrates an example for $T = 2$ subframes. The reason of using T subframes and different b_k is to resolve the bin collisions. The intuition is that, with a sufficient number of subframes, a sufficient number of devices are hashed to a singleton bin in at least one subframe with high probability.

Signaling for timing synchronization: The pilots consists of C_3 OFDM symbols, which are used to estimate of the delay of each device. Each symbol consists of B samples. For each OFDM symbol, we assign pseudonoise sequences on the frequencies such that the time-domain samples are Gaussian distributed. The synchronization can be achieved by performing the correlation between the received signal and the pilots.

3.4. Asynchronous Neighbor Discovery Algorithm

We first describe a robust bin detection that achieves two goals: (i) It can distinguish whether a frequency bin is a zeroton, a singleton, or a multiton bin; (ii) For singleton bins, it can detect the device index reliably. Then we describe the overall asynchronous neighbor discovery scheme.

3.4.1. Robust Singleton Detection

We focus on a certain device k that is hashed to a singleton bin $b_k = b$. The frequency value is given by $\boldsymbol{Y}_b = \left[\bar{\boldsymbol{Y}}_b^\dagger, \tilde{\boldsymbol{Y}}_b^\dagger, \dot{\boldsymbol{Y}}_b^\dagger\right]^\dagger$.

3.4.1.1. Channel Phase Estimation. We want to reliably estimate the phase of A_k. The underlying reason is that if the phase estimate is accurate enough, then we can further estimate the device index.

We use the first C_0 symbols in each subframe to estimate the phase of A_k as

$$\hat{\theta} = \angle\left(\frac{1}{C_0}\sum_{c=0}^{C_0-1}\bar{Y}_b^c\right). \tag{3.12}$$

Suppose device k transmits at a singleton frequency bin and C_0 is large enough, we can obtain an accurate estimate of the channel phase.

3.4.1.2. User Index Estimation. With the phase estimation $\hat{\theta}_k$, we can compensate the phase of A_k for each bin and try to decode the device index information. It can be seen that for singleton bins, the random transformation $\tilde{g}_k^c \to \mathrm{Re}\left\{\tilde{Y}_b^c e^{-\iota\hat{\theta}_k}\right\}$ is equivalent to a binary-input additive white Gaussian noise (BI-AWGN) channel. It will be shown that with a proper choice of the symbol number C_0, a large proportion of the signal strength can be preserved with high probability.

In order to robustly estimate the information bits $(k_0, \cdots, k_{\lceil \log N\rceil - 1})$, we apply error control codes to code over the bits. Instead of transmitting $\lceil \log N\rceil$ binary symbols, we transmit $C_1 = \lceil \log N\rceil / R$ binary symbols, where the symbols are the coded bits with code rate R. In particular, we construct $C_1 = \lceil \log N\rceil / R$ symbols with $\tilde{g}_k^c = (-1)^{r_{k,c}}$,

$c = 0 \cdots, C_1 - 1$, and

$$[r_{k,0}, \cdots, r_{k,C_1-1}] = [k_0, \cdots, k_{\lceil \log N \rceil - 1}]G, \tag{3.13}$$

where the operation is over the binary field and $G \in \mathbb{F}_2^{\lceil \log N \rceil \times C_1}$ is a generator matrix of an error-control code with rate R. We can apply the low-complexity capacity approaching codes [**60**]. For each bin b, we first perform hard decoding on $\text{Re}\left\{\tilde{Y}_b^c e^{-\iota\hat{\theta}_k}\right\}$ for each symbol c, further decode the sequence and then obtain the estimated index information. We focus on index estimation from singleton bins, which allows us to subsequently apply the well-studied point-to-point capacity approaching codes. Performing index estimation on a zeroton or multiton bin may produce false alarms. We will need a singleton verification step to prevent these false alarms.

3.4.1.3. Singleton Verification. Suppose a device has no collision on its frequency tone, its index information can be reliably estimated. This, however, is not true for multiton and zeroton bins. We need to provide a mechanism to verify if the estimated index comes from a singleton bin. We generate C_2 symbols in each subframe with $\dot{g}_k^c$ being i.i.d. Rademacher variables, i.e., $\mathsf{P}\{\dot{g}_k^c = \pm 1\} = 1/2$.

We consider the analysis on a fixed bin b. First, we claim bin b is a zeroton if the energy of the frequency bin value is low, i.e.,

$$||\dot{\boldsymbol{Y}}_b||_2^2 \leq \eta, \tag{3.14}$$

where η is some threshold constant. Suppose $\hat{k}$ is the estimated index from bin b. We perform the following validation process. We estimate the nonzero signal as

$$\dot{A}_{\hat{k}} = \frac{1}{C_2}\dot{\boldsymbol{g}}_{\hat{k}}^{\dagger}\dot{\boldsymbol{Y}}_b. \tag{3.15}$$

Then we claim that $\hat{k}$ is a correct estimate only if it passes the energy threshold test,

$$||\dot{\boldsymbol{Y}}_b - \dot{A}_{\hat{k}}\dot{\boldsymbol{g}}_{\hat{k}}||_2^2 \leq \eta. \tag{3.16}$$

The above validation scheme is similar to that used for sparse DFT and sparse WHT in [**56**,**61**]. The singleton verification approach proved to work for signal amplitudes lying in a known alphabet, whereas we show that it can effectively identify the singletons for *arbitrary* signal amplitudes that are bounded away from zero.

3.4.2. Overall Framework

Sparse OFDM effectively achieves random access over both frequency and time. The device index can be reliably estimated whenever it is hashed to a singleton frequency bin regardless of the delay. Once a device index is estimated, its contribution to the connected bins can be canceled, which may result in more singleton bins. This successive cancellation framework is similar to that proposed in [**12**]. There are, however, two challenges. First, the frequency bin values due to (3.6) depends on both the transmit frequency and its random delay. We need to estimate the delay in order to perform successive cancellation. Second, the residual error of channel estimation may propagate due to the successive cancellation process. For reliable neighbor recovery, we need to characterize the error propagation effects.

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

Input: Bin values $\boldsymbol{Y} = [\bar{\boldsymbol{Y}}, \tilde{\boldsymbol{Y}}, \dot{\boldsymbol{Y}}]$, 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: Estimated active device index i.
if $||\dot{\boldsymbol{Y}}||^2 < \eta$ **then**
  Declare zeroton and return $i \leftarrow \emptyset$.
end if
Phase estimation: $\hat{\theta} \leftarrow \text{phase}\left(\mathbf{1}^T \bar{\boldsymbol{Y}}/C_0\right)$.
Index location: $\boldsymbol{Z} \leftarrow \text{Re}\{\tilde{\boldsymbol{Y}} e^{-\iota\hat{\theta}}\}$,
$\hat{k} \leftarrow \text{Decoder}(\boldsymbol{Z})$.
Singleton verification:
$\dot{A}_{\hat{k}} \leftarrow \dot{\boldsymbol{g}}_{\hat{k}}^{\dagger} \dot{\boldsymbol{Y}}/C_2$.
if $||\dot{\boldsymbol{Y}} - \dot{A}_{\hat{k}} \dot{\boldsymbol{g}}_{\hat{k}}||_2^2 \leq \eta$ **then**
  Return $i \leftarrow \hat{k}$.
else
  Return $i \leftarrow \emptyset$.
end if

The first T subframes are used for bin detection. Denote the set of samples in the synchronization subframe as

$$\mathcal{I} = \{(B+M)CT, (B+M)CT+1, \cdots, (B+M)CT + BC_3 - 1\}. \tag{3.17}$$

Define the decision statistic as

$$\mathcal{T}(m) = \sum_{i \in \mathcal{I}} y_{i+m} s_{k,i}^*. \tag{3.18}$$

Assume no noise and correct delay estimate, $\mathcal{T}(m) = |a_k|^2 BC_3$. We estimate the delay of device k as

$$\hat{m}_k = \arg \max_{m=0,\cdots,M} |\mathcal{T}(m)|. \tag{3.19}$$

Algorithm 2 Asynchronous Neighbor Discovery via Sparse OFDM

Output: Detected active devices $\hat{\mathcal{K}}$.
Initialize: Set $\mathcal{B}$ as the set of unprocessed bins. Set $\mathcal{L} = \emptyset$.
Global singleton estimation:
for $b \in \mathcal{B}$ **do**
  $\hat{k} \leftarrow$ Robust-Bin-Detect $\left(\bar{\boldsymbol{Y}}_b, \tilde{\boldsymbol{Y}}_b, \dot{\boldsymbol{Y}}_b\right)$.
  if $\hat{k} \neq \emptyset$ **then**
    $\hat{\mathcal{K}} \leftarrow \hat{\mathcal{K}} \cup \{\hat{k}\}$
    $\mathcal{L} \leftarrow \mathcal{L} \cup \{\hat{k}\}$.
    $\mathcal{B} \leftarrow \mathcal{B} \backslash b$.
    Estimate $\hat{m}_{\hat{k}}$ and $\hat{a}_{\hat{k}}$ according to (3.19) and (3.20).
  end if
end for
Successive cancellation:
for $k \in \mathcal{L}$ **do**
  $\mathcal{L} \leftarrow \mathcal{L} \backslash k$.
  for Every bin $b \in \mathcal{B}$ that is connected with k **do**
    $\bar{\boldsymbol{Y}}_b \leftarrow \bar{\boldsymbol{Y}}_b - \hat{a}_k \exp\left(\frac{\iota 2\pi b(M-\hat{m}_k)}{B}\right) \mathbf{1}$.
    $\tilde{\boldsymbol{Y}}_b \leftarrow \tilde{\boldsymbol{Y}}_b - \hat{a}_k \exp\left(\frac{\iota 2\pi b(M-\hat{m}_k)}{B}\right) \tilde{\boldsymbol{g}}_k$.
    $\dot{\boldsymbol{Y}}_b \leftarrow \dot{\boldsymbol{Y}}_b - \hat{a}_k \exp\left(\frac{\iota 2\pi b(M-\hat{m}_k)}{B}\right) \dot{\boldsymbol{g}}_k$.
    $\hat{k} \leftarrow$ Robust-Bin-Detect $\left(\bar{\boldsymbol{Y}}_b, \tilde{\boldsymbol{Y}}_b, \dot{\boldsymbol{Y}}_b\right)$.
    if $\hat{k} \neq \emptyset$ **then**
      $\hat{\mathcal{K}} \leftarrow \hat{\mathcal{K}} \cup \{\hat{k}\}$
      $\mathcal{L} \leftarrow \mathcal{L} \cup \{\hat{k}\}$.
      $\mathcal{B} \leftarrow \mathcal{B} \backslash b$.
    end if
  end for
end for

Given an estimate of delay $\hat{m}_k$, the channel coefficient is estimated to be

$$\hat{a}_k = \frac{1}{C} \boldsymbol{g}_k^\dagger \boldsymbol{Y}_b e^{-\frac{\iota 2\pi b_k (M-\hat{m}_k)}{B}}. \tag{3.20}$$

The frequency value of the connected unprocessed bin b' is then updated according to

$$\boldsymbol{Y}_{b'} \leftarrow \boldsymbol{Y}_{b'} - \hat{a}_k \exp\left(\frac{\iota 2\pi b'(M-\hat{m}_k)}{B}\right)\boldsymbol{g}_k. \tag{3.21}$$

3.5. Proof of Theorem 7 (the Synchronous Case)

We prove Theorem 7 in the asymptotic regime as K increases without bound. Here is an outline of the proof using $O(\cdot)$ notation and a rigorous proof is provided in the Section 3.5.1 to 3.5.3. We choose some $B = O(K)$ and $T = O(1)$. In the codeword structure, the number of symbols is chosen as $C_0 = C_1 = C_2 = O(\log N)$. Due to synchronicity, there is no need to estimate the delays, so we set $C_3 = 0$. The total transmission length is $L = TBC = O(K \log N)$. For each subframe, we need to perform B-point FFT for $C = O(\log N)$ symbols. The total computational complexity is $O(K(\log K)(\log N))$ (FFT) operations. For each bin, the index estimation involves $O(\log N)$ operations and the phase estimation involves $O(\log N)$ operations. Since there are $O(K)$ bins, the total complexity due to phase estimation and index estimation is $O(K \log N)$. Since there are $T = O(1)$ subframes, the total computational complexity is $O(K(\log K)(\log N))$.

Neighbor discovery is said to fail if there is any miss or false alarm. The analysis for the bin detection error probability follows exactly as in [**62**] if the channel coefficient is known or is constrained to lie in a finite alphabet. The key challenge here is that the channel coefficient is an arbitrary unknown value and the residual estimation error may propagate through the successive cancellation process. We will prove that neighbor discovery fails with probability $O(1/K)$ for some $T = O(1)$, $B = O(\log K)$ and $C =$

$O(\log N)$. Specifically, the parameters are chosen as

$$T \geq 3 \tag{3.22}$$

$$B = \beta_0 K \tag{3.23}$$

$$C_0 = \lceil \log N \rceil \tag{3.24}$$

$$C_1 = \lceil \log N \rceil / R \tag{3.25}$$

$$C_2 = \beta_1 \lceil \log N \rceil, \tag{3.26}$$

where $\beta_0 \geq 2T(T-1)$, R is the code rate of a low-complexity binary-symmetric channel (BSC) capacity-approaching codes such that the transmission of $\log N$ bits under an SNR of $\underline{a}^2/(32\sigma^2/B)$ succeeds with probability higher than $1 - 1/N^2$ [**60**], and β_1 is some constant that will be specified later (Theorem 10). Given the parameter setting, the total number of OFDM symbols in each subframe is $C = (1 + \beta_1 + 1/R)\lceil \log N \rceil$. The correctness of neighbor discovery is established based on the following claims, whose proofs are provided in subsequent subsections.

Claim 1. *Let $\mathcal{G}$ denote the ensemble of bipartite graphs induced by sparse OFDM that consist of only trees and unicyclic components, and the largest component has fewer than $\beta_2 \log K$ left nodes, where β_2 is some constant depending on β_0. Given the parameter setting* (3.22)-(3.26)*, the induced bipartite graph $\mathsf{P}\{G \in \mathcal{G}\} \geq 1 - \gamma_0/K$, where γ_0 is some constant depending on β_0.*

Claim 2. *Given $G \in \mathcal{G}$ and the parameter setting* (3.22)-(3.26)*, the residual error of channel estimation are Gaussian variables with zero mean and variance bounded by $\beta_3\sigma^2/B$, where $\beta_3 = 8\beta_2/\beta_1$.*

Claim 3. *Given $G \in \mathcal{G}$ and the parameter setting* (3.22)-(3.26)*, the robust bin detection Algorithm 1 fails to identify a zeroton, a singleton, or a multiton with probability no greater than γ_1/K^2 for some γ_1.*

Suppose the robust bin detection does not make any error and $G \in \mathcal{G}$, then every device will be detected based on the successive cancellation process in Algorithm 2 [**63**]. Therefore, neighbor discovery fails only if either $G \notin \mathcal{G}$ or the robust bin detection makes an error throughout the detection process, denoted as error event E_b. The error probability of neighbor discovery is upper bounded as

$$\mathsf{P}_s = \mathsf{P}\{E_b \cup (G \notin \mathcal{G})\} \tag{3.27}$$

$$\leq \mathsf{P}\{E_b | G \in \mathcal{G}\} + P\{G \notin \mathcal{G}\}. \tag{3.28}$$

Every time a device is recovered, Algorithm 1 is performed on its connected bins, which is at most $T = O(1)$. Throughout the detection process, Algorithm 1 runs for at most KT times. By the union bound and the result of Claim 3,

$$\mathsf{P}\{E_b | G \in \mathcal{G}\} \leq \gamma_1 KT/K^2 = \gamma_1 T/K. \tag{3.29}$$

Combining Claim 1, (3.28) and (3.29), neighbor discovery fails with probability less than $(\gamma_0 + \gamma_1 T)/K$. Therefore, given the choice of T, B and C, neighbor discovery

fails with probability less than ϵ for $K \geq (\gamma_0 + \gamma_1 T)/\epsilon$. The required codelength is $BTC = \beta_0\beta_2 TK\lceil \log N \rceil$.

3.5.1. Proof of Claim 1

Consider the bipartite graph of sparse OFDM for multiple subframes in Fig. 3.2(b). The left nodes are randomly connected with a set of B bins in each subframe. If two node are connected with a common frequency bin, they would cause interference to each other in the successive peeling process due to the propagation of channel estimation error. We consider another bipartite graph, referred to as worst-interference bipartite graph where the frequency bins for T subframes collapse into one set of B bins. In the worst-interference graph, left node i is connected with bin j if it is connected with bin j at least once in some subframe. Fig. 3.2(c) illustrates an example of the worst-interference bipartite graph.

In the worst-interference bipartite graph, each device is randomly hashed to at most T out of B bins. The well-established results in random hypergraph show that with $B \geq 2T(T-1)K$ bins, the worst-interference bipartite graph consists of only trees and unicyclic components, and the largest component has fewer than $\beta_2 \log K$ left nodes with probability $1-\gamma_0/K$, where β_2 is some constant depending on β_0 [**59**,**64**]. The intuition is that as B gets larger, the bipartite graph becomes sparser and consists of a large number of isolated components. Since the number of paths from left node i to left node j in the original bipartite graph (e.g., Fig. 3.2(b)) must be fewer than that in the worst-interference bipartite graph (e.g., Fig. 3.2(c)), Claim 1 is proved.

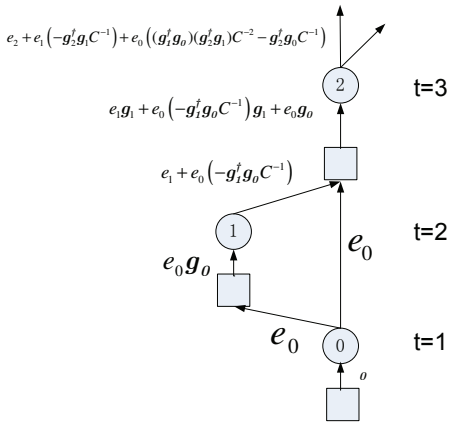

Figure 3.4. Error propagation graph for device 2.

3.5.2. Proof of Claim 2

We make use of the error propagation graph proposed in [**59**] to characterize the residual estimation error of channel estimation. For completeness, we describe the idea of error propagation graph in the following.

An error propagation graph for device k is a subgraph induced by the recovery algorithm, which contains the signal nodes that are estimated in the iterations before device k is recovered, and have paths to node k. Fig. 3.4 illustrates the the error propagation graph for device 2.

Define the channel estimation error of device k as

$$p_k = A_k - \hat{A}_k, \tag{3.30}$$

where A_k is given by (3.6) and $\hat{A}_k = \boldsymbol{g}^\dagger \boldsymbol{Y}/C$ is the estimate according to (3.20) with $M = m = 0$. Let b_k be the measurement bin used to recover the device k. Define the

point error of A_k as

$$e_k = -\frac{1}{C}\boldsymbol{g}_k^\dagger \boldsymbol{W}_{b_k}. \tag{3.31}$$

We will keep track of p_k using the error propagation graph. The estimation error can be calculated recursively according to some message passing rules over the graph. In particular, let p_k be the channel estimation error propagated from signal node k and $\boldsymbol{q}_j$ be the error vector of length-C propagated from the measurement bin j. The errors can be calculated according to the following rules:

$$p_k = e_k + \left(-C^{-1}\boldsymbol{g}_k^\dagger \boldsymbol{q}_{\text{in}(k)}\right) \tag{3.32}$$

$$\boldsymbol{q}_j = \sum_{k \in \text{in}(j)} p_k \boldsymbol{g}_k, \tag{3.33}$$

where $\text{in}(k)$ denotes the indices of the measurement bins (signal nodes) incoming to signal node (measurement bin) k. Since we use one singleton bin to decode the device index, the input message to a signal node k is from one measurement bin $\text{in}(k)$.

Let $S(t)$ denote the signal indices that are recovered in the t-th iteration. We first show that (3.32) and (3.33) hold for any $k \in S(1) \cup S(2)$ and then show that they hold for any k by induction. Consider the estimation of A_k, $k \in S(1)$. The frequency value of bin b_k and the residual estimation error are given by

$$\boldsymbol{Y}_{b_k} = A_k \boldsymbol{g}_k + \boldsymbol{W}_{b_k} \tag{3.34}$$

$$p_k = e_k. \tag{3.35}$$

Consider the estimation for A_k, $k \in S(2)$. With the successive cancellation process, for $k \in S(1)$, the updated measurement vector of b_k and the estimation error becomes

$$\boldsymbol{Y}_{b_k} = A_k \boldsymbol{g}_k + \boldsymbol{W}_{b_k} + \sum_{\ell \in S(1):\ell \text{ connected with } b_k} e_\ell \boldsymbol{g}_\ell \tag{3.36}$$

$$p_k = e_k + \sum_{\ell \in S(1):\ell \text{ connected with } b_k} e_\ell \left(-C^{-1} \boldsymbol{g}_k^\dagger \boldsymbol{g}_\ell\right) \tag{3.37}$$

$$= e_k + \left(-C^{-1} \boldsymbol{g}_k^\dagger \boldsymbol{q}_{b_k}\right), \tag{3.38}$$

where $\boldsymbol{q}_{b_k} = \sum_{\ell \in \text{ in } (b_k)} e_\ell \boldsymbol{g}_\ell$. It is important to note that $|\boldsymbol{g}_k^\dagger \boldsymbol{g}_\ell C^{-1}| \leq 1$ for every realization of the design vectors.

Suppose message passing rules (3.32) and (3.33) hold for $k \in S(t-1)$. For $k \in S(t)$, with successive cancellation, the updated frequency value is

$$\boldsymbol{Y}_{b_k} = A_k \boldsymbol{g}_k + \boldsymbol{W}_{b_k} + \sum_{\ell \in \text{in}(b_k)} p_\ell \boldsymbol{g}_\ell. \tag{3.39}$$

The channel estimation error is

$$p_k = A_k - \frac{1}{C} \boldsymbol{g}_k^\dagger \boldsymbol{Y}_{b_k} \tag{3.40}$$

$$= e_k - \sum_{\ell \in \text{in}(b_k)} C^{-1} p_\ell \boldsymbol{g}_k^\dagger \boldsymbol{g}_\ell \tag{3.41}$$

$$= e_k - C^{-1} \boldsymbol{g}_k^\dagger \boldsymbol{q}_{b_k}, \tag{3.42}$$

where (3.42) follows from the definition of $\boldsymbol{q}_{b_k}$. In the error propagation graph, $b_k = \text{in}(k)$ and the message passing rules are thus proved to hold for any k by induction.

By the error message passing rules (3.32) and (3.33), the estimation error of A_k, $k \in S(t)$, is calculated as

$$p_k = e_k + \sum_{\ell \in \cup_{j=1}^{t-1} S(j) \cap D(k)} \left(\sum_{p=1}^{P(\ell,k)} d_{\ell,p} \right) e_\ell, \tag{3.43}$$

where $D(k)$ be the connected subgraph of the bipartite graph containing node k, $\mathcal{P}(\ell, k)$ is the number of paths from node ℓ to node k in $D(k)$, and $d_{\ell,p}$ is some coefficient depending on both the design parameters $\{\boldsymbol{g}_k\}$ and the path, which satisfies $|d_{\ell,p}| \leq 1$.

Fig. 3.4 illustrates an example. The number of paths from node 0 to node 2 is $\mathcal{P}(0,2) = 2$, with the corresponding coefficients being $d_{0,1} = -\boldsymbol{g}_2^\dagger \boldsymbol{g}_0 / C$ and $d_{0,2} = \boldsymbol{g}_1^\dagger \boldsymbol{g}_0 \boldsymbol{g}_2^\dagger \boldsymbol{g}_1 / C^2$. The number of paths from node 1 to node 2 is $\mathcal{P}(1,2) = 1$, with the coefficients being $d_{1,1} = -\boldsymbol{g}_2^\dagger \boldsymbol{g}_1 / C$.

Suppose $G \in \mathcal{G}$, the frequency value for any singleton bin b_k for device k can be written as

$$\boldsymbol{Y}_{b_k} = A_k \boldsymbol{g}_k + \boldsymbol{W}_{b_k} + \boldsymbol{V}_{b_k}, \tag{3.44}$$

where $\boldsymbol{V}_{b_k}$ is the interference on bin b_k due to the channel estimation residual errors. Mathematically, it can be calculated as

$$\boldsymbol{V}_{b_k} = \sum_{\ell \in \cup_{j=1}^{t-1} S(j) \cap D(k)} \sum_{p=1}^{P(\ell,k)} e_\ell d_{\ell,p} \boldsymbol{g}_{\ell_p}, \tag{3.45}$$

where $\ell_p \in \text{in}(b_k)$ depends on the path from ℓ to bin b_k. The point error e_ℓ is given by (3.31). It is easy to see that every point error is independent of the design parameters $\{\boldsymbol{g}_k\}$ of all the devices and is distributed according to $\mathcal{CN}(0, 2\sigma^2/BC)$.

Suppose the bipartite graph belongs to $\mathcal{G}$, $P(\ell, k)$ is less than or equal to 2. Moreover, by Claim 1, the number of left nodes in each component is less than $\beta_2 \log K$. Conditioned on the design parameters g_ℓ of the previously identified devices, each entry of $\boldsymbol{V}_{b_k}$ is Gaussian variable with zero mean and variance bounded by $8\beta_2 \log K\sigma^2/BC = \beta_3\sigma^2/B$, where $\beta_3 = 8\beta_2/\beta_1$.

3.5.3. Proof of Claim 3

Let $E_{b,0}, E_{b,1}, E_{b,2}$ denote the failure of robust bin detection for a zeroton, singleton and multiton, respectively. Suppose $G \in \mathcal{G}$, we will show that with a proper choice of the threshold η in Algorithm 2, the error probabilities can be bounded as $\mathsf{P}\{E_{b,i}\} = O(1/K^2)$, $i = 0, 1, 2$. Then the bin detection error is less than γ_1/K^2 for some γ_1.

As described in the proof of Claim 2, the frequency values of a bin b can be written as

$$\boldsymbol{Y}_b = \sum_{k\in\mathcal{K}:b_k=b} A_k\boldsymbol{g}_k + \boldsymbol{W}_b + \boldsymbol{V}_b, \tag{3.46}$$

where the sum is over the set of active devices that are hashed to frequency bin b and not yet recovered, and $\boldsymbol{V}_b$ is due to the residual channel estimation errors from the recovered devices. The detection error depends on $\boldsymbol{V}_b$ and hence on the number of devices that cause interference. Let $\boldsymbol{Z}_b = \boldsymbol{W}_b + \boldsymbol{V}_b$ denote the interference plus noise. We set the energy thresholds as

$$\eta = C_2\tau_0/\sqrt{B} \tag{3.47}$$

where τ_0 is some constant that will be specified later.

3.5.3.1. Zeroton Error Detection. The zeroton error $E_{b,0}$ occurs only if $||\dot{\boldsymbol{Y}}_b||$ is greater than the threshold η. Thus,

$$\mathsf{P}\{E_{b,0}\} = \mathsf{P}\left\{||\dot{\boldsymbol{Z}}_b||_2^2 \geq \eta\right\} \tag{3.48}$$

$$\leq C_2 \mathsf{P}\left\{|\dot{Z}_{b,c}|^2 \geq \frac{\eta}{C_2}\right\}. \tag{3.49}$$

Let $\underline{\boldsymbol{g}}$ denote the set of design parameters $\boldsymbol{g}_k$ of all the previously identified devices. Conditioned on $\underline{\boldsymbol{g}}$, each entry of $\boldsymbol{Z}_b$ is distributed according to $\mathcal{CN}(0, 2\sigma_z^2)$, where $\sigma_z \leq (1+\beta_3/2)\sigma^2/B$. The probability of passing the energy threshold is

$$\mathsf{P}\left\{|\dot{Z}_{b,c}|^2 \geq \frac{\eta}{C_2}\middle|\underline{\boldsymbol{g}}\right\} \leq 2\mathsf{P}\left\{|\mathrm{Re}\{\dot{Z}_{b,c}\}|^2 \geq \frac{\eta}{2C_2}\middle|\underline{\boldsymbol{g}}\right\} \tag{3.50}$$

$$\leq 4e^{-\frac{\eta}{4\sigma_z^2 C_2}} \tag{3.51}$$

$$\leq 4e^{-\frac{B\eta}{(4+2\beta_3)\sigma^2 C_2}}. \tag{3.52}$$

With fixed B and C_2, there exists τ_0 such that $C_2 4e^{-\frac{B\eta}{(4+2\beta_3)\sigma^2 C_2}} \leq 1/K^2$. Moreover, (3.52) holds for every realization of $\underline{\boldsymbol{g}}$, by averaging $\underline{\boldsymbol{g}}$, we have

$$\mathsf{P}\left\{||\dot{\boldsymbol{Z}}_b||_2^2 \geq \eta\right\} \leq C_2 \mathsf{P}\left\{|\dot{Z}_{b,c}|^2 \geq \frac{\eta}{C_2}\right\} \tag{3.53}$$

$$\leq \frac{1}{K^2}. \tag{3.54}$$

Combining (3.49) and (3.54), we have

$$\mathsf{P}\{E_{b,0}\} \leq \frac{1}{K^2}. \tag{3.55}$$

3.5.3.2. Singleton Error Detection. Suppose device k is hashed to a singleton bin b. Let $E_{b,1}$ denote the bin detection error. A singleton detection error occurs due to three events: (1) $E_{b,1,0} = \left\{||\dot{\boldsymbol{Y}}_b||_2^2 < \eta\right\}$; (2) $E_{b,1,1} = \left\{||\dot{\boldsymbol{Y}}_b - \dot{A}_k\dot{\boldsymbol{g}}_k||_2^2 > \eta\right\}$; (3) $E_{b,1,2} = \left\{\hat{k}_b \neq k\right\}$. Thus $E_{b,1} \subseteq E_{b,1,0} \cup E_{b,1,1} \cup E_{b,1,2}$.

By large deviation, with high probability $||\dot{\boldsymbol{Y}}_b||$ is concentrated around $C_2(|a_k|^2+2\sigma_z^2)$. Since τ_0 is some fixed constant, $2\sigma_z^2\tau_0$ is smaller than $|a_k|^2$ for large enough K. Following a similar derivation in [**56**], it can be shown that

$$\mathsf{P}\left\{E_{b,1,0}\right\} = O\left(\frac{1}{K^2}\right). \tag{3.56}$$

We next upper bound the probability of $E_{b,1,1}$. Let $\dot{A}_k = \dot{\boldsymbol{g}}_k^\dagger\dot{\boldsymbol{Y}}_b/C_2$. We have

$$\dot{\boldsymbol{Y}}_b - \dot{A}_k\dot{\boldsymbol{g}}_k = \left(\boldsymbol{I} - \frac{1}{C_2}\dot{\boldsymbol{g}}_k\dot{\boldsymbol{g}}_k^\dagger\right)\dot{\boldsymbol{Z}}_b. \tag{3.57}$$

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

$$\boldsymbol{Q} = \boldsymbol{U}\Lambda\boldsymbol{U}^\dagger \tag{3.58}$$

$$= \boldsymbol{U}\mathrm{diag}\{0, 1, \cdots, 1\}\boldsymbol{U}^\dagger \tag{3.59}$$

where $\boldsymbol{U}$ is a unitary matrix. Then we have

$$||\dot{\boldsymbol{Y}}_b - \dot{A}_k \dot{\boldsymbol{g}}_k||_2^2 = ||\boldsymbol{Q}\dot{\boldsymbol{Z}}_b||_2^2 \tag{3.60}$$

$$= ||\Lambda \boldsymbol{U}^\dagger \dot{\boldsymbol{Z}}_b||_2^2 \tag{3.61}$$

$$= \sum_{c=1}^{C_2-1} |Z'_c|^2, \tag{3.62}$$

where $\boldsymbol{Z}' = \boldsymbol{U}^\dagger \dot{\boldsymbol{Z}}_b$ and (3.62) is due to $\Lambda = \text{diag}(0, 1, \cdots, 1)$. Since $||\dot{\boldsymbol{Z}}_b||_2^2 = ||\boldsymbol{Z}'||_2^2$, we have

$$\mathsf{P}\left\{||\dot{\boldsymbol{Y}}_b - \dot{A}_k \dot{\boldsymbol{g}}_k||_2^2 \geq \eta\right\} = \mathsf{P}\left\{\sum_{c=1}^{C_2-1} |Z'_c|^2 \geq \eta\right\} \tag{3.63}$$

$$\leq \mathsf{P}\left\{||\dot{\boldsymbol{Z}}_b||_2^2 \geq \eta\right\} \tag{3.64}$$

$$\leq \frac{1}{K^2}, \tag{3.65}$$

where (3.65) follows from (3.54). Therefore, we have

$$\mathsf{P}\left\{E_{b,1,1}\right\} \leq \frac{1}{K^2}. \tag{3.66}$$

We next bound the error probability of $E_{b,1,2}$. We first show that the phase compensation is accurate with high probability. Second, we show that the interference only causes a slight SNR degradation 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 $\theta = \angle A_k$ as (3.12). Let $\bar{Z} = \sum_{c=0}^{C_0-1} \left(\bar{W}_b^c + \bar{V}_b^c\right)/C_0$. Then the estimated phase is calculated as

$$\hat{\theta} = \angle\left(A_k + \bar{Z}\right). \tag{3.67}$$

From the geometric interpretation, the maximum phase offsets occurs when the noise is orthogonal to the measurement. We choose a small θ_0 such that $\theta_0 < \frac{\pi}{3}$ and $\sin\theta_0 > \theta_0/2$, then

$$\mathsf{P}\left\{|\hat{\theta} - \theta| > \theta_0\right\} \leq \mathsf{P}\left\{\arcsin\frac{|\bar{Z}|}{|A_k|} > \theta_0\right\} \tag{3.68}$$

$$\leq \mathsf{P}\left\{|\bar{Z}| > \underline{a}\sin\theta_0\right\} \tag{3.69}$$

$$\leq \mathsf{P}\left\{|\bar{Z}| > \frac{\underline{a}\theta_0}{2}\right\} \tag{3.70}$$

$$\leq \mathsf{P}\left\{|\mathrm{Re}\{Z\}| > \frac{\underline{a}\theta_0}{4}\right\} + \mathsf{P}\left\{|\mathrm{Im}\{Z\}| > \frac{\underline{a}\theta_0}{4}\right\} \tag{3.71}$$

$$\leq 4\exp\left(-\frac{\underline{a}^2\theta_0^2}{32\sigma_z^2}\right), \tag{3.72}$$

where (3.72) is due to $\mathsf{P}\{\mathcal{N}(0,1) > x\} \leq e^{-x^2/2}$.

Therefore, given $B = \beta_0 K$, with high probability $1 - e^{-\Omega(K)}$,

$$\mathrm{Re}\left\{A_k e^{-\iota\hat{\theta}}\right\} = \underline{a}\cos(\hat{\theta} - \theta) \tag{3.73}$$

$$\geq \underline{a}\cos\theta_0 \tag{3.74}$$

$$\geq \underline{a}/2. \tag{3.75}$$

We consider the corruption of signal strength from interference $\boldsymbol{V}$. Since $\tilde{V}_b^c$ is Gaussian distributed with variance less than or equal to $\beta_3\sigma^2/B$, $\mathsf{P}\left\{|\mathrm{Re}\{\tilde{V}_b^c\}| \geq \underline{a}/4\right\} \leq e^{-\Omega(K)}$. Combining (3.75), we have $\mathrm{Re}\left\{A_k e^{-\iota\hat{\theta}}\tilde{g}_k^c + \tilde{V}_b^c\right\} = c\tilde{g}_k^c$ with $c \geq \underline{a}/4$ for all $c = 0, \cdots, C_1 - 1$ with probability higher than $1 - C_1 e^{-\Omega(K)}$. Conditioned on this, the device index transmission corrupted by noise $\boldsymbol{W}_b$ can be regarded as transmission over BSC channel with an SNR at least $(\underline{a}/4)^2/(2\sigma^2/B) = \underline{a}^2/(32\sigma^2/B)$. Since the error-control code with rate R used to encode the device index information $(k_0, \cdots, k_{\lceil \log N \rceil - 1})$ is chosen such that the index can be recovered correctly with probability at least $1 - 1/N^2$, the singleton error occurs with probability

$$\mathsf{P}\left\{E_{b,1,2}\right\} \leq \frac{1}{N^2}. \tag{3.76}$$

By (3.56), (3.66) and (3.76), we conclude

$$\mathsf{P}\left\{E_{b,1}\right\} \leq \frac{\gamma'}{N^2}, \tag{3.77}$$

for some γ'.

3.5.3.3. Multiton Error Detection. It is proved in Appendix B.1 that

$$\mathsf{P}\left\{E_{b,2}\right\} \leq 1/K^2. \tag{3.78}$$

Therefore, combining (3.55), (3.77) and (3.78), we conclude that the robust bin detection correctly identify the zeroton, singleton or multiton with probability higher than $1 - \gamma_1/K^2$ for some γ_1.

3.6. Proof of Theorem 8

In the asynchronous neighbor discovery case, we choose the parameters according to (3.22)–(3.26). The number of OFDM symbols used for synchronization is $C_3 = \beta_3 \lceil \log(K+M) \rceil$, where β_3 are specified later. In the codeword structure, the number of symbols in each subframe is $C = O(\log N)$. The total transmission length in transmit symbols is thus

$$L = T(B+M)C + BC_3 \tag{3.79}$$

$$= O\left((K+M)\log N + K\log(K+M)\right). \tag{3.80}$$

The FFT operation and channel estimation involve the same number of operations as the synchronous. Different from the synchronous, each device needs to estimate its delay once. The complexity of delay estimation is $MKC_3 = O(MK\log(K+M))$ (corresponding to M times auto-correlations). A total of K devices need to estimate their delays. The total computational complexity is thus $O\left(K(\log K)(\log N)\right) + O\left(K^2 M \log(K+M)\right)$.

The following lemma shows that for each device the delay estimate is correct with high probability.

Lemma 6. *Suppose the conditions specified in Theorem 8 hold. Suppose the bipartite graph $G \in \mathcal{G}$ and the parameters are chosen according to* (3.22)–(3.26)*, there exists some positive β_3 such that $C_3 = \beta_3 \lceil \log(K+M) \rceil$ OFDM symbols are used for timing synchronization and the delay of a device estimated according to* (3.19) *is correct with probability $O(1/K^2)$.*

Proof. See Appendix B.2. □

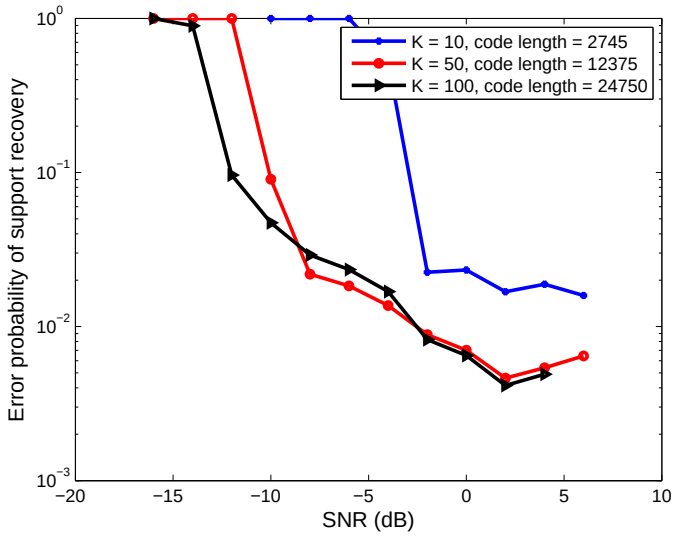

Figure 3.5. Error probability of support recovery in the case of synchronous transmission. The device population is $N = 2^{38}$.

By the union bound and Lemma 6, the delay of each device can be correctly estimated with probability $1-O(1/K)$. Conditioned on that the device delays are correctly detected, the residual errors of channel estimation can be exactly characterized as the synchronous case. The proof for correct asynchronous neighbor discovery follows that for synchronous case.

3.7. Simulation Results

3.7.1. Synchronous neighbor discovery

We simulate the error probability of asynchronous neighbor discovery via sparse OFDM. The total number of devices is $N = 2^{38}$. The frame consists of $T = 3$ subframes. The number of measurement bins is $B = \lceil 1.5K \rceil$, where K is the number of active devices. The number of OFDM symbols for phase estimation singleton verification are set as $C_0 = 6$

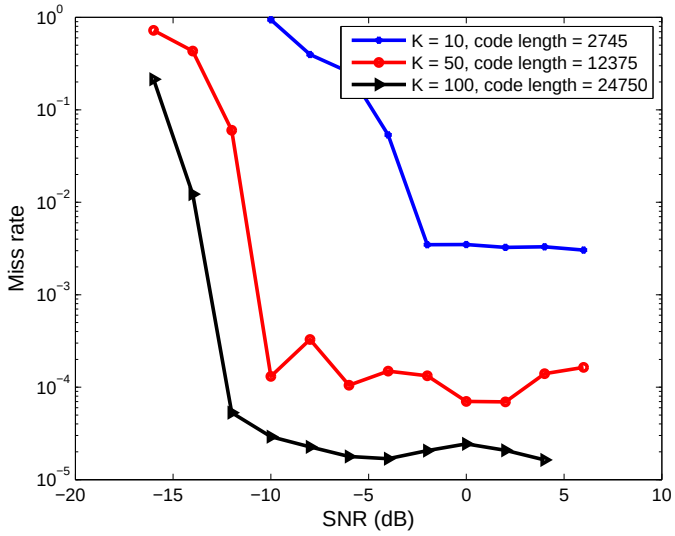

Figure 3.6. Rate of missed detection in the case of synchronous transmission. The device population is $N = 2^{38}$.

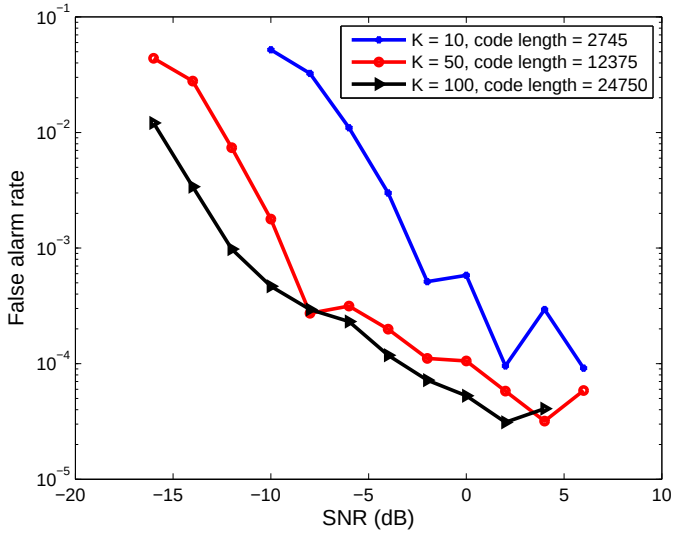

Figure 3.7. Rate of false alarm in the case of synchronous transmission. The device population is $N = 2^{38}$.

and $C_2 = 6$, respectively. We adopt a rate $R = 0.9$ random LDPC code as subcode. The

number of OFDM symbols carrying device index information is thus $C_1 = \log N/R$. The number of OFDM symbols used in synchronization is $C_3 = 8$.

Fig. 3.5 shows the error probability of support recovery for asynchronous neighbor discovery. In each simulation, if there exists missed detection or false alarm, it claims to have an error. Fig. 3.6 and Fig. 3.7 show the miss and false alarm rates, respectively. We define the missed detection (false alarm) rate as the average number of misses (false alarms) in each simulation normalized by the number of devices.

Simulation shows that under SNR of 6 dB, in order to achieve miss detection and false alarm rate of 10^{-4}, the transmission length required to identify $K = 100$ out of 2^{38} devices is around 25000 samples. In the case of a 20 MHz channel bandwidth, the transmission time is approximately 1.25 ms.

3.7.2. Discrete delay

We simulate the error probability of asynchronous neighbor discovery via sparse OFDM. The system parameters are the same as in the synchronous setting. The device population is $N = 2^{38}$ and the maximum delay in terms of transmit samples is $M = 20$ still. The delay of each device transmission is assumed to be multiples of T_s. Fig. 3.8 shows the error probability of support recovery for asynchronous neighbor discovery. Fig. 3.9 and Fig. 3.10 show the missed detection rate and false alarm rate, respectively. As in the synchronous setting, the error probability is low under moderate SNR, which confirms our theoretical analysis. In order to achieve a similar error performance, the required transmission length is more than that of synchronous setting. As the number of active devices K increases, the increase of transmission length due to delay becomes less.

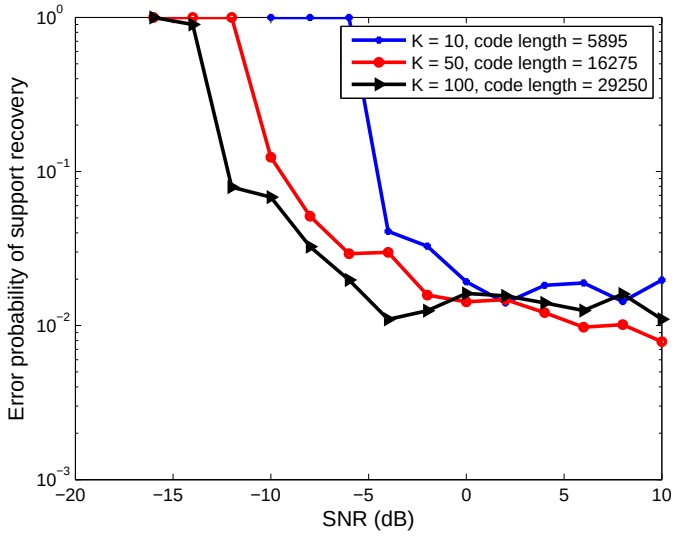

Figure 3.8. Error probability of support recovery in the case of discrete delay. The device population is $N = 2^{38}$ and the maximum delay is $M = 20$.

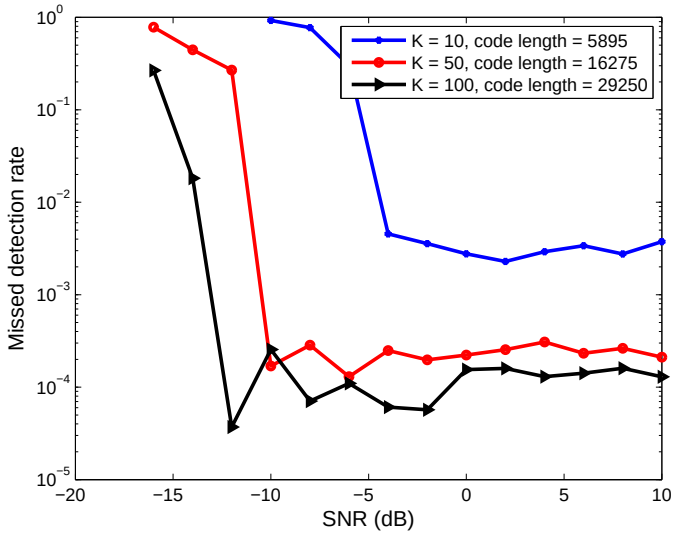

Figure 3.9. Rate of missed detection in the case of discrete delay. The device population is $N = 2^{38}$ and the maximum delay is $M = 20$.

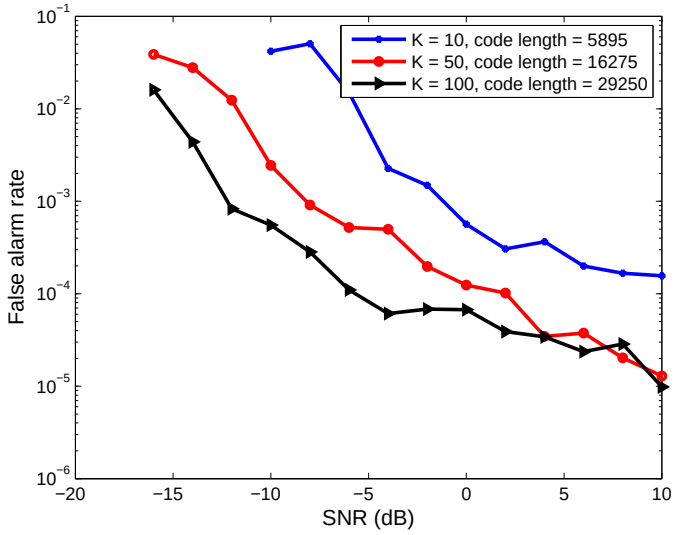

Figure 3.10. Rate of false alarm in the case of discrete delay. The device population is $N = 2^{38}$ and the maximum delay is $M = 20$.

3.7.3. Comparison with random access

3.7.3.1. 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:

$$\mathsf{P}_{\text{miss,aloha}} = \left(1 - (1-p)^{K-1}p\right)^{N_s}. \tag{3.81}$$

Setting $p = 1/K$ minimizes $\mathsf{P}_{\text{miss,aloha}}$. Suppose each slot consists of (only) 25 symbols.

3.7.3.2. 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 to transmit is

$$\mathsf{P}_{\text{miss,csma}} = \left(1 - \mathsf{P}\left\{T_1 < \min\{T_2, \cdots, T_K\}\right\}\right)^{N_s}. \tag{3.82}$$

In order to reliably transmit the device index $\log N$ bits, the number of symbols required in each frame is at least $\log N / \log(1 + \text{SNR})$. Therefore, depending on the achieved missed detection rate, the total number of symbols required is $N_s \log N / \log(1 + \text{SNR})$. Under SNR = -4 dB, it can be seen from Fig. 3.9 and Fig. 3.10 that sparse OFDM can achieve missed detection and false alarm rate low than 10^{-4}. The advantage of sparse OFDM over random access becomes more obvious as the number of active devices increases. For example, when $K = 50$, the transmission length of sparse OFDM is around 16000, while slotted ALOHA and CSMA requires more than 35000 symbols to achieve a missed detection rate of 10^{-4}. Sparse OFDM can effectively reduce the transmission length by over 50%. Moreover, the rate reduction is even greater for larger K and a lower error probability requirement.

3.8. Conclusion

We have proposed a low-complexity asynchronous neighbor discovery scheme for very large networks with applications to the Internet of Things. The scheme, referred to as sparse OFDM, applies the recently developed sparse Fourier transform to compressed neighbor discovery. Compared with random access schemes, sparse OFDM requires much

shorter transmission length by exploiting the multiaccess nature of the channel and the multiuser detection gain. Sparse OFDM adopts well-established point-to-point capacity approaching codes and involves low complexity. It provides practical physical layer capability for multipacket reception and it would be a useful next step to extend this technique to the design of asynchronous neighbor discovery network protocols.

CHAPTER 4

Conclusion and Future Work

In this thesis, a novel many-user paradigm has been proposed, where the number of users in the system scale with the blocklength. The motivation is to model emerging communication systems with massive access. Important applications include the Internet of Things, machine-to-machine communications and sensor networks, where the number of users is comparable or even exceeds the blocklength.

As a first step towards the many-user information theory, we have studied in Chapter 2 the Gaussian many-access channel, which consists of a single receiver and many transmitters. The transmitters access the channel with random on-off patterns. The many-access channel can be used to model the uplink transmission in the Internet of Things. A new notion of channel capacity has been defined and the symmetric capacity of the Gaussian many-access channel is derived. One achievability scheme is to first detect the active users and then decode their messages.

One insight from the study of the fundamental limit of the many-access channel is that user identification, also known as neighbor discovery, is a crucial step to achieve the capacity. In Chapter 3, a low-complexity asynchronous neighbor discovery scheme, termed as sparse OFDM, has been proposed. In conventional OFDM, a large number of narrow subcarriers are needed to schedule all the devices to transmit at the same time. Instead, sparse OFDM divides the spectrum into a small number of sparsely spaced subcarriers and randomly assign the subcarriers to the users. Sparse OFDM can provide reliable

neighbor discovery for asynchronous transmissions. Moreover, it is of low complexity and low overhead.

We conclude the thesis by highlighting the future research directions.

Many-user source coding: Shannon's lossless source coding problem is concerned with the minimum number of bits needed for representing a discrete memoryless source. The key to the answer is the asymptotic equipartition property (AEP) of the i.i.d. sequence $X_1, \ldots, X_n$ with distribution P_X as $n \to \infty$, which states that there are approximately $2^{nH(X)}$ typical sequences, whose probabilities are similar, and which collectively almost exhaust all the probability. Thus, the minimum compression ratio is essentially $H(X)$ bits per symbol.

With multiple correlated sources, the classical distributed lossless source coding problem is solved by understanding the joint AEP of correlated sequences [**65**]. It is interesting to develop a theory of distributed many-source coding, for the situation of k (correlated) sources, where k is so large that it is comparable to the length of the source sequences n. The central question is what is the minimum number of bits it takes, as n becomes large, to encode the k_n sources distributedly, so that they can be reconstructed by a decoder.

Finte block length analysis: The asymptotic results of many-user information theory will be used to approximate the performance of systems with large, albeit finite block-length and number of users. Finite blocklength analysis of the capacity will provide more insight by analyzing the back-off from capacity for a fixed blocklength and an error probability [**38**].

Finite-blocklength analysis of some multiuser channels has been carried out in [**66**–**69**], where the dispersion terms quickly become intractable as the number of users increases.

In the many-user paradigm, with blocklength n and k_n users, it will be interesting to investigate the first-order scaling of the maximum achievable message length as a function of the blocklength and the error probabilities.

Neighbor discovery for symbol-asynchronous models: In our work of asychronous neighbor discovery, we assume a frame-asynchronous symbol-synchronous model, where the delay of each device is in the unit of symbol interval. This assumption allows successful estimation of the delay such that successive cancellation can be applied. In practice, however, it is hard to achieve symbol synchronicity due to arbitrary propagation delay or clock mismatch. It will be interesting to study the performance of neighbor discovery for the symbol-asynchronous model and the performance tradesoff between the delay and transmission length.

References

[1] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "A first look at cellular machine-to-machine traffic: large scale measurement and characterization," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 1, 2012, pp. 65–76.

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[3] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, 1999.

[4] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.

[5] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *submitted to IEEE Trans. Inf. Theory*, 2016. [Online]. Available: https://arxiv.org/abs/1607.01048

[6] 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.

[7] S. A. Borbash, A. Ephremides, and M. J. McGlynn, "An asynchronous neighbor discovery algorithm for wireless sensor networks," *Ad Hoc Networks*, vol. 5, no. 7, pp. 998–1016, 2007.

[8] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili, "Neighbor discovery in wireless networks and the coupon collector's problem," in *Proc. ACM Mobicom*, Beijing, China, 2009, pp. 181–192.

[9] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A. C.-C. Hsu, "Overload control for machine-type-communications in lte-advanced system," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 38–45, 2012.

[10] 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.

[11] 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.

[12] S. Pawar and K. Ramchandran, "Computing a k-sparse n-length discrete Fourier transform using at most 4k samples and O (k log k) complexity," in *Proc. IEEE Int. Symp. Inform. Theory*, Istanbul, 2013, pp. 464–468.

[13] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," in *Proc. IEEE Int. Symp. Inform. Theory*, 1997, p. 150.

[14] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, 1996.

[15] S.-C. Chang and E. Weldon, "Coding for t-user multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 684–691, 1979.

[16] X. Chen and D. Guo, "Gaussian many-access channels: Definition and symmetric capacity," in *Proc. IEEE Information Theory Workshop*, Sevilla, Spain, 2013, pp. 1–5.

[17] ——, "Many-access channels: The Gaussian case with random user activities," in *Proc. IEEE Int. Symp. Information Theory*, Honolulu, HI, June 2014, pp. 3127–3131.

[18] S. Shahi, D. Tuninetti, and N. Devroye, "On the capacity of strong asynchronous multiple access channels with a large number of users," in *Proc. IEEE Int. Symp. Information Theory*, Barcelona, Spain, July 2016.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley-interscience, 2006.

[20] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Information Theory*, 1971, pp. 23–52.

[21] H. Liao, "A coding theorem for multiple access communications," in *Proc. IEEE Int. Symp. Information Theory*, Asilomar,CA, 1972.

[22] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, 1985.

[23] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[24] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[25] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[26] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[27] ——, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[28] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, 2009.

[29] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, 2010.

[30] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, 2010.

[31] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.

[32] K. R. Rad, "Nearly sharp sufficient conditions on exact sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4672–4679, 2011.

[33] 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.

[34] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[35] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified shannon theoretic approach," in *Proc. IEEE Information Theory Workshop*, Sevilla, 2013, pp. 1–5.

[36] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2010.

[37] R. Arratia and L. Gordon, "Tutorial on large deviations for the binomial distribution," *Bulletin of mathematical biology*, vol. 51, no. 1, pp. 125–131, 1989.

[38] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite block-length regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[39] E. MolavianJazi and J. N. Laneman, "On the second-order cost of TDMA for Gaussian multiple access," in *Proc. IEEE Int. Symp. Information Theory*, Honolulu, HI, June 2014, pp. 266–270.

[40] 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.

[41] 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.

[42] R. Xie, H. Yin, X. Chen, and Z. Wang, "Many access for small packets based on precoding and sparsity-aware recovery," *arXiv preprint arXiv:1510.06454*, 2015.

[43] C. MacGillivray, V. Turner, and D. Lund, "Worldwide Internet of Things (IoT) 2013–2020 forecast: Billions of things, trillions of dollars," *Market Analysis*, vol. 243661, 2013.

[44] M. J. McGlynn and S. A. Borbash, "Birthday protocols for low energy deployment and flexible neighbor discovery in ad hoc wireless networks," in *Proc. ACM MobiHoc*, Long Beach, CA, 2001, pp. 137–145.

[45] 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.

[46] S. Vasudevan, J. Kurose, and D. Towsley, "On neighbor discovery in wireless networks with directional antennas," in *Proc. IEEE INFOCOM*, vol. 4, Miami, FL, March 2005, pp. 2502–2512.

[47] S. Chen, A. Russell, R. Jin, Y. Qin, B. Wang, and S. Vasudevan, "Asynchronous neighbor discovery on duty-cycled mobile devices: Integer and non-integer schedules," in *Proc. ACM MobiHoc*, Hangzhou, China, 2015, pp. 47–56.

[48] J. Ni, R. Srikant, and X. Wu, "Coloring spatial point processes with applications to peer discovery in large wireless networks," *IEEE/ACM Trans. Networking*, vol. 19, no. 2, pp. 575–588, 2011.

[49] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, "Flashlinq: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Trans. Networking*, vol. 21, no. 4, pp. 1215–1228, 2013.

[50] R. De Gaudenzi, O. del Rio Herrero, G. Acar, and E. G. Barrabés, "Asynchronous contention resolution diversity aloha: Making crdsa truly asynchronous," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6193–6206, 2014.

[51] F. Clazzer, F. Lazaro, G. Liva, and M. Marchese, "Detection and combining techniques for asynchronous random access with time diversity," *arXiv preprint arXiv:1604.06221*, 2016.

[52] E. Sandgren, F. Brännström *et al.*, "On frame asynchronous coded slotted aloha: Asymptotic, finite length, and delay analysis," *arXiv preprint arXiv:1606.03242*, 2016.

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

[54] X. Chen and D. Guo, "Robust sublinear complexity Walsh Hadamard transform with arbitrary sparse support," in *Proc. IEEE Int. Symp. Information Theory*, Hong Kong, June 2015.

[55] 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.

[56] 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.

[57] W. Zeng, S. Vasudevan, X. Chen, B. Wang, A. Russell, and W. Wei, "Neighbor discovery in wireless networks with multipacket reception," in *Proc. ACM MobiHoc*, Paris, France, 2011, p. 3.

[58] J. Jeon and A. Ephremides, "Neighbor discovery in a wireless sensor network: Multipacket reception capability and physical-layer signal processing," *J. Commun. Networks*, vol. 14, no. 5, pp. 566–577, 2012.

[59] 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.

[60] 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.

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

[62] 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.

[63] 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.

[64] 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.

[65] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information theory*, vol. 19, no. 4, pp. 471–480, 1973.

[66] J. Scarlett and V. Y. Tan, "Second-order asymptotics for the gaussian mac with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6700–6718, 2015.

[67] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, 2014.

[68] E. M. Jazi and J. N. Laneman, "Multiaccess communication in the finite blocklength regime," in *Proc. Workshop on Information Theory and its Applications*, 2012.

[69] ——, "Simpler achievable rate regions for multiaccess with finite block length," in *Proc. IEEE Int. Symp. Inform. Theory*, 2012, pp. 36–40.

[70] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.

[71] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *Proc. Annual Allerton Conference on Commun., Control, and Computing*, Monticello, IL, 2010, pp. 1334–1339.

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

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

APPENDIX A

Appendix for Chapter 2

A.1. Proof of Lemma 1

To upper bound the input-output mutual information of the white Gaussian noise channel, it sufficies to identify the power constraint on the input signal $\underline{\boldsymbol{s}}\boldsymbol{X}$ based on the power constraint (2.2) on $\underline{\boldsymbol{s}}$ and the structure of the binary vector $\boldsymbol{X}$.

According to the distribution of $\boldsymbol{X}$, we can obtain the marginal distribution of X_i, $i = 1, \cdots, M\ell_n$, as $\mathsf{P}\{X_i = 0\} = 1 - \frac{\alpha_n}{M}$ and $\mathsf{P}\{X_i = 1\} = \frac{\alpha_n}{M}$. Therefore, $\mathsf{E}\{X_i\} = \frac{\alpha_n}{M}$ and

$$\mathsf{E}\{X_i X_j\} = \begin{cases} \frac{\alpha_n}{M} & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j, i, j \in I(\ell) \quad \text{for some } \ell \\ \left(\frac{\alpha_n}{M}\right)^2 & \text{otherwise} \end{cases} . \tag{A.1}$$

where we let the indices corresponding to transmitter ℓ be $I(\ell) = \{(\ell-1)M+1, \cdots, \ell M\}$, $\ell = 1, \cdots, \ell_n$. Thus, the covariance matrix $\boldsymbol{K} = \mathsf{E}\left\{(\boldsymbol{X} - \mathsf{E}\boldsymbol{X})(\boldsymbol{X} - \mathsf{E}\boldsymbol{X})^T\right\}$ can be calculated as

$$\boldsymbol{K}_{ij} = \begin{cases} \frac{\alpha_n}{M}\left(1 - \frac{\alpha_n}{M}\right) & i = j, \\ -\left(\frac{\alpha_n}{M}\right)^2 & i \neq j, i, j \in I(\ell) \text{ for some } \ell, \\ 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

Let $tr(\cdot)$ find the trace of a matrix. The power constraint on the codewords induces the power constraint on $\underline{\boldsymbol{s}}\boldsymbol{X}$ as

$$tr\left(\underline{\boldsymbol{s}}\boldsymbol{K}\underline{\boldsymbol{s}}^T\right) = tr\left(\boldsymbol{K}\underline{\boldsymbol{s}}^T\underline{\boldsymbol{s}}\right) \tag{A.3}$$

$$= \sum_{i=1}^{M\ell_n}\sum_{j=1}^{M\ell_n}\sum_{k=1}^{n} K_{ij}s_{ki}s_{kj} \tag{A.4}$$

$$= \sum_{k=1}^{n}\left[\frac{\alpha_n}{M}\left(1-\frac{\alpha_n}{M}\right)\sum_{i=1}^{M\ell_n} s_{ki}^2 - \left(\frac{\alpha_n}{M}\right)^2\sum_{\ell=1}^{\ell_n}\sum_{i\neq j, i,j\in I(\ell)} s_{ki}s_{kj}\right] \tag{A.5}$$

$$= \sum_{k=1}^{n}\left[\frac{\alpha_n}{M}\sum_{i=1}^{M\ell_n} s_{ki}^2 - \left(\frac{\alpha_n}{M}\right)^2\sum_{\ell=1}^{\ell_n}\sum_{i\in I(\ell)}\sum_{j\in I(\ell)} s_{ki}s_{kj}\right] \tag{A.6}$$

$$\leq \frac{n\alpha_n}{M}\sum_{i=1}^{M\ell_n}\frac{1}{n}\sum_{k=1}^{n} s_{ki}^2 \tag{A.7}$$

$$\leq k_n nP, \tag{A.8}$$

where (A.7) is due to

$$\sum_{i\in I(\ell)}\sum_{j\in I(\ell)} s_{ki}s_{kj} = \left(\sum_{i\in I(\ell)} s_{ki}\right)^2 \geq 0, \tag{A.9}$$

and the last inequality is due to the power constraint $\frac{1}{n}\sum_{k=1}^{n} s_{ki}^2 \leq P$.

Since $\boldsymbol{X} \to \underline{\boldsymbol{s}}\boldsymbol{X} \to \boldsymbol{Y}$ forms a Markov chain, we can obtain an upper bound of $I(\boldsymbol{X};\boldsymbol{Y})$ as

$$I(\boldsymbol{X};\boldsymbol{Y}) \leq I(\underline{\boldsymbol{s}}\boldsymbol{X};\boldsymbol{Y}) \tag{A.10}$$

$$\leq \max_{tr(\underline{\boldsymbol{s}}\boldsymbol{K}\underline{\boldsymbol{s}}^T)\leq k_n n P} I(\underline{\boldsymbol{s}}\boldsymbol{X};\boldsymbol{Y}) \tag{A.11}$$

$$\leq \frac{n}{2}\log(1+k_n P), \tag{A.12}$$

where (A.12) follows by the results on parallel Gaussian channels [**19**, Chapter 10].

A.2. Proof of Lemma 2

Conditioned on $E=0$, $H\left(\boldsymbol{X}|E=0,\boldsymbol{Y},1\left\{\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\right\}\right)=0$. Therefore, we can obtain

$$H\left(\boldsymbol{X}|E,\boldsymbol{Y},1\left\{\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\right\}\right) =$$

$$H(\boldsymbol{X}|E=1,\boldsymbol{Y},\boldsymbol{X}\notin\mathcal{B}_M^{\ell_n}(\delta,k_n))P\{E=1,\boldsymbol{X}\notin\mathcal{B}_M^{\ell_n}(\delta,k_n)\}$$

$$+H(\boldsymbol{X}|E=1,\boldsymbol{Y},\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n))P\{E=1,\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\}. \tag{A.13}$$

We upper bound the first term on the right hand side of (A.13) as follows: $\boldsymbol{X}$ can take at most $(M+1)^{\ell_n}$ values and $||\boldsymbol{X}||_0$ follows the binomial distribution $\mathrm{Bin}(\ell_n,\alpha_n)$ with mean $\ell_n\alpha_n=k_n$, then $\mathsf{P}\{\boldsymbol{X}\notin\mathcal{B}_M^{\ell_n}(\delta,k_n)\}$ can be upper bounded by $e^{-c(\delta)k_n}$ [**37**], where $c(\delta)$ is

some constant depending on δ by the large deviations for binomial distribution. Then

(A.14)
$$H(\boldsymbol{X}|E=1,\boldsymbol{Y},\boldsymbol{X}\notin\mathcal{B}_M^{\ell_n}(\delta,k_n))P\{E=1,\boldsymbol{X}\notin\mathcal{B}_M^{\ell_n}(\delta,k_n)\}\le e^{-c(\delta)k_n}\ell_n\log(M+1)$$

(A.15)
$$\le_n \log M.$$

For the second term on the RHS of (A.13), $P\{E=1,\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n)\}\le \mathsf{P}_e^{(n)}$ and

(A.16)
$$H(\boldsymbol{X}|E=1,\boldsymbol{Y},\boldsymbol{X}\in\mathcal{B}_M^{\ell_n}(\delta,k_n))\le\log|\mathcal{B}_M^{\ell_n}(\delta,k_n)|.$$

The cardinality of $\mathcal{B}_M^{\ell_n}(\delta,k_n)$ is

(A.17)
$$|\mathcal{B}_M^{\ell_n}(\delta,k_n)|=\sum_{j=1}^{(1+\delta)k_n}\binom{\ell_n}{j}M^j$$

(A.18)
$$\le(1+\delta)k_nM^{(1+\delta)k_n}\max_{1\le j\le(1+\delta)k_n}\binom{\ell_n}{j}.$$

If $(1+\delta)k_n\ge\frac{\ell_n}{2}$, then

(A.19)
$$\max_{1\le j\le(1+\delta)k_n}\binom{\ell_n}{j}\le 2^{\ell_n}$$

(A.20)
$$\le\exp(2(1+\delta)k_n\log 2).$$

If $(1+\delta)k_n<\frac{\ell_n}{2}$, then

(A.21)
$$\max_{1\le j\le(1+\delta)k_n}\binom{\ell_n}{j}\le\binom{\ell_n}{(1+\delta)k_n}$$

(A.22)
$$\le\exp(\ell_nH_2((1+\delta)\alpha_n)).$$

We further upper bound $H_2((1+\delta)\alpha_n)$ in terms of $H_2(\alpha_n)$. By the mean value theorem, there exists some γ_n' in between α_n and $(1+\delta)\alpha_n$ such that

$$H_2((1+\delta)\alpha_n) - H_2(\alpha_n) = \delta\alpha_n \log\frac{1-\gamma_n'}{\gamma_n'}, \tag{A.23}$$

where $\log\frac{1-x}{x}$ is the first order derivative of $H_2(x)$. Since $\log\frac{1-x}{x}$ is decreasing in x, we have

$$H_2((1+\delta)\alpha_n) - H_2(\alpha_n) \leq \delta\alpha_n \log\frac{1-\alpha_n}{\alpha_n} \leq \delta H_2(\alpha_n). \tag{A.24}$$

As a result,

(A.25)

$$\log|\mathcal{B}_M^{\ell_n}(\delta, k_n)| \leq \log((1+\delta)k_n) + (1+\delta)k_n \log M + 2(1+\delta)k_n \log 2 + (1+\delta)\ell_n H_2(\alpha_n).$$

For large enough n, we have $\log((1+\delta)k_n) \leq (1+\delta)k_n$. Then

$$H(\boldsymbol{X}|E=1, \boldsymbol{X} \in \mathcal{B}_M^{\ell_n}(\delta, k_n), \boldsymbol{Y}) \leq_n 4(k_n \log M + k_n + \ell_n H_2(\alpha_n)). \tag{A.26}$$

Combining (A.13), (A.15) and (A.26) yields the lemma.

A.3. Derivation of (2.78)

We will derive the closed-form expression of (2.74), which is calculated as

(A.27)
$$m_{\lambda,\rho}(w_1,w_2)=\int_{\mathbb{R}}\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a)\left(\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a)\middle|\boldsymbol{S}_{A^*}^a\right\}\right)^{\rho}\right\}dy$$

(A.28)
$$=\int_{\mathbb{R}}\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a)\left(\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a)\middle|\boldsymbol{S}_{A^*\backslash A_1}^a\right\}\right)^{\rho}\right\}dy$$

(A.29)
$$=\int_{\mathbb{R}}\mathsf{E}\left\{\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a)\middle|\boldsymbol{S}_{A^*\backslash A_1}^a\right\}\left(\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a)\middle|\boldsymbol{S}_{A^*\backslash A_1}^a\right\}\right)^{\rho}\right\}dy$$

where (A.28) follows because $A\cap A^*=A^*\backslash A_1$.

Let $Z_1=\sum_{k\in A_1}S_k^a$, $Z_2=\sum_{k\in A_2}S_k^a$ and $Z_3=\sum_{k\in A^*\backslash A_1}S_k^a$. Since $|A_1|=w_1$ and $|A_2|=w_2$, we have $Z_1\sim\mathcal{N}(0,v_1)$, $Z_2\sim\mathcal{N}(0,v_2)$, $Z_3\sim\mathcal{N}(0,v_3)$, where $v_1=w_1P'$, $v_2=w_2P'$ and $v_3=(|A^*|-w_1)P'$.

We can write

(A.30)
$$\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a)|\boldsymbol{S}_{A^*\backslash A_1}^a\right\}=\mathsf{E}\left\{\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-Z_3-Z_2)^2}{2}}\right)^{\lambda}\middle|Z_3\right\}$$

(A.31)
$$=\int_{\mathbb{R}}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-Z_3-z_2)^2}{2}}\right)^{\lambda}\frac{1}{\sqrt{2\pi v_2}}e^{-\frac{z_2^2}{2v_2}}dz_2$$

(A.32)
$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{\lambda}\sqrt{\frac{t_3}{v_2}}e^{\frac{\mu_3^2}{2t_3}}e^{-\frac{\lambda(y-Z_3)^2}{2}}\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi t_3}}e^{-\frac{(z_2-\mu_3)^2}{2t_3}}dz_2$$

(A.33)
$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{\lambda}\sqrt{\frac{t_3}{v_2}}e^{\frac{\mu_3^2}{2t_3}}e^{-\frac{\lambda(y-Z_3)^2}{2}},$$

where $\frac{1}{t_3}=\lambda+\frac{1}{v_2}$ and $\mu_3=\lambda(y-Z_3)t_3$.

Similarly,

(A.34)

$$\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a)|\boldsymbol{S}_{A^*\setminus A_1}^a\right\} = \mathsf{E}\left\{\left(\frac{1}{\sqrt{2\pi}}\right)^{1-\lambda\rho} e^{-\frac{(1-\lambda\rho)(y-Z_3-Z_1)^2}{2}}\middle| Z_3\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{1-\lambda\rho} \int_{\mathbb{R}} e^{-\frac{(1-\lambda\rho)(y-Z_3-z_1)^2}{2}} \frac{1}{\sqrt{2\pi v_1}} e^{-\frac{z_1^2}{2v_1}} dz_1 \tag{A.35}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{1-\lambda\rho} \sqrt{\frac{t_4}{v_1}} e^{\frac{\mu_4^2}{2t_4}} e^{-\frac{(1-\lambda\rho)(y-Z_3)^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t_4}} e^{-\frac{(z_1-\mu_4)^2}{2t_4}} dz_1 \tag{A.36}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{1-\lambda\rho} \sqrt{\frac{t_4}{v_1}} e^{\frac{\mu_4^2}{2t_4}} e^{-\frac{(1-\lambda\rho)(y-Z_3)^2}{2}}, \tag{A.37}$$

where $\frac{1}{t_4} = 1 - \lambda\rho + \frac{1}{v_1}$ and $\mu_4 = (1-\lambda\rho)(y - Z_3)t_4$.

Then

$$\left(\mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a)|\boldsymbol{S}_{A^*\setminus A_1}^a\right\}\right)^{\rho} \mathsf{E}\left\{p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a)|\boldsymbol{S}_{A^*\setminus A_1}^a\right\} =$$

$$\frac{1}{\sqrt{2\pi}} \left(\sqrt{\frac{t_3}{v_2}}\right)^{\rho} \sqrt{\frac{t_4}{v_1}} e^{\frac{\rho\mu_3^2}{2t_3} + \frac{\mu_4^2}{2t_4} - \frac{(y-Z_3)^2}{2}}. \tag{A.38}$$

Plugging μ_3, t_3, μ_4 and t_4 yields $\frac{\mu_3^2}{t_3} = \frac{\lambda^2 v_2 (y-Z_3)^2}{1+\lambda v_2}$ and $\frac{\mu_4^2}{t_4} = \frac{(1-\lambda\rho)^2 (y-Z_3)^2 v_1}{1+(1-\lambda\rho)v_1}$. Let $t_0 = \frac{1}{\sqrt{2\pi}} \left(\sqrt{\frac{t_3}{v_2}}\right)^{\rho} \sqrt{\frac{t_4}{v_1}}$ and $\frac{1}{t_5} = 1 - \frac{\rho\lambda^2 v_2}{1+\lambda v_2} - \frac{(1-\lambda\rho)^2 v_1}{1+(1-\lambda\rho)v_1}$. We have

$$\int_{\mathbb{R}} \mathsf{E}\left\{ \left(\mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{\lambda}(y|\boldsymbol{S}_A^a) | \boldsymbol{S}_{A^*\setminus A_1}^a \right\} \right)^{\rho} \mathsf{E}\left\{ p_{Y|\boldsymbol{S}_A}^{1-\lambda\rho}(y|\boldsymbol{S}_{A^*}^a) | \boldsymbol{S}_{A^*\setminus A_1}^a \right\} \right\} dy$$

$$= t_0 \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi v_3}} e^{-\frac{z_3^2}{2v_3}} e^{\frac{\rho\lambda^2 v_2 (y-z_3)^2}{2(1+\lambda v_2)} + \frac{(1-\lambda\rho)^2 (y-z_3)^2 v_1}{2(1+(1-\lambda\rho)v_1)} - \frac{(y-z_3)^2}{2}} dz_3 dy \tag{A.39}$$

$$= t_0 \int_{\mathbb{R}} \sqrt{\frac{t_5}{v_3}} e^{-\frac{z_3^2}{2v_3}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t_5}} e^{-\frac{(y-z_3)^2}{2t_5}} dy dz_3 \tag{A.40}$$

$$= t_0 \int_{\mathbb{R}} \sqrt{\frac{t_5}{v_3}} e^{-\frac{z_3^2}{2v_3}} dz_3 \tag{A.41}$$

$$= \left(\sqrt{\frac{t_3}{v_2}}\right)^{\rho} \sqrt{\frac{t_4 t_5}{v_1}} \tag{A.42}$$

$$= (1+\lambda v_2)^{-\rho/2} \left(\frac{1+\lambda v_2}{1+\lambda(1-\lambda\rho)v_2 + \lambda\rho(1-\lambda\rho)v_1} \right)^{1/2}. \tag{A.43}$$

Therefore, $m_{\lambda,\rho}(w_1, w_2)$ is given by (2.76).

A.4. Proof of Lemma 3

We first establish the following two lemmas that will be useful in the proof.

Lemma 7. *Suppose* (2.18) *holds, i.e.,* $\lim_{\ell\to\infty} \ell e^{-\delta k_\ell} = 0$ *for every* $\delta > 0$*, then for every constant* $\bar{w} \geq 0$*,*

$$\lim_{\ell\to\infty} \frac{\ell}{k_\ell} H_2\left(\frac{\bar{w}}{\ell}\right) = 0. \tag{A.44}$$

Proof. The case of $\bar{w} = 0$ is trivial. Suppose $\bar{w} > 0$. Since $\bar{w}/\ell \to 0$,

$$\frac{\ell}{k_\ell} H_2\left(\frac{\bar{w}}{\ell}\right) = \frac{\ell}{k_\ell}\left(\frac{\bar{w}}{\ell}\log\frac{\ell}{\bar{w}} - \left(1-\frac{\bar{w}}{\ell}\right)\log\left(1-\frac{\bar{w}}{\ell}\right)\right) \tag{A.45}$$

$$\leq_n \frac{\ell}{k_\ell}\left(\frac{\bar{w}}{\ell}\log\frac{\ell}{\bar{w}} + \left(1-\frac{\bar{w}}{\ell}\right)\frac{2\bar{w}}{\ell}\right) \tag{A.46}$$

$$\leq \frac{\bar{w}}{k_\ell}\left(\log\ell - \log\bar{w} + 2\right). \tag{A.47}$$

Since $\ell e^{-\delta k_\ell} \to 0$ for every $\delta > 0$, we have $\ell \leq_\ell e^{\delta k_\ell}$, so that $\log\ell \leq_\ell \delta k_\ell$. This implies $(\log\ell)/k_\ell \to 0$, so that the right hand side of (A.47) vanishes. □

Lemma 8. *Suppose* (2.18) *holds for every* $\delta > 0$. *Let* $A > 0$, $B > 0$ *and* $\bar{w} \geq 1$ *be constants. Let* $\{a_\ell\}$ *and* $\{b_\ell\}$ *be two sequences that satisfy* $b_\ell \leq a_\ell$, $\lim_{\ell\to\infty} \frac{k_\ell}{a_\ell} = a \in [0,\infty)$, *and* $\lim_{\ell\to\infty} \frac{k_\ell}{b_\ell} = b \in (0,\infty)$. *Let* A_ℓ *be a sequence that satisfies* $\liminf_{\ell\to\infty} A_\ell = A$. *Define* $h_\ell(\cdot)$ *on* $[0, a_\ell]$ *as*

$$h_\ell(w) = A_\ell \log(1+Bw) - \frac{a_\ell}{k_\ell} H_2\left(\frac{w}{a_\ell}\right). \tag{A.48}$$

Let w_ℓ^* *achieve the global minimum of* $h_\ell(\cdot)$ *restricted to* $[\bar{w}, b_\ell]$. *For large enough* ℓ, *either* $w_\ell^* = \bar{w}$ *or* $w_\ell^* \in [cb_\ell, b_\ell]$, *where*

$$c = \min\left\{\frac{bA}{64(1+Aa)}, 1\right\}. \tag{A.49}$$

Proof. The function $h_\ell(w)$ is equal to the difference of two concave functions. Its first two derivatives on $(0, a_\ell)$ are:

$$h_\ell'(w) = \frac{A_\ell B}{1+Bw} + \frac{1}{k_\ell}\log\frac{w}{a_\ell - w} \tag{A.50}$$

and

$$h''_\ell(w) = \frac{a_\ell}{k_\ell w(a_\ell - w)} - \frac{A_\ell B^2}{(1+Bw)^2} \tag{A.51}$$

$$= \frac{a_\ell g_\ell(w)}{k_\ell w(a_\ell - w)(1+Bw)^2}, \tag{A.52}$$

where

$$g_\ell(w) = (B^2 + k_\ell A_\ell B^2/a_\ell)w^2 + (2B - k_\ell A_\ell B^2)w + 1. \tag{A.53}$$

Due to (2.18), $k_\ell \to \infty$ as $\ell \to \infty$. For large enough ℓ, $g_\ell(0) = 1$, $g_\ell(1) = -A_\ell B^2 k_\ell + A_\ell B^2 k_\ell/a_\ell + (B+1)^2 < 0$, and $g_\ell(a_\ell) = (Ba_\ell + 1)^2 > 0$. Moreover, the minimum of the quadratic function $g_\ell(w)$ is achieved at:

$$v_\ell = \frac{k_\ell A_\ell B - 2}{2B(1 + k_\ell A_\ell/a_\ell)}\,. \tag{A.54}$$

Since $\frac{1}{2}k_\ell A_\ell B \geq_n 2$, we have $k_\ell A_\ell B - 2 \geq_n \frac{1}{2}k_\ell A_\ell B$. Also, $A_\ell k_\ell/a_\ell \leq_\ell 1 + 2Aa$. We have

$$\frac{v_\ell}{b_\ell} \geq_n \frac{\frac{1}{2}\frac{k_\ell}{b_\ell}A_\ell B}{2B(1 + A_\ell \frac{k_\ell}{a_\ell})} \tag{A.55}$$

$$\geq_n \frac{\frac{1}{2}\left(\frac{1}{2}b\right)\left(\frac{1}{2}A\right)}{2(2+2Aa)} \tag{A.56}$$

$$= \frac{bA}{32(1+Aa)}. \tag{A.57}$$

Note that $b_\ell \to \infty$ and (A.57) implies $v_\ell \to \infty$. For large enough ℓ, since $h''_\ell(w) < 0$ for every $w \in [\bar{w}, v_\ell]$, $h_\ell(w)$ is concave over $[\bar{w}, v_\ell]$. Since $v_\ell/b_\ell \geq_n 2c$, we have either $w^*_\ell = \bar{w}$ or $w^*_\ell \in [cb_\ell, b_\ell]$ for large enough ℓ. □

The general idea for proving Lemma 3 is to divide $\mathcal{W}^{(\ell)}$ into two regions based on whether the error probabily is dominated by false alarms or miss detections, and to lower bound $h_{\lambda,\rho}(w_1, w_2)$ given by (2.78) for (w_1, w_2) in those two regions separately. It is crucial to note that Lemma 3 claims the existence of a *uniform* lower bound of $h_{\lambda,\rho}(w_1, w_2)$, i.e., ℓ^* is such that for all $\ell \geq \ell^*$, $h_{\lambda,\rho}(w_1, w_2) \geq c_0$ *regardless* of (w_1, w_2), which in general depend on ℓ.

Define

$$\phi_\ell = \frac{n(\ell)}{k_\ell} = \frac{2\ell H_2(\alpha_\ell)}{k_\ell \log(1 + k_\ell P')}, \tag{A.58}$$

which can be regarded as the identification cost per active user. Let

$$\phi = \lim_{\ell \to \infty} \phi_\ell, \tag{A.59}$$

which may be ∞. As $\phi \geq 0$, we prove the cases of $\phi > 0$ and $\phi = 0$ separately.

A.4.1. The case of $\phi > 0$

In this case, by (2.41), the signature length is $n_0 = (1 + \epsilon)\, \phi_\ell k_\ell$. As we shall see, if the number of false alarms $w_2 = |A \backslash A^*|$ is small, the error probability is dominated by miss detections; whereas for relatively large w_2, the error probability is dominated by false alarms.

Define the following positive constant:

$$\bar{w} = \max\left\{ \frac{4}{P'} e^{(8+4\epsilon)/\phi}, 1 \right\}. \tag{A.60}$$

We will derive lower bounds of $h_{\lambda,\rho}(w_1, w_2)$ for the cases of $0 \leq w_2 \leq \bar{w}$ and $\bar{w} < w_2 \leq (1+\delta_\ell)k_\ell$ separately.

A.4.1.1. The case of $0 \leq w_2 \leq \bar{w}$. Recall that $\rho \in [0,1]$ and $\lambda \in [0,\infty)$ can be chosen arbitrarily to yield a lower bound. We shall always choose them to satisfy $0 \leq \lambda\rho \leq 1$. This implies that

$$\log\left(1+\lambda(1-\lambda\rho)w_2P' + \lambda\rho(1-\lambda\rho)w_1P'\right) \geq \frac{1}{2}\log\left(1+\lambda(1-\lambda\rho)w_2P'\right) + \frac{1}{2}\log\left(1+\lambda\rho(1-\lambda\rho)w_1P'\right). \tag{A.61}$$

In this case, a lower bound of $h_{\lambda,\rho}(w_1, w_2)$ can be splitted into two parts as

$$h_{\lambda,\rho}(w_1, w_2) \geq g^1_{\lambda,\rho}(w_1) + g^2_{\lambda,\rho}(w_2), \tag{A.62}$$

where

$$g^1_{\lambda,\rho}(w_1) = \frac{n_0}{4k_\ell}\log\left(1+\lambda\rho(1-\lambda\rho)w_1P'\right) - \frac{|A^*|}{k_\ell}H_2\left(\frac{w_1}{|A^*|}\right) \tag{A.63}$$

and

$$g^2_{\lambda,\rho}(w_2) = \frac{n_0}{4k_\ell}\log\left(1+\lambda(1-\lambda\rho)w_2P'\right) - \frac{(1-\rho)n_0}{2k_\ell}\log\left(1+\lambda w_2P'\right) - \frac{\rho\ell}{k_\ell}H_2\left(\frac{w_2}{\ell}\right). \tag{A.64}$$

Note that $g^1_{\lambda,\rho}(0) = g^2_{\lambda,\rho}(0) = 0$. However, since $(w_1, w_2) \in \mathcal{W}^{(\ell)}$, they cannot be 0 simultaneously. In the following, we lower bound $g^1_{\lambda,\rho}(w_1)$ for $w_1 \geq 1$ and $g^2_{\lambda,\rho}(w_2)$ for $w_2 \geq 1$. Then $h_{\lambda,\rho}(w_1, w_2)$ can be lower bounded by the minimum of the two lower bounds of $g^1_{\lambda,\rho}(w_1)$ and $g^2_{\lambda,\rho}(w_2)$.

Choose $\lambda = 2/3$ and $\rho = 3/4$. We have

$$g^2_{2/3,3/4}(w_2) = \frac{n_0}{4k_\ell}\log\left(1+\frac{w_2P'}{3}\right) - \frac{n_0}{8k_\ell}\log\left(1+\frac{2w_2P'}{3}\right) - \frac{3\ell}{4k_\ell}H_2\left(\frac{w_2}{\ell}\right). \tag{A.65}$$

Since $(1+x)^r \le 1+rx$ for $r \in [0,1]$, we have

$$\log(1+rx) \ge r\log(1+x) \tag{A.66}$$

for $x \ge 0$ and the equality is achieved only if $x = 0$. Letting $r = 1/2$, $x = 2w_2P'/3$, we can see that for $w_2 > 0$,

$$\log\left(1+\frac{w_2P'}{3}\right) > \frac{1}{2}\log\left(1+\frac{2w_2P'}{3}\right). \tag{A.67}$$

Define a positive constant

$$\epsilon' = \min_{1\le w_2\le \bar{w}} \frac{\phi}{8}\left[\log\left(1+\frac{w_2P'}{3}\right) - \frac{1}{2}\log\left(1+\frac{2w_2P'}{3}\right)\right]. \tag{A.68}$$

By Lemma 7, $\frac{\ell}{k_\ell}H_2(\bar{w}/\ell)$ vanishes as ℓ increases. We can find some $\ell_0 > 2\bar{w}$ such that for all $\ell \ge \ell_0$, $\phi_\ell > \phi/2$ and $\frac{3\ell}{4k_\ell}H_2(\bar{w}/\ell) \le \frac{\epsilon'}{2}$.

For every $\ell \ge \ell_0$, we have $H_2(w_2/\ell) \le H_2(\bar{w}/\ell)$ for $1 \le w_2 \le \bar{w}$ and thus $g^2_{2/3,3/4}(w_2)$ is lower bounded as

$$g^2_{2/3,3/4}(w_2) \ge \frac{\phi_\ell}{4}\left[\log\left(1+\frac{w_2P'}{3}\right) - \frac{1}{2}\log\left(1+\frac{2w_2P'}{3}\right)\right] - \frac{3\ell}{4k_\ell}H_2(\bar{w}/\ell) \tag{A.69}$$

$$\ge \epsilon' - \frac{\epsilon'}{2} \tag{A.70}$$

$$= \frac{\epsilon'}{2}. \tag{A.71}$$

Meanwhile,

$$g^1_{2/3,3/4}(w_1) = \frac{(1+\epsilon)\phi_\ell}{4}\log\left(1+\frac{w_1 P'}{4}\right) - \frac{|A^*|}{k_\ell}H_2\left(\frac{w_1}{|A^*|}\right). \tag{A.72}$$

When $w_1 \geq 1$, we shall invoke Lemma 8 to show that the minimum of the RHS of (A.72) is achieved at either $w_1 = 1$ or some value close to k_ℓ. Define

$$a = \min\left\{\frac{\phi}{16}\log\left(1+\frac{P'}{4}\right), 1\right\} \tag{A.73}$$

We consider the following three cases separately:

$$\text{case a): } 1 \leq |A^*| \leq ak_\ell, 1 \leq w_1 \leq |A^*| \tag{A.74}$$

$$\text{case b): } ak_\ell \leq |A^*| \leq (1+\delta_\ell)k_\ell, ak_\ell/2 \leq w_1 \leq |A^*| \tag{A.75}$$

$$\text{case c): } ak_\ell \leq |A^*| \leq (1+\delta_\ell)k_\ell, 1 \leq w_1 \leq ak_\ell/2. \tag{A.76}$$

For every $\ell \geq \ell_0$, $g^1_{2/3,3/4}(w_1)$ in case a) is lower bounded as

$$g^1_{2/3,3/4}(w_1) \geq \frac{\phi_\ell}{4}\log\left(1+\frac{P'}{4}\right) - a \tag{A.77}$$

$$\geq \frac{\phi}{8}\log\left(1+\frac{P'}{4}\right) - a \tag{A.78}$$

$$\geq \frac{\phi}{16}\log\left(1+\frac{P'}{4}\right). \tag{A.79}$$

In case b), $g^1_{2/3,3/4}(w_1)$ is lower bounded as

$$g^1_{2/3,3/4}(w_1) \geq \frac{(1+\epsilon)\phi_\ell}{4}\log\left(1+\frac{ak_\ell P'}{8}\right) - (1+\delta_\ell), \tag{A.80}$$

which grows without bound as ℓ increases.

In case c), $w_1/|A^*| \leq 1/2$. Since $H_2(\cdot)$ is increasing on $[0, 1/2]$, by (A.72),

$$g^1_{2/3,3/4}(w_1) \geq \frac{(1+\epsilon)\phi_\ell}{4} \log\left(1 + \frac{w_1 P'}{4}\right) - \frac{(1+\delta_\ell)k_\ell}{k_\ell} H_2\left(\frac{w_1}{ak_\ell}\right) \tag{A.81}$$

$$\geq \frac{2}{a}\left[\frac{(1+\epsilon)a\phi_\ell}{8} \log\left(1 + \frac{w_1 P'}{4}\right) - \frac{ak_\ell}{k_\ell} H_2\left(\frac{w_1}{ak_\ell}\right)\right]. \tag{A.82}$$

Applying Lemma 8 with $A_\ell = (1+\epsilon)a\phi_\ell/8$, $B = P'/4$, $a_\ell = ak_\ell$, $\bar{w} = 1$ and $b_\ell = ak_\ell/2$, we conclude that there exists ℓ_1 such that for all $\ell \geq \ell_1$, the RHS of (A.82) restricted to $w_1 \in [1, ak_\ell/2]$ achieves the minimum either at 1 or on $[cak_\ell/2, ak_\ell/2]$ for some $c \in (0, 1]$. Moreover, $H_2\left(\frac{1}{ak_\ell}\right)$ vanishes as ℓ increases. There exists some ℓ_2 such that for all $\ell \geq \ell_2$, $H_2\left(\frac{1}{ak_\ell}\right) \leq \frac{\phi}{32} \log\left(1 + \frac{P'}{4}\right)$ and $\phi_\ell \geq \phi/2$.

For every $\ell \geq \max\{\ell_1, \ell_2\}$, if the minimum of the RHS of (A.82) is achieved at 1, then $g^1_{2/3,3/4}(w_1)$ in case c) is lower bounded as

$$g^1_{2/3,3/4}(w_1) \geq \frac{\phi_\ell}{4} \log\left(1 + \frac{P'}{4}\right) - 2H_2\left(\frac{1}{ak_\ell}\right) \tag{A.83}$$

$$\geq \frac{\phi}{8} \log\left(1 + \frac{P'}{4}\right) - 2H_2\left(\frac{1}{ak_\ell}\right) \tag{A.84}$$

$$\geq \frac{\phi}{16} \log\left(1 + \frac{P'}{4}\right). \tag{A.85}$$

For every $\ell \geq \max\{\ell_1, \ell_2\}$, if the minimum of the RHS of (A.82) is achieved on $[cak_\ell/2, ak_\ell/2]$, then then $g^1_{2/3,3/4}(w_1)$ in case c) is lower bounded as

$$g^1_{2/3,3/4}(w_1) \geq \frac{\phi_\ell}{4} \log\left(1 + \frac{cak_\ell P'}{8}\right) - 2, \tag{A.86}$$

which grows without bound as ℓ increases.

By (A.79), (A.80), (A.85) and (A.86), it concludes that for all $\ell \geq \max\{\ell_0, \ell_1, \ell_2\}$, $g^1_{2/3,3/4}(w_1) \geq \frac{\phi}{16}\log\left(1+\frac{P'}{4}\right)$ for all $1 \leq w_1 \leq |A^*|$ and for all $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$. Combining the lower bound of $g^1_{2/3,3/4}(w_2)$ given by (A.71), we conclude that for all $\ell \geq \max(\ell_0, \ell_1, \ell_2)$ and for all $(w_1, w_2) \in \mathcal{W}^{(\ell)}$ with $0 \leq w_2 \leq \bar{w}$, $h_{2/3,3/4}(w_1, w_2)$ can be uniformly lower bounded as

$$h_{2/3,3/4}(w_1, w_2) \geq \min\left\{\frac{\epsilon'}{2}, \frac{\phi}{16}\log\left(1+\frac{P'}{4}\right)\right\}. \tag{A.87}$$

A.4.1.2. The case of $\bar{w} < w_2 \leq (1+\delta_\ell)k_\ell$. Letting $\lambda = 1/2$ and $\rho = 1$ in (2.78), and using the fact that $w_1 \geq 0$ and $|A^*|/k_\ell \leq 2$, we have

$$h_{1/2,1}(w_1, w_2) \geq \frac{n_0}{2k_\ell}\log\left(1+\frac{w_2 P'}{4}\right) - \frac{\ell}{k_\ell}H_2\left(\frac{w_2}{\ell}\right) - \frac{|A^*|}{k_\ell}H_2\left(\frac{w_1}{|A^*|}\right) \tag{A.88}$$

$$\geq \frac{(1+\epsilon)\phi_\ell}{2}\log\left(1+\frac{w_2 P'}{4}\right) - \frac{\ell}{k_\ell}H_2\left(\frac{w_2}{\ell}\right) - 2. \tag{A.89}$$

Applying Lemma 8 with $A_\ell = (1+\epsilon)\phi_\ell/2$, $B = P'/4$, $a_\ell = \ell$ and $b_\ell = (1+\delta_\ell)k_\ell$, we can conclude that there exists some ℓ_3 such that for all $\ell \geq \ell_3$, the minimum of the RHS of (A.89) restricted to $[\bar{w}, (1+\delta_\ell)k_\ell]$ is achieved either at $\bar{w}$ or on $[ck_\ell, (1+\delta_\ell)k_\ell]$, for some $c \in (0, 1]$. Moreover, by Lemma 7, there exists some ℓ_4 such that for all $\ell \geq \ell_4$, $\frac{\ell}{k_\ell}H_2(\bar{w}/\ell) \leq 1$ and $\phi_\ell > \phi/2$.

For every $\ell \geq \max\{\ell_3, \ell_4\}$, if the minimum of the RHS of (A.89) is achived at $\bar{w}$, then $h_{1/2,1}(w_1, w_2)$ is uniformly lower bounded as

$$h_{1/2,1}(w_1, w_2) \geq \frac{\phi}{4}\log\left(1+\frac{\bar{w}P'}{4}\right) - 2 \tag{A.90}$$

$$\geq \epsilon. \tag{A.91}$$

For every $\ell \geq \max\{\ell_3, \ell_4\}$, if the minimum of the RHS of (A.89) is achieved on $[ck_\ell, (1+\delta_\ell)k_\ell]$, we consider two cases:

$$\text{case a): } \ell > 2(1+\delta_\ell)k_\ell \tag{A.92}$$

$$\text{case b): } \ell \leq 2(1+\delta_\ell)k_\ell \tag{A.93}$$

In case a), $w_2/\ell < 1/2$. Since $H_2(\cdot)$ is increasing on $[0, 1/2]$, by (A.89), we have

$$h_{1/2,1}(w_1, w_2) \geq \frac{(1+\epsilon)\phi_\ell}{2} \log\left(1 + \frac{ck_\ell P'}{4}\right) - \frac{\ell}{k_\ell} H_2\left(\frac{(1+\delta_\ell)k_\ell}{\ell}\right) - 2 \tag{A.94}$$

$$\geq \frac{(1+\epsilon)\phi_\ell}{2} \log\left(1 + \frac{ck_\ell P'}{4}\right) - (1+\delta_\ell)\frac{\ell}{k_\ell} H_2\left(\frac{k_\ell}{\ell}\right) - 2 \tag{A.95}$$

$$= \frac{\phi_\ell}{2}\left[(1+\epsilon)\log\left(1 + \frac{ck_\ell P'}{4}\right) - (1+\delta_\ell)\log(1+k_\ell P')\right] - 2, \tag{A.96}$$

where (A.95) follows from (A.24), and (A.96) is due to (A.58). By (2.44), $\delta_\ell \log(1+k_\ell P')$ vanishes as k_ℓ increases. Moreover,

$$\lim_{k_\ell \to \infty} \log\left(1 + \frac{ck_\ell P'}{4}\right) - \log(1+k_\ell P') = \log(c/4). \tag{A.97}$$

Thus, the RHS of (A.96) grows without bound (uniformly for (w_1, w_2)) as ℓ increases.

In case b), by (A.89), we have

$$h_{1/2,1}(w_1, w_2) \geq \frac{(1+\epsilon)\phi_\ell}{2} \log\left(1 + \frac{ck_\ell P'}{4}\right) - \frac{\ell}{k_\ell} - 2 \tag{A.98}$$

$$\geq \frac{(1+\epsilon)\phi_\ell}{2} \log\left(1 + \frac{ck_\ell P'}{4}\right) - 5, \tag{A.99}$$

which grows without bound (uniformly for (w_1, w_2)) as ℓ increases.

By (A.91), (A.96) and (A.99), we conclude that for all $\ell \geq \max\{\ell_3, \ell_4\}$,

$$h_{1/2,1}(w_1, w_2) \geq \epsilon \tag{A.100}$$

uniformly for all $0 \leq w_1 \leq |A^*|$, $\bar{w} \leq w_2 \leq (1+\delta_\ell)k_\ell$, and $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$.

Combining (A.87) and (A.100), we conclude that Lemma 3 holds for the case of $\phi > 0$ with $\ell^* = \max\{\ell_0, \ell_1, \ell_2, \ell_3, \ell_4\}$.

A.4.2. The case of $\phi = 0$

In this case, $n_0 = \epsilon k_\ell$ by (2.41). We let $\lambda = 3/5$, $\rho = 5/6$. Note that (A.62) - (A.64) remain true in this case.

Consider first $g^2_{3/5,5/6}(w_2)$. By (A.66), we have

$$\log\left(1 + \frac{3w_2P'}{10}\right) \geq \frac{1}{2}\log\left(1 + \frac{3w_2P'}{5}\right). \tag{A.101}$$

Thus,

$$g^2_{3/5,5/6}(w_2) = \frac{\epsilon}{4}\log\left(1 + \frac{3w_2P'}{10}\right) - \frac{\epsilon}{12}\log\left(1 + \frac{3w_2P'}{5}\right) - \frac{5\ell}{6k_\ell}H_2\left(\frac{w_2}{\ell}\right) \tag{A.102}$$

$$\geq \frac{\epsilon}{24}\log\left(1 + \frac{3w_2P'}{5}\right) - \frac{5\ell}{6k_\ell}H_2\left(\frac{w_2}{\ell}\right). \tag{A.103}$$

Applying Lemma 8 with $A_\ell = \epsilon/20$, $B = 3P'/5$, $\bar{w} = 1$, $a_\ell = \ell$ and $b_\ell = (1+\delta_\ell)k_\ell$, we conclude that there exists some ℓ_5 such that for all $\ell \geq \ell_5$, the minimum of the RHS of (A.103) restricted to $w_2 \in [1, (1+\delta_\ell)k_\ell]$ is achieved at either 1 or on $[ck_\ell, (1+\delta_\ell)k_\ell]$ for some $c \in (0, 1]$. Moreover, by Lemma 7, there exists some ℓ_6 such that for all $\ell \geq \ell_6$, $\frac{5\ell}{6k_\ell}H_2\left(\frac{1}{\ell}\right) \leq \frac{\epsilon}{48}\log\left(1 + \frac{3P'}{5}\right)$.

For every $\ell \geq \max\{\ell_5, \ell_6\}$, if the minimum of the RHS of (A.103) is achieved at 1, then $g^2_{3/5,5/6}(w_2)$ is lower bounded as

$$g^2_{3/5,5/6}(w_2) \geq \frac{\epsilon}{24}\log\left(1+\frac{3P'}{5}\right) - \frac{5\ell}{6k_\ell}H_2\left(\frac{1}{\ell}\right) \tag{A.104}$$

$$\geq \frac{\epsilon}{48}\log\left(1+\frac{3P'}{5}\right). \tag{A.105}$$

For every $\ell \geq \max\{\ell_5, \ell_6\}$, if the minimum of the RHS of (A.103) is achieved on $[ck_\ell, (1+\delta_\ell)k_\ell]$, we consider two cases:

$$\text{case a): } \ell > 2(1+\delta_\ell)k_\ell \tag{A.106}$$

$$\text{case b): } \ell \leq 2(1+\delta_\ell)k_\ell. \tag{A.107}$$

In case a), $w_2/\ell < 1/2$. Since $H_2(\cdot)$ is increasing on $[0, 1/2]$, we have

$$g^2_{3/5,5/6}(w_2) \geq \frac{\epsilon}{24}\log\left(1+\frac{3ck_\ell P'}{5}\right) - \frac{5\ell}{6k_\ell}H_2\left(\frac{(1+\delta_\ell)k_\ell}{\ell}\right) \tag{A.108}$$

$$\geq \frac{\epsilon}{24}\log\left(1+\frac{3ck_\ell P'}{5}\right) - (1+\delta_\ell)\frac{5\ell}{6k_\ell}H_2\left(\frac{k_\ell}{\ell}\right) \tag{A.109}$$

$$= \frac{\epsilon}{24}\log\left(1+\frac{3ck_\ell P'}{5}\right) - (1+\delta_\ell)\frac{5\phi_\ell}{12}\log\left(1+k_\ell P\right) \tag{A.110}$$

$$= \left[\frac{\epsilon}{24} - (1+\delta_\ell)\frac{5\phi_\ell}{12}\frac{\log\left(1+k_\ell P\right)}{\log\left(1+\frac{3ck_\ell P'}{5}\right)}\right]\log\left(1+\frac{3ck_\ell P'}{5}\right). \tag{A.111}$$

where (A.109) is due to (A.24). Since $\phi_\ell \to 0$, we have

$$(1+\delta_\ell)\frac{5\phi_\ell}{12}\frac{\log\left(1+k_\ell P\right)}{\log\left(1+\frac{3ck_\ell P'}{5}\right)} \to 0. \tag{A.112}$$

The right hand side of (A.111) thus grows without bound (uniformly for all w_2) as ℓ increases.

In the case b), we have

$$g^2_{3/5,5/6}(w_2) \geq \frac{\epsilon}{24}\log\left(1+\frac{3ck_\ell P'}{5}\right) - \frac{5\ell}{6k_\ell} \tag{A.113}$$

$$\geq \frac{\epsilon}{24}\log\left(1+\frac{3ck_\ell P'}{5}\right) - \frac{10}{3}. \tag{A.114}$$

which grows without bound (uniformly for all w_2) as k_ℓ increases.

By (A.105), (A.111) and (A.114), we conclude that for all $\ell \geq \max\{\ell_5, \ell_6\}$,

$$g^2_{3/5,5/6}(w_2) \geq \frac{\epsilon}{48}\log\left(1+\frac{3P'}{5}\right) \tag{A.115}$$

holds uniformly for all $1 \leq w_2 \leq (1+\delta_\ell)k_\ell$.

Consider next $g^1_{3/5,5/6}(w_1)$.

$$g^1_{3/5,5/6}(w_1) = \frac{\epsilon}{4}\log\left(1+\frac{w_1 P'}{4}\right) - \frac{|A^*|}{k_\ell}H_2\left(\frac{w_1}{|A^*|}\right). \tag{A.116}$$

Define

$$a = \min\left\{\frac{\epsilon}{8}\log\left(1+\frac{P'}{4}\right), 1\right\}. \tag{A.117}$$

We consider the following three cases:

$$\text{case a): } 1 \le |A^*| \le ak_\ell, 1 \le w_1 \le |A^*| \tag{A.118}$$

$$\text{case b): } ak_\ell \le |A^*| \le (1+\delta_\ell)k_\ell, ak_\ell/2 \le w_1 \le |A^*| \tag{A.119}$$

$$\text{case c): } ak_\ell \le |A^*| \le (1+\delta_\ell)k_\ell, 1 \le w_1 \le ak_\ell/2. \tag{A.120}$$

In case a), $g^1_{3/5,5/6}(w_1)$ is uniformly lower bounded as

$$g^1_{3/5,5/6}(w_1) \ge \frac{\epsilon}{4}\log\left(1+\frac{P'}{4}\right) - a \tag{A.121}$$

$$\ge \frac{\epsilon}{8}\log\left(1+\frac{P'}{4}\right). \tag{A.122}$$

In case b), $g^1_{3/5,5/6}(w_1)$ is uniformly lower bounded as

$$g^1_{3/5,5/6}(w_1) \ge \frac{\epsilon}{4}\log\left(1+\frac{ak_\ell P'}{8}\right) - (1+\delta_\ell), \tag{A.123}$$

which grows without bound as k_ℓ increases.

In case c), $w_1/|A^*| \le 1/2$. Since $H_2(\cdot)$ is increasing on $[0, 1/2]$, we have

$$g^1_{3/5,5/6}(w_1) \ge \frac{\epsilon}{4}\log\left(1+\frac{w_1P'}{4}\right) - (1+\delta_\ell)H_2\left(\frac{w_1}{ak_\ell}\right) \tag{A.124}$$

$$\ge \frac{\epsilon}{4}\log\left(1+\frac{w_1P'}{4}\right) - \frac{2}{a}\frac{ak_\ell}{k_\ell}H_2\left(\frac{w_1}{ak_\ell}\right). \tag{A.125}$$

Applying Lemma 8 with $A_\ell = a\epsilon/8$, $B = P'/4$, $a_\ell = ak_\ell$, $\bar{w} = 1$ and $b_\ell = ak_\ell/2$, we conclude that there exists some ℓ_7 such that for all $\ell \ge \ell_7$, the RHS of (A.125) restricted

to $w_1 \in [1, ak_\ell/2]$ achieves minimum either at 1 or on $[cak_\ell/2, ak_\ell/2]$ for some $c \in (0, 1]$. Moreover, there exists some ℓ_8 such that for all $\ell \geq \ell_8$, $H_2\left(\frac{1}{ak_\ell}\right) \leq \frac{\epsilon}{16}\log\left(1+\frac{P'}{4}\right)$.

For every $\ell \geq \max\{\ell_7, \ell_8\}$, if the minimum of the RHS of (A.125) is achieved at $w_1 = 1$, then $g^1_{3/5,5/6}(w_1)$ in case c) is lower bounded as

$$g^1_{3/5,5/6}(w_1) \geq \frac{\epsilon}{4}\log\left(1+\frac{P'}{4}\right) - 2H_2\left(\frac{1}{ak_\ell}\right) \tag{A.126}$$

$$\geq \frac{\epsilon}{8}\log\left(1+\frac{P'}{4}\right). \tag{A.127}$$

For every $\ell \geq \max\{\ell_7, \ell_8\}$, if the minimum is achieved on $[cak_\ell/2, ak_\ell/2]$, then $g^1_{3/5,5/6}(w_1)$ in case c) is uniformly lower bounded as

$$g^1_{3/5,5/6}(w_1) \geq \frac{\epsilon}{4}\log\left(1+\frac{ack_\ell P'}{8}\right) - 2, \tag{A.128}$$

which grows without bound as k_ℓ increases.

By (A.122), (A.123), (A.127) and (A.128), it concludes that for all $\ell \geq \max\{\ell_7, \ell_8\}$,

$$g^1_{3/5,5/6}(w_1) \geq \frac{\epsilon}{8}\log\left(1+\frac{P'}{4}\right) \tag{A.129}$$

holds uniformly for all $1 \leq w_1 \leq |A^*|$. Combining the lower bound of $g^2_{3/5,5/6}(w_2)$ given by (A.115), we conclude that for all $\ell \geq \max\{\ell_5, \ell_6, \ell_7, \ell_8\}$, and all $1 \leq |A^*| \leq (1+\delta_\ell)k_\ell$,

$$h_{2/3,3/4}(w_1, w_2) \geq \min\left\{\frac{\epsilon}{48}\log\left(1+\frac{3P'}{5}\right), \frac{\epsilon}{8}\log\left(1+\frac{P'}{4}\right)\right\} \tag{A.130}$$

holds uniformly for all all $(w_1, w_2) \in \mathcal{W}^{(\ell)}$. Consequently, Lemma 3 is established for the case of $\phi = 0$. Combining the results of Appendix A.4.1 and Appendix A.4.2 proves Lemma 3.

A.5. Proof of Lemma 4

The lemma was proved for $k_n = o(n)$ in [**16**]. In this thesis, we prove the achievability result for $k_n = O(n)$. Throughout the proof, we focus on the case where k_n grows without bound as n increases, because the case of bounded k_n was included in [**16**].

Let $f(\gamma, \rho)$ be defined as (2.102). Choosing $\rho = 1$, we have

$$f(\gamma, 1) = \frac{1}{2}\log\left(1 + \frac{\gamma k_n P'}{2}\right) - \frac{(1-\epsilon)\gamma}{2}\log(1 + k_n P') - \frac{k_n}{n}H_2(\gamma). \tag{A.131}$$

Denote $c_n = k_n/n$ and $c = \limsup_{n\to\infty} c_n$. By differentiating $f(\gamma, 1)$ with respect to γ, we have

$$\frac{df(\gamma,1)}{d\gamma} = \frac{k_n P'}{4 + 2\gamma k_n P'} - \frac{1-\epsilon}{2}\log(1 + k_n P') + \frac{k_n}{n}\log\frac{\gamma}{1-\gamma}, \tag{A.132}$$

and

$$\frac{d^2 f(\gamma,1)}{d\gamma^2} = \frac{c_n}{\gamma(1-\gamma)} - \frac{(k_n P')^2}{2(2 + \gamma k_n P')^2}. \tag{A.133}$$

Note that $k_n = O(n)$, k_n is increasing without bound and $\gamma \geq 1/k_n$. Evidently,

$$8c_n \leq_n k_n P'^2/4 \tag{A.134}$$

$$\leq \frac{1}{4}(k_n P')^2\gamma. \tag{A.135}$$

Therefore, for sufficiently large n,

$$8c_n k_n P'\gamma + 8c_n \leq \frac{1}{2}(k_n P')^2\gamma \tag{A.136}$$

holds *uniformly* for all $\gamma \in [1/k_n, 1]$. Thus, for sufficiently large n,

$$\frac{d^2 f(\gamma,1)}{d\gamma^2} = \frac{(1+2c_n)\gamma^2(k_nP')^2 - (k_nP')^2\gamma + 8c_nk_nP'\gamma + 8c_n}{2(2+\gamma k_nP')^2\gamma(1-\gamma)} \tag{A.137}$$

$$\leq \frac{(1+2c_n)\gamma^2(k_nP')^2 - (k_nP')^2\gamma + \frac{1}{2}(k_nP')^2\gamma}{2(2+\gamma k_nP')^2\gamma(1-\gamma)} \tag{A.138}$$

$$= \frac{[(1+2c_n)\gamma - 1/2]\,(k_nP')^2}{2(2+\gamma k_nP')^2(1-\gamma)} \tag{A.139}$$

$$\leq \frac{[(1+4c)\gamma - 1/2]\,(k_nP')^2}{2(2+\gamma k_nP')^2(1-\gamma)} \tag{A.140}$$

holds uniformly for all γ.

We pick the constant $\gamma' = \frac{1/2}{1+4c}$. Since $0 \leq c < \infty$, we have $0 < \gamma' \leq 1/2$. By (A.140), for sufficiently large n, $\frac{d^2 f(\gamma,1)}{d\gamma^2} < 0$ holds uniformly for all $1/k_n \leq \gamma \leq \gamma'$. It means $f(\gamma,1)$ is concave over $\gamma \in [1/k_n, \gamma']$. Therefore, there exists some N_0 such that for all $n \geq N_0$,

$$\min_{1/k_n \leq \gamma \leq 1} f(\gamma,1) = \min\left\{ f(1/k_n,1),\ \min_{\gamma' \leq \gamma \leq 1} f(\gamma,1) \right\}. \tag{A.141}$$

If the minimum is achieved at $\gamma = 1/k_n$, we have

$$f(1/k_n,1) = \frac{1}{2}\log\left(1+\frac{P'}{2}\right) - \frac{(1-\epsilon)}{2k_n}\log(1+k_nP') - \frac{k_n}{n}H_2\left(\frac{1}{k_n}\right). \tag{A.142}$$

Since $(1/k_n)\log(1+k_nP')$ and $\frac{k_n}{n}H_2(1/k_n)$ vanishes as k_n increases, there exists N_1 such that for all $n \geq N_1$,

$$f(1/k_n,1) \geq \frac{1}{4}\log\left(1+\frac{P'}{2}\right). \tag{A.143}$$

If the minimum is achieved on $[\gamma', 1]$, we can lower bound $f(\gamma, 1)$ as

$$f(\gamma,1) \geq \frac{1}{2}\log\left(1+\frac{\gamma' k_n P'}{2}\right) - \frac{(1-\epsilon)}{2}\log(1+k_n P') - \frac{k_n}{n}. \tag{A.144}$$

Since $\log\left(1+\gamma' k_n P'/2\right) - \log(1+k_n P')$ and k_n/n converge to some constants, the lower bound given by (A.144) grows without bound as n increases.

In summary, combining (A.141), (A.143) and (A.144), it concludes that for all $n \geq \max\{N_0, N_1\}$ and all $|A^*|$, the error exponent is lower bounded

$$E_r \geq \min_{1/k_n \leq \gamma \leq 1} f(\gamma,1) \tag{A.145}$$

$$\geq \frac{1}{4}\log\left(1+\frac{P'}{2}\right). \tag{A.146}$$

The lemma is thus established.

A.6. Proof of Theorem 5

Unlike the case of unbounded k_n, there is a nonvanishing probability that the number of active users is zero. Let A^* denote the set of active users and $\mathcal{E}_d$ denote the event of detection error. Given an increasing sequence s_n satisfying the conditions specified in Theorem 5. The overall error probability can be calculated as

$$\mathsf{P}\left\{\mathcal{E}_d\right\} \leq \mathsf{P}\left\{|A^*| > s_n\right\} + \mathsf{P}\left\{\mathcal{E}_d | 1 \leq |A^*| \leq s_n\right\} + \mathsf{P}\left\{\mathcal{E}_d | |A^*| = 0\right\}. \tag{A.147}$$

By the Chernoff bound for binomial distribution [**37**], the probability that the number of active users is greater than s_n is calculated as

$$\mathsf{P}\left\{|A^*| > s_n\right\} \leq \exp\left(-k_n(s_n/k_n - 1)^2/3\right), \tag{A.148}$$

which vanishes as s_n grows without bound.

Note that the sequence s_n satisfies $\ell_n e^{-\delta s_n} \to 0$ for every $\delta > 0$ and

$$\lim_{n\to\infty} \frac{2s_n H_2(s_n/\ell_n)}{n\log(1+s_nP)} < 1, \tag{A.149}$$

which are the regularity conditions for unbounded k_n as specified in Case 1) of Theorem 1. The error probability $\mathsf{P}\left\{\mathcal{E}_d | 1 \leq |A^*| \leq s_n\right\}$ vanishes by following exactly the same as the analysis for the case of unbounded k_n (i.e., Case 1)) by treating s_n as an unbounded k_n.

We consider the identification error when $|A^*| = 0$. If no user is active, the received signal in the first n_0 channel uses is purely noise, i.e., $\boldsymbol{Y}^a = \boldsymbol{Z}^a$. By the user identification rule (2.43) with k_n replaced by s_n, a detection error occurs if at least one user is claimed to be active. The detection error probability can be calculated as

$$\mathsf{P}\left\{\mathcal{E}_d | |A^*| = 0\right\} \leq \sum_{w=1}^{(1+\delta_n)s_n} \binom{\ell_n}{w} \mathsf{P}\left\{\left|\left|\boldsymbol{Z}^a - \sum_{i=1}^{w} \boldsymbol{S}_i^a\right|\right|^2 \leq ||\boldsymbol{Z}^a||^2\right\}. \tag{A.150}$$

Let $\bar{\boldsymbol{S}} = \sum_{i=1}^{w} \boldsymbol{S}_i^a$. The entries of $\bar{\boldsymbol{S}}$ are i.i.d. according to $\mathcal{N}(0, wP')$. We have

$$\mathsf{P}\left\{\left|\left|\boldsymbol{Z}^a - \sum_{i=1}^{w} \boldsymbol{S}_i^a\right|\right|^2 \leq ||\boldsymbol{Z}^a||^2\right\} = \mathsf{P}\left\{\sum_{i=1}^{n_0} Z_i^a \bar{S}_i \geq \frac{1}{2}||\bar{\boldsymbol{S}}||^2\right\} \tag{A.151}$$

$$= \mathsf{E}\left\{\mathsf{P}\left\{\sum_{i=1}^{n_0} Z_i^a \bar{S}_i \geq \frac{1}{2}||\bar{\boldsymbol{S}}||^2\right\} \middle| \bar{\boldsymbol{S}}\right\}. \tag{A.152}$$

Conditioned on $\bar{\boldsymbol{S}}$, $\sum_{i=1}^{n_0} Z_i^a \bar{S}_i \sim \mathcal{N}(0, ||\bar{\boldsymbol{S}}||^2)$. Therefore,

$$\mathsf{E}\left\{\mathsf{P}\left\{\sum_{i=1}^{n_0} Z_i^a \bar{S}_i \geq \frac{1}{2}||\bar{\boldsymbol{S}}||^2\right\}\Big|\bar{\boldsymbol{S}}\right\} \leq \mathsf{E}\left\{\mathsf{Q}\left(\frac{||\bar{\boldsymbol{S}}||}{2}\right)\right\} \tag{A.153}$$

$$\leq \mathsf{E}\left\{e^{-\frac{||\bar{\boldsymbol{S}}||^2}{8}}\right\} \tag{A.154}$$

$$= (1 + wP'/4)^{-\frac{n_0}{2}} \tag{A.155}$$

where (A.154) is due to $\mathsf{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2}) du \leq e^{-x^2/2}$, and (A.155) follows because $||\bar{\boldsymbol{S}}||^2/wP$ is chi-squared distributed with n_0 degrees of freedom and $\mathsf{E}\left\{e^{tX}\right\} = (1-2t)^{-n/2}$ for a chi-squared distributed variable X with n degrees of freedom.

Combining (A.150), (A.152) and (A.155), the detection error probability for $|A^*| = 0$ can be upper bounded as

$$\mathsf{P}\left\{\mathcal{E}_d||A^*| = 0\right\} \leq \sum_{w=1}^{(1+\delta_n)s_n} \exp\left(\ell_n H_2(w/\ell_n) - \frac{n_0}{2}\log(1 + wP'/4)\right). \tag{A.156}$$

Let θ_n be given by (2.11) with k_n replaced by s_n and define $\theta = \lim_{n\to\infty} \theta_n$. By the choice of the signature length given by (2.84), $n_0 \geq_n \delta n$, where $\delta = \min(\epsilon, \theta(1+\epsilon)/2)$. For a large enough n, the error probability can be further upper bounded as

$$\mathsf{P}\left\{\mathcal{E}_d||A^*| = 0\right\} \leq \sum_{w=1}^{(1+\delta_n)s_n} \exp\left(-s_n h(w)\right), \tag{A.157}$$

where

$$h(w) = \frac{\delta n}{2s_n}\log(1 + wP'/4) - \frac{\ell_n}{s_n} H_2\left(\frac{w}{\ell_n}\right). \tag{A.158}$$

Note that $s_n = O(n)$. Applying Lemma 8 with $\ell = n$, $\bar{w} = 1$, $A_n = \delta n/(2s_n)$, $k_n = s_n$, $a_n = \ell_n$ and $b_n = (1+\delta_n)s_n$, we conclude that for large enougn n, the minimum of $h(w)$ restricted to $[1, (1+\delta_n)s_n]$ is achieved either at 1 or $[cs_n, (1+\delta_n)s_n]$ for some $0 < c \leq 1$.

As long as s_n satisfies the conditions as specified in Theorem 5, $\frac{\ell_n}{s_n} H_2\left(1/\ell_n\right)$ vanishes as n increases by Lemma 7. For large enough n, if the minimum of $h(w)$ is achieved at $w = 1$, $h(w)$ is uniformly lower bounded by some constant $c_0 > 0$. If the minimum of $h(w)$ is achieved on $[cs_n, (1+\delta_n)s_n]$, it implies that $h(w)$ grows without bound. It concludes that there exists some N_0, such that for all $n \geq N_0$, $h(w)$ is uniformly lower bounded by c_0 for all $1 \leq w \leq (1+\delta_n)s_n$.

By (A.157), there exists some N_0 and $c_0 > 0$ such that for all $n \geq N_0$,

$$\mathsf{P}\left\{\mathcal{E}_d | |A^*| = 0\right\} \leq (1+\delta_n) s_n e^{-c_0 s_n}. \tag{A.159}$$

Therefore, $\mathsf{P}\left\{\mathcal{E}_d | |A^*| = 0\right\}$ vanishes as the blocklength n increases. Since the three terms on the RHS of (A.147) all vanish, the overall detection error probability also vanishes.

A.7. Proof of Lemma 5

Since the users adopt Gaussian random codes, by treating the other users as interference, the first user to be decoded effectively sees Gaussian noise with variance $1+(k_n-1)P$. In order to prove the lemma, we show that the error probability of *any* $(\lceil \exp(v(n)) \rceil, n)$ code for the first user, where the message length $v(n)$ is given by (2.123), is lower bounded by some positive constant.

Let $\mathsf{P}_m(v(n), n)$ denote the average error probability for the first user achieved by the best channel code of blocklength n with message length $v(n)$, where each codeword

satisfies the *maximal* power constraint (2.2). Let $\mathsf{P}_e(v(n), n)$ denote the average error probability for the first user achieved by the best channel code of blocklength n with message length $v(n)$, where each codeword satisfies the *equal* power constraint, i.e., each codeword lies on a power-sphere $\sum_{i=1}^{n} s_{ki} = nP$. According to [**70**, eq. (83)], we have

$$\mathsf{P}_m(v(n-1), n-1) \geq \mathsf{P}_e(v(n-1), n). \tag{A.160}$$

We will lower bound $\mathsf{P}_e(v(n-1), n)$ in order to show that $\mathsf{P}_m(v(n), n)$ is strictly bounded away from zero for $v(n)$ given by (2.123).

Let $\lambda > 1$ be an arbitrary constant. Following the notations in [**71**, eq. (13)], let the decoding threshold be $\gamma = (n-1)(1-\lambda\epsilon)C$, $P'_{\boldsymbol{Y}}$ be the distribution of n i.i.d. Gaussian random variables with zero mean and variance $1+k_nP$, $P_{\boldsymbol{Y}|\boldsymbol{X}=[\sqrt{P},\cdots,\sqrt{P}]}$ be the distribution of n i.i.d. Gaussian random variables with mean $\sqrt{P}$ and variance $1+(k_n-1)P$, and $\beta_{1-\epsilon_n}\left(P_{\boldsymbol{Y}|\boldsymbol{X}=[\sqrt{P},\cdots,\sqrt{P}]}, P'_{\boldsymbol{Y}}\right)$, where $\beta_\alpha(P, P')$ is the minimum error probability of the binary hypothesis test under hypothesis P' if the error probability under hypothesis P is not larger than $1-\alpha$. The error probability $\mathsf{P}_e(v(n-1), n)$ is lower bounded as (see also [**71**, eq. (88)])

$$\begin{aligned}\mathsf{P}_e(v(n-1), n) \geq \mathsf{P}\left\{\frac{1}{2(1+Q)}\sum_{i=1}^{n} Q(1-Z_i^2) + 2\sqrt{Q}Z_i \leq -\lambda\epsilon nC - (1-\lambda\epsilon)C\right\} \\ - e^{-(\lambda-1)(n-1)\epsilon C}.\end{aligned} \tag{A.161}$$

We will follow a similar step as in [**71**] to further calculate the RHS of (A.161). Let $X_i = -Q(1-Z_i^2) - 2\sqrt{Q}Z_i$, where Z_i are i.i.d. standard Gaussian random variables.

Then $EX_i = 0$. By recalling Rozovsky's large deviation result [**71**, Theorem 5], we have

$$\mathsf{P}\left\{\sum_{i=1}^{n} X_i > x\sqrt{S}\right\} \geq \mathsf{Q}(x)e^{-\frac{A_1 T x^3}{S^{3/2}}}\left(1 - \frac{A_2 T x}{S^{3/2}}\right), \tag{A.162}$$

where A_1, A_2 are some universal constants, $S = \sum_{i=1}^{n} E|X_i|^2$, and $T = \sum_{i=1}^{n} E|X_i|^3$ which is equivalent to (2.126).

Then the first term in (A.161) can be calculated as

(A.163)
$$\mathsf{P}\left\{\frac{1}{2(1+Q)}\sum_{i=1}^{n} Q(1-Z_i^2) + 2\sqrt{Q}Z_i \leq -\lambda\epsilon nC - (1-\lambda\epsilon)C\right\} = \mathsf{P}\left\{\sum_{i=1}^{n} X_i \geq x\sqrt{S}\right\},$$

where $x = \frac{2(\lambda\epsilon n + 1 - \lambda\epsilon)C(1+Q)}{\sqrt{S}}$.

We can derive that $S = 2nQ(2+Q)$. Since $Q = \frac{P}{1+(k_n-1)P} \to 0$ as n increases, we have

$$E|X_i|^3 = O\left(Q^{3/2}\right). \tag{A.164}$$

Moreover, since $k = an$, we have $T = O\left(nQ^{3/2}\right)$ and therefore T tends to zero as n increases.

APPENDIX B

Appendix for Chapter 3

B.1. Proof of Multiton Error

We rewrite the bin values as follows,

$$\dot{\boldsymbol{Y}}_b = \sum_{k\in\mathcal{K}:b_k=b} A_k \dot{\boldsymbol{g}}_k + \dot{\boldsymbol{Z}}_b, \tag{B.1}$$

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

Suppose the incorrect estimate index from bin b is k_0. We have

$$\dot{A}_{k_0} = \sum_{k\in\mathcal{K}:b_k=b} \frac{1}{C_2} A_k \dot{\boldsymbol{g}}_{k_0}^\dagger \dot{\boldsymbol{g}}_k + \frac{1}{C_2} \dot{\boldsymbol{g}}_{k_0}^\dagger \dot{\boldsymbol{Z}}_b. \tag{B.2}$$

Thus,

$$\dot{\boldsymbol{Y}}_b - \dot{A}_{k_0}\dot{\boldsymbol{g}}_{k_0} = \sum_{k\in\mathcal{K}:b_k=b} A_k \left(\boldsymbol{I} - \frac{\dot{\boldsymbol{g}}_{k_0}\dot{\boldsymbol{g}}_{k_0}^\dagger}{C_2}\right)\dot{\boldsymbol{g}}_k + \left(\boldsymbol{I} - \frac{\dot{\boldsymbol{g}}_{k_0}\dot{\boldsymbol{g}}_{k_0}^\dagger}{C_2}\right)\dot{\boldsymbol{Z}}_b. \tag{B.3}$$

Let

$$\boldsymbol{S} = \sum_{k\in\mathcal{K}:b_k=b} A_k \dot{\boldsymbol{g}}_k \tag{B.4}$$

and $\boldsymbol{Q} = \boldsymbol{I} - \frac{1}{C_2}\dot{\boldsymbol{g}}_{k_0}\dot{\boldsymbol{g}}_{k_0}^{\dagger}$, then the first term in (B.3) can be written as

$$\sum_{k\in\mathcal{K}:b_k=b} A_k \left(\boldsymbol{I} - \frac{\dot{\boldsymbol{g}}_{k_0}\dot{\boldsymbol{g}}_{k_0}^{\dagger}}{C_2}\right)\dot{\boldsymbol{g}}_k = \boldsymbol{Q}\boldsymbol{S}. \tag{B.5}$$

The multiton bin cannot be distinguished when there exists some k_0 such that $||\boldsymbol{Y}_b - \dot{A}_{k_0}\dot{\boldsymbol{g}}_{k_0}||_2^2 \leq \eta$. In the following, we upper bound the error probability assuming $k_0 \notin \{k \in \mathcal{K} : b_k = b\}$. A similar analysis can be carried out for the case of $k_0 \in \{k \in \mathcal{K} : b_k = b\}$. The intuition is that if k_0 is the true device index, $\boldsymbol{Y}_b - \dot{A}_{k_0}\dot{\boldsymbol{g}}_{k_0}$ consists of noise and is of low power. For ease of notation, we write $\sum_{k\in\mathcal{K}:b_k=b}$ as $\sum_k$ in the following. The error probability can be upper bounded as

$$\begin{aligned}\mathsf{P}\left\{E_{b,1,2}\right\} \leq \mathsf{P}\left\{||\boldsymbol{Y}_b - \dot{A}_{k_0}\dot{\boldsymbol{g}}_{k_0}||_2^2 \leq \eta \middle| ||\boldsymbol{Q}\boldsymbol{S}||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\} + \\ \mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}||_2^2 \leq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\}.\end{aligned} \tag{B.6}$$

B.1.0.1. Bounding the first item of (B.6)**.** As in (3.59), $\boldsymbol{Q}$ can be written as $\boldsymbol{Q} = \boldsymbol{U}\Lambda\boldsymbol{U}^{\dagger}$, where $\Lambda = \mathrm{diag}\{0, 1, \cdots, 1\}$. We can write (B.3) as

$$||\dot{\boldsymbol{Y}}_b - \dot{A}_{k_0}\boldsymbol{g}_{k_0}||_2^2 = ||\boldsymbol{Q}\boldsymbol{S} + \boldsymbol{Q}\dot{\boldsymbol{Z}}_b||_2^2 \tag{B.7}$$

$$= ||\Lambda\boldsymbol{U}^{\dagger}\boldsymbol{S} + \Lambda\boldsymbol{U}^{\dagger}\dot{\boldsymbol{Z}}_b||_2^2 \tag{B.8}$$

$$= ||\boldsymbol{S}' + \Lambda\boldsymbol{Z}'||_2^2, \tag{B.9}$$

where $\boldsymbol{S}' = \Lambda\boldsymbol{U}^{\dagger}\boldsymbol{S}$ and $\boldsymbol{Z}' = \boldsymbol{U}^{\dagger}\dot{\boldsymbol{Z}}_b$.

By the triangular inequality $||\Lambda \boldsymbol{Z}'||_2 + ||\boldsymbol{S}' + \Lambda \boldsymbol{Z}'||_2 \geq ||\boldsymbol{S}'||_2$, we have

$$\mathsf{P}\left\{||\boldsymbol{Y}_b - \dot{A}_{k_0}\dot{\boldsymbol{g}}_{k_0}||_2^2 \leq \eta \middle| ||\boldsymbol{Q}\boldsymbol{S}||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\}$$

$$= \mathsf{P}\left\{||\boldsymbol{S}' + \Lambda \boldsymbol{Z}'||_2 \leq \sqrt{\eta} \middle| ||\boldsymbol{S}'||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\} \tag{B.10}$$

$$\leq \mathsf{P}\left\{||\boldsymbol{S}'||_2 - ||\Lambda \boldsymbol{Z}'||_2 \leq \sqrt{\eta} \middle| ||\boldsymbol{S}'||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\}. \tag{B.11}$$

For large enough K, $\frac{(C_2-1)\sum_k |A_k|^2}{2} \geq 4\eta$. Therefore, for large enough K,

$$\mathsf{P}\left\{||\boldsymbol{S}'||_2 - ||\Lambda \boldsymbol{Z}'||_2 \leq \sqrt{\eta} \middle| ||\boldsymbol{S}'||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\}$$

$$\leq \mathsf{P}\left\{||\Lambda \boldsymbol{Z}'||_2 \geq \sqrt{\eta} \middle| ||\boldsymbol{S}'||_2^2 \geq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\} \tag{B.12}$$

$$\leq 1/K^2, \tag{B.13}$$

where (B.13) is due to (3.65).

B.1.0.2. Bounding the second item of (B.6)**.** We introduce the definition of sub-Gaussian variable, which will be used in the proof.

Definition 7. *X is σ-subGaussian if there exists $\sigma > 0$ such that*

$$\mathsf{E}\left\{\exp(tX)\right\} \leq \exp(\sigma^2 t^2/2), \quad \forall t \geq 0. \tag{B.14}$$

Definition 8 (subGaussian norm)**.** *The subGaussian norm of the random variable X is defined as*

$$||X||_{\phi_2} = \sup_{p \geq 1} p^{-1/2} \left(\mathsf{E}|X|^p\right)^{1/p}. \tag{B.15}$$

The second item of (B.6) is derived in the following steps. We first show that the real and imaginary parts of $\boldsymbol{S}$ are subGaussian variables. Then, we show that $||\boldsymbol{QS}||_2^2$ are concentrated around $(C_2-1)\sum_k |A_k|^2$ with high probability using the large deviation results on subGaussian variables. In the following, we denote X_R and X_I as the real and imaginary component of X, respectively.

Lemma 9. *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 $\boldsymbol{S}$ defined as* (B.4)*, respectively. Then $S_{c,R}$ are i.i.d. $\sqrt{\sum_k A_{k,R}^2}$-subGaussian random variables with $\mathsf{E}S_{c,R} = 0$ and the subGaussian norm satisfies $||S_{c,R}||_{\phi_2} \leq 2\sqrt{\sum_k A_{k,R}^2}$. Similarly, $S_{c,I}$ are i.i.d. $\sqrt{\sum_k A_{k,I}^2}$-subGaussian random variables with $\mathsf{E}S_{c,I} = 0$ and the subGaussian norm satisfies $||S_{c,I}||_{\phi_2} \leq 2\sqrt{\sum_k A_{k,I}^2}$.*

Proof. Since g_k^c is Rademacher variable, it is 1-subGaussian with mean zero. Thus, $\mathsf{E}S_{c,R} = 0$. Moreover, g_k^c are independent across k. According to Lemma 10 and Lemma 11, $S_{c,R} = \sum_k A_{k,R}\dot{g}_k^c$ is $\sqrt{\sum_k A_{k,R}^2}$-subGaussian. $\{S_{c,R}\}_{c=0}^{C_2-1}$ are independent, because $\dot{g}_k^c$ are independent across $c = 0, \cdots, C_2-1$.

We know that $\mathsf{E}\{\exp(tS_{c,R})\} \leq \exp\left(\sum_k A_{k,R}^2 t^2/2\right)$. Let $X = S_{c,R}/\sqrt{2\sum_k A_{k,R}^2}$. Then $\mathsf{E}\{\exp(tX)\} \leq \exp(t^2/4)$. According to Theorem 9, for all $p \geq 1$, $(\mathsf{E}|X|^p)^{1/p} \leq \sqrt{2p}$, which yields

$$(\mathsf{E}|S_{c,R}|^p)^{1/p} \leq 2\sqrt{\sum_k A_{k,R}^2 p}. \tag{B.16}$$

The subGaussian norm of $S_{c,R}$ is upper bounded as $||S_{c,R}||_{\phi_2} \leq 2\sqrt{\sum_k A_{k,R}^2}$. The statement for the imaginary parts follow similarly. □

In order to apply Theorem 10 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

$$||\boldsymbol{Q}||_F^2 = tr\{\boldsymbol{Q}\boldsymbol{Q}^\dagger\} \tag{B.17}$$

$$= tr\{\boldsymbol{Q}\} \tag{B.18}$$

$$= C_2 - 1. \tag{B.19}$$

where (B.18) follows by $\boldsymbol{Q}\boldsymbol{Q}^\dagger = \boldsymbol{Q}$, and (B.19) follows because the sum of the eigenvalues of $\boldsymbol{Q}$ is $C_2 - 1$.

Since the largest eigenvalue of $\boldsymbol{Q}$ is 1, the operator norm of $\boldsymbol{Q}$ is calculated as

$$||\boldsymbol{Q}|| = \max_{\boldsymbol{x}\neq 0} \frac{||\boldsymbol{Q}\boldsymbol{x}||_2}{||\boldsymbol{x}||_2} = 1. \tag{B.20}$$

Conditioned on $\boldsymbol{g}_{k_0}$,

$$\mathsf{E}\left\{||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 \middle| \boldsymbol{g}_{k_0}\right\} = \mathsf{E}\left\{\boldsymbol{S}_R^\dagger \boldsymbol{Q}\boldsymbol{Q}^\dagger \boldsymbol{S}_R \middle| \boldsymbol{g}_{k_0}\right\} \tag{B.21}$$

$$= \mathsf{E}\left\{\boldsymbol{S}_R^\dagger \boldsymbol{Q}\boldsymbol{S}_R \middle| \boldsymbol{g}_{k_0}\right\} \tag{B.22}$$

$$= \mathsf{E}\{S_{c,R}^2\} tr\{\boldsymbol{Q}\} \tag{B.23}$$

$$= (C_2 - 1)\sum_k A_{k,R}^2, \tag{B.24}$$

where (B.23) follows because $\{S_{c,R}\}_{c=0}^{C_2-1}$ are i.i.d. distributed. Similarly, $\mathsf{E}\left\{||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 \middle| \boldsymbol{g}_{k_0}\right\} = (C_2 - 1)\sum_k A_{k,I}^2$.

Applying Theorem 10 with $\boldsymbol{Z} = \boldsymbol{S}_R$, $\boldsymbol{A} = \boldsymbol{Q}$ with $||\boldsymbol{Q}|| = 1$, $||\boldsymbol{Q}||_F^2 = C_2 - 1$ and $K = 2\sqrt{\sum_k A_{k,R}^2}$ yields

$$\mathsf{P}\left\{\left|||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 - (C_2-1)\sum_k A_{k,R}^2\right| > t\middle|\boldsymbol{g}_{k_0}\right\} \leq$$

$$2\exp\left(-c\min\left(\frac{t^2}{16(C_2-1)(\sum_k A_{k,R}^2)^2}, \frac{t}{4\sum_k A_{k,R}^2}\right)\right). \tag{B.25}$$

Letting $t = (C_2-1)\sum_k A_{k,R}^2/2$, we have

$$\mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 \leq \frac{(C_2-1)\sum_k A_{k,R}^2}{2}\middle|\boldsymbol{g}_{k_0}\right\} \leq 2\exp\left(-\frac{c}{64}(C_2-1)\right). \tag{B.26}$$

Since (B.26) holds for all $\boldsymbol{g}_{k_0}$, we have

$$\mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 \leq \frac{(C_2-1)\sum_k A_{k,R}^2}{2}\right\} \leq 2\exp\left(-\frac{c}{64}(C_2-1)\right). \tag{B.27}$$

Similarly, we have

$$\mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}_I||_2^2 \leq \frac{(C_2-1)\sum_k A_{k,I}^2}{2}\right\} \leq 2\exp\left(-\frac{c}{64}(C_2-1)\right). \tag{B.28}$$

Since $\boldsymbol{Q}$ is a real-valued matrix, $||\boldsymbol{Q}\boldsymbol{S}||_2^2 = ||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 + ||\boldsymbol{Q}\boldsymbol{S}_I||_2^2$. Moreover, $\sum_k |A_k|^2 = \sum_k A_{k,R}^2 + A_{k,I}^2$. Combining (B.27) and (B.28), we have

$$\mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}||_2^2 \leq \frac{(C_2-1)\sum_k |A_k|^2}{2}\right\}$$

$$\leq \mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}_R||_2^2 \leq \frac{(C_2-1)\sum_k A_{k,R}^2}{2}\right\} + \mathsf{P}\left\{||\boldsymbol{Q}\boldsymbol{S}_I||_2^2 \leq \frac{(C_2-1)\sum_k A_{k,I}^2}{2}\right\} \tag{B.29}$$

$$\leq 4\exp\left(-\frac{c}{64}(C_2-1)\right). \tag{B.30}$$

Combining (B.6), (B.13) and (B.30), there exists some large enough β_1 such that $C_2 = \beta_1 \lceil \log N \rceil$ and

$$\mathsf{P}\left\{E_{b,1,2}\right\} \leq \frac{1}{N^2}. \tag{B.31}$$

B.2. Proof of Lemma 6

We focus on 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 can be written as

$$y_i = a_k s_{k,i} + z_i, \tag{B.32}$$

where the time-domain samples of the pilots $s_{k,i} \sim \mathcal{CN}(0,1)$ and $z_i \sim \mathcal{CN}(0, 2\sigma_z^2)$, where the variance is bounded as

$$(K-1)\underline{a}^2 + 2\sigma^2 \leq 2\sigma_z^2 \leq (K-1)\bar{a}^2 + 2\sigma^2. \tag{B.33}$$

Let $\mathcal{I}$ be given by (3.17). The number of samples contained in $\mathcal{I}$ is $|\mathcal{I}| = BC_3$. According to (3.18), The test metric is calculated as

$$\mathcal{T}(m) = \sum_{i \in \mathcal{I}} a_k s_{i+m} s_i^* + z_{i+m} s_i^*. \tag{B.34}$$

When the delay is correctly estimated, the test metric given by (3.18) is calculated as

$$\mathcal{T}(0) = BC_3 a_k + \sum_{i \in \mathcal{I}} z_i s_i^*. \tag{B.35}$$

By the central limit theorem, $\mathcal{T}(0)$ is distributed according to $\mathcal{CN}\left(BC_3 a_k, KBC_3\sigma_0^2\right)$ and $\mathcal{T}(m)$ is distributed according to $\mathcal{CN}\left(0, KBC_3\sigma_m^2\right)$, where σ_0 and σ_m are some bounded constants independent of K.

We define a threshold $\bar{\mathcal{T}} = \underline{a}BC_3/2$. If $|\mathcal{T}(0)| > \bar{\mathcal{T}}$ and $|\mathcal{T}(m)| < \bar{\mathcal{T}}$ for all $m = 1, \cdots, M$, the delay can be correctly estimated. Therefore, the error probability can be upper bounded as

$$P_e \leq \sum_{m=1}^{M} \mathsf{P}\{|\mathcal{T}(m)| \geq \bar{\mathcal{T}}\} + \mathsf{P}\{|\mathcal{T}(0)| \leq \bar{\mathcal{T}}\}. \tag{B.36}$$

The first term can be upper bounded as

$$\mathsf{P}\{|\mathcal{T}(m)| \geq \bar{\mathcal{T}}\} \leq \mathsf{P}\{|\mathrm{Re}\{\mathcal{T}(m)\}| \geq \bar{\mathcal{T}}/2\} + \mathsf{P}\{|\mathrm{Im}\{\mathcal{T}(m)\}| \geq \bar{\mathcal{T}}/2\} \tag{B.37}$$

$$\leq 4\mathsf{P}\left\{\mathcal{N}\left(0, \frac{KBC_3\sigma_m^2}{2}\right) \leq \frac{\underline{a}BC_3}{4}\right\} \tag{B.38}$$

$$\leq 4e^{-\frac{\underline{a}^2 BC_3}{16\sigma_m^2 K}}. \tag{B.39}$$

Let ϕ be the phase of a_k. The second term in (B.36) can be upper bounded as

$$\mathsf{P}\{|\mathcal{T}(0)| \leq \bar{\mathcal{T}}\} = \mathsf{P}\{||a_k|BC_3 + e^{-j\phi}\sum_{i\in\mathcal{I}} z_i s_i^*| \leq \bar{\mathcal{T}}\} \tag{B.40}$$

$$\leq \mathsf{P}\left\{\mathcal{N}\left(|a_k|BC_3, \frac{KBC_3\sigma_0^2}{2}\right) \leq \frac{\underline{a}BC_3}{2}\right\} \tag{B.41}$$

$$\leq e^{-\frac{\underline{a}^2 BC_3}{4\sigma_0^2 K}}. \tag{B.42}$$

Combining (B.36), (B.39) and (B.42), given that $B = \beta_0 K$, there exists some β_3 such that $C_3 = \beta_3\lceil\log(K+M)\rceil$ and the error probability of delay estimation is less than $1/K^2$.

B.3. Auxiliary Results on Sub-Gaussian Variables

The following lemmas and theorem are established in [**72**].

Lemma 10. *Suppose X is σ-subGaussian, then aX is $a\sigma$-subGaussian.*

Lemma 11. *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.*

Theorem 9 (Characterization of subGaussian variables)**.** *Let $\mathsf{E}X = 0$. The following are equivalent:*

(1) $\mathsf{E}(e^{tX}) \leq e^{\frac{t^2}{4}}$.

(2) $\forall t > 0$, $\mathsf{P}\{|X| > t\} \leq 2\exp(-t^2)$.

(3) $\forall p \geq 1$, $(\mathsf{E}|X|^p)^{1/p} \leq \sqrt{2p}$.

Proof. (1) $\Rightarrow$ (2)

$$\mathsf{P}\{X > t\} \leq \frac{\mathsf{E}\exp(\lambda X)}{\exp(\lambda t)} \tag{B.43}$$

$$\leq \exp\left(\frac{\lambda^2}{4} - \lambda t\right) \tag{B.44}$$

$$\leq \exp(-t^2). \tag{B.45}$$

Similarly, we have $\mathsf{P}\{X < -t\} \leq \exp(-t^2)$.

(2) $\Rightarrow$ (3)

$$\mathsf{E}|X|^p = \mathsf{E}\left\{\int_0^{|X|} pt^{p-1}dt\right\} \tag{B.46}$$

$$= \mathsf{E}\left\{\int_0^{\infty} pt^{p-1}1\{|X| \geq t\}dt\right\} \tag{B.47}$$

$$= \int_0^{\infty} pt^{p-1}\mathsf{P}\{|X| \geq t\}dt \tag{B.48}$$

$$\leq \int_0^{\infty} pt^{p-1}2\exp(-t^2)dt \tag{B.49}$$

$$= \int_0^{\infty} pu^{p/2-1}\exp(-u)du \tag{B.50}$$

$$= p\Gamma(p/2) \tag{B.51}$$

$$= 2\Gamma(p/2+1) \tag{B.52}$$

$$\leq 2(p/2)^{p/2}. \tag{B.53}$$

Therefore, $(\mathsf{E}|X|^p)^{1/p} \leq 2^{1/p}(p/2)^{1/2}$. Since $p \geq 1$, $(\mathsf{E}|X|^p)^{1/p} \leq \sqrt{2p}$.

The proof of (3) $\Rightarrow$ (1) is omitted. We only need the result of (1) $\Rightarrow$ (3) in the thesis. $\square$

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

Theorem 10 (Hanson-Wright inequality [**73**])**.** *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* $||Z_i||_{\phi_2} \leq K$*. Let* $\boldsymbol{A}$ *be an* $n \times n$ *matrix. Then, for every* $t \geq 0$*,*

$$\mathsf{P}\left\{|\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{Z} - \mathsf{E}\left\{\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{Z}\right\}| > t\right\} \leq 2\exp\left[-c\min\left(\frac{t^2}{K^4||\boldsymbol{A}||_F^2}, \frac{t}{K^2||\boldsymbol{A}||}\right)\right],$$

where the operator norm of $\boldsymbol{A}$ *is* $||\boldsymbol{A}|| = \max_{\boldsymbol{x}\neq 0} \frac{||\boldsymbol{A}\boldsymbol{x}||_2}{||\boldsymbol{x}||_2}$ *and the Frobenius norm of* $\boldsymbol{A}$ *is* $||\boldsymbol{A}||_F = (\sum_{i,j} |A_{i,j}|^2)^{1/2}$.