NORTHWESTERN UNIVERSITY

Virtual Full Duplex Wireless Networks

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

Lei Zhang

EVANSTON, ILLINOIS

August 2012

ABSTRACT

Virtual Full Duplex Wireless Networks

Lei Zhang

A novel paradigm is proposed in this thesis for designing the physical and medium access control (MAC) layers of wireless ad hoc or peer-to-peer networks formed by half-duplex radios. A node equipped with such a radio cannot simultaneously transmit and receive useful signals at the same frequency. Unlike in conventional designs, where a node's transmission frames are scheduled away from its reception, each node transmits its signal through an assigned on-off duplex mask (or signature) over every frame interval, and receive a signal through each of its own off-slots. This is called *rapid on-off-division duplex (RODD)*. Over the period of a single frame, every node can transmit a message to some or all of its peers, and may simultaneously receive a message from each peer. Thus RODD achieves virtual full-duplex communication using half-duplex radios without complicated scheduling at the frame level.

This treatise consists of four parts, which are presented in Chapters 2 - 5, respectively. As a first step toward quantifying the advantage of on-off signaling, Chapter 2 studies the capacity of scalar discrete-time Gaussian channels subject to duty cycle constraint as well

as average transmit power constraint. A unique discrete input distribution is shown to achieve the channel capacity. In many situations, numerically optimized on-off signaling can achieve much higher rate than Gaussian signaling over a deterministic schedule of frame transmissions.

To further explore the advantages of RODD in wireless networks with half-duplex constraint, Chapter 3 evaluates the throughput of RODD, which is found to be significantly larger than that of ALOHA under some general settings. RODD is especially efficient in the case that the dominant traffic is mutual broadcast, i.e., all nodes wish to broadcast information to and receive information from their respective one-hop peers.

Chapter 4 proposes a novel solution to the mutual broadcast problem in wireless networks by applying RODD signaling. Decoding can be viewed as a compressed sensing or sparse recovery problem. In the case that each message consists of a small number of bits, an iterative message-passing algorithm based on belief propagation is developed. The proposed scheme achieves several times the rate of slotted ALOHA and CSMA with the same packet error rate (1%).

In Chapter 5, RODD signaling derived from Reed-Muller codes is used to carry out peer discovery in wireless networks. To identify its peers out of a large network address space, each node solves a compressed sensing problem using a chirp decoding algorithm. The algorithm is scalable to networks of virtually any size of practical interest due to its sub-linear complexity. The new scheme allows all nodes to simultaneously discover their respective one-hop peers within a single frame transmission, which entails significantly less overhead than conventional random-access discovery schemes.

In summary, this thesis proposes RODD signaling, which achieves virtual full-duplex communication in wireless networks, and contributes to the understanding of its theory and applications.

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 enlightening guidance towards students demonstrate the qualities of a great scholar and professor. I could not have imagined a better advisor and mentor in my graduate study.

I would like to thank Professor Randall Berry and Professor Aggelos Katsaggelos, for serving in my thesis committee and giving me insightful comments.

I am indebted to many colleagues in the Communications and Networking Laboratory at Northwestern University: Changxin Shi, Mingguang Xu, Binnan Zhuang, Hang Zhou, Jun Luo, Yan Zhu, Ka Hung Hui, Suvarup Saha, Fei Teng, Kai Shen, Hui Li, Jieying Chen, and Ning Wen. I am also grateful to Jialue Fan for our precious friendship. All of them have made my Ph.D. life full of fun.

Last, and most importantly, I want to express my heartfelt appreciation to my parents, Wei Zhang and Meimei Jiang. Their selfless love and endless support is in the end what makes this thesis possible. To them I dedicate this dissertation.

To my parents

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Despite decades of advances in wireless and networking technologies, to design a functional and reliable mobile ad hoc or peer-to-peer network remains enormously challenging [**3**]. The main roadblocks include the difficult nature of the wireless medium and the mobility of wireless terminals, among others. A crucial constraint on wireless systems is the half-duplex nature of affordable radios, which prevents a radio from receiving any useful signal at the same time and over the same frequency band within which it is transmitting [**65**]. The physical reason is that during transmission, a radio's own signal picked up by its receive antenna is typically orders of magnitude stronger than the signals from its peers, such that the desired signals are obliterated due to noise and the limited dynamic range of the radio frequency (RF) circuits. The half-duplex constraint has far-reaching consequences in the design of wireless networks: The uplink and downlink transmissions in any cellular-type network are separated using time-division duplex (TDD) or frequency-division duplex (FDD); standard designs of wireless ad hoc networks schedule transmission frames of a node away from the time and frequency slot over which the node receives data [**89**].

In this thesis, the half-duplex constraint is addressed at a fundamental level, which is that the received signal of a half-duplex node is viewed as erasures during periods of its own active transmission. We recognize that, it is neither necessary nor efficient to separate the transmission slots and listening slots of a node in the timescale of a frame of

hundreds or thousands of symbols as in TDD. We propose a novel technique referred to as *rapid on-off-division duplex (RODD)*. The key idea is to let each node transmit according to a unique *on-off duplex mask (or signature)* over a frame of symbols or slots, so that the node can receive useful signals from its peers during the off-slots interleaved between its on-slot transmissions. Importantly, all nodes may send (error-control-coded) information simultaneously over a frame interval, as long as the masks of peers are sufficiently different, so that a node receives enough signals during its off-slots to decode information from its peers. Over the period of a single frame, every node simultaneously broadcasts a message to some or all other peers, and may receive a message from each peer at the same time. Thus, the virtual full-duplex communication is enabled by using half-duplex radios.

Switching the carrier on and off at the timescale of one or several symbols is feasible, thanks to the sub-nanosecond response time of RF circuits. In fact, on-off signaling over submillisecond slots is used by time-division multiple-access (TDMA) cellular systems such as GSM. Time-hopping impulse radio transmits on and off at nanosecond intervals [**86**], which is orders of magnitude faster than needed by RODD (in microseconds). Moreover, receiving signals during one's own off-slots avoids self-interference and circumvents the dynamic range issue which plagues other full-duplex schemes, such as code-division duplex (CDD) [**7**, **49**].

The signaling of RODD is quite different from that of TDD and FDD. It is important to note that FDD and TDD suffice in cellular networks is because uplink and downlink transmissions are clearly separable. In peer-to-peer networks, however, one node's transmission (downlink) is its peer's reception (uplink), so that there is no absolute separation of the notions of uplink and downlink. The prevalence of FDD and TDD in current ad

hoc networks is in part inherited from the more mature technologies of wired and cellular networks, and due to the difficulty of separating superposed signals. Advances in multiuser detection and decoding (e.g., [**34**]) and recent progress in sparse recovery have enabled new technologies that break away from the model of packet collisions, and hence set the stage for RODD.

Wireless networks using RODD have unique advantages: (1) RODD enables virtual full-duplex transmission and greatly simplifies the design of higher-layer protocols. In particular, "scheduling" is carried out in a microscopic timescale over the slots, so that there is no need to separate transmitting and listening frames; (2) RODD signaling takes full advantage of the superposition and broadcast nature of the wireless medium. As we shall see, the throughput of a RODD-based network is greater than that of ALOHA-type random access, and is more than twice as large as that of slotted ALOHA in many cases; (3) RODD signaling is particularly efficient when the traffic is predominantly peer-to-peer broadcast, such as in mobile systems used in local advertising, spontaneous social networks, emergency situations or on battlefield; (4) Communication overhead usually comes as an afterthought in network design, whereas RODD enables extremely efficient exchange of a small amount of state information amongst neighbors; (5) Because nodes simultaneously transmit, the channel-access delay is typically smaller and more stable than in conventional reservation or scheduling schemes.

1.1. Related Work

There have been numerous works on the design of physical and MAC layers for wireless networks (see the surveys [**47**, **66**, **73**] and references therein). Two major challenges

need to be addressed: One is the half-duplex constraint; the other is the broadcast and superposition nature of the wireless medium, so that simultaneous transmissions interfere with each other at a receiver.

1.1.1. State of the Art

State-of-the-art designs either schedule nodes orthogonally ahead of transmissions, or apply an ALOHA-type random access scheme, or use a mixture of random access and scheduling reservation [**58**]. Typically, the collision model is assumed, where if multiple nodes simultaneously transmit, their transmissions fail due to collision at the receiver. Under such a model, random access leads to poor efficiency (e.g., ALOHA's efficiency is less than $1/e$). On the other hand, scheduling node transmissions is often difficult and subject to the hidden terminal and exposed terminal problems.

Despite the half-duplex constraint, it is neither necessary nor efficient to separate a node's transmission slots and listening slots in the timescale of a frame. In fact, time-sharing can fall considerably short of the theoretical optimum. In particular, non-transmission can be regarded as an additional symbol for signaling (besides 0 and 1), whose positions can be used to communicate information (see also [**46**, **54**, **55**]).

Several recent works on the implementation of physical and MAC layers break away from the collision model and single-user transmission. For example, superposition coding for degraded broadcast channels has been implemented using software-defined radios [**24**]. Analog network coding has also been implemented based on 802.11 technology [**41**], where, when two senders transmit simultaneously, their packets collide, or more precisely, superpose at the receiver, so that if the receiver already knows the content of one of the packets,

it can cancel the interference and decode the other packet. Similar ideas have been proven feasible in some other contexts to achieve interference cancellation in unmanaged ZigBee networks [**33**], ZigZag decoding for 802.11 in [**25**], and interference alignment and cancellation in [**26**].

1.1.2. Relationship to CDD, TDD, and Time-Hopping Impulse Radio

Rapid on-off-division duplex is related to code-division duplex, which was proposed in the context of code-division multiple access (CDMA) [**7**]. In CDD, orthogonal (typically antipodal) spreading sequences are allocated to uplink and downlink communications, so that a receiver ideally cancels self-interference by matched filtering with its own receive spreading sequence. Despite the claimed higher spectral efficiency than that of TDD and FDD in [**49**], CDD is not used in practice because it is difficult to maintain orthogonality due to channel impairments and suppress self-interference which is orders of magnitude stronger than the desired signal. In RODD, the desired signal is sifted through the off-slots of the transmission frame, so that the leakage of the transmit energy into the received signal is kept to the minimum. RODD can be viewed as CDD using on-off sequences without spreading.

RODD can also be viewed as (very fast) TDD with irregular symbol-level transition between transmit and receive slots as well as coding over many slots. Although on-off signaling can in principle be applied to the frequency domain, it would be much harder to implement sharp band-pass filters to remove self-interference.

The RODD signaling also has some similarities to that of time-hopping impulse radio [**72**, **85**]. Both schemes transmit a sequence of randomly spaced pulses. There are

crucial differences: Each on-slot (or pulse) in RODD spans one or a few data symbols (in microseconds), whereas each pulse in impulse radio is a baseband monocycle of a nanosecond or so duration. Moreover, impulse radio is carrier-free and spreads the spectrum by many orders of magnitude, whereas RODD uses a carrier and is not necessarily spread-spectrum.

1.1.3. Relationship to Other Full-Duplex Schemes

Recently, it has been proposed in the literature that full-duplex communication with half-duplex radios can be achieved based on interference cancellation. The key technique is to let the receive chain of a node remove the self-interference caused by the known signal from its transmit chain, so that reception can be concurrent with transmission. The idea is not new (see, e.g. [**15**, **43**, **63**, **64**]), but has only been successfully implemented in a laboratory environment in the past years [**16**, **21**, **38**, **68**]. Two groups's work has received much attention. One group uses a balanced/unbalanced transformer to negate the transmitted signal for analog cancellation at the receiver, followed by subsequent digital cancellation. It is reported that up to 73 dB self-interference is successfully removed in a controlled laboratory environment [**38**]. (This outperforms the earlier beamforming idea in [**16**] from the same group.) The other group uses a combination of transmit/receive antenna separation and analog and optional digital self-interference cancellation. They report that up to 80 dB self-interference can be removed [**68**].

Comparing with RODD which has to introduce off-slots in a frame to achieve *virtual* full-duplex communication, interference-based full duplex scheme would be more efficient

if the self-interference can be completely removed. However, there may exist space limitations for adequate antenna separation. And analog cancellation is hard with multiple transmit antennas, because it is not easy to separate several self-interference signals from their superposition for cancellation. Also, self-interference cancellation is unlikely to be feasible when the power of its own transmission is around the noise level. In such cases, RODD is a more viable solution. In fact, RODD and interference-based full duplex scheme can be combined together: Off-slots are introduced to avoid self-interference, whereas during on-slots the self-interference can be removed or at least suppressed to yield more useful received signals.

1.2. System Model

We start with a physical-layer model for RODD in wireless networks with perfect synchronization. Consider a network with N nodes, indexed by $1, \ldots, N$. Suppose all transmissions are over the same frequency band. Let time be divided into slots of equal length, and one or a few symbols can be transmitted over each slot, where in the latter case we regard the transmit signal as a vector symbol. Let each frame consist of M slots and the on-off signature of node n be denoted as $\boldsymbol{S}_n = [s_{1n}, \ldots, s_{Mn}]^\top$. During slot m, node j may transmit a symbol if $s_{mj} = 1$, whereas if $s_{mj} = 0$, node j listens to the channel and emits no energy. The physical link between any pair of nodes can be modeled as a fading channel. Let the path loss satisfy a power law with exponent α. Let Δ denote the duration of a slot and $p_j(t)$ denote the waveform for a single slot of node j (which may include multipath components). Let d_{nj} denote the distance between nodes n and j, h_{nj} denote the fading coefficient and X_{mj} denote the transmitted symbol of node j at

time slot m. Let us also assume that the signaling of each node is subject to unit average power constraint, i.e.,

$$\sum_{m=1}^{M} s_{mn}|x_{mn}|^2 \leq M \tag{1.1}$$

for each codeword $(x_{1n}, \ldots, x_{Mn})$. The received signal of node n over a single frame is described by

$$Y_n(t) = \sum_{j \neq n} \sqrt{\gamma_j} d_{nj}^{-\alpha/2} h_{nj} \sum_{m=1}^{M} (1 - s_{mn}) s_{mj} X_{mj} \mathbf{1}_{\{t \in [(m-1)\Delta, m\Delta]\}} p_j(t - (m-1)\Delta - \tau_{nj}) + W_n(t) \tag{1.2}$$

where τ_{nj} denotes the relative delay from node j to node n, $W_n(t)$ denotes additive white Gaussian noise (AWGN) of unit spectral density and γ_j essentially denotes the signal-to-noise ratio (SNR) of node j over each active slot in absence of fading and path loss. Here $\mathbf{1}_{\{t \in [a,b]\}}$ denotes a rectangular waveform on the interval $[a, b]$. The received signal of node n over its own off-slots is the noisy superposition of the signals from other nodes over those slots.

The SNR of the link from node j to node n can be regarded as $\gamma_{nj} = \gamma_j \, d_{nj}^{-\alpha} |h_{nj}|^2$. We say node j is a (one-hop) *neighbor* or *peer* of node n if γ_{nj} exceeds a given threshold.[1] Let the set of neighbors (or peers) of n be denoted as ∂n, which is also called its *neighborhood*. We are only interested in communication over links between neighbors. Suppose the propagation delay from nodes in the neighborhood can be ignored compared with the duration of each on/off slot, i.e., $\tau_{nj} \approx 0$, the discrete-time counterpart of model (1.2)

[1]The neighbor relationship is not necessarily reciprocal because $\gamma_j |h_{nj}|^2$ and $\gamma_n |h_{jn}|^2$ need not be identical.

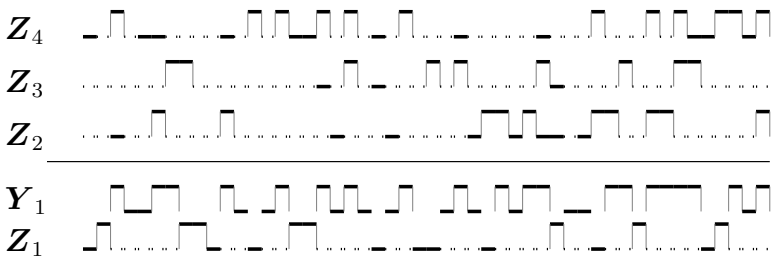

Figure 1.1. RODD signaling of four nodes.

with perfect intersymbol interference (ISI) cancellation is

$$Y_{mn} = (1 - s_{mn}) \sum_{j \in \partial n} \sqrt{\gamma_j} d_{nj}^{-\alpha/2} h_{nj} s_{mj} X_{mj} + V_{mn} \tag{1.3}$$

where Y_{mn} denotes the received signal of node n during each slot $m \in \{1, \ldots, M\}$ and V_{mn} consists of the additive noise W_{mn} as well as the aggregate interference caused by non-neighbors.

Note that (1.2) and (1.3) model the half-duplex constraint at a fundamental level: If node n transmits during a slot, then its received signal during that slot is erased. Fig. 1.1 illustrates a snapshot of RODD signals of four nodes taken over 50 slots. Here $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_4$ represent the transmitted signals of node 1 through node 4, respectively, where the solid lines represent on-slots and the dotted lines represent off-slots. The received signal of node 1 through its off-slots is $\boldsymbol{Y}_1$, which is the superposition of $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$, and $\boldsymbol{Z}_4$ with erasures at its own on-slots (represented by blanks). That is, RODD forms fundamentally a *multiaccess channel with erasure*.

1.3. Design Issues

1.3.1. Synchronization

Synchronicity has been studied extensively in the context of ad hoc and sensor networks. One possible shortcut, if applicable, is to have all nodes globally synchronized using the global positioning system (GPS) or via listening to base stations in an existing cellular network. Alternatively, various distributed algorithms for reaching consensus [**69**, **70**, **76**] can be used to achieve local synchronicity, i.e., the timing fluctuates over the network, but is a smooth function geographically. Local synchronicity can also be achieved using a common reference, such as a strong beacon signal. In a RODD system, it suffices to have all communicating peers be approximately symbol-synchronized, as long as the timing difference (including the propagation delay) is much smaller than the symbol interval. For instance, if neighbors are within 300 meters, the propagation delay is at most 1 microsecond, which is much smaller than the bit or pulse interval of a typical MANET. More pronounced propagation delays can also be explicitly addressed in the physical model, but this is out of the scope of this thesis.

In order to decode the information from neighbors, it is necessary to acquire their timing (or relative delay) regardless of whether RODD or any other physical- and MAC-layer technology is used. Timing acquisition and decoding are generally easier if the frames arriving at a receiver are synchronous locally within each neighborhood, although synchronization is not a necessity. Whether synchronizing the nodes is worthwhile is a challenging question, which is not discussed further in this thesis.

1.3.2. Signature Distribution and Neighbor Discovery

In this thesis, it is assumed that each node has complete knowledge of the signatures of all nodes. It is, however, not necessary to directly distribute the set of duplex masks to each node in the network. It suffices to let nodes generate their signatures using the same pseudo-random number generator or some other deterministic function with their respective unique network interface address (NIA) as the seed. In principle, every node can reconstruct all signatures by enumerating all NIAs.

Before establishing data links, a node needs to acquire the identities or NIAs of its neighbors. This is called *neighbor discovery (or peer discovery)*. By applying RODD signaling, all nodes simultaneously send their on-off signatures and make measurements through their respective off-slots. Therefore, all nodes can simultaneously discover their respective neighbors, i.e., virtual full-duplex discovery is achievable,

References [**52**, **53**] have pointed out that to identify a small number of neighbors out of a large collection of nodes based on the signal received over a linear channel is fundamentally a *compressed sensing (or sparse recovery)* problem, for which a small number of measurements (channel uses) suffice [**13**, **19**].[2] Using *pseudo-random on-off signatures* for neighbor discovery was proposed in [**52**, **53**] along with a group testing algorithm. The key observation is that, from one node's viewpoint, for each slot with (essentially) no energy received, any node who would have transmitted a pulse during that slot cannot be a neighbor. A node basically goes through every off-slot and eliminates nodes incompatible with the measurement; the surviving nodes are then regarded as neighbors. Using

[2]Several authors have studied user activity problem in cellular networks using multiuser detection techniques [**4**,**5**,**50**]. These works assume channel coefficients are known to the receiver, which is not the case in most networks.

random signatures requires only noncoherent energy detection and has been shown to be effective and efficient at moderate SNRs. The disadvantage, however, is that the system is not scalable to accommodate a very large address space (beyond 20-bit NIAs), because the discovery complexity is linear in the node population. In Chapter 5, we propose a new scheme with deterministic signatures which overcomes the scalability problem and has better performance.

1.4. Outline and Contributions

In this thesis, we study both theoretic limitations and applications of RODD in wireless ad hoc and peer-to-peer networks.

As a first step toward quantifying the advantage of on-off signaling, Chapter 2 answers a basic question of what is the optimal signaling for a discrete-time scalar AWGN channel with duty cycle constraint as well as average transmission power constraint. The duty cycle constraint can be regarded as a requirement on the minimum fraction of nontransmissions or zero symbols in each codeword. A unique *discrete* input distribution is shown to achieve the channel capacity. In many situations, numerically optimized on-off signaling can achieve much higher rate than Gaussian signaling over a deterministic transmission schedule. This is in part because the positions of nontransmissions in a codeword can convey information. The results suggest that, under the duty cycle constraint, departing from the usual paradigm of intermittent frame transmissions may yield substantial gain.

To further explore the advantages of RODD in wireless networks with half-duplex constraint, Chapter 3 presents a study of network capacity in the scenario that the traffic is mutual broadcast, i.e., all nodes wish to broadcast information to and receive information

from their respective peers simultaneously. The throughput of a fully-connected, synchronized, RODD-based network is studied under the assumption that each node has complete knowledge of the duplex masks of all nodes in the network. Numerical results demonstrate that the throughput of RODD evaluated under some general settings is significantly larger than that of ALOHA.

In Chapter 4, we study the mutual broadcast problem as an important application of RODD in wireless networks. The defining feature of our scheme is to let all nodes send their messages at the same time, where each node broadcasts an on-off codeword (selected from its unique codebook according to its message). Decoding can be viewed as a problem of compressed sensing (or sparse support recovery) based on linear measurements. In the case that each message consists of a small number of bits, an iterative message-passing algorithm based on belief propagation is developed, and its performance is characterized using a state evolution formula in the limit where each node has a large number of peers. In a network consisting of Poisson distributed nodes with the same transmit power, numerical results demonstrate that the proposed scheme achieves several times the rate of slotted ALOHA and CSMA with the same packet error rate (1%).

Chapter 5 proposes a novel scheme using RODD signaling for the problem of neighbor discovery in wireless networks, namely, each node wishes to discover and identify the NIAs of those nodes within its single hop. The key technique is to assign each node a unique on-off signature derived from a second-order Reed-Muller code and let all nodes simultaneously transmit their signatures. To identify its neighbors out of a large network address space, each node solves a compressed sensing problem using a chirp decoding algorithm. The decoding complexity is sublinear in the NIA space, which is in principle

scalable to billions of nodes with 48-bit IEEE 802.11 MAC addresses. A network of over one million Poisson distributed nodes (with 20-bit NIAs) is studied numerically, where each node has 30 neighbors on average, and the channel between each pair of nodes is subject to path loss and Rayleigh fading. Within a single frame of 4,096 symbols, nodes can discover their respective neighbors with on average 99.8% accuracy at 11 dB SNR. The new scheme is much more efficient than conventional random-access discovery, where nodes have to retransmit over many frames with random delays to be successfully discovered.

Chapter 6 concludes this thesis, and also discusses some future research directions.

CHAPTER 2

Capacity of Gaussian Channels with Duty Cycle Constraint

In many wireless communication systems, a radio is designed to transmit actively only for a fraction of the time, which is known as its *duty cycle*. For example, the ultra-wideband system in [**39**] transmits short bursts of signals to trade bandwidth for power savings. The physical half-duplex constraint also requires a radio to stop transmission over a frequency band from time to time if it wishes to receive useful signals over the same band. Thus wireless relays are subject to duty cycle constraint, so do cognitive radios which have to listen to the channel frequently to avoid causing interference to primary users. The *de facto* standard solution under duty cycle constraint is to transmit packets intermittently.

This chapter studies the fundamental question of what is the optimal signaling for a Gaussian channel with duty cycle constraint as well as average transmission power constraint. An important observation is that the signaling in nontransmission periods can be regarded as transmission of a special *zero* signal. We make a simplistic and idealized assumption that the analog waveform corresponding to each transmitted symbol is exactly of the span of one symbol interval. Practical pulse shaping filters, however, would introduce higher duty cycle in continuous time than its discrete-time counterpart. In order to alleviate such impact in practice, designs for pulse shaping filters need to be taken into consideration. In this work, however, we restrict our focus on the discrete-time model, where the duty cycle constraint is equivalent to a requirement on the minimum

fraction of zero symbols in each transmitted codeword. The mathematical model of the AWGN channel and input constraints is described in Section 2.1.

Determining the capacity of a channel subject to various input constraints is a classical problem. It is well-known that Gaussian signaling achieves the capacity of a Gaussian channel with average input power constraint only. In addition, Zamir [**90**] shows that the mutual information rate achievable using a white Gaussian input never incurs a loss of more than half a bit per sample with respect to the power constrained capacity. Furthermore, Smith [**77**] investigated the capacity of a scalar AWGN channel under both peak power constraint and average power constraint. The input distribution that achieves the capacity is shown to be discrete with a finite number of probability mass points. The discreteness of capacity-achieving distributions for various channels, including quadrature Gaussian channels, and Rayleigh-fading channels is also established in [**2**, **32**, **36**, **42**, **74**, **75**]. Chan [**14**] studied the capacity-achieving input distribution for conditional Gaussian channels which form a general channel model for many practical communication systems.

The main results of this chapter are summarized in Section 2.2. Because all costs associated with the constraints can be decomposed into per-letter costs, the optimal input distribution is independent and identically distributed (i.i.d.). In Section 2.3, We use a similar approach as in [**77**] and [**14**] to show that the capacity-achieving input distribution for an AWGN channel with duty cycle constraint and average power constraints is discrete. Unlike in [**77**] and [**14**], the optimal distribution has an infinite number of probability mass points, whereas only a finite number of the points are found in every bounded interval. This allows efficient numerical optimization of the input distribution.

Numerical results in Section 2.4 demonstrate that using a numerically optimized discrete signaling achieves higher rates than using Gaussian signaling over a deterministic transmission schedule. For example, if the radio is allowed to transmit no more than half the time, i.e., the duty cycle is no greater than 50%, a near-optimal discrete input achieves 50% higher rate at 10 dB SNR. This suggests that, compared to intermittently transmitting frames using Gaussian or Gaussian-like signaling, it is more efficient to disperse nontransmission symbols within each frame to form codewords, which results in a form of *on-off* signaling.

One of the reasons for the superiority of on-off signaling is that the positions of nontransmission symbols can be used to convey information, the impact of which is particularly significant in case of low SNR or low duty cycle. This has been observed in the past. For example, as shown in [**54**] (see also [**46**, **55**]), time sharing or time-division duplex (TDD) can fall considerably short of the theoretical limits in a relay network: The capacity of a cascade of two noiseless binary bit pipes through a half-duplex relay is 1.14 bits per channel use, which far exceeds the 0.5 bit achieved by TDD and even the 1 bit upper bound on the rate of binary signaling.

2.1. System Model

Consider digital communication systems where coded data are mapped to waveforms for transmission. Usually there is a collection of pulse waveforms, where each pulse represents a symbol (or letter) from a discrete alphabet. We view nontransmission over a symbol interval as transmitting the all zero waveform. In other words, a symbol interval

of nontransmission is simply regarded as transmitting a special symbol "0," which carries no energy.

As far as the capacity-achieving input is concerned it suffices to consider the baseband discrete-time model for the AWGN channel. The received signal over a block of n symbols can be described by

$$Y_i = X_i + N_i \tag{2.1}$$

where $i = 1, \ldots, n$, X_i denotes the transmitted symbol at time i and $N_1, \ldots, N_n$ are independent standard Gaussian random variables. For simplicity, we assume no inter-symbol interference is at receiver. Each symbol modulates a continuous-time pulse waveform for transmission. Under the assumption that the width of all pulses is exactly of one symbol interval, the duty cycle is equal to the fraction of nonzero symbols in a codeword.

Let $1 - q$ denote the maximum duty cycle allowed. In this chapter, we require every codeword $(x_1, x_2, \cdots, x_n)$ to satisfy

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(x_i \neq 0\right) \leq 1 - q \tag{2.2}$$

where $\mathbf{1}(\cdot)$ is the indicator function. In addition, we consider the usual average input power constraint,

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq \gamma. \tag{2.3}$$

In many wireless systems, the transmitter's activity is constrained in the frequency domain as well as in the time domain. In principle, the results in this chapter also apply to the more general model where the duty cycle constraint is on the time-frequency plane.

2.2. Main Results

Let μ denote the distribution of the channel input X. The set of distributions with duty cycle no greater than $1-q$ and power constraint γ is denoted by

$$\Lambda(\gamma, q) = \left\{\mu : \mu(\{0\}) \geq q, \, \mathsf{E}_\mu\left\{X^2\right\} \leq \gamma\right\}. \tag{2.4}$$

It should be understood that μ is a probability measure defined on the Borel algebra on the real number set, denoted by $\mathcal{B}(\mathbb{R})$.

Theorem 2.1. *The capacity of the additive white Gaussian noise channel* (2.1) *with duty cycle no greater than* $1-q$ *and the average power no greater than* γ *is*

$$C(\gamma, q) = \max_{\mu \in \Lambda(\gamma, q)} I(\mu)\,. \tag{2.5}$$

In particular, the following properties hold:

a) the maximum of (2.5) *is achieved by a unique (capacity-achieving) distribution* $\mu_0 \in \Lambda(\gamma, q)$*;*

b) μ_0 *is symmetric about 0 and its second moment is exactly equal to* γ*; and*

c) μ_0 *is discrete with an infinite number of probability mass points, whereas the number of probability mass points in any bounded interval is finite.*

The proof of Theorem 2.1 is relegated to Section 2.3. Property (b) suggests that the capacity-achieving input always exhausts the power budget. Property (c) indicates that the capacity-achieving input can be well approximated by some discrete inputs with finite alphabet, which can be computed using numerical methods. The achievable rate of numerically optimized input distribution is studied in Section 2.4.

2.3. Proof of Theorem 2.1

This section is devoted to a proof of Theorem 2.1. The conditional probability density function (pdf) of the output given the input of the AWGN channel (2.1) is

$$p_{Y|X}(y|x) = \phi(y - x) \tag{2.6}$$

where

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \tag{2.7}$$

is the standard Gaussian pdf.

The capacity of the AWGN channel is achieved by an i.i.d. process and the duty cycle constraint reduces to a per symbol cost constraint. For given input distribution μ, the pdf of the output exists and is expressed as

$$p_Y(y;\mu) = \int p_{Y|X}(y|x)\,\mu(\mathrm{d}x) = \mathsf{E}_\mu\left\{\phi(y - X)\right\}\,. \tag{2.8}$$

Denote the relative entropy $D\left(p_{Y|X}(\cdot|x)\|p_Y(\cdot;\mu)\right)$ by $d(x;\mu)$, which is expressed as

$$d(x;\mu) = \int_{-\infty}^{\infty} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y;\mu)} \mathrm{d}y\,. \tag{2.9}$$

The mutual information $I(\mu) = I(X;Y)$ is then

$$I(\mu) = \int d(x;\mu)\,\mu(\mathrm{d}x) = \mathsf{E}_\mu\left\{d(X;\mu)\right\}. \tag{2.10}$$

The capacity of the AWGN channel under per-letter duty cycle constraint and power constraint is evidently given by the supremum of the mutual information $I(\mu)$ where

$\mu \in \Lambda(\gamma, q)$. The achievability and converse of this result can be established using standard techniques in information theory.

The proof of property (a) is presented in Section 2.3.1. Now suppose μ_0 is the unique capacity-achieving distribution, property (b) is established as follows. Since the mirror reflection of μ_0 about 0 is evidently also a maximizer of (2.5), the uniqueness requires that μ_0 be symmetric. Note that linear scaling of the input to increase its power maintains its duty cycle and cannot reduce the mutual information, as the receiver can add noise to maintain the same SNR. By the uniqueness of the maximizer μ_0, the power constraint must be binding, i.e., the second moment of μ_0 must be equal to γ. In order to prove property (c), we first establish a sufficient and necessary condition for μ_0 in Section 2.3.2 and then apply it to show the discreteness of μ_0 in Section 2.3.3.

2.3.1. Existence and Uniqueness of μ_0

Let $\mathcal{P}$ denote the collection of all Borel probability measures defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is a topological space with the topology of weak convergence [**78**]. We first establish the following lemma.

Lemma 2.1. *$\Lambda(\gamma, q)$ is compact in the topological space $\mathcal{P}$.*

Proof. According to [**78**], the topology of weak convergence on $\mathcal{P}$ is metrizable. Therefore, by Prokhorov's theorem [**62**], in order to prove that $\Lambda(\gamma, q)$ is compact in $\mathcal{P}$, it suffices to show that it is both tight and closed.

For any $\epsilon > 0$, there exits an $a_\epsilon > 0$, such that for all $\mu \in \Lambda_\gamma$,

$$\mu(|X| > a_\epsilon) \leq \frac{\mathsf{E}_\mu\{X^2\}}{a_\epsilon^2} \leq \frac{\gamma}{a_\epsilon^2} < \epsilon \tag{2.11}$$

by Chebyshev's inequality. Choose $K_\epsilon = [-a_\epsilon, a_\epsilon]$, then K_ϵ is compact in $\mathbb{R}$ and $\mu(K_\epsilon) \geq 1 - \epsilon$ for all $\mu \in \Lambda(\gamma, q)$, thus $\Lambda(\gamma, q)$ is tight.

Let $B_m = \left[-\frac{1}{m}, \frac{1}{m}\right]$ for $m = 1, 2, \ldots$. Let $\{\mu_n\}_{n=1}^\infty$ be a convergent sequence in $\Lambda(\gamma, q)$ with limit μ_0. Since $\mu_n(B_m) \geq q$ for every m, n, we have [**78**, Section 3.1]

$$q \leq \limsup_{n\to\infty} \mu_n(B_m) \leq \mu_0(B_m), \tag{2.12}$$

and hence

$$\mu_0(\{0\}) = \mu_0\left(\bigcap_{m=1}^\infty B_m\right) = \lim_{m\to\infty} \mu_0(B_m) \geq q. \tag{2.13}$$

Moreover, let $f(x) = x^2$ which is continuous and bounded below. By weak convergence [**78**, Section 3.1], we have

$$\mathsf{E}_{\mu_0}\left\{X^2\right\} = \int f \mathrm{d}\mu_0 \leq \liminf_{n\to\infty} \int f \mathrm{d}\mu_n \leq \gamma. \tag{2.14}$$

Therefore, $\mu_0 \in \Lambda(\gamma, q)$, i.e., $\Lambda(\gamma, q)$ is closed, and the compactness of $\Lambda(\gamma, q)$ then follows.

□

Since the mutual information $I(\mu)$ is continuous on $\mathcal{P}$ [**87**, Theorem 9], it must achieve its maximum on the compact set $\Lambda(\gamma, q)$. Hence the capacity-achieving distribution μ_0 exists.

According to [**87**, Corollary 2], the mutual information $I(\mu)$ is strictly concave. It is easy to see that $\Lambda(\gamma, q)$ is convex. Hence the capacity-achieving distribution μ_0 must be unique.

2.3.2. Sufficient and Necessary Conditions

We denote the finite-power set as

$$\Lambda(q) = \cup_{0\leq\gamma<\infty}\Lambda(\gamma, q). \tag{2.15}$$

Let $\phi(\cdot)$ defined in (2.7) be extended to the complex plane. The relative entropy $d(x;\mu)$ defined in (2.9) can be extended to the complex plane $\mathbb{C}$ and has the following property:

Lemma 2.2. *For any $\mu \in \Lambda(q)$ and $z \in \mathbb{C}$,*

$$d(z;\mu) = \int_{-\infty}^{\infty} \phi(y-z) \log \frac{\phi(y-z)}{p_Y(y;\mu)} \mathrm{d}y \tag{2.16}$$

is a holomorphic function of z on $\mathbb{C}$. Consequently, $d(x;\mu)$ is a continuous function of x on $\mathbb{R}$.

Proof. It can be shown that $\int_{-\infty}^{\infty} \phi(y-z)\log\phi(y-z)\mathrm{d}y$ is a constant, thus a holomorphic function of z on $\mathbb{C}$. Therefore, it remains to prove that

$$\xi(z) = \int_{-\infty}^{\infty} \phi(y-z)\log p_Y(y;\mu)\mathrm{d}y \tag{2.17}$$

is a holomorphic function of z on $\mathbb{C}$.

First, by Jensen's inequality, we have

$$p_Y(y;\mu) = \mathsf{E}_\mu\left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-X)^2}{2}}\right\} \tag{2.18}$$

$$\geq \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\mathsf{E}_\mu\left\{(y-X)^2\right\}} \tag{2.19}$$

$$= e^{-\frac{1}{2}y^2-ay-b} \tag{2.20}$$

where $a = -\mathsf{E}_\mu\{X\}$ and $b = \frac{1}{2}\left(\mathsf{E}_\mu\{X^2\} + \log(2\pi)\right)$ are real numbers due to the fact that $\mu \in \Lambda(q)$. Thus, $p_Y(y;\mu) \in [e^{-\frac{1}{2}y^2-ay-b}, 1]$, i.e.,

$$|\log P_Y(y;\mu)| \leq \frac{1}{2}y^2 + ay + b. \tag{2.21}$$

As a result, we have

$$|\phi(y-z)\log p_Y(y;\mu)| \leq \frac{1}{\sqrt{2\pi}}\left|e^{-\frac{(y-z)^2}{2}}\right|\left(\frac{1}{2}y^2 + ay + b\right) \tag{2.22}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\mathrm{Re}(z))^2-\mathrm{Im}^2(z)}{2}}\left(\frac{1}{2}y^2 + ay + b\right), \tag{2.23}$$

which is integrable. (Here $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ represent the real and imaginary parts of z, respectively.) It follows that $\xi(z)$ given by (2.17) exists for any $\mu \in \Lambda(q)$ and $z \in \mathbb{C}$.

Suppose U is an open and bounded subset of $\mathbb{C}$. There exists an $r > 0$ such that $|\mathrm{Re}(z)| \leq r$ and $|\mathrm{Im}(z)| \leq r$ for all $z \in U$. It is easy to check that

$$e^{-\frac{(y-\mathrm{Re}(z))^2}{2}} \leq e^{-\frac{y^2}{2}+|yr|} \tag{2.24}$$

$$\leq e^{-\frac{y^2}{2}+yr} + e^{-\frac{y^2}{2}-yr} \tag{2.25}$$

$$= e^{\frac{r^2}{2}}\left[e^{-\frac{1}{2}(y-r)^2} + e^{-\frac{1}{2}(y+r)^2}\right]. \tag{2.26}$$

Combining (2.22) and (2.26) yields that

$$|\phi(y-z)\log p_Y(y;\mu)| \leq \frac{e^{r^2}}{\sqrt{2\pi}}\left[e^{-\frac{1}{2}(y-r)^2}+e^{-\frac{1}{2}(y+r)^2}\right]\left(\frac{1}{2}y^2+ay+b\right), \tag{2.27}$$

which is integrable. Therefore, the integral $\int_{-\infty}^{\infty}\phi(y-z)\log p_Y(y;\mu)\mathrm{d}y$ is uniformly convergent for all $z \in U$. Moreover, $\phi(y-z)\log p_Y(y;\mu)$ is a holomorphic function of z on U for each $y \in \mathbb{R}$. According to the differentiation lemma [**48**], $\xi(z)$ is a holomorphic function of z on U. It then follows that it is holomorphic on the whole complex plane $\mathbb{C}$. Lemma 2.2 is thus established. □

Let $F(\mu)$ be a real-valued function defined on the convex set $\Lambda(q)$ and $\mu_0 \in \Lambda(q)$. Define the weak derivative of $F(\mu)$ at μ_0 as

$$F'_{\mu_0}(\mu) = \lim_{\theta\to 0^+}\frac{F\left((1-\theta)\mu_0+\theta\mu\right)-F(\mu_0)}{\theta} \tag{2.28}$$

whenever the limit exists. The following result, which finds its parallel in [**2**, **14**, **36**] gives the weak derivative of the mutual information function $I(\mu)$.

Lemma 2.3. *Let $\mu_0, \mu \in \Lambda(q)$, the weak derivative of the mutual information function $I(\mu)$ at μ_0 is*

$$I'_{\mu_0}(\mu) = \int d(x;\mu_0)\,\mu(\mathrm{d}x) - I(\mu_0). \tag{2.29}$$

Proof. Define $\mu_\theta = (1-\theta)\mu_0 + \theta\mu$ for all $\theta \in (0,1]$. It can be shown that

$$\begin{aligned}
&\frac{1}{\theta}\left(I(\mu_\theta) - I(\mu_0)\right) \\
&= \frac{1}{\theta}\int \left(d(x;\mu_\theta) - d(x;\mu_0)\right)\, \mu_\theta(\mathrm{d}x) + \frac{1}{\theta}\left(\int d(x;\mu_0)\, \mu_\theta(\mathrm{d}x) - I(\mu_0)\right) \qquad (2.30)\\
&= -\frac{1}{\theta}\int_{-\infty}^{\infty} p_Y(y;\mu_\theta) \log \frac{p_Y(y;\mu_\theta)}{p_Y(y;\mu_0)} \mathrm{d}y + \int d(x;\mu_0)\, \mu(\mathrm{d}x) - I(\mu_0). \qquad (2.31)
\end{aligned}$$

Therefore, it suffices to show that

$$\lim_{\theta\to 0^+} \int_{-\infty}^{\infty} \frac{1}{\theta} p_Y(y;\mu_\theta) \log \frac{p_Y(y;\mu_\theta)}{p_Y(y;\mu_0)} \mathrm{d}y = 0. \qquad (2.32)$$

In the remainder of this proof, we find a function independent of θ that dominates the integrand so that dominated convergence theorem can be used to establish (2.32) by exchanging the order of the limit and the integral therein.

Lemma 2.4. *Let $\theta, a, b \in (0,1]$. Define*

$$f(\theta) = \frac{(1-\theta)a + \theta b}{\theta} \log \frac{(1-\theta)a + \theta b}{a}\,, \qquad (2.33)$$

then

$$|f(\theta)| \le b + a - b\log b - b\log a\,. \qquad (2.34)$$

Proof. It is easy to check that $f(1) = b\log\frac{b}{a}$, $f(0^+) = b - a$ and

$$f'(\theta) = \frac{b-a}{\theta} - \frac{a}{\theta^2}\log\left(1 - \theta + \frac{b}{a}\theta\right). \qquad (2.35)$$

Define $g(\theta) = \theta(b-a) - a\log\left(1-\theta+\frac{b}{a}\theta\right)$ for $\theta \in (0,1]$, then we have

$$g'(\theta) = \frac{\theta(b-a)^2}{(1-\theta)a+\theta b} \geq 0. \tag{2.36}$$

Since $g(0^+) = 0$, $g(\theta) \geq 0$ for all $\theta \in (0,1]$. According to (2.35), we have $f'(\theta) = \frac{g(\theta)}{\theta^2} \geq 0$. It follows that for all $\theta \in (0,1]$,

$$b-a = f(0^+) \leq f(\theta) \leq f(1) = b\log\frac{b}{a}, \tag{2.37}$$

and hence

$$|f(\theta)| \leq \max\left\{|b-a|, \left|b\log\frac{b}{a}\right|\right\} \tag{2.38}$$

$$\leq b + a - b\log b - b\log a. \tag{2.39}$$

Lemma 2.4 is thus established. □

Applying Lemma 2.4 with $a = p_Y(y;\mu_0)$ and $b = p_Y(y;\mu)$, we have

$$\left|\frac{1}{\theta}p_Y(y;\mu_\theta)\log\frac{p_Y(y;\mu_\theta)}{p_Y(y;\mu_0)}\right| \leq p_Y(y;\mu) + p_Y(y;\mu_0)$$

$$- p_Y(y;\mu)\log p_Y(y;\mu) - p_Y(y;\mu)\log p_Y(y;\mu_0) \tag{2.40}$$

where the right hand side is an integrable function of y by the result that $-\int_{-\infty}^{\infty} p_Y(y;\mu_2)\log p_Y(y;\mu_1)\mathrm{d}y < \infty$ for any $\mu_1, \mu_2 \in \Lambda(q)$. In fact, as in the proof of Lemma 2.2

(see (2.21)), there exist $a, b \in \mathbb{R}$ such that $|\log p_Y(y;\mu_1)| \leq \frac{1}{2}y^2 + ay + b$. Therefore,

$$\int_{-\infty}^{\infty} |p_Y(y;\mu_2) \log p_Y(y;\mu_1)| \mathrm{d}y \leq \int_{-\infty}^{\infty} p_Y(y;\mu_2) \left(\frac{1}{2}y^2 + ay + b\right) \mathrm{d}y \tag{2.41}$$

$$= \frac{1}{2}\mathsf{E}_{\mu_2}\left\{X^2\right\} + a\mathsf{E}_{\mu_2}\left\{X\right\} + b + \frac{1}{2} \tag{2.42}$$

$$< \infty \tag{2.43}$$

due to the assumption that $\mu_2 \in \Lambda(q)$.

Therefore, the dominated convergence theorem provides that

$$\lim_{\theta \to 0^+} \frac{1}{\theta} \int_{-\infty}^{\infty} p_Y(y;\mu_\theta) \log \frac{p_Y(y;\mu_\theta)}{p_Y(y;\mu_0)} \mathrm{d}y = \int_{-\infty}^{\infty} \lim_{\theta \to 0^+} \frac{1}{\theta} p_Y(y;\mu_\theta) \log \frac{p_Y(y;\mu_\theta)}{p_Y(y;\mu_0)} \mathrm{d}y \tag{2.44}$$

$$= \int_{-\infty}^{\infty} \left(p_Y(y;\mu) - p_Y(y;\mu_0)\right) \mathrm{d}y \tag{2.45}$$

$$= 0. \tag{2.46}$$

Lemma 2.3 is thus proved. □

We establish the following sufficient and necessary condition for the optimal input distribution.

Lemma 2.5. *Let*

$$f_\lambda(x;\mu) = d(x;\mu) - I(\mu) - \lambda(x^2 - \gamma). \tag{2.47}$$

Then $\mu_0 \in \Lambda(\gamma, q)$ achieves the capacity if and only if there exists $\lambda \geq 0$ such that $\lambda \mathsf{E}_{\mu_0}\left\{X^2 - \gamma\right\} = 0$ and $\mathsf{E}_\mu\left\{f_\lambda(X;\mu_0)\right\} \leq 0$ for all $\mu \in \Lambda(q)$.

Proof. Define the Lagrangian

$$J(\mu) = I(\mu) - \lambda \mathsf{E}_\mu \left\{X^2 - \gamma\right\} \tag{2.48}$$

where λ is the Lagrange multiplier. Since $\Lambda(q)$ is a convex set and $I(\mu) < \infty$ on $\Lambda(q)$, μ_0 is capacity-achieving if and only if there exists $\lambda \geq 0$ such that the following conditions hold [**51**]:

(i) $\lambda \mathsf{E}_{\mu_0} \{X^2 - \gamma\} = 0$;

(ii) for all $\mu \in \Lambda(q)$, $J(\mu_0) \geq J(\mu)$.

Due to concavity of $I(\mu)$, $J(\mu)$ is also concave. Condition (ii) is then equivalent to that the weak derivative $J'_{\mu_0}(\mu) \leq 0$ for all $\mu \in \Lambda(q)$.

By Lemma 2.3, the linearity of $\mathsf{E}_\mu \{X^2 - \gamma\}$ with respect to (w.r.t.) μ and Condition (i), $J'_{\mu_0}(\mu)$ can be easily calculated as

$$J'_{\mu_0}(\mu) = \mathsf{E}_\mu \{f_\lambda(X; \mu_0)\}. \tag{2.49}$$

Therefore, Condition (ii) is equivalent to $\mathsf{E}_\mu \{f_\lambda(X; \mu_0)\} \leq 0$ for all $\mu \in \Lambda(q)$. Thus Lemma 2.5 follows. □

We call $x \in \mathbb{R}$ a point of increase of a measure μ if $\mu(O) > 0$ for every open subset O of $\mathbb{R}$ containing x. Let S_μ be the set of points of increase of μ. Based on Lemma 2.5, we derive another sufficient and necessary condition for the optimal input distribution, which will be used to prove Property (c) of Theorem 2.1 in Section 2.3.3.

Lemma 2.6. *Let*

$$g_\lambda(x;\mu) = q f_\lambda(0;\mu) + (1-q) f_\lambda(x;\mu). \tag{2.50}$$

Then $\mu_0 \in \Lambda(\gamma, q)$ achieves the capacity if and only if there exists $\lambda \geq 0$ such that for every $x \in \mathbb{R}$,

$$g_\lambda(x;\mu_0) \leq 0\,. \tag{2.51}$$

Furthermore, $g_\lambda(x;\mu_0) = 0$ for every $x \in S_{\mu_0}\backslash\{0\}$.

Proof. The necessity part is shown as follows. Suppose μ_0 achieves the capacity, then by Lemma 2.5, there exists $\lambda \geq 0$ such that $\lambda \mathsf{E}_{\mu_0}\{X^2 - \gamma\} = 0$ and $\mathsf{E}_\mu\{f_\lambda(X;\mu_0)\} \leq 0$ for all $\mu \in \Lambda(q)$. For any $x \in \mathbb{R}\backslash\{0\}$, choose μ such that $\mu(\{0\}) = q$ and $\mu(\{x\}) = 1-q$, so by the fact that $\mu \in \Lambda(q)$, we have

$$0 \geq \mathsf{E}_\mu\{f_\lambda(X;\mu_0)\} = q f_\lambda(0;\mu_0) + (1-q) f_\lambda(X;\mu_0). \tag{2.52}$$

Due to the continuity of $d(x;\mu_0)$ by Lemma 2.2, $f_\lambda(x;\mu_0)$ is also continuous so that (2.52) holds for all $x \in \mathbb{R}$, i.e., $g_\lambda(x;\mu_0) \leq 0$ for every $x \in \mathbb{R}$.

To finish proving the necessity, it suffices to show that $g_\lambda(x;\mu_0) = 0$ for all $x \in S_{\mu_0}\backslash\{0\}$. Evidently, $g_\lambda(0;\mu_0) = f_\lambda(0;\mu_0)$ and by (2.10) and $\lambda \mathsf{E}_{\mu_0}\{X^2 - \gamma\} = 0$,

$$\int f_\lambda(x;\mu_0)\,\mu_0(\mathrm{d}x) = 0\,. \tag{2.53}$$

Hence,

$$\int_{\mathbb{R}\backslash\{0\}} g_\lambda(x;\mu_0)\,\mu_0(\mathrm{d}x) = \int g_\lambda(x;\mu_0)\,\mu_0(\mathrm{d}x) - g_\lambda(0;\mu_0)\mu_0(\{0\}) \tag{2.54}$$

$$\geq q f_\lambda(0;\mu_0) + (1-q)\int f_\lambda(x;\mu_0)\,\mu_0(\mathrm{d}x) - q f_\lambda(0;\mu_0) \tag{2.55}$$

$$= 0. \tag{2.56}$$

Since $g_\lambda(x;\mu_0) \leq 0$ for every $x \in \mathbb{R}$, (2.56) implies that on $\mathbb{R}\backslash\{0\}$, $g_\lambda(x;\mu_0) = 0$ μ_0-almost surely, so that $g_\lambda(x;\mu_0) = 0$ for all $x \in S_{\mu_0}\backslash\{0\}$ follows immediately.

The sufficiency part of Lemma 2.6 is established as follows. Suppose $g_\lambda(x;\mu_0) \leq 0$ for every $x \in \mathbb{R}$. By integrating $g_\lambda(x;\mu_0)$ w.r.t. μ_0, we have

$$q g_\lambda(0;\mu_0) \geq \int g_\lambda(x;\mu_0)\,\mu_0(\mathrm{d}x) \tag{2.57}$$

$$= q g_\lambda(0;\mu_0) - (1-q)\lambda \mathsf{E}_{\mu_0}\left\{X^2 - \gamma\right\} \tag{2.58}$$

$$\geq q g_\lambda(0;\mu_0) \tag{2.59}$$

where (2.58) is due to (2.10) and $g_\lambda(0;\mu_0) = f_\lambda(0;\mu_0)$, and (2.59) follows from $\mathsf{E}_{\mu_0}\{X^2\} \leq \gamma$ since $\mu_0 \in \Lambda(\gamma, q)$. Hence, $\lambda \mathsf{E}_{\mu_0}\{X^2 - \gamma\} = 0$ due to the fact that $q < 1$. Furthermore, for any $\mu \in \Lambda(q)$, by integrating $g_\lambda(x;\mu_0)$ w.r.t. μ, we have

$$q g_\lambda(0;\mu_0) \geq \int g_\lambda(x;\mu_0)\,\mu(\mathrm{d}x) \tag{2.60}$$

$$= q f_\lambda(0;\mu_0) + (1-q)\mathsf{E}_\mu\left\{f_\lambda(X;\mu_0)\right\}. \tag{2.61}$$

Because $g_\lambda(0;\mu_0) = f_\lambda(0;\mu_0)$, we have $\mathsf{E}_\mu\{f_\lambda(X;\mu_0)\} \leq 0$. Together with $\lambda \mathsf{E}_{\mu_0}\{X^2 - \gamma\} = 0$ and Lemma 2.5, this implies that μ_0 must be capacity-achieving. □

2.3.3. Discreteness of μ_0

With Lemma 2.6 established, we now prove Property (c) in Theorem 2.1.

Let $\lambda \geq 0$ satisfy condition (2.51) and $d(z;\mu)$ be defined in (2.16). We extend functions $f_\lambda(x;\mu)$ in Lemma 2.5 and $g_\lambda(x;\mu)$ in Lemma 2.6 to be defined on the whole complex plane $\mathbb{C}$ as (2.47) and (2.50), respectively, with x replaced by $z \in \mathbb{C}$. By Lemma 2.2, $d(z;\mu)$ is a holomorphic function of z on $\mathbb{C}$, hence so is $g_\lambda(z;\mu)$. According to Lemma 2.6, each element in the set $S_{\mu_0}\backslash\{0\}$ is a zero of the function $g_\lambda(z;\mu_0)$.

Next we show that for any bounded interval L of $\mathbb{R}$, $S_{\mu_0} \bigcap L$ is a finite set. Suppose, to the contrary, $S_{\mu_0} \bigcap L$ is infinite, then it has a limit point in $\mathbb{R}$ by the Bolzano-Weierstrass Theorem [**48**] and hence, $g_\lambda(z;\mu_0) = 0$ on the whole complex plane $\mathbb{C}$ by the Identity Theorem [**67**]. Then, by (2.9), (2.47) and (2.50), for every $x \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} \phi(y-x) r(y) \mathrm{d}y = 0 \tag{2.62}$$

where

$$r(y) = \log p_Y(y;\mu_0) + \lambda y^2 + c \tag{2.63}$$

and $c = \frac{1}{2}\log(2\pi e) + I(\mu_0) - \frac{q}{1-q} d(0) - \lambda(\gamma+1)$ is a constant.

As in the proof of Lemma 2.2, there exist $a, b \in \mathbb{R}$ such that $|\log p_Y(y;\mu_0)| \leq \frac{1}{2}y^2 + ay + b$. As a result, there exist some $\alpha, \beta > 0$ such that $|r(y)| \leq \alpha y^2 + \beta$. Since the convolution of $r(y)$ and the Gaussian density is equal to the zero function by (2.62), $r(y)$ must be the zero function according to [**14**, Corollary 9]. This requires the capacity-achieving output distribution $p_Y(y;\mu_0)$ be Gaussian, which cannot be true unless X is Gaussian,

which contradicts the assumption that X has a probability mass at 0. Therefore, $S_{\mu_0} \bigcap L$ must be a finite set for any bounded interval L, which further implies that S_{μ_0} is at most countable.

Finally, we show that S_{μ_0} is countably infinite. Suppose, to the contrary, $S_{\mu_0} = \{x_i\}_{i=1}^N$ is a finite set with $\mu_0(\{x_i\}) = p_i$ and $|x_i| \leq B_1$ for all $i = 1, 2, \ldots, N$. For any $y > B_1$,

$$p_Y(y; \mu_0) = \sum_{i=1}^{N} p_i \phi(y - x_i) \leq e^{-\frac{(y-B_1)^2}{2}} \ . \tag{2.64}$$

For any $\epsilon > 0$, choose $B_2 > 0$ such that $\int_{-B_2}^{B_2} \phi(x) \mathrm{d}x > 1 - \epsilon$. By (2.9), (2.47), (2.50) and (2.51), for any $x > B_1 + B_2$, we have

$$0 \geq -\int_{-\infty}^{\infty} \phi(y - x) \log p_Y(y; \mu_0) \mathrm{d}y - \lambda x^2 - (c + \lambda) \tag{2.65}$$

$$\geq \int_{x-B_2}^{x+B_2} \phi(y - x) \frac{1}{2} (y - B_1)^2 \mathrm{d}y - \lambda x^2 - (c + \lambda) \tag{2.66}$$

$$= \int_{B_2}^{B_2} \phi(t) \frac{1}{2} (x - B_1 + t)^2 \mathrm{d}t - \lambda x^2 - (c + \lambda) \tag{2.67}$$

$$\geq \frac{1}{2} (x - B_1)^2 (1 - \epsilon) - \lambda x^2 - (c + \lambda). \tag{2.68}$$

For (2.65) to hold for large x, λ must satisfy $\lambda \geq \frac{1}{2}$.

To finish the proof, it suffices to show that $\lambda < \frac{1}{2}$ for any $\gamma > 0$, so that contradiction arises, which implies that S_{μ_0} must be countably infinite. For fixed $q \in (0, 1)$, denote the Lagrange multiplier in (2.51) as $\lambda(\gamma)$. Denote $C_G(\gamma) = \frac{1}{2} \log(1 + \gamma)$, which is the channel capacity of a Gaussian channel with the average power constraint only. By the envelope theorem [**51**], $\lambda(\gamma)$ is the derivative of $C(\gamma, q)$ w.r.t. γ. Since $C(0, q) = C_G(0) = 0$ and the derivative of $C_G(\gamma)$ at $\gamma = 0$ is $\frac{1}{2}$, we have $\lambda(0) \leq \frac{1}{2}$, otherwise we could find

a small enough γ such that $C(\gamma, q)$ would exceed $C_G(\gamma)$ which is obviously impossible. Next we show that $C(\gamma, q)$ is strictly concave for $\gamma \geq 0$. Suppose μ_1 and μ_2 are the capacity-achieving input distributions of (2.5) for different power constraints γ_1 and γ_2, respectively. Due to Property (b) in Theorem 2.1, μ_1 and μ_2 must be different. Define $\mu_\theta = \theta\mu_1 + (1-\theta)\mu_2$ for $\theta \in (0,1)$. It is easy to see that μ_θ satisfies that the duty cycle is no greater than $1-q$ and the average input power is no greater than $\theta\gamma_1 + (1-\theta)\gamma_2$. Now we have

$$C(\theta\gamma_1 + (1-\theta)\gamma_2, q) \geq I(\mu_\theta) \tag{2.69}$$

$$> \theta I(\mu_1) + (1-\theta) I(\mu_2) \tag{2.70}$$

$$= \theta C(\gamma_1, q) + (1-\theta) C(\gamma_2, q), \tag{2.71}$$

where (2.70) is due to the strict concavity of $I(\mu)$. Therefore, the strict concavity of $C(\gamma, q)$ for $\gamma \geq 0$ follows, which implies that $\lambda(\gamma) < \lambda(0) = \frac{1}{2}$ for all $\gamma > 0$.

2.4. Numerical results

One implication of Theorem 2.1 is that directly computing the capacity-achieving input distribution requires solving an optimization problem with infinite variables which is prohibitive. Assuming any upper bound on the number of probability mass points, however, a numerical optimization over the mutual information can yield a suboptimal input distribution and a lower bound on the channel capacity. As we increase the number of mass points, the lower bound can be further refined. We take this approach to numerically compute a good approximation of the channel capacity by optimizing over a sufficient number of probability mass points.

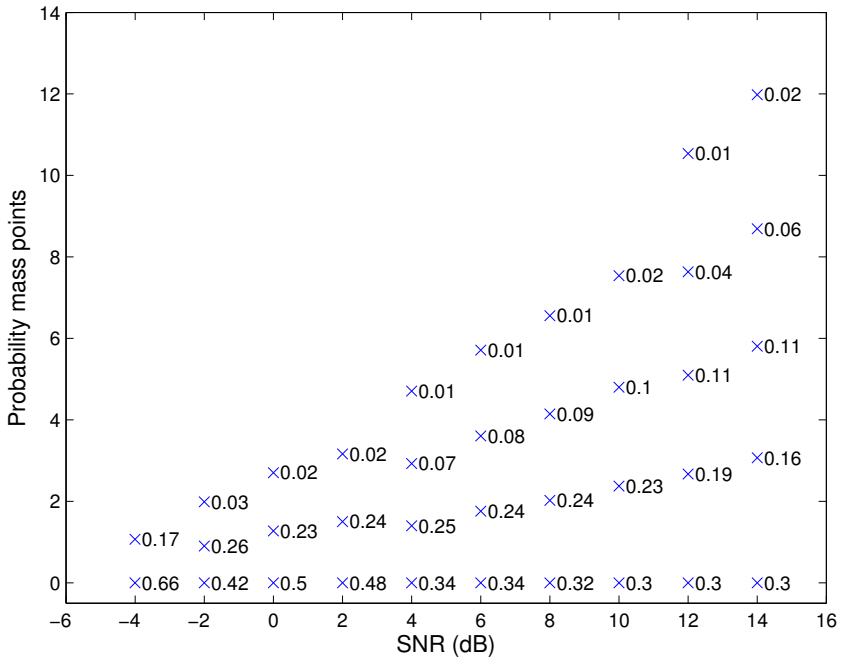

Figure 2.1. Suboptimal input distribution for $P(X = 0) \geq q = 0.3$.

Given the duty cycle and power constraints, we first numerically optimize the mutual information by a 3-point input distribution (including a mass at 0), then increase the number of probability mass points by 2 at a time to improve the mutual information, until the improvement is less than 10^{-3}.

First consider the case that the duty cycle is no greater than 70%, i.e., $P(X = 0) \geq q = 0.3$. For different SNRs, the mass points of the near-optimal input distribution with finite support along with the corresponding probability masses are shown in Fig. 2.1. Due to symmetry, only the positive half of the input distribution is plotted. We can see that as the SNR increases, more masses are put on higher-amplitude points, whereas the probability mass at zero achieves its lower bound 0.3 eventually.

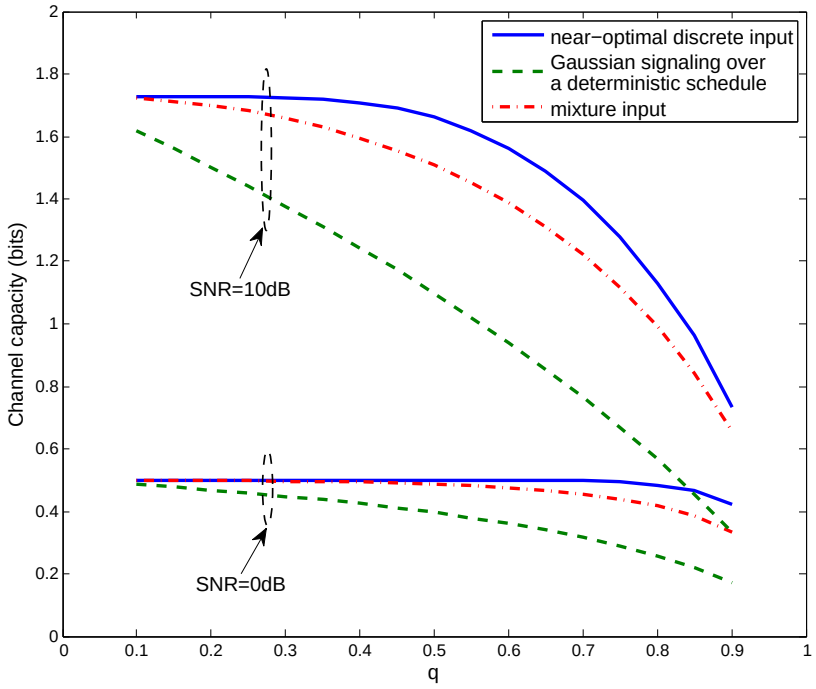

Figure 2.2. Achievable rates under duty cycle constraint for 0 dB and 10 dB SNRs.

In Fig. 2.2, we compare the rate achieved by the near-optimal input distribution and the rate achieved by a conventional scheme using Gaussian signaling over a deterministic schedule, which is $(1-q)$ times the Gaussian channel capacity without duty cycle constraint. It is shown in the figure that there is substantial gain for both 0 dB and 10 dB SNRs by using discrete input over Gaussian signaling with a deterministic schedule. For example, when the SNR is 10 dB, given the duty cycle is no more than 50%, the discrete input distribution achieves 50% higher rate. Hence departing from the usual paradigm of intermittent frame transmissions may yield significant gain.

We also plot in Fig. 2.2 the achievable rate by a superposition coding, where the input distribution is a mixture of Gaussian and a point mass at 0. We first decode the support of the input to find out the positions of nonzero symbols, and then the Gaussian codeword

conditioned on the support. It is shown in the figure that the near-optimal discrete input achieves higher rate compared with the mixture input.

2.5. Summary

In this chapter we have studied the impact of duty cycle constraint on the capacity of AWGN channels. The optimal distribution is discrete and has a finite number of probability mass points in any bounded interval. This allows efficient numerical optimization of the input distribution. The numerical results show that under the duty cycle constraint, using on-off signaling inside each frame instead of the usual paradigm of intermittent frame transmissions may yield substantial gain. The results in this chapter have been published in part in [**91**].

CHAPTER 3

Network Capacity with Half-Duplex Constraint

To further quantify the advantages and potentials of the RODD technology, in this chapter, we present theoretic results on the capacity of simple RODD network models and a comparison with ALOHA-type random access scheme.

The traffic we consider here is mutual broadcast, i.e., all nodes wish to broadcast information to and receive information from its neighbors. An important example of mutual broadcast is the network state information exchange. Many advanced wireless transmission techniques require knowledge of the state of communicating parties, such as the power, modulation format, beamforming vector, code rate, acknowledgment (ACK), queue length, etc. Conventional schemes often treat such network state information similarly as data, so that exchange of such information require a substantial amount of overhead and, in ad hoc networks, often many retransmissions. In a highly mobile network, the overhead easily dominates the data traffic [**3**]. By creating a virtual full-duplex channel, RODD is particularly suitable for nodes to efficiently broadcast local state information to their respective neighbors. One potential application of this idea is to assist distributed scheduling by letting each node choose whether to transmit based on its own state and the states of its neighbors [**37**]. Another application is distributed interference management by exchanging interference prices as studied in [**71**].

The remainder of this chapter is organized as follows. Mathematical models of a network of nodes with RODD signaling is presented in Section 3.1. Assuming mutual

broadcast traffic, the throughput of a fully-connected, synchronized, RODD-based network is studied in Section 3.2. Section 3.3 summarizes this chapter.

3.1. Network Models

Consider a wireless network consisting of N nodes, indexed by $1, \ldots, N$. Suppose all transmissions are over the same frequency band. Suppose for simplicity each slot is of one symbol interval and all nodes are perfectly synchronized over each frame of M slots. Let the binary on-off duplex mask of node n over slots 1 through M be denoted by $\boldsymbol{S}_n = [s_{1n}, \ldots, s_{Mn}]^\top$. During slot m, node n may transmit a symbol if $s_{mn} = 1$, whereas if $s_{mn} = 0$, the node listens to the channel and emits no energy.

3.1.1. The Fading Channel Model

As described by model (1.3), RODD forms fundamentally a multiaccess channel with erasure. Denote the SNR of the link from node j to node n by $\gamma_{nj} = \gamma_j \, d_{nj}^{-\alpha} |h_{nj}|^2$. Model (1.3) can be rewritten as

$$Y_{mn} = (1 - s_{mn}) \sum_{j \in \partial n} \sqrt{\gamma_{nj}} s_{mj} X_{mj} + V_{mn} \tag{3.1}$$

We simply assume that V_{mn} are i.i.d. Gaussian random variables with zero mean and unit variance.

3.1.2. A Deterministic Model

It is instructive to consider a simplification of model (1.3) by assuming noiseless reception and non-coherent energy detection. That is, as long as some neighbor transmits

energy during an off-slot of node n, a "1" is observed in the slot, whereas if no neighbor emits energy during the slot, a "0" is observed. This can be described as an *inclusive-or* multiaccess channel (referred to as OR-channel) with erasure:

$$Y_{mn} = (1 - s_{mn}) \left(\vee_{j \in \partial n} (s_{mj} X_{mj}) \right) \tag{3.2}$$

for $m = 1, \ldots, M$, where the binary inputs X_{mj} and outputs Y_{mn} take values from $\{0, 1\}$. Since the output is a deterministic function of the inputs, (3.2) belongs to the family of *deterministic models*, which have been found to be a very effective tool in understanding multiuser channels (see, e.g, [**8**, **22**]). Despite its simplicity, it captures the superposition nature of the physical channel, while ignoring the effect of noise and interference, although those impairments can also be easily included in the model.

3.2. Throughput Results

Suppose each node has a message to broadcast to all its neighbors by transmitting a frame over M slots. An M-slot frame is regarded as being successful for a given node if the messages from all its neighbors are decoded correctly; otherwise the frame is in error. A rate tuple for the N nodes is achievable if there exists a code using which the nodes can transmit at their respective rates with vanishing error probability in the limit where the frame length $M \to \infty$.

The achievable rates obviously depends on the network topology and the duplex masks. Although carefully designed duplex masks can carry information (as discussed in Chapter 2), it is simply assumed that the elements s_{mn} of the duplex masks are i.i.d. Bernoulli random variables with $\mathsf{P}(s_{mn} = 1) = q$. Suppose every node has complete knowledge of

the duplex masks of all peers. For simplicity, we consider a symmetric network of N nodes who are neighbors of each other, where the gain between every pair of nodes is identical.

We assume that each node encodes its information independently. In the simplest scenario, all nodes use randomly generated i.i.d. codebooks dependent on the parameters (N, M, q) but independent of the duplex masks otherwise. Such a code is called a *signature-independent code*. Alternatively, nodes may use *signature-dependent codes*, where the codebooks may depend on the on-off activity pattern $\boldsymbol{A}_m = [s_{m1}, s_{m2}, \ldots, s_{mN}]$ in every slot m.

In case all messages are of the same number of bits, the rate tuple collapses to a single number. The maximum achievable such rate by using signature-independent (resp. signature-dependent) codes is called the *symmetric rate* (resp. *symmetric capacity*).

In Section 3.2.1 we first describe the region of rate tuples when signature-independent codes or signature-dependent codes are used. The results are then applied to derive the symmetric rate and the symmetric capacity for the deterministic channel and the Gaussian multiacess channel in Sections 3.2.2 and 3.2.3, respectively, and the asymmetric rate for Gaussian multiaccess channel in Section 3.2.4.

3.2.1. Capacity Region

For each node n in the network, denote the alphabets of its transmit symbols and receive symbols as $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively. Suppose node n chooses an index w_n uniformly from the set $\mathcal{W}_n = \{1, 2, \ldots, 2^{MR_n}\}$ and sends the corresponding M-length codeword over the channel according to its encoding function $f_n : \mathcal{W}_n \to \mathcal{X}_n^M$. Assume the distribution of messages $\boldsymbol{w} = (w_1, \ldots, w_N)$ over the product set $\prod_{n=1}^{N} \mathcal{W}_n$ is uniform, i.e., the messages

are independent and equally likely. Denote the receive signal and the decoding function at node n as $\boldsymbol{Y}_n$ and $g_n : \mathcal{Y}_n^M \to \prod_{i\neq n} \mathcal{W}_i$, respectively. We define the average probability of error in an M-length frame as follows:

$$\mathsf{P}_{\mathrm{e}}^{(M)} = \frac{1}{2^{M\sum_{n=1}^{N} R_n}} \sum_{\boldsymbol{w}\in\prod_{n=1}^{N}\mathcal{W}_n} \sum_{n=1}^{N} \mathsf{P}\left\{g_n(\boldsymbol{Y}_n) \neq \boldsymbol{w}\backslash w_n \middle| \boldsymbol{w} \text{ sent}\right\}, \tag{3.3}$$

where $\boldsymbol{w}\backslash w_n \in \prod_{i\neq n} \mathcal{W}_j$ represents the subset of $\boldsymbol{w}$ excluding w_n. A rate tuple $(R_1, \ldots, R_N)$ is achievable if $\mathsf{P}_{\mathrm{e}}^{(M)} \to 0$ as $M \to \infty$. And the capacity region is the closure of the set of achievable rate tuples.

The on-off pattern $\boldsymbol{A}_m = [s_{m1}, s_{m2}, \ldots, s_{mN}]$ can be viewed as the user activity. In the case that the signature-independent codes are used, the user activity information is not utilized by encoders to generate transmit symbols; while when the signature-dependent codes are used, we can view that the user activity information is revealed at both encoders and decoders.

Let $n \in \mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n = \mathcal{N}\backslash\{n\}$. Let $\mathcal{S}_n^c$ denote the complement of $\mathcal{S}_n$ in $\mathcal{N}_n$. Let $R(\mathcal{S}_n) = \sum_{i\in\mathcal{S}_n} R_i$ and $X(\mathcal{S}_n) = \{X_i : i \in \mathcal{S}_n\}$ with $X_i \in \mathcal{X}_i$. Denote random variables $Y_n \in \mathcal{Y}_n$ and $\boldsymbol{A}$ as the receive symbol at node n and the user activity, respectively. For any given pattern $\boldsymbol{a}$ with n zero entries, the probability that $\boldsymbol{A} = \boldsymbol{a}$ is $q^{N-n}(1-q)^n$.

We establish the following result describing the capacity region in both cases where signature-independent or signature-dependent codes are used.

Proposition 3.1. *The capacity region is the closure of the convex hull of all rate tuples* $(R_1, R_2, \ldots, R_N)$ *satisfying*

$$R(\mathcal{S}_n) \leq I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \text{ for all } n \in \mathcal{N} \text{ and } \mathcal{S}_n \subseteq \mathcal{N}_n \tag{3.4}$$

for some product distribution $\prod_{n=1}^{N} p_{X_n}(x_n)$ *on* $\prod_{n=1}^{N} \mathcal{X}_n$ *when the signature-independent codes are used. In the case that the signature-dependent codes are used, the capacity region is the closure of the convex hull of all rate tuples satisfying* (3.4) *for some product distribution* $\prod_{n=1}^{N} p_{X_n|\boldsymbol{A}}(x_n|\boldsymbol{a})$ *on* $\prod_{n=1}^{N} \mathcal{X}_n$ *for any given user activity* $\boldsymbol{A} = \boldsymbol{a}$.

We can view node n as the receiver and all other nodes as transmitters. The rest of the proof of Proposition 3.1 then follows similar steps as in the multiple access channel [**17**].

Proof. We first prove the case that the signature-independent codes are used. Since the user activity $\boldsymbol{A}$ is not available at encoders, it can be viewed as another output of the channel besides $\boldsymbol{Y}_n$. Thus, according to the results of the multiple access channel [**17**], the capacity region here is the closure of the convex hull of all rate tuples $(R_1, R_2, \ldots, R_N)$ satisfying

$$R(\mathcal{S}_n) \leq I\left(X(\mathcal{S}_n); Y_n, \boldsymbol{A} \middle| X(\mathcal{S}_n^c)\right) \tag{3.5}$$

$$= I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \tag{3.6}$$

for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$, where (3.6) is due to the independence between $X(\mathcal{S}_n)$ and $\boldsymbol{A}$.

In the case that the signature-dependent codes are used, i.e., user activity $\boldsymbol{A}$ is available at both encoders and decoders, for any activity pattern $\boldsymbol{A} = \boldsymbol{a}$, it follows directly from

the results of the multiple access channel [**17**] that the rate tuple $(R_1, \ldots, R_N)$ satisfying

$$R(\mathcal{S}_n) \leq I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) \tag{3.7}$$

for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$ is achievable. Thus, (3.4) can be achieved by time sharing.

To show the converse, we first prove the conditional version of Fano's inequality and data processing inequality [**17**]. Define $w(\mathcal{S}_n) = \{w_i : i \in \mathcal{S}_n\}$ and

$$E_n = \begin{cases} 1, & g_n(\boldsymbol{Y}_n) \neq \boldsymbol{w} \backslash w_n; \\ 0, & \text{otherwise.} \end{cases} \tag{3.8}$$

It is easy to see that

$$H\left(E_n, w(\mathcal{N}_n) | \boldsymbol{Y}_n, \boldsymbol{A}\right) = H\left(w(\mathcal{N}_n) | \boldsymbol{Y}_n, \boldsymbol{A}\right) + H\left(E_n | \boldsymbol{Y}_n, \boldsymbol{A}, w(\mathcal{N}_n)\right) \tag{3.9}$$

$$= H\left(E_n | \boldsymbol{Y}_n, \boldsymbol{A}\right) + H\left(w(\mathcal{N}_n) | \boldsymbol{Y}_n, \boldsymbol{A}, E_n\right) \tag{3.10}$$

$$\leq 1 + \mathsf{P}_{\mathrm{e}}^{(M)} M R(\mathcal{N}_n) \tag{3.11}$$

$$\triangleq M\epsilon_M \tag{3.12}$$

where $\epsilon_M \to 0$ as $\mathsf{P}_{\mathrm{e}}^{(M)} \to 0$. Since $H\left(E_n \middle| \boldsymbol{Y}_n, \boldsymbol{A}, w(\mathcal{N}_n)\right) = 0$, from (3.9) and (3.12) we have

$$H\left(w(\mathcal{S}_n) \middle| \boldsymbol{Y}_n, \boldsymbol{A}\right) \leq H\left(w(\mathcal{N}_n) \middle| \boldsymbol{Y}_n, \boldsymbol{A}\right) \leq M\epsilon_M. \tag{3.13}$$

Define $\boldsymbol{X}(\mathcal{S}_n) = \{\boldsymbol{X}_i : i \in \mathcal{S}_n\}$. For any $\mathcal{S}_n \subseteq \mathcal{N}_n$, we have

$$
\begin{aligned}
& I\left(w(\mathcal{S}_n), \boldsymbol{X}(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) \\
&= I\left(\boldsymbol{X}(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) + I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{X}(\mathcal{S}_n), \boldsymbol{A}\right) && (3.14) \\
&= I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) + I\left(\boldsymbol{X}(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{N}_n), \boldsymbol{A}\right) && (3.15) \\
&\geq I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right). && (3.16)
\end{aligned}
$$

Since $I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{X}(\mathcal{S}_n), \boldsymbol{A}\right) = 0$ due to the conditional independence of $w(\mathcal{S}_n)$ and $\boldsymbol{Y}_n$ given $w(\mathcal{S}_n^c)$, $\boldsymbol{X}(\mathcal{S}_n)$ and $\boldsymbol{A}$, from (3.14) and (3.16), we have

$$
I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) \leq I\left(\boldsymbol{X}(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right). \tag{3.17}
$$

Let $X_m(\mathcal{S}_n)$ denote the set of transmit symbols in slot m from nodes in the set $\mathcal{S}_n$. We can now bound the sum rate $R(\mathcal{S}_n)$ as

$$
\begin{aligned}
MR(\mathcal{S}_n) &= H(w(\mathcal{S}_n)) && (3.18) \\
&\leq I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n, \boldsymbol{A}\right) + M\epsilon_M && (3.19) \\
&= H\left(w(\mathcal{S}_n) \middle| \boldsymbol{A}\right) - H\left(w(\mathcal{S}_n) \middle| \boldsymbol{Y}_n, \boldsymbol{A}\right) + M\epsilon_M && (3.20) \\
&\leq H\left(w(\mathcal{S}_n) \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) - H\left(w(\mathcal{S}_n) \middle| w(\mathcal{S}_n^c), \boldsymbol{Y}_n, \boldsymbol{A}\right) + M\epsilon_M && (3.21) \\
&= I\left(w(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) + M\epsilon_M && (3.22) \\
&\leq I\left(\boldsymbol{X}(\mathcal{S}_n); \boldsymbol{Y}_n \middle| w(\mathcal{S}_n^c), \boldsymbol{A}\right) + M\epsilon_M && (3.23) \\
&\leq \sum_{m=1}^{M} I\left(X_m(\mathcal{S}_n); Y_{mn} \middle| X_m(\mathcal{S}_n^c), \boldsymbol{A}\right) + M\epsilon_M && (3.24)
\end{aligned}
$$

where

(3.19) follows from (3.13),

(3.21) follows from the fact that since $w(\mathcal{S}_n)$ and $w(\mathcal{S}_n^c)$ are independent, so are $\boldsymbol{X}(\mathcal{S}_n)$ and $\boldsymbol{X}(\mathcal{S}_n)$ given $\boldsymbol{A}$, and hence $H\left(w(\mathcal{S}_n)\middle|\boldsymbol{A}\right) = H\left(w(\mathcal{S}_n)\middle|w(\mathcal{S}_n^c), \boldsymbol{A}\right)$, and by conditioning, $H\left(w(\mathcal{S}_n)\middle|\boldsymbol{Y}_n, \boldsymbol{A}\right) \geq H\left(w(\mathcal{S}_n)\middle|w(\mathcal{S}_n^c), \boldsymbol{Y}_n, \boldsymbol{A}\right)$,

(3.23) follows from (3.17),

(3.24) follows from the chain rule and removing conditioning.

Hence, we have

$$R(\mathcal{S}_n) \leq \frac{1}{M}\sum_{m=1}^{M} I\left(X_m(\mathcal{S}_n); Y_{mn}\middle|X_m(\mathcal{S}_n^c), \boldsymbol{A}\right) + \epsilon_M \tag{3.25}$$

By introducing a new time-sharing random variable Q, the rest of the proof of converse is the same as in the multiple access channel, thus is omitted here. □

3.2.2. The Deterministic Model

Consider the OR-channel described by (3.2). A node's codeword is basically erased by its own signature mask before transmission.

Proposition 3.2. *The symmetric rate of the OR-channel* (3.2) *is*

$$R = \max_{p\in[0,1]} \frac{1}{N-1}\sum_{k=1}^{N-1}\binom{N-1}{k} q^k(1-q)^{N-k}H_2(p^k) \tag{3.26}$$

where $H_2(p) = -p\log p - (1-p)\log(1-p)$ *is the binary entropy function.*

Proof. We prove by using Proposition 3.1. In the case that the signature-independent codes are used, there exists product distribution $\prod_{n=1}^{N} P_{X_n}(x_n)$ on $\prod_{n=1}^{N}\mathcal{X}_n$ such that for

all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$,

$$|\mathcal{S}_n| R \leq I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \tag{3.27}$$

where $|\mathcal{S}_n|$ represents the cardinality of $\mathcal{S}_n$. Here $P_{X_n}(x_n)$ represents the probability mass function of random variable $X_n \in \{0, 1\}$, and it is assumed that $P_{X_n}(0) = p \in [0, 1]$ since each node encodes its message independently without knowledge of user activity information. Next we will evaluate (3.27) for the special case that $\mathcal{S}_n = \mathcal{N}_n$ to show that the symmetric rate is upper bounded by (3.26). We compete the proof by showing that the symmetric rate given by (3.26) satisfy (3.27), thus is achievable.

For any $\boldsymbol{a} = [s_1, \ldots, s_N]$, denote $\boldsymbol{a} \cdot \mathcal{S}_n = \sum_{i \in \mathcal{S}_n} s_i$. It follows that

$$\begin{aligned}
&I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) \\
&= H(Y_n \big| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}) - H(Y_n \big| X(\mathcal{N}_n), \boldsymbol{A} = \boldsymbol{a}) && (3.28) \\
&= H(Y_n \big| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}) && (3.29) \\
&= (1 - s_n) p^{\boldsymbol{a} \cdot \mathcal{S}_n^c} H_2\left(p^{\boldsymbol{a} \cdot \mathcal{S}_n}\right) && (3.30)
\end{aligned}$$

where (3.29) is due to the deterministic nature of the model and (3.30) is due to the property of the OR-channel with erasure. Consider the special case that $\mathcal{S}_n = \mathcal{N}_n$. By averaging over all realizations of $\boldsymbol{A}$, it follows from (3.27) and (3.30) that

$$\begin{aligned}
R &\leq \frac{1}{N-1} I\left(X(\mathcal{N}_n); Y_n \middle| \boldsymbol{A}\right) && (3.31) \\
&\leq \max_{p \in [0,1]} \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} H_2(p^k) && (3.32)
\end{aligned}$$

where the equality is achieved by random codebooks with i.i.d. Bernoulli $(1-p_*)$ entries where p_* maximizes (3.32).

Next we show that the Bernoulli codebooks designed above satisfy the condition (3.27) for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. In fact, it can be shown that

$$p^{t_1-t_2} H_2\left(p^{t_2}\right) \geq \frac{t_2}{t_1} H_2\left(p^{t_1}\right) \tag{3.33}$$

for any $t_1 \geq t_2 > 0$ and $p \in [0,1]$. Therefore, (3.30) can be lower bounded as

$$I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) \geq \begin{cases} 0, & \boldsymbol{a} \cdot \mathcal{N}_n = 0; \\ (1-s_n) \frac{\boldsymbol{a} \cdot \mathcal{S}_n}{\boldsymbol{a} \cdot \mathcal{N}_n} H_2\left(p_*^{\boldsymbol{a} \cdot \mathcal{N}_n}\right), & \text{otherwise.} \end{cases} \tag{3.34}$$

By averaging over all realizations of $\boldsymbol{A}$, it follows from (3.34) that for any $\mathcal{S}_n$ with $|\mathcal{S}_n| = l \leq N-1$,

$$\begin{aligned}
&\frac{1}{l} I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \\
&\geq \frac{1}{l} \sum_{k=1}^{N-1} q^k (1-q)^{N-k} H_2(p_*^k) \sum_{k_1} \frac{k_1}{k} \binom{l}{k_1} \binom{N-1-l}{k-k_1} && (3.35) \\
&= \sum_{k=1}^{N-1} q^k (1-q)^{N-k} H_2(p_*^k) \frac{1}{k} \sum_{k_1} \binom{l-1}{k_1-1} \binom{N-1-l}{k-k_1} && (3.36) \\
&= \sum_{k=1}^{N-1} q^k (1-q)^{N-k} H_2(p_*^k) \frac{1}{k} \binom{N-2}{k-1} && (3.37) \\
&= \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} H_2(p_*^k) && (3.38) \\
&= R && (3.39)
\end{aligned}$$

where k_1 in (3.35) and (3.36) satisfies $\max\{0, k+l+1-N\} \leq k_1 \leq \min\{l, k\}$, and (3.37) is due to the fact that [**27**, Page 5]

$$\sum_{k_1} \binom{l-1}{k_1-1} \binom{N-1-l}{k-k_1} = \binom{N-2}{k-1}. \tag{3.40}$$

Therefore, (3.27) holds for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. Proposition 3.2 is thus established. □

Proposition 3.3. *The symmetric capacity of the OR-channel* (3.2) *is*

$$C = \frac{1}{N-1} \left[(1-q) - (1-q)^N \right]. \tag{3.41}$$

Proof. The proof follows the similar steps as in Proposition 3.2. In the case that the signature-dependent codes are used, according to Proposition 3.1, there exists product distribution $\prod_{n=1}^{N} P_{X_n|\boldsymbol{A}}(x_n|\boldsymbol{a})$ on $\prod_{n=1}^{N} \mathcal{X}_n$ for any given user activity $\boldsymbol{A} = \boldsymbol{a}$ such that for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$,

$$|\mathcal{S}_n| C \leq I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right). \tag{3.42}$$

Assumed that for each $n \in \mathcal{N}$, $P_{X_n|\boldsymbol{A}}(0|\boldsymbol{a}) = p_n(\boldsymbol{a}) \in [0,1]$, which is a function of $\boldsymbol{a}$. Similarly as in (3.30), we have

$$I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) = (1-s_n) \prod_{i \in \mathcal{S}_n^c} p_i^{s_i}(\boldsymbol{a}) \, H_2\left(\prod_{j \in \mathcal{S}_n} p_j^{s_j}(\boldsymbol{a}) \right). \tag{3.43}$$

Consider a special case that $\mathcal{S}_n = \mathcal{N}_n$. From (3.43), we have

$$I\left(X(\mathcal{N}_n); Y_n \middle| \boldsymbol{A} = \boldsymbol{a}\right) \leq (1-s_n) 1(\boldsymbol{a} \cdot \mathcal{N} \neq 0). \tag{3.44}$$

By averaging over all realizations of $\boldsymbol{A}$, it follows from (3.42) and (3.44) that

$$C \leq \frac{1}{N-1}(1-q)\sum_{k=1}^{N-1}\binom{N-1}{k}q^k(1-q)^{N-1-k} \tag{3.45}$$

$$= \frac{1}{N-1}\left[(1-q)-(1-q)^N\right] \tag{3.46}$$

where the equality is achieved by the following multiplexing scheme: whenever the user activity is $\boldsymbol{a}$, each node uses random codebook with i.i.d. Bernoulli $1-p(\boldsymbol{a})$ entries with $p(\boldsymbol{a}) = 2^{-1/\boldsymbol{a}\cdot\mathcal{N}}$.

Next we show that the Bernoulli codebooks with the choice of $p(\boldsymbol{a})$ satisfy the condition (3.42). In fact, similarly as in (3.34), it follows from (3.43) that

$$I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}=\boldsymbol{a}\right) = (1-s_n)\prod_{i\in\mathcal{S}_n^c} p^{s_i}(\boldsymbol{a})\, H_2\left(\prod_{j\in\mathcal{S}_n} p^{s_j}(\boldsymbol{a})\right) \tag{3.47}$$

$$\geq \begin{cases} 0, & \boldsymbol{a}\cdot\mathcal{N}_n = 0; \\ (1-s_n)\frac{\boldsymbol{a}\cdot\mathcal{S}_n}{\boldsymbol{a}\cdot\mathcal{N}_n}, & \text{otherwise.} \end{cases} \tag{3.48}$$

By averaging over all realizations of $\boldsymbol{A}$, it follows from (3.48) and (3.40) that for any $\mathcal{S}_n$ with $|\mathcal{S}_n| = l \leq N-1$,

$$\frac{1}{l} I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \geq \frac{1}{l}\sum_{k=1}^{N-1} q^k(1-q)^{N-k}\sum_{k_1}\frac{k_1}{k}\binom{l}{k_1}\binom{N-1-l}{k-k_1} \tag{3.49}$$

$$= \frac{1}{N-1}\left[(1-q)-(1-q)^N\right] \tag{3.50}$$

$$= C. \tag{3.51}$$

Therefore, (3.42) holds for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. Proposition 3.3 is thus established. $\square$

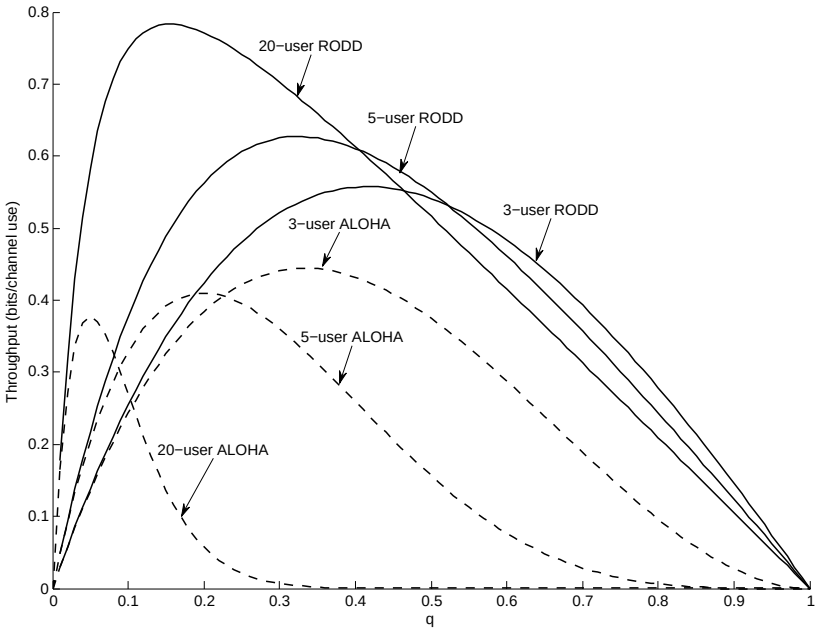

Figure 3.1. Comparison of the throughput of RODD and ALOHA over OR-channel.

The symmetric capacity is higher than the symmetric rate because there is gain to adapt the codebooks to the signatures. Basically the codebook entries at each slot are generated as independent Bernoulli random variables whose mean value depends on the number of transmitting nodes in the slot (a.k.a. the weight of $\boldsymbol{A}_m$). The parameters of the Bernoulli variables can be optimized for achieving the capacity. For example, suppose $N = 3$, then there are 8 different on-off activity patterns. By symmetry, we only consider node 1. If the pattern is [1 0 0], [0 1 0] or [0 0 1], node 1 uses random codebook with i.i.d. Bernoulli entries with parameter 1/2; if the pattern is [0 1 1], [1 0 1] or [1 1 0], node 1 uses random codebook with i.i.d. Bernoulli entries with parameter $1 - 1/\sqrt{2}$; otherwise, node 1 transmits all-zero codeword.

We next compare the throughput of a RODD-based scheme with that of ALOHA-type random access schemes over the same channel (3.2), where the throughput is defined as

the sum rate of all nodes. During each frame interval (or contention period), every node in ALOHA independently chooses either to transmit (with probability q) or to listen (with probability $1-q$) and the choices are independent across contention periods. A node successfully broadcasts its message to all other nodes if the frame is the only transmission during a given frame interval. It is easy to see that the throughput of the system with ALOHA is $Nq(1-q)^{N-1}$, which achieves the maximum $(1-1/N)^{N-1}$ with $q = 1/N$. [1]

For three different node populations ($N = 3, 5, 20$), the comparison between RODD and ALOHA is shown in Fig. 3.1. The sum symmetric rate achieved by signature-independent codes is plotted for RODD. Clearly, the maximum throughput of RODD is much higher than that of ALOHA, where the gap increases as the number of nodes increases. In fact the throughput of RODD exceeds that of ALOHA for all values of q. In case of a large number of nodes, the throughput of ALOHA approaches $1/e$. On the other hand, with $p = 1 - 2^{-\frac{1}{(N-1)q}}$, the total throughput achieved by using RODD signaling approaches $1-q$ as $N \to \infty$, which is also the asymptotic sum capacity of RODD achieved by signature-dependent codes.

The reason for the inferior performance of ALOHA is largely due to packet retransmissions after collision. Even if multi-packet reception is allowed, the throughput of ALOHA is still far inferior compared to RODD signaling due to the half-duplex constraint. This is in part because, in the case of broadcast traffic studied here, if two nodes simultaneously and successfully transmit their packets to all other nodes, they still have to exchange their messages using at least two additional transmissions.

[1] One conceivable protocol is, after n nodes have succeeded, to let the remaining $N-n$ nodes contend for transmission. This improves the throughput of ALOHA slightly, but the advantage of RODD remains true for every $N > 3$.

3.2.3. The Gaussian Multiaccess Channel

Consider now a (non-fading) Gaussian multiaccess channel described by (3.1), where $d_{nj} = 1$, $h_{nj} = 1$ for all n, j. For simplicity, let all nodes be of the same SNR, $\gamma_j = \gamma$. Thus, the SNR of the link from node j to node n is $\gamma_{nj} = \gamma$. Recall that the average power of each transmitted codeword is assumed to be 1. Since each node only transmits over about qM slots, the average SNR during each active slot is essentially γ/q.

It is easy to see that the throughput of ALOHA over the Gaussian channel is

$$\frac{N}{2} q(1-q)^{N-1} \log(1 + \frac{\gamma}{q}). \tag{3.52}$$

Similar to the results for the deterministic model, we can show that the symmetric rate and the symmetric capacity for the Gaussian multiaccess channel are achieved with Gaussian codebooks by signature-independent codes and signature-dependent codes, respectively.

Proposition 3.4. *The symmetric rate of the non-fading Gaussian multiaccess channel described by* (3.1) *is*

$$R = \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} g\left(\frac{k\gamma}{q}\right) \tag{3.53}$$

where $g(x) = \frac{1}{2}\log(1+x)$.

Proof. We prove by using Proposition 3.1 and follow similar steps as in Proposition 3.2. In the case that the signature-independent codes are used, according to Proposition 3.1, there exists product distribution $\prod_{n=1}^{N} p_{X_n}(x_n)$ on $\prod_{n=1}^{N} \mathcal{X}_n$ such that for all

$n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$,

$$|\mathcal{S}_n| R \leq I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right). \tag{3.54}$$

Here $p_{X_n}(x_n)$ satisfies that $\mathsf{E}\{X_n^2\} = 1/q$ since the average power of each transmitted codeword is assumed to be 1.

Consider a special case that $\mathcal{S}_n = \mathcal{N}_n$. For any on-off activity pattern $\boldsymbol{a}$, we have

$$\begin{aligned}
&I\left(X(\mathcal{N}_n); Y_n \middle| \boldsymbol{A} = \boldsymbol{a}\right) \\
&= h(Y_n \big| \boldsymbol{A} = \boldsymbol{a}) - h(Y_n \big| X(\mathcal{N}_n), \boldsymbol{A} = \boldsymbol{a}) && (3.55)\\
&= h(Y_n \big| \boldsymbol{A} = \boldsymbol{a}) - \frac{1}{2}\log(2\pi e) && (3.56)\\
&\leq \frac{1}{2}(1 - s_n) \log\left(2\pi e\left(1 + \frac{(\boldsymbol{a} \cdot \mathcal{N}_n)\gamma}{q}\right)\right) + \frac{1}{2} s_n \log(2\pi e) - \frac{1}{2}\log(2\pi e) && (3.57)\\
&= (1 - s_n) g\left(\frac{(\boldsymbol{a} \cdot \mathcal{N}_n)\gamma}{q}\right) && (3.58)
\end{aligned}$$

where (3.57) is because Y_n is a Gaussian random variable with unit average power if $s_n = 1$, otherwise the average power of Y_n is $1 + (\boldsymbol{a} \cdot \mathcal{N}_n)\gamma/q$. By averaging over all realizations of $\boldsymbol{A}$, it follows from (3.54) and (3.58) that

$$R \leq \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} g\left(\frac{k\gamma}{q}\right) \tag{3.59}$$

where the equality is achieved by using random Gaussian codebooks.

Next we show that random Guassian codebooks satisfy the condition (3.54), thus achieve the symmetric rate in (3.59). In fact, for any on-off activity pattern $\boldsymbol{a}$, similarly

as in (3.58), we have

$$
\begin{aligned}
& I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}=\boldsymbol{a}\right) \\
&= h(Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}=\boldsymbol{a}) - h(Y_n \middle| X(\mathcal{N}_n), \boldsymbol{A}=\boldsymbol{a}) && (3.60)\\
&= \frac{1}{2}(1-s_n)\log\left(2\pi e\left(1+\frac{(\boldsymbol{a}\cdot\mathcal{S}_n)\gamma}{q}\right)\right)+\frac{1}{2}s_n\log(2\pi e)-\frac{1}{2}\log(2\pi e) && (3.61)\\
&= (1-s_n)g\left(\frac{(\boldsymbol{a}\cdot\mathcal{S}_n)\gamma}{q}\right). && (3.62)
\end{aligned}
$$

It follows that for any $\mathcal{S}_n$ with $|\mathcal{S}_n| = l \le N-1$,

$$
\begin{aligned}
& \frac{1}{l} I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \\
&= \frac{1}{l}\sum_{k=0}^{N-1} q^k(1-q)^{N-k}\sum_{k_1} g\left(\frac{k_1\gamma}{q}\right)\binom{l}{k_1}\binom{N-1-l}{k-k_1} && (3.63)\\
&\ge \sum_{k=1}^{N-1} q^k(1-q)^{N-k} g\left(\frac{k\gamma}{q}\right)\sum_{k_1}\frac{1}{k}\binom{l-1}{k_1-1}\binom{N-1-l}{k-k_1} && (3.64)\\
&= \frac{1}{N-1}\sum_{k=1}^{N-1}\binom{N-1}{k} q^k(1-q)^{N-k} g\left(\frac{k\gamma}{q}\right) && (3.65)
\end{aligned}
$$

where (3.64) is due to the fact that

$$
\frac{1}{t_1} g(ct_1) \le \frac{1}{t_2} g(ct_2) \tag{3.66}
$$

for any $t_1 \ge t_2 > 0$ and $c > 0$, and (3.65) is due to (3.40). Therefore, (3.54) holds for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. Proposition 3.4 is thus established. □

Proposition 3.5. *The symmetric capacity of the non-fading Gaussian multiaccess channel described by* (3.1) *is*

$$C = \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} g(u_k) \tag{3.67}$$

where $g(x) = \frac{1}{2}\log(1+x)$, $u_k = \max\{(N-k)v - 1, 0\}$ *and* v *is chosen to satisfy*

$$\frac{1}{N} \sum_{k=1}^{N-1} \binom{N}{k} q^n (1-q)^{N-k} u_k = \gamma\,. \tag{3.68}$$

Proof. The proof follows similar steps as in Proposition 3.5. In the case that the signature-dependent codes are used, according to Proposition 3.1, there exists product distribution $\prod_{n=1}^{N} p_{X_n|\boldsymbol{A}}(x_n|\boldsymbol{a})$ on $\prod_{n=1}^{N} \mathcal{X}_n$ for any given user activity $\boldsymbol{A} = \boldsymbol{a}$ such that for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$,

$$|\mathcal{S}_n| C \le I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right). \tag{3.69}$$

Let $\gamma(\boldsymbol{a})$ denote the average total transmit power of all nodes when the user activity is $\boldsymbol{a}$. Since the average power of each transmitted codeword is 1, $\gamma(\boldsymbol{a})$ satisfies

$$\sum_{\boldsymbol{a}} \mathsf{P}\left(\boldsymbol{A} = \boldsymbol{a}\right) \gamma(\boldsymbol{a}) = \gamma N. \tag{3.70}$$

Consider a special case that $\mathcal{S}_n = \mathcal{N}_n$. Similarly as in (3.58), we have

$$I\left(X(\mathcal{N}_n); Y_n \middle| \boldsymbol{A} = \boldsymbol{a}\right) \le (1 - s_n) g\left(\gamma(\boldsymbol{a})\right). \tag{3.71}$$

By averaging over all realizations of $\boldsymbol{A}$ and summing over $n = 1, \ldots, N$, it follows from (3.69) and (3.71) that

$$N(N-1)C \leq \max_{\gamma(\boldsymbol{a})} \sum_{\boldsymbol{a}} \mathsf{P}\left(\boldsymbol{A} = \boldsymbol{a}\right) \left(N - \boldsymbol{a} \cdot \mathcal{N}\right) g\left(\gamma(\boldsymbol{a})\right). \tag{3.72}$$

The optimal power allocation $\gamma^*(\boldsymbol{a})$ to solve the maximization problem in (3.72) under the constraint (3.70) is similar as the waterfilling in parallel Gaussian channels [**17**, **82**], which can be expressed as

$$\gamma^*(\boldsymbol{a}) = \begin{cases} 0, & \boldsymbol{a} \cdot \mathcal{N} = 0; \\ \max\left\{\left(N - \boldsymbol{a} \cdot \mathcal{N}\right) v - 1, 0\right)\}, & \text{otherwise.} \end{cases} \tag{3.73}$$

and v is chosen to satisfy (3.70). Denote $u_k = \gamma^*(\boldsymbol{a}) = \max\left\{\left(N-k\right) v - 1, 0\right)\}$ when $\boldsymbol{a} \cdot \mathcal{N} = k \geq 1$, then according to (3.70), u_k satisfies

$$\frac{1}{N} \sum_{k=1}^{N-1} \binom{N}{k} q^n (1-q)^{N-k} u_k = \gamma\,. \tag{3.74}$$

It follows from (3.72) that

$$C \leq \frac{1}{N(N-1)} \sum_{k=1}^{N} \binom{N}{k} q^k (1-q)^{N-k} (N-k) g(u_k) \tag{3.75}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} \binom{N-1}{k} q^k (1-q)^{N-k} g(u_k) \tag{3.76}$$

the equality is achieved by the following multiplexing scheme: whenever the user activity is $\boldsymbol{a}$ with $\boldsymbol{a} \cdot \mathcal{N} = k \geq 1$, each node uses random codebook with i.i.d. Gaussian entries with average power u_k/k.

Now we show that the codebooks designed above satisfy the condition (3.69). In fact, for any on-off activity pattern $\boldsymbol{a}$ with $\boldsymbol{a} \cdot \mathcal{N} = k \geq 1$, similarly as in (3.62), we have

$$I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) = (1 - s_n) g\left(\frac{(\boldsymbol{a} \cdot \mathcal{S}_n) u_k}{k}\right). \tag{3.77}$$

By following similar steps as in (3.65), it can be proved that for $\mathcal{S}_n$ with $|\mathcal{S}_n| = l \leq N-1$,

$$\frac{1}{l} I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) \geq \frac{1}{N-1} \sum_{k=1}^{N-1} \binom{N-1}{k} q^k (1-q)^{N-k} g\left(u_k\right) = C. \tag{3.78}$$

Therefore, (3.69) holds for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. Proposition 3.5 is thus established. □

The case of signature-dependent codes can be regarded as allocating different powers to different on-off activity patterns in a parallel Gaussian multiaccess model. And the allocated power only depends on the number of transmitting nodes in each slot. As is shown in Fig. 3.2, the throughput of RODD with signature-independent codes is higher than that of ALOHA for all number of nodes and every value of q. The more nodes in the network, the more advantage of RODD signaling.

3.2.4. The Achievable Asymmetric Rates

In many applications, the amount of data different nodes transmit/broadcast can be very different. In random access schemes, nodes with more data will contend for more resources. The data rate, transmit power and modulation format of a RODD-based codebook can be adapted to the amount of data to be transmitted.

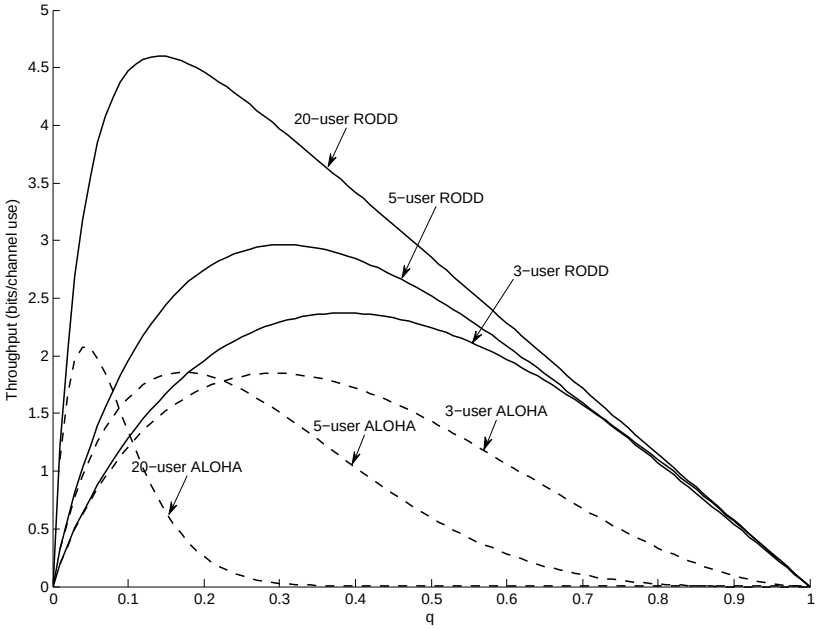

Figure 3.2. Comparison of the throughput of RODD and ALOHA over Gaussian multiaccess channel at SNR $\gamma = 20$ dB.

Suppose the elements s_{mn} of node n's signature are i.i.d. Bernoulli random variables with $\mathsf{P}(s_{mn} = 1) = q_n$. Here we study the asymmetric rate region of RODD achieved by signature-independent codes under the fading channel model described in (3.1).

Proposition 3.6. *The rate tuple* $(R_1, \ldots, R_N)$ *is achievable over the fading channel* (3.1) *if*

$$R_k \leq \min_{i \neq k} \sum_{\mathcal{A} \subseteq \mathcal{N} \setminus \{i\}, k \in \mathcal{A}} \frac{\gamma_{ik}}{q_k h^i_{\mathcal{A}}} g(h^i_{\mathcal{A}}) \prod_{j \in \mathcal{A}} q_j \prod_{l \in \mathcal{N} \setminus \mathcal{A}} (1 - q_l) \tag{3.79}$$

for all $k \in \mathcal{N} = \{1, 2, \ldots, N\}$ *and* $h^i_{\mathcal{A}}$ *is defined as* $h^i_{\mathcal{A}} = \sum_{j \in \mathcal{A}} \frac{\gamma_{ij}}{q_j}$.

Proof. Suppose each node uses signature-independent codes with random Gaussian codebooks, i.e., X_n is a Gaussian random variable with zero mean and variance γ/q_n since the average power of each transmitted codeword is assumed to be 1. In order to prove

Proposition 3.6, it suffices to show that the rate tuple expressed in (3.79) falls inside the capacity region (3.4) in Proposition 3.1.

For any on-off activity pattern $\boldsymbol{a}$, similarly as in (3.62), we have

$$I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A} = \boldsymbol{a}\right) = (1 - s_n) g\left(\sum_{k \in \mathcal{S}_n} \frac{s_k \gamma_{nk}}{q_k}\right). \tag{3.80}$$

By averaging over all realizations of $\boldsymbol{A}$, it follows that

$$\begin{aligned} I\left(X(\mathcal{S}_n); Y_n \middle| X(\mathcal{S}_n^c), \boldsymbol{A}\right) &= \sum_{\mathcal{A} \subseteq \mathcal{N}_n} \prod_{j \in \mathcal{A}} q_j \prod_{l \notin \mathcal{A}} (1 - q_l) g\left(\sum_{k \in \mathcal{S}_n \cap \mathcal{A}} \frac{\gamma_{nk}}{q_k}\right) \\ &\geq \sum_{\mathcal{A} \subseteq \mathcal{N}_n} \prod_{j \in \mathcal{A}} q_j \prod_{l \notin \mathcal{A}} (1 - q_l) g\left(h^n_{\mathcal{S}_n}\right) \sum_{k \in \mathcal{S}_n \cap \mathcal{A}} \frac{\gamma_{nk}}{q_k h^n_{\mathcal{S}_n}} \qquad (3.81) \\ &= \sum_{k \in \mathcal{S}_n} \sum_{\mathcal{A} \subseteq \mathcal{N}_n, k \in \mathcal{A}} \frac{\gamma_{nk}}{q_k h^n_{\mathcal{S}_n}} g(h^i_{\mathcal{A}}) \prod_{j \in \mathcal{A}} q_j \prod_{l \notin \mathcal{A}} (1 - q_l) \\ &\geq \sum_{k \in \mathcal{S}_n} R_k \qquad (3.82) \end{aligned}$$

where (3.81) is due to (3.66) and (3.82) is due to (3.79). Therefore, 3.1 holds for all $n \in \mathcal{N}$ and $\mathcal{S}_n \subseteq \mathcal{N}_n$. Proposition 3.6 is thus established. □

In the special case that $\gamma_{ij} = \gamma$ for all $i \neq j$ and $q_i = q$ for all $i \in \mathcal{N}$, it follows from (3.79) that

$$\begin{aligned} R_k &\leq \sum_{\mathcal{A} \subseteq \mathcal{N}_i, k \in \mathcal{A}} \frac{1}{|\mathcal{A}|} q^{|\mathcal{A}|} (1 - q)^{N - |\mathcal{A}|} g\left(\frac{|\mathcal{A}|\gamma}{q}\right) \qquad (3.83) \\ &= \sum_{i=1}^{N-1} \frac{1}{i} \binom{N-2}{i-1} q^i (1 - q)^{N-1} g\left(\frac{i\gamma}{q}\right) \qquad (3.84) \\ &= \frac{1}{N-1} \sum_{i=1}^{N-1} \binom{N-1}{i} q^i (1 - q)^{N-i} g\left(\frac{i\gamma}{q}\right), \qquad (3.85) \end{aligned}$$

which is the same expression as the symmetric rate of the non-fading Gaussian multiaccess channel described in (3.53).

3.3. Summary

This chapter evaluates the network capacity of RODD under two simple channel models, namely, the deterministic-OR channel and Gaussian channel. The traffic is assumed to be mutual broadcast from each node to all other nodes in the network. The throughput of RODD is analyzed and shown by numerical results to be significantly larger than that of ALOHA. The results in this chapter have been published in part in [**31**].

CHAPTER 4

Virtual Full-Duplex Mutual Broadcast of Short Messages

Consider a frequent situation in wireless peer-to-peer networks, where every node wishes to broadcast messages to all its peers, and also wishes to receive the messages from its peers. Such mutual broadcast traffic can be dominant in many applications, such as messaging or video conferencing of multiple parties in a spontaneous social network, or on an incident scene or a battlefield. Wireless mutual broadcast is also critical to efficient network resource allocation, where messages are exchanged between nodes about their demands and local states, such as queue length, channel quality, code and modulation format, and request for certain resources and services.

To achieve reliable mutual broadcast using a usual packet-based scheme, nodes have to repeat their packets a number of times interleaved with random delays, so that all peers can hear each other after enough retransmissions. This leads to the ubiquitous random channel access solution. A closer examination of the half-duplex constraint, however, reveals that a node does not need to transmit an entire packet before listening to the channel. An alternative solution by using RODD signaling is conceivable. If nodes activate different sets of on-slots, then nodes can all transmit information during a frame and receive useful signals within the same frame, and decode messages from peers as long as sufficiently strong error-control codes are applied. In fact, reliable mutual broadcast can be achieved using a single frame interval.

In this chapter, we focus on a special use of RODD signaling and a special case of mutual broadcast, where each node has a small number of bits to send to its peers. The goal here is to provide a practical algorithm for encoding and decoding the short messages to achieve reliable and efficient mutual broadcast. Decoding is in fact a problem of support recovery based on linear measurements, since the received signal is basically a noisy superposition of peers' codewords selected from their respective codebooks. There are many algorithms developed in the compressed sensing (or sparse recovery) literature to solve the problem, the complexity of which is often polynomial in the size of the codebook (see, e.g., [**10**, **18**, **20**, **59**]). In this chapter, an iterative message-passing algorithm based on belief propagation (BP) with linear complexity is developed. Numerical results show that the proposed RODD scheme significantly outperforms slotted-ALOHA with multi-packet reception capability and CSMA in terms of data rate.

The remainder of this chapter is organized as follows. After the system model is presented in Section 4.1, Section 4.2 studies the conventional random-access schemes, namely slotted-ALOHA with multi-packet reception capability and CSMA. In Section 4.3, the proposed coding scheme for mutual broadcast is described. The message-passing decoding algorithm is developed and analyzed in Section 4.4. Numerical comparisons are presented in Section 4.5. Section 4.6 summarizes the chapter.

4.1. Channel and Network Models

4.1.1. Linear Channel Model

Let $\Phi = \{X_i\}_i$ denote the set of nodes on the plane. We refer to a node by its location X_i. Suppose all transmissions use the same single carrier frequency. Let time be slotted and all

nodes be perfectly synchronized.[1] Let $\varpi_i \in \{1, \ldots, 2^l\}$ denote the message node X_i wishes to broadcast. In discrete-time baseband, let $\boldsymbol{S}_i(\varpi_i)$ denote the signature (codeword) transmitted by node X_i, whose entries take values in $\{-1, 0, +1\}$. Let U_{0i} denote the complex-valued coefficient of the wireless link from X_i to X_0. The signal received by node X_0, if it could listen over the entire frame, is then described by

$$\widetilde{\boldsymbol{Y}} = \sqrt{\gamma} \sum_{X_i \in \Phi \setminus \{X_0\}} U_{0i} \boldsymbol{S}_i(\varpi_i) + \widetilde{\boldsymbol{W}} \tag{4.1}$$

where $\widetilde{\boldsymbol{W}}$ is noise consisting of i.i.d. circularly symmetric complex Gaussian entries with zero mean and unit variance, and γ denotes the nominal SNR. Denote the set of neighbors of X_0 by $\mathcal{N}(X_0)$. For simplicity, if we further assume that transmissions from non-neighbors, if any, are accounted for as part of the additive Gaussian noise, (4.1) can be rewritten as

$$\widetilde{\boldsymbol{Y}} = \sqrt{\gamma} \sum_{X_i \in \mathcal{N}(X_0)} U_{0i} \boldsymbol{S}_i(\varpi_i) + \overline{\boldsymbol{W}} \tag{4.2}$$

where each element in $\overline{\boldsymbol{W}}$ is assumed to be circularly symmetric complex Gaussian with variance σ^2. The variance accounts for noise and interference from non-neighbors and depends on the network topology. It will be derived in Section 4.3.

4.1.2. Network Model

Consider a network with nodes distributed across the plane according to a homogeneous Poisson point process (p.p.p.) with intensity λ. The number of nodes in any region of area A is a Poisson random variable with mean λA. Without loss of generality, we assume node

[1] See Introduction for a discussion of synchronization issues. In [**6**], cyclic codes are proposed to resolve the user delays in a multiaccess channel.

X_0 is located at the origin and focus on its performance, which should be representative of any node in the network.

Poisson point process is the most frequently used model to study wireless networks (see [**9**] and references therein). The RODD signaling and the mutual broadcast scheme proposed in this chapter are of course not limited to homogeneous Poisson distributed networks. The homogeneous p.p.p. model is assumed in this chapter to facilitate analysis and comparison of RODD and competing technologies.

4.1.3. Propagation Model and Neighborhood

It is assumed that the large-scale signal attenuation over distance follows power law with some path-loss exponent $\alpha > 2$, and the small-scale fading of a link is modeled by a Rayleigh random variable with mean equal to 1. There are different ways to define the neighborhood of a node. For concreteness, we say that nodes X_i and X_j are neighbors of each other if the channel gain between them is no less than a certain threshold, denoted by θ. Link reciprocity is regarded as given.

For any pair of nodes $X_i, X_j \in \Phi$, let $R_{ij} = |X_i - X_j|$ and G_{ij} denote the distance and the small-fading gain between them in a given frame, respectively. Then the channel gain from X_j to X_i is $G_{ij}R_{ij}^{-\alpha}$. The neighborhood of a node depends on the instantaneous fading gains. Specifically, we denote the set of neighbors of node X_i as

$$\mathcal{N}(X_i) = \left\{ X_j \in \Phi : G_{ij}R_{ij}^{-\alpha} \geq \theta, j \neq i \right\}. \tag{4.3}$$

The channel coefficient U_{ij} should satisfy $|U_{ij}|^2 = G_{ij}R_{ij}^{-\alpha}$, where its phase is assumed to be uniformly distributed on $[0, 2\pi)$ independent of everything else. Assuming the Poisson

network model introduced in Section 4.1.2, the distribution of the amplitude of channel coefficient U_{0i} in (4.1) for an arbitrary neighbor $X_i \in \mathcal{N}(X_0)$ is derived in the following.

Without loss of generality, we drop the indices 0 and i, and use R and G to denote the distance and the fading gain, respectively. Since the two nodes are assumed to be neighbors, G and R satisfy $GR^{-\alpha} \geq \theta$, i.e., $R \leq (G/\theta)^{1/\alpha}$. Under the assumption that all nodes form a p.p.p., for given G, this arbitrary neighbor X_i is uniformly distributed in a disc centered at node X_0 with radius $(G/\theta)^{1/\alpha}$. Therefore, the conditional distribution of R given G can be expressed as

$$\mathsf{P}(R \leq r \big| G) = \min\left\{1, r^2 \left(\frac{\theta}{G}\right)^{\frac{2}{\alpha}}\right\}. \tag{4.4}$$

Now for every $u \geq \sqrt{\theta}$, by (4.4) we have

$$\mathsf{P}(GR^{-\alpha} \geq u^2) = \mathsf{E}_G\left\{\mathsf{P}\left(R \leq \left(\frac{G}{u^2}\right)^{\frac{1}{\alpha}} \Big| G\right)\right\} \tag{4.5}$$

$$= \mathsf{E}_G\left\{\left(\frac{G}{u^2}\right)^{\frac{2}{\alpha}} \left(\frac{\theta}{G}\right)^{\frac{2}{\alpha}}\right\} \tag{4.6}$$

$$= \frac{\theta^{\frac{2}{\alpha}}}{u^{\frac{4}{\alpha}}}. \tag{4.7}$$

Hence the pdf of $|U_{0i}|$ of neighbor X_i is

$$p(u) = \begin{cases} \frac{4}{\alpha}\frac{\theta^{2/\alpha}}{u^{4/\alpha+1}}, & u \geq \sqrt{\theta}; \\ 0, & \text{otherwise.} \end{cases} \tag{4.8}$$

In fact, coefficient vector $\mathcal{G}_i = (G_{ji})_j$ for all $j \neq i$ can be regarded as a mark of node X_i, so that $\tilde{\Phi} = \{(X_i, \mathcal{G}_i)\}_i$ is a marked p.p.p.. Denote

$$\hat{\Phi} = \tilde{\Phi} \backslash (X_0, \mathcal{G}_0) \tag{4.9}$$

given that $(X_0, \mathcal{G}_0)$ is at the origin. By the Slivnyak-Meche theorem [**9**], $\hat{\Phi}$ is also a marked p.p.p. with intensity λ. By the Campbell's theorem [**9**], the average number of neighbors of X_0 can be obtained as:

$$c = \mathsf{E}_{\hat{\Phi}} \left\{ \sum_{(X_i, \mathcal{G}_i) \in \hat{\Phi}} 1\left(G_{0i} R_{0i}^{-\alpha} \geq \theta\right) \right\} \tag{4.10}$$

$$= 2\pi\lambda \int_0^\infty \int_0^\infty 1\left(g r^{-\alpha} \geq \theta\right) r e^{-g} \mathrm{d}r \mathrm{d}g \tag{4.11}$$

$$= \frac{2}{\alpha} \pi \lambda \theta^{-2/\alpha} \Gamma\left(\frac{2}{\alpha}\right) \tag{4.12}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $\Gamma(\cdot)$ is the Gamma function.

4.2. Random-Access Schemes

In this section we describe two random access schemes, namely slotted ALOHA and CSMA, and provide lower bounds on the error probability for a given number of symbol transmissions. The results will be used in Section 4.5 to compare with the performance of our proposed sparse recovery scheme.

Let L denote the total number of bits encoded into a frame, which includes an l-bit message and a few additional bits which identify the sender. This is in contrast to broadcast via sparse recovery, where the signature itself identifies the sender (and carries the message). Each broadcast period consists of a number of frames to allow for

retransmissions. Without loss of generality, we still consider the typical node X_0 at the origin. An error event is defined as that node X_0 cannot correctly recover the message from one specific neighbor. The corresponding error probabilities achieved by slotted ALOHA and CSMA are denote by $\mathsf{P}_{\mathrm{a}}^{\mathrm{e}}$ and $\mathsf{P}_{\mathrm{c}}^{\mathrm{e}}$, respectively.

4.2.1. Slotted ALOHA

In slotted ALOHA, suppose each node chooses independently with the same probability p to transmit in every frame interval. A message is assumed to be decoded correctly if the signal-to-interference-plus-noise ratio (SINR) in the corresponding frame transmission is no smaller than a threshold δ (multi-packet reception is possible only if $\delta < 1$). Over the additive white noise channel with SINR δ, in order to send L bits reliably through the channel, the number of symbols in a frame must exceed

$$\frac{L}{\log_2(1+\delta)}\,. \tag{4.13}$$

Let X denote one specific neighbor of node X_0 and G denote the fading coefficient between them. Suppose the mark of X is denoted by $\mathcal{G}$. Given that $(X,\mathcal{G}) \in \hat{\Phi}$ where $\hat{\Phi}$ is given by (4.9), denote $\hat{\Phi}_1 = \hat{\Phi}\backslash\{(X,\mathcal{G})\}$, which is also a marked p.p.p. with intensity λ. For a given realization of (X,G) and $\hat{\Phi}_1$, define $\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X,G,\hat{\Phi}_1)$ as the probability that the received SINR from X to X_0 is no less than the threshold δ conditioning on that X transmits in a given frame. In any given frame, the probability of the event that X transmits, X_0 listens, and the transmission is successful is thus $p(1-p)\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X,G,\hat{\Phi}_1)$. Therefore, the probability that the message from X has not been successfully received by

X_0 after M_f consecutive frame intervals can be expressed as

$$\mathsf{P}_{\mathrm{a}}^{\mathrm{e}} = \mathsf{E}\left\{\left(1-p(1-p)\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X,G,\hat{\Phi}_1)\right)^{M_f}\right\} \tag{4.14}$$

where the expectation is over the joint distribution of $(X, G, \hat{\Phi}_1)$. Due to the convexity of function $(\max\{0, 1-z\})^n$, $z \geq 0, n \in \{1, 2, \dots\}$, $\mathsf{P}_{\mathrm{a}}^{\mathrm{e}}$ in (4.14) can be lower bounded as

$$\mathsf{P}_{\mathrm{a}}^{\mathrm{e}} \geq \left(\max\left\{0, 1-p(1-p)\mathsf{E}\left\{\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X,G,\hat{\Phi}_1)\right\}\right\}\right)^{M_f}. \tag{4.15}$$

The expectation of $\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X, G, \hat{\Phi}_1)$ can be calculated using the known Laplace transform of the distribution of the interference [**9**]. For a given number of symbol transmissions M_a, the lower bound on $\mathsf{P}_{\mathrm{a}}^{\mathrm{e}}$ is presented in the following result.

Proposition 4.1. *Consider an arbitrary neighbor X of node X_0. The probability that X_0 cannot successfully receive the message from X after M_a symbol transmissions is lower bounded as follows:*

$$\mathsf{P}_a^e \geq \left(\max\left\{0, 1-\frac{1}{\pi}p(1-p)\left(\frac{\theta}{\delta}\right)^b \sin\left(\frac{b\pi}{2}\right)\Gamma(1-b)\int_{-\infty}^{\infty}|\omega|^{b-1}e^{-\lambda p\frac{b\pi^2}{\sin(b\pi)}(\iota\omega)^b-\iota\frac{\omega}{\gamma}}\mathrm{d}\omega\right\}\right)^{n_a} \tag{4.16}$$

where $\iota = \sqrt{-1}$, $b = 2/\alpha$ and $n_a = M_a \log_2(1+\delta)/L$.

Proof. Let $\hat{\Phi}_1^p$ be an independent thinning of $\hat{\Phi}_1$ with retention probability p to represent the transmitting nodes. It is easy to see that $\hat{\Phi}_1^p$ is an independent marked p.p.p. with intensity λp. Denote

$$I = \sum_{(X_i,\mathcal{G}_i)\in\hat{\Phi}_1^p} G_{0i}|X_{0i}|^{-\alpha}, \tag{4.17}$$

then we have

$$\mathsf{E}\left\{\mathsf{P}_{\mathrm{a}}^{\mathrm{s}}(X,\mathcal{G},\hat{\Phi}_1)\right\} = \mathsf{E}\left\{\mathsf{E}\left\{1\left(\frac{\gamma G|X|^{-\alpha}}{\gamma I+1}\geq\delta\right)\,\middle|\,\hat{\Phi}_1^p\right\}\right\} \tag{4.18}$$

$$= \mathsf{E}\left\{\mathsf{P}\left\{G|X|^{-\alpha}\geq\delta\left(I+\frac{1}{\gamma}\right)\middle|\hat{\Phi}_1^p\right\}\right\} \tag{4.19}$$

$$= \mathsf{E}\left\{\left(\frac{\theta}{\delta}\right)^{\frac{2}{\alpha}}\left(I+\frac{1}{\gamma}\right)^{-\frac{2}{\alpha}}\right\} \tag{4.20}$$

$$= \left(\frac{\theta}{\delta}\right)^{\frac{2}{\alpha}}\int_{-\infty}^{\infty}\left|i+\frac{1}{\gamma}\right|^{-\frac{2}{\alpha}}p_I(i)\mathrm{d}i \tag{4.21}$$

where (4.20) is derived from (4.7) and p_I is the pdf of interference I.

According to [**9**], the Laplace transform of p_I can be expressed as

$$\mathcal{L}_{p_I}(s) = \exp\left\{-\lambda p s^{2/\alpha}\frac{2\pi^2}{\alpha\sin(2\pi/\alpha)}\right\}. \tag{4.22}$$

Therefore, the Fourier transform[2] of p_I can be obtained by replacing s in (4.22) by $\iota\omega$ with $\iota=\sqrt{-1}$ as

$$\mathcal{F}_{p_I}(\omega) = \exp\left\{-\lambda p(\iota\omega)^{2/\alpha}\frac{2\pi^2}{\alpha\sin(2\pi/\alpha)}\right\}. \tag{4.23}$$

Since the Fourier transform of $|x|^a$ for $-1<a<0$ is

$$\mathcal{F}_{|x|^a}(\omega) = -\frac{2\sin(a\pi/2)\Gamma(a+1)}{|\omega|^{a+1}}, \tag{4.24}$$

[2]The reasons to work with Fourier transform in lieu of Laplace transform are: 1) The inverse Fourier transform here is easier to calculate; 2) the Fourier transform of $|i+\frac{1}{\gamma}|^{-\frac{2}{\alpha}}$ has a closed form.

the Fourier transform of

$$q_I(i) = |i + \frac{1}{\gamma}|^{-\frac{2}{\alpha}} \tag{4.25}$$

for $\alpha > 2$ can be expressed as

$$\mathcal{F}_{q_I}(\omega) = e^{\iota\omega/\gamma} \frac{2\sin(\pi/\alpha)\Gamma(1-2/\alpha)}{|\omega|^{1-2/\alpha}}\,. \tag{4.26}$$

Since the integral in (4.21) can be viewed as the Fourier transform of $p_I(i)q_I(i)$ at $\omega = 0$, it can be calculated as the convolution of $\mathcal{F}_{p_I}(\omega)$ and $\mathcal{F}_{q_I}(\omega)$ at $\omega = 0$ [**61**]. Therefore, by (4.23) and (4.26), we have

$$\int_{-\infty}^{\infty} \left| i + \frac{1}{\gamma} \right|^{-\frac{2}{\alpha}} p_I(i)\mathrm{d}i = \frac{1}{2\pi} \mathcal{F}_{p_I}(\omega) * \mathcal{F}_{q_I}(\omega) \Big|_{\omega=0} \tag{4.27}$$

where $*$ is the convolution operator. Therefore, according to (4.15), (4.21) and (4.27), the error probability $\mathsf{P}_{\mathrm{a}}^{\mathrm{e}}$ can be lower bounded as

$$\mathsf{P}_{\mathrm{a}}^{\mathrm{e}} \geq \left(\max\left\{ 0, 1 - \frac{1}{2\pi} p(1-p) \left(\frac{\theta}{T}\right)^{\frac{2}{\alpha}} \mathcal{F}_{p_I}(\omega) * \mathcal{F}_{q_I}(\omega) \Big|_{\omega=0} \right\} \right)^{M_f}. \tag{4.28}$$

According to (4.13), the number of frames in a period of M_a symbol intervals should satisfy

$$M_f \geq M_a \log_2(1+\delta)/L\,. \tag{4.29}$$

Therefore, (4.16) in Proposition 4.1 follows by combining (4.28) and (4.29). □

Although (4.16) appears to be complicated, computing it only involves a straightforward single-variable integral (the outcome of the integral is in fact real-valued).

In the slotted ALOHA scheme, despite repeated transmissions, a given link may still fail to deliver the message due to the half-duplex constraint (the receiver happens to transmit during the same frame) and consistently weak received SINR due to random interference from other links.

4.2.2. CSMA

As an improvement over ALOHA, CSMA lets nodes use a brief contention period to negotiate a schedule in such a way that nodes in a small neighborhood do not transmit data simultaneously. We analyze the performance of CSMA by using the Matérn hard core model [**9**]. To be specific, consider the following generic scheme: Each node senses the channel continuously; if the channel is busy, the node remains silent and disables its timer; as soon as the channel becomes available, the node starts its timer with a random offset, and waits till the timer expires to transmit its frame. Clearly, the node whose timer expires first in its neighborhood captures the channel and transmits its frame.

Mathematically, let $\{T_i\}$ be i.i.d. random variables with uniform distribution on $[0, 1]$, which represent the timer offsets for all nodes $\{X_i\}$ in Φ, respectively. By viewing T_i as a mark of node X_i we redefine $\tilde{\Phi} = \{(X_i, \mathcal{G}_i, T_i)\}_i$, which is still a marked p.p.p. with intensity λ. The medium access indicators $\{e_i\}_i$ are additional dependent marks of the nodes in Φ defined as follows:

$$E_i = \mathbf{1}\left(T_j > T_i,\ \forall X_j \in \mathcal{N}(X_i)\right). \tag{4.30}$$

The probability of $T_j = T_i$ for $j \neq i$ is zero. Node X_i will transmit its frame if and only if $E_i = 1$.

The same as in the slotted ALOHA case, a message is assumed to be decoded correctly if SINR in the corresponding frame transmission is no smaller than δ. It follows that the number of symbols in a frame must exceed (4.13).

Let X be one specific neighbor of X_0. Define G and $\mathcal{G}$ as in Section 4.2.1. Given that $(X, \mathcal{G}) \in \hat{\Phi}$, denote $\hat{\Phi}_1 = \hat{\Phi} \backslash \{(X, \mathcal{G})\}$. For a given realization of (X, G) and $\hat{\Phi}_1$, define $\mathsf{P}_{\mathrm{c}}^{\mathrm{s}}(X, G, \hat{\Phi}_1)$ as the probability that node X transmits its frame and the received SINR from X is no less than the threshold δ. Therefore, the probability that the message from X has not been successfully received after M_f consecutive frame intervals can be expressed as

$$\mathsf{P}_{\mathrm{c}}^{\mathrm{e}} = \mathsf{E}\left\{\left(1 - \mathsf{P}_{\mathrm{c}}^{\mathrm{s}}(X, G, \hat{\Phi}_1)\right)^{M_f}\right\} \tag{4.31}$$

$$\geq \left(\max\left\{0, 1 - \mathsf{E}\left\{\mathsf{P}_{\mathrm{c}}^{\mathrm{s}}(X, G, \hat{\Phi}_1)\right\}\right\}\right)^{M_f} \tag{4.32}$$

where the expectation is over the joint distribution of $(X, G, \hat{\Phi}_1)$, and (4.32) is due to the convexity of function $(\max\{0, 1 - z\})^n$, $z \geq 0, n = 0, 1, \ldots$.

For any given number of symbol transmissions M_c, the lower bound on error probability $\mathsf{P}_{\mathrm{c}}^{\mathrm{e}}$ is given by the following result.

Proposition 4.2. *Consider an arbitrary neighbor X of node X_0. The probability that X_0 cannot successfully receive the message from X after M_c symbol transmissions is lower bounded as follows:*

$$\mathsf{P}_c^e \geq \left(\max\left\{0, 1 - \frac{1}{c^2}\left(\frac{\theta\gamma}{\delta}\right)^{\frac{2}{\alpha}}\left(e^{-c} + c - 1\right)\right\}\right)^{n_c} \tag{4.33}$$

where c is defined in (4.12) *and $n_c = M_c \log_2(1 + \delta)/L$.*

Proof. Denote G_i as the fading coefficient from node $X_i \in \hat{\Phi}_1$ to node X_0. Define the following indicators for node X

$$F_1 = \mathbf{1}\left(T_0 > T\right) \tag{4.34}$$

$$F_2 = \mathbf{1}\left(T_i > T,\ \forall X_i \in \hat{\Phi}_1 \text{ with } G_i|X_i - X|^{-\alpha} \geq \theta\right) \tag{4.35}$$

$$F_3 = \mathbf{1}\left(\gamma G|X|^{-\alpha} \geq \delta\right) \tag{4.36}$$

where $F_1 = 1$ if and only if the timer of X expires before that of X_0, $F_2 = 1$ if and only if X's timer expires sooner than those of all its neighbors excluding X_0, $F_3 = 1$ if and only if the received SNR from node X to node X_0 exceeds the threshold δ. In order for the transmission to be successful, we must have $F_1 = F_2 = F_3 = 1$. That is

$$\mathsf{E}\left\{\mathsf{P}_{\mathrm{c}}^{\mathrm{s}}(X, \mathcal{G}, \hat{\Phi}_1)\right\} \leq \mathsf{E}\left\{F_1 F_2 F_3\right\}. \tag{4.37}$$

Conditioned on $T = \varsigma$, we express the indicator F_2 as the value of some extremal shot-noise [**9**, Section 2.4]. For fixed ς, define the indicator of the event that X_i is a neighbor of X and it has a timer smaller than ς:

$$L(X, X_i, G_i, T_i) = \mathbf{1}\left(G_i|X_i - X|^{-\alpha} \geq \theta \text{ and } T_i < \varsigma\right) \tag{4.38}$$

for all $(X_i, \mathcal{G}_i, T_i) \in \hat{\Phi}_1$. Define the extremal shot-noise at node X as

$$Z_{\hat{\Phi}_1}(X) = \max_{(X_i, \mathcal{G}_i, T_i) \in \hat{\Phi}_1} L(X, X_i, G_i, T_i). \tag{4.39}$$

Note that $Z_{\hat{\Phi}_1}(X)$ takes only two values 0 or 1 and consequently

$$\mathsf{E}\left\{F_2\middle|T=\varsigma\right\}=\mathsf{P}\left\{Z_{\hat{\Phi}_1}(X)\leq 0\middle|T=\varsigma\right\}. \tag{4.40}$$

By [**9**, Proposition 2.4.2], (4.40) can be further calculated as

$$\begin{aligned}\mathsf{E}\left\{F_2\middle|T=\varsigma\right\}&=\exp\left\{-\lambda\int_{\mathbb{R}^2}\int_0^\infty\int_0^1 1\left(L(X,x,g,t)=1\right)e^{-g}\mathrm{d}t\mathrm{d}g\mathrm{d}x\right\} &(4.41)\\ &=\exp\left\{-2\pi\lambda\varsigma\int_0^\infty\int_0^\infty 1\left(gr^{-\alpha}\geq\theta\right)re^{-g}\mathrm{d}g\mathrm{d}r\right\} &(4.42)\\ &=e^{-c\varsigma} &(4.43)\end{aligned}$$

where c is the average number of neighbors defined in (4.12).

Therefore, according to (4.37), we have

$$\begin{aligned}\mathsf{E}\left\{\mathsf{P}_{\mathrm{c}}^{\mathrm{s}}(X,\mathcal{G},\hat{\Phi}_1)\right\}&\leq\left(\frac{\theta\gamma}{\delta}\right)^{\frac{2}{\alpha}}\int_0^1(1-\varsigma)e^{-c\varsigma}\mathrm{d}\varsigma &(4.44)\\ &=\frac{1}{c^2}\left(\frac{\theta\gamma}{\delta}\right)^{\frac{2}{\alpha}}\left(e^{-c}+c-1\right) &(4.45)\end{aligned}$$

where (4.44) is derived from the the uniform distribution of T_0, (4.7) and (4.40).

According to (4.13), the number of frames in a period of M_c symbol intervals should satisfy

$$M_f\geq M_c\log_2(1+\delta)/L\,. \tag{4.46}$$

Therefore, Proposition 4.2 follows by combining (4.45) and (4.46). □

As a by-product, by averaging over ς in (4.43), which is uniformly distributed on $[0, 1]$, the probability that a given node captures the channel to transmit in each slot can be calculated as $(1 - e^{-c})/c$.

In contrast to slotted ALOHA, frame loss due to the half-duplex constraint is eliminated through contention. However, a given link may still fail to deliver the message after repeated transmissions because the received SINR is consistently weak due to random interference outside the neighborhood.

4.3. Encoding for Mutual Broadcast

In contrast to random-access schemes, where many retransmissions are needed to achieve a desired error performance, we next describe a unique signaling that allow all message exchanges to finish within one (longer) frame of transmission. The key idea is that each node broadcasts a codeword consisting of on-slots and off-slots. A node transmits only during its on-slots, and listens to its peers through its own off-slots.

Suppose each node X_i is assigned a unique codebook of 2^l on-off signatures (codewords) of length M_s, denoted by $\{\boldsymbol{S}_i(1), \ldots, \boldsymbol{S}_i(2^l)\}$. For simplicity, let each element of each signature be generated randomly and independently, which is 0 with probability $1-q$ and 1 and -1 with probability $q/2$ each. Node X_i broadcasts its l-bit message (or information index) $\varpi_i \in \{1, \ldots, 2^l\}$ by transmitting the codeword $\boldsymbol{S}_i(\varpi_i)$.

In each symbol slot, those transmitting nodes in $\hat{\Phi}$ defined in (4.9) form an independent thinning of $\hat{\Phi}$ with retention probability q, denoted by $\hat{\Phi}_q$. $\hat{\Phi}_q$ is still a marked p.p.p. but with intensity λq. Thus, the sum power from all transmitting non-neighbors of node X_0

in each time slot is derived as

$$\mathsf{E}_{\hat{\Phi}_q}\left\{\sum_{(X_i,\mathcal{G}_i)\in\hat{\Phi}_q}\gamma G_{0i}R_{0i}^{-\alpha}1\left(G_{0i}R_{0i}^{-\alpha}<\theta\right)\right\}$$

$$=2\pi\lambda q\gamma\int_0^\infty\int_0^\infty gr^{-\alpha}1\left(gr^{-\alpha}<\theta\right)re^{-g}\mathrm{d}r\mathrm{d}g \tag{4.47}$$

$$=2\pi\lambda q\gamma\int_0^\infty r^{-\alpha+1}\left[1-(\theta r^\alpha+1)e^{-\theta r^\alpha}\right]\mathrm{d}r \tag{4.48}$$

$$=\frac{4\pi\lambda q\gamma\theta}{\alpha-2}\int_0^\infty re^{-\theta r^\alpha}\mathrm{d}r \tag{4.49}$$

$$=\frac{4}{\alpha(\alpha-2)}\pi\lambda q\gamma\theta^{1-2/\alpha}\Gamma\left(\frac{2}{\alpha}\right). \tag{4.50}$$

Therefore, the variance of each element of $\overline{\boldsymbol{W}}$ in (4.2) is

$$\sigma^2=\frac{4}{\alpha(\alpha-2)}\pi\lambda q\gamma\theta^{1-2/\alpha}\Gamma\left(\frac{2}{\alpha}\right)+1. \tag{4.51}$$

The signal received by the typical node X_0, if it could listen over the entire frame, is described by (4.2). Suppose $|\mathcal{N}(X_0)|=K$ and the neighbors of X_0 are indexed by $1,2,\ldots,K$. The total number of signatures of all neighbors is $N=2^lK$. Due to the half-duplex constraint, however, node X_0 can only listen during its off-slots, the number of which has binomial distribution, denoted by $M\sim\mathcal{B}(M_s,1-q)$, whose expected value is $\mathsf{E}\{M\}=M_s(1-q)$. Let the matrix $\underline{\boldsymbol{S}}\in\mathbb{R}^{M\times N}$ consist of columns of the signatures from all neighbors of node X_0, observable during the M off-slots of node X_0, and then normalized by $\sqrt{M_sq(1-q)}$ so that the expected value of the l_2 norm of each column in $\underline{\boldsymbol{S}}$ is equal to 1. Based on (4.2), the M-vector observed through all off-slots of node X_0 can be expressed as

$$\boldsymbol{Y}=\sqrt{\gamma_s}\underline{\boldsymbol{S}}\boldsymbol{X}+\boldsymbol{W} \tag{4.52}$$

where

$$\gamma_s = \gamma M_s q(1-q)/\sigma^2, \tag{4.53}$$

$\boldsymbol{W}$ is Gaussian noise consisting of i.i.d. entries of zero mean and unit variance, and $\boldsymbol{X}$ is an N-vector indicating which K signatures are selected to form the sum in (4.2) as well as the signal strength for each neighbor. Precisely, $X_{(j-1)2^l+i} = U_j \mathbf{1}(w_j = i)$ for $1 \le j \le K$ and $1 \le i \le 2^l$. For example, consider $K = 3$ neighbors with $l = 2$ bits of information each, where $w_1 = 3, w_2 = 2, w_3 = 1$, then vector $\boldsymbol{X}$ is expressed as

$$\boldsymbol{X} = [0 \;\; 0 \;\; U_1 \;\; 0 \;\; 0 \;\; U_2 \;\; 0 \;\; 0 \;\; U_3 \;\; 0 \;\; 0 \;\; 0]. \tag{4.54}$$

The sparsity of $\boldsymbol{X}$ is exactly 2^{-l}, which is very small for large l. The average system load is defined as $\beta = \mathsf{E}\{N\}/\big(M_s(1-q)\big) = 2^l c/\big(M_s(1-q)\big)$.

In general, the decoding problem node X_0 faces is to identify, out of a total of $N = 2^l K$ signatures from all its neighbors, which K signatures were selected. This requires every node to know the codebooks of all neighbors. One solution is to let the codebook of each node be generated using a pseudo-random number generator using its NIA as the seed, so that it suffices to acquire all neighbors' NIAs. This, in turn, is a neighbor discovery problem, which shall be studied in Chapter 5, where the discovery scheme uses similar on-off signalling and also solves a compressed sensing problem.

4.4. Sparse Recovery Decoding via Message Passing

The problem of recovering the support of the sparse input $\boldsymbol{X}$ based on the observation $\boldsymbol{Y}$ has been intensively studied in the compressed sensing literature. In this section, we

develop an iterative message-passing algorithm based on belief propagation, and characterize its performance in a certain limit. The reasons for the choice of message passing algorithm include: 1) It is one of the most competitive decoding schemes in terms of error performance; and 2) the complexity is only linear in the vector to be estimated.

4.4.1. The Factor Graph

Belief propagation belongs to a general class of message-passing algorithms for statistical inference on graphical models, which has demonstrated empirical success in many applications including error-control codes, neural networks, and multiuser detection in CDMA systems.

In order to apply BP to the coded mutual broadcast problem, we construct a Forney-style bipartite factor graph to represent the model (4.52). Here, we separate the real and imaginary parts in (4.52) as

$$\boldsymbol{Y}^{(1)} = \sqrt{\gamma_s}\underline{\boldsymbol{S}}\boldsymbol{X}^{(1)} + \boldsymbol{W}^{(1)},\ \boldsymbol{Y}^{(2)} = \sqrt{\gamma_s}\underline{\boldsymbol{S}}\boldsymbol{X}^{(2)} + \boldsymbol{W}^{(2)} \tag{4.55}$$

where the superscripts (1) and (2) represent the real and imaginary parts respectively, $\boldsymbol{W}^{(i)}, i = 1, 2$ consists of i.i.d. Gaussian random variables with zero mean and variance 1/2. The message passing algorithm we shall develop based on (4.55) is not optimal, but such separation facilitates approximation and computation, which will be discussed in Section 4.4.2. Since two parts in (4.55) share the same factor graph, we treat one of them and omit the superscripts:

$$y_\mu = \sqrt{\gamma_s}\sum_{k=1}^{N} s_{\mu k}x_k + w_\mu \tag{4.56}$$

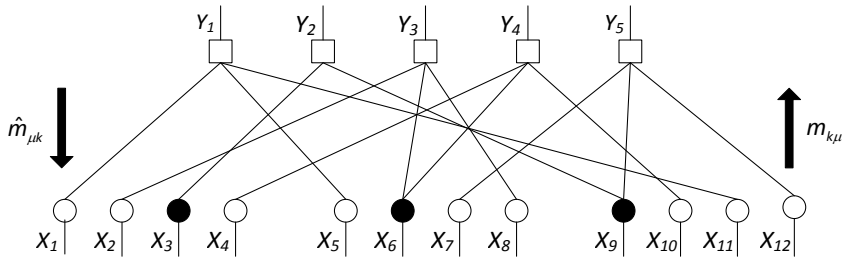

Figure 4.1. The Forney-style factor graph of coded mutual broadcast.

where $\mu \in 1, 2, \ldots, M$ and $k \in 1, 2, \ldots, N$ index the measurements and the input "symbols," respectively. For simplicity, we ignore the dependence of the symbols $\{X_k\}$ for now, which shall be addressed toward the end of this section. Each X_k then corresponds to a symbol node and each Y_μ corresponds to a measurement node, where the joint distribution of all $\{X_k\}$ and $\{Y_\mu\}$ are decomposed into a product of $M + N$ factors, one corresponding to each node. For every (μ, k), symbol node k and measurement node μ are connected by an edge if $s_{\mu k} \neq 0$. A simple example is shown in Fig. 4.1 for 5 measurements and 3 neighbors each with 4 messages, i.e., $M = 5$, $K = 3$, and $N = 3 \times 4 = 12$. The actual message chosen by each neighbor for broadcast is marked by a dark node, i.e., $w_1 = 3, w_2 = 2, w_3 = 1$.

4.4.2. The Message-Passing Algorithm

In general, an iterative message-passing algorithm involves two steps in each iteration, where a message (or belief, which shall be distinguished from an information message) is first sent from each symbol node to every measurement node it is connected to, and then a new set of messages are computed and sent in the reverse direction, and so forth. The algorithm performs exact inference within finite number of iterations if there are no

loops in the graph (the graph becomes a tree if it remains connected), while it provides in general a good approximation for loopy graphs as the one in the current problem.

For convenience, let $\partial\mu$ (resp. ∂k) denote the subset of symbol nodes (resp. measurement nodes) connected directly to measurement node μ (resp. symbol node k), called its neighborhood.[3] $|\partial\mu|$ (resp. $|\partial k|$) represents the cardinality of the neighborhood of measurement node μ (resp. symbol node k). Also, let $\partial\mu\backslash k$ denote the neighborhood of measurement node μ excluding symbol node k and let $\partial k\backslash\mu$ be similarly defined.

The message-passing algorithm given as Algorithm 1, decodes the information indexes $w_1, \ldots, w_K$, and is ready for implementation. The key steps are described as follows. The superscripts $i = 1, 2$ in Algorithm 1 represent the real and imaginary parts, respectively. Here, $\mathsf{E}\left\{X \,\middle|\, Y = y; \sigma^2\right\}$ and $\mathsf{var}\left\{X \,\middle|\, Y = y; \sigma^2\right\}$ represents the conditional mean and variance of the input given the Gaussian channel output $Y = X + W$ with $W \sim \mathcal{N}(0, \sigma^2)$ is equal to y. Mathematically, assume X has cumulative distribution function $P_X(x)$, then for $n = 1, 2, \ldots,$

$$\mathsf{E}\left\{X^n \,\middle|\, Y = y; \sigma^2\right\} = \frac{\int x^n e^{-\frac{(y-x)^2}{2\sigma^2}} \mathrm{d}P_X(x)}{\int e^{-\frac{(y-x)^2}{2\sigma^2}} \mathrm{d}P_X(x)}, \tag{4.57}$$

and

$$\mathsf{var}\left\{X \,\middle|\, Y = y; \sigma^2\right\} = \mathsf{E}\left\{X^2 \,\middle|\, Y = y; \sigma^2\right\} - \left(\mathsf{E}\left\{X \,\middle|\, Y = y; \sigma^2\right\}\right)^2 \tag{4.58}$$

where $\int \cdot \, \mathrm{d}P_X(x)$ in (4.57) denotes the Riemann-Stieltjes integral.

[3]This is to be distinguished from the notion of neighborhood in the wireless network defined in Section 4.1.3.

Algorithm 1 Message-Passing Decoding Algorithm

Input: $\underline{\boldsymbol{S}}, \boldsymbol{Y}, \gamma_s, M_s, q$.
Initialization:
$z_{\mu k}^{0,i} \leftarrow y_\mu^i/(\sqrt{\gamma_s} s_{\mu k})$ for all $s_{\mu k} \neq 0$ and $i = 1, 2$.
Initialize $\tau^{0,i}$ to a large positive number for $i = 1, 2$.
Main iterations:
for $t = 1$ to $T - 1$ **do**
  for all μ, k with $s_{\mu k} \neq 0$ and $i = 1, 2$ **do**
    $m_{k\mu}^{t,i} \leftarrow \mathsf{E}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu \in \partial k \setminus \mu} z_{\nu k}^{t-1,i}}{|\partial k|-1}; \frac{\tau^{t-1,i}}{|\partial k|-1} \right\}$.
    $(\sigma_{k\mu}^{t,i})^2 \leftarrow \mathsf{var}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu \in \partial k \setminus \mu} z_{\nu k}^{t-1,i}}{|\partial k|-1}; \frac{\tau^{t-1,i}}{|\partial k|-1} \right\}$.
    $z_{\mu k}^{t,i} \leftarrow \frac{1}{\sqrt{\gamma_s} s_{\mu k}} \left(y_\mu^i - \sqrt{\gamma_s} \sum_{j \in \partial \mu \setminus k} s_{\mu j} m_{j\mu}^{t,i} \right)$.
  end for
  $\tau^{t,i} \leftarrow \frac{1}{\sum_\mu |\partial \mu|} \sum_\mu |\partial \mu| \sum_{j \in \partial \mu} (\sigma_{j\mu}^{t,i})^2 + \frac{1}{2\gamma_s} M_s q(1-q)$ for $i = 1, 2$.
end for
$m_k^i \leftarrow \mathsf{E}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu \in \partial k} z_{\nu k}^{T-1,i}}{|\partial k|-1}; \frac{\tau^{T-1,i}}{|\partial k|-1} \right\}$ for all k, $i = 1, 2$.
Output: $\hat{w}_k = \arg\max_{j=1,\ldots,2^l} \left| m_{(k-1)2^l+j}^1 + \sqrt{-1}\, m_{(k-1)2^l+j}^2 \right|$, $k = 1, \ldots, K$.

In the following, we derive Algorithm 1 starting from (4.56) which is valid for both real and imaginary parts in (4.55). It is a simplification of the original iterative BP algorithm, which iteratively computes the marginal *a posteriori* distribution of all symbols given the measurements, assuming that the graph is free of cycles. For each $k \in \partial \mu$ (hence $\mu \in \partial k$), let $\left\{ V_{k\mu}^t(x) \right\}$ represent the message from symbol node k to measurement node μ at the t-th iteration and $\left\{ U_{\mu k}^t(x) \right\}$ represent the message in the reverse direction. Each message is basically the belief (in terms of a probability density (or mass) function) the algorithm has accumulated about the corresponding symbol based on the measurements on the subgraph traversed so far, assuming it is a tree. Let $p_X(x)$ denote the *a priori*

probability density function of X. In the t-th iteration, we compute

$$V_{k\mu}^{t}(x) \propto p_X(x) \prod_{\nu \in \partial k \backslash \mu} U_{\nu k}^{t-1}(x) \tag{4.59a}$$

for all (k, μ) with $s_{\mu k} \neq 0$, and then

$$U_{\mu k}^{t}(x) \propto \int_{(x_j)_{\partial \mu \backslash k}} \exp\left[-\left(y_\mu - \sqrt{\gamma_s} s_{\mu k} x - \sqrt{\gamma_s} \sum_{j \in \partial \mu \backslash k} s_{\mu j} x_j\right)^2 \right] \left(\prod_{j \in \partial \mu \backslash k} V_{j\mu}^{t}(x_j) \mathrm{d}x_j \right) \tag{4.59b}$$

where $\int_{(x_j)_{\partial \mu \backslash k}}$ denotes integral over all x_j with $j \in \partial \mu \backslash k$, and $V(x) \propto u(x)$ means that $V(x)$ is proportional to $u(x)$ with proper normalization such that $\int_{-\infty}^{\infty} V(x) \mathrm{d}x = 1$. In case X is a discrete random variable, the integral shall be replaced by a sum over the alphabet of X. In this problem, X follows a mixture of discrete and continuous distributions, so the expectation can be decomposed as an integral and a sum.

The complexity of computing the integral in (4.59b) is exponential in $|\partial \mu| = \mathcal{O}(qN)$, which is in general infeasible for the problem at hand. However, as $qN \gg 1$, the computation carried out at each measurement node admits a good approximation by using the central limit theorem. A similar technique has been used in the CDMA detection problem, for fully-connected bipartite graph in [**40**, **80**, **81**], and for a graph with large node degrees in [**30**].

To streamline (4.59a) and (4.59b), we introduce $m_{k\mu}^{t}$ and $(\sigma_{k\mu}^{t})^2$ for all (μ, k) pairs with $s_{\mu k} \neq 0$ to represent the mean and variance of a random variable with distribution $V_{k\mu}^{t}(x)$. Using Gaussian approximation, one can reduce the message-passing algorithm to

iteratively computing the following messages:

$$m_{k\mu}^t = \mathsf{E}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu\in\partial k\backslash\mu} z_{\nu k}^{t-1}}{|\partial k|-1}; \frac{\tau^{t-1}}{|\partial k|-1} \right\} \tag{4.60a}$$

$$(\sigma_{k\mu}^t)^2 = \mathsf{var}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu\in\partial k\backslash\mu} z_{\nu k}^{t-1}}{|\partial k|-1}; \frac{\tau^{t-1}}{|\partial k|-1} \right\} \tag{4.60b}$$

$$z_{\mu k}^t = \frac{1}{\sqrt{\gamma_s} s_{\mu k}} \left(y_\mu - \sqrt{\gamma_s} \sum_{j\in\partial\mu\backslash k} s_{\mu j} m_{j\mu}^t \right) \tag{4.60c}$$

$$\tau^t = \frac{1}{\sum_\mu |\partial\mu|} \sum_\mu |\partial\mu| \sum_{j\in\partial\mu} (\sigma_{j\mu}^t)^2 + \frac{1}{2\gamma_s} M_s q(1-q) \tag{4.60d}$$

where (4.60a) and (4.60b) calculate the conditional expectation and variance, respectively. At the T-th iteration, the approximated posterior mean of x_k can be expressed as

$$m_k = \mathsf{E}\left\{ X \,\middle|\, Y = \frac{\sum_{\nu\in\partial k} z_{\nu k}^{T-1}}{|\partial k|-1}; \frac{\tau^{T-1}}{|\partial k|-1} \right\} . \tag{4.61}$$

The detailed derivation of (4.60) from (4.59) is as follows. Denote $\Delta_{\mu k} = \sum_{j\in\partial\mu\backslash k} s_{\mu j} x_j$. The key to the simplification is to recognize that $\Delta_{\mu k}$ is approximately Gaussian. To be precise, if $\{x_j\}_{j\in\partial\mu\backslash k}$ were independent (conditioned on the observations traversed so far on the graph), then, by central limit theorem, $\Delta_{\mu k}$ converges weakly to a Gaussian random variable, whose mean is

$$v_{\mu k}^t = \sum_{j\in\partial\mu\backslash k} s_{\mu j} m_{j\mu}^t \tag{4.62}$$

and variance is

$$(\sigma_{\mu k}^t)^2 = \sum_{j\in\partial\mu\backslash k} s_{\mu j}^2 (\sigma_{j\mu}^t)^2 . \tag{4.63}$$

Using the preceding Gaussian approximation, (4.59b) can be calculated by a change of probability measure as

$$U_{\mu k}^t(x) \propto \int_{-\infty}^{\infty} \exp\left[-\left(y_\mu - \sqrt{\gamma_s} s_{\mu k} x - \sqrt{\gamma_s}\Delta\right)^2\right] \cdot \frac{1}{\sqrt{2\pi(\sigma_{\mu k}^t)^2}} \exp\left[-\frac{1}{2(\sigma_{\mu k}^t)^2}\left(\Delta - v_{\mu k}\right)^2\right] \mathrm{d}\Delta \tag{4.64}$$

$$\propto \exp\left[-\frac{1}{2\tau_{\mu k}^t}(x - z_{\mu k}^t)^2\right] \tag{4.65}$$

where $z_{\mu k}^t$ is defined in (4.60c) and

$$\tau_{\mu k}^t = \sum_{j \in \partial\mu \backslash k} (\sigma_{j\mu}^t)^2 + \frac{1}{2\gamma_s s_{\mu k}^2}\,. \tag{4.66}$$

Using law of large numbers, we further approximate $\tau_{\mu k}^t$ by its average over all (μ, k) pairs with $s_{\mu k} \neq 0$, i.e., $\tau_{\mu k}^t$ is replaced by

$$\tau^t = \frac{1}{\sum_\mu |\partial\mu|} \sum_k \sum_{\mu \in k} \sum_{j \in \partial\mu \backslash k} (\sigma_{j\mu}^t)^2 + \frac{1}{2\gamma_s s_{\mu k}^2} \tag{4.67}$$

$$\approx \frac{1}{\sum_\mu |\partial\mu|} \sum_\mu |\partial\mu| \sum_{j \in \partial\mu} (\sigma_{j\mu}^t)^2 + \frac{1}{2\gamma_s} M_s q(1-q) \tag{4.68}$$

as shown in (4.60d).

Now we have $U_{\mu k}^t \sim \mathcal{N}(z_{\mu k}^t, \tau^t)$, so it is easy to see that

$$\prod_{\nu \in \partial k \backslash \mu} U_{\nu k}^t \sim \mathcal{N}\left(\frac{\sum_{\nu \in \partial k \backslash \mu} z_{\nu k}^t}{|\partial k| - 1}, \frac{\tau^t}{|\partial k| - 1}\right)\,. \tag{4.69}$$

According to (4.59a), we view $V_{k\mu}^{t+1}(x)$ as the conditional distribution $p_{X|Y}\left(x \,\middle|\, \frac{\sum_{\nu\in\partial k\setminus\mu} z_{\nu k}^t}{|\partial k|-1}\right)$, where Y is the output of a Gaussian channel with noise $W \sim \mathcal{N}\left(0, \frac{\tau^t}{|\partial k|-1}\right)$. Therefore, by definition, $m_{k\mu}^{t+1}$ and $(\sigma_{k\mu}^{t+1})^2$ can be expressed as the conditional mean and conditional variance as in (4.60a) and (4.60b).

It is time consuming to compute (4.60a) and (4.60b) for all (μ, k) pairs with $s_{\mu k} \neq 0$, especially in the case of large matrix $\underline{\boldsymbol{S}}$. We can use the following two approximation techniques to further decrease the computational complexity. First, $|\partial k|$ in (4.60a), (4.60b) and (4.61) is replaced by its mean value $M_s q(1-q)$. Second, we use interpolation and extrapolation to further reduce the computation complexity of (4.60a), (4.60b) and (4.61). Specifically, in each iteration t, we only compute the conditional mean and variance for some chosen y's, i.e., we choose $y_1^t < y_2^t < \cdots < y_n^t$ which is a partition of an interval depending on τ^{t-1}, compute

$$a_j^t = \mathsf{E}\left\{X \,\middle|\, Y = y_j^t; \frac{\tau^{t-1}}{M_s q(1-q)-1}\right\}, \tag{4.70}$$

$$b_j^t = \mathsf{var}\left\{X \,\middle|\, Y = y_j^t; \frac{\tau^{t-1}}{M_s q(1-q)-1}\right\} \tag{4.71}$$

for $j = 1, 2, \ldots, n$, and then use those values to calculate (4.60a), (4.60b) by interpolation or extrapolation. To be more precise, for any pair μ, k with $s_{\mu k} \neq 0$, suppose y_j^t and y_{j+1}^t are chosen to be the closest to $y = \frac{\sum_{\nu\in\partial k\setminus\mu} z_{\nu k}^{t-1,i}}{|\partial k|-1}$, then $m_{k\mu}^t$ and $(\sigma_{k\mu}^t)^2$ can be approximated by

$$m_{k\mu}^t = a_j^t + \frac{y - y_j^t}{y_{j+1}^t - y_j^t}(a_{j+1}^t - a_j^t), \tag{4.72}$$

$$(\sigma_{k\mu}^t)^2 = b_j^t + \frac{y - y_j^t}{y_{j+1}^t - y_j^t}(b_{j+1}^t - b_j^t). \tag{4.73}$$

We now revisit the assumption that $\boldsymbol{X}$ has independent elements. In fact, $\boldsymbol{X}$ consists of K sub-vectors of length 2^l, where the entries of each sub-vector are all zero except for one position corresponding to the transmitted message. After obtaining the approximated posterior mean $\widetilde{m}_k$ by incorporating both real and imaginary parts calculated from (4.61), Algorithm 1 outputs the position of the element with the largest magnitude in each of the K sub-vectors of $[\widetilde{m}_1, \ldots, \widetilde{m}_N]$. In fact the factor graph Fig. 4.1 can be modified to include K additional nodes, each of which puts a constraint on one sub-vector. Slight improvement over Algorithm 1 can be obtained by carrying out message passing on the modified graph.

4.4.3. Performance Characterization of Algorithm 1

The iterative decoding dynamics of Algorithm 1 can be quantified in the large-system limit, where $K, M \to \infty$ and the average system load β and all other parameters are held constant. Numerical results in Section 4.5 suggest that the large-system analysis provides a reasonable approximation for the performance of systems of moderate size.

For simplicity, an error event is defined as that node X_0 cannot correctly recover the message from one specific neighbor. Let $\mathsf{P}_{\mathrm{s}}^{\mathrm{e}}$ denote the probability of decoding error averaged over all realizations of all possible messages, signatures and noise. The main result of the large-system analysis is summarized in the following proposition.

Proposition 4.3. *In the large-system limit, with sufficient number of iterations of message-passing decoding described in Algorithm 1, the decoding error probability* P_s^e *is expressed*

by

$$\mathsf{P}_s^e = 1 - \frac{8}{\alpha}(\gamma_s \eta_f \theta)^{\frac{2}{\alpha}} \int_{\sqrt{\gamma_s \eta_f \theta}}^{\infty} \int_0^{\infty} \left(1 - e^{-r^2}\right)^{2^l - 1} e^{-r^2 - x^2} I_0(2xr) r x^{-\frac{4}{\alpha} - 1} \mathrm{d}r \mathrm{d}x \tag{4.74}$$

given that

$$\eta^{-1} = 1 + 2\beta\gamma_s \mathsf{mmse}(P_X, \eta\gamma_s) \tag{4.75}$$

has a unique fixed-point η_f, where $I_0(\cdot)$ is a modified Bessel function of the first kind, and

$$\mathsf{mmse}(P_X, a) = \mathsf{E}\left\{\left(X - \mathsf{E}\left\{X \middle| \sqrt{a}X + W\right\}\right)^2\right\} \tag{4.76}$$

denotes the minimum mean-square error of estimating the input X with the prior distribution $P_X(x)$ in Gaussian noise $W \sim \mathcal{N}\left(0, \frac{1}{2}\right)$.

Proof. It is shown in [**28**, **30**] that when matrix $\underline{\boldsymbol{S}}$ is sparse and satisfies certain conditions, estimating each input symbol given the entire observation $\boldsymbol{Y}^{(i)}, i = 1, 2$ in (4.55) using belief propagation is asymptotically equivalent, in terms of performance, to estimating the same input under a scalar Gaussian channel with some degradation in the SNR, where the degradation factor is determined from an iterative formula. Mathematically, with a fixed number of iterations t, for every i, k and x, the posterior distribution converges in probability as the system becomes large:

$$p^{\mathrm{bp}}_{X_k^{(i)}|\boldsymbol{Y}^{(i)}, \underline{\boldsymbol{S}}}\left(x \middle| \boldsymbol{Y}^{(i)}, \underline{\boldsymbol{S}}\right) \xrightarrow{p} p_{X_k^{(i)}|Z_k^{(i)}}\left(x \middle| Z_k^{(i)}\right) \tag{4.77}$$

where $X_k^{(i)}$ is a random variable representing the real or imaginary part of the k-th input symbol in (4.55), $Z_k^{(i)}$ is the output of a Gaussian channel with input $X_k^{(i)}$, channel gain $\sqrt{\eta^t \gamma_s}$ and noise variance 1/2, i.e., $Z_k^{(i)} \sim \mathcal{N}\left(\sqrt{\eta^t \gamma_s} X_k^{(i)}, \frac{1}{2}\right)$, and $\eta^t, t = 1, 2, \ldots,$ are

determined by the following recursion:

$$\frac{1}{\eta^{t+1}} = 1 + 2\beta\gamma_s \mathsf{mmse}(\eta^t \gamma_s) \tag{4.78}$$

where $\mathsf{mmse}(\cdot)$ is defined in (4.76). Moreover, if $\eta^{-1} = 1 + 2\beta\gamma_s \mathsf{mmse}(\eta\gamma_s)$ has a unique fixed-point η_f, it can be shown that $\lim_{t\to\infty} \eta^t = \eta_f$. If there exist more than one fixed-points, message passing algorithm finds the smallest one [**28**, **30**]. But we have not seen multiple fixed points during our numerical simulation.

Now after a sufficient number of iterations in the message passing algorithm, the channel model (4.55), from the view point of each symbol is approximately equivalent to

$$Z^{(1)} = \sqrt{\gamma_s \eta_f} X^{(1)} + W^{(1)},\ Z^{(2)} = \sqrt{\gamma_s \eta_f} X^{(2)} + W^{(2)} \tag{4.79}$$

where the superscripts (1) and (2) represent the real and imaginary parts respectively, $W^{(i)}, i = 1, 2$ are i.i.d. Gaussian random variables with zero mean and variance $\frac{1}{2}$. By combining two equations in (4.79), we have the corresponding complex model

$$Z = \sqrt{\gamma_s \eta_f} X + W \tag{4.80}$$

where X has the prior distribution $X = 0$ with probability $1 - 2^{-l}$ and $X = |U|e^{\iota\vartheta}$ with probability 2^{-l} with $|U|$ distributed as in (4.8) and ϑ uniformly distributed in $[0, 2\pi)$, W is circularly symmetric complex Gaussian with zero mean and unit variance.

The following lemma regarding the conditional expectation of the input given the output of a complex Gaussian channel is useful. A similar result regarding a real Gaussian channel can be found in [**29**].

Lemma 4.1. *Let $Z = aX + W$ with $a > 0$, where X, W are circularly symmetric complex random variables and further W is Gaussian with zero mean and unit variance. Then the magnitude of the conditional expectation $\mathsf{E}\{X|\, Z\}$ is a non-decreasing function of $|Z|$.*

Proof. For any $z \in \mathbb{C}$, define

$$q_n(z) = \mathsf{E}\left\{X^n e^{-|z-aX|^2}\right\}, n = 0, 1, \ldots . \tag{4.81}$$

Let $Z = re^{\iota\varphi}$, where $r \geq 0$. Since X and $Xe^{-\iota\varphi}$ has the same distribution, we have

$$q_n(re^{\iota\varphi}) = \mathsf{E}\left\{(X^n e^{-|re^{\iota\varphi}-aX|^2}\right\} \tag{4.82}$$

$$= e^{\iota n\varphi}\mathsf{E}\left\{(Xe^{-\iota\varphi})^n e^{-|r-aXe^{-\iota\varphi}|^2}\right\} = e^{\iota n\varphi} q_n(r)\,. \tag{4.83}$$

Thus $|\mathsf{E}\{X|\, Z\}|$ can be expressed as

$$|\mathsf{E}\{X|\, Z\}| = \left|\frac{q_1(re^{\iota\varphi})}{q_0(re^{\iota\varphi})}\right| = \left|\frac{q_1(r)}{q_0(r)}\right|, \tag{4.84}$$

which shows that it is indeed a function of $|Z|$. Next we will prove that $|\mathsf{E}\{X|\, Z\}|$ is non-decreasing as $|Z|$ increases. Equivalently, it suffices to show that the derivative of $q_1^2(r)/q_0^2(r)$ with respect to r is non-negative.

Denote $X = te^{j\phi}$ with $t \geq 0$. ϕ is uniformly distributed in $[0, 2\pi)$ due to the circular symmetry of X. From (4.81), it is easy to see that for $n = 0, 1, \ldots$

$$q_n(r) = \mathsf{E}\left\{t^n \cos(n\phi) e^{-|r-ate^{j\phi}|^2}\right\}. \tag{4.85}$$

Furthermore, we have

$$\frac{\mathrm{d}\, q_n(r)}{\mathrm{d}r} = -\mathsf{E}\left\{t^n \cos(n\phi) e^{-|r-ate^{j\phi}|^2} \left(2r - 2at\cos\phi\right)\right\} \tag{4.86}$$

$$= 2a\mathsf{E}\left\{t^{n+1}\cos(n\phi)\cos\phi e^{-|r-ate^{j\phi}|^2}\right\} - 2rq_n(r)\,. \tag{4.87}$$

Therefore, the derivative of $q_1^2(r)/q_0^2(r)$ with respect to r can be calculated as follows:

$$\frac{2q_0^2(r)q_1(r)\frac{\mathrm{d}\, q_1(r)}{\mathrm{d}r} - 2q_1^2(r)q_0(r)\frac{\mathrm{d}\, q_0(r)}{\mathrm{d}r}}{q_0^4(r)}$$

$$= \frac{4aq_1(r)}{q_0^3(r)}\left(\mathsf{E}\left\{e^{-|r-ate^{j\phi}|^2}\right\}\mathsf{E}\left\{t^2\cos^2\phi e^{-|r-ate^{j\phi}|^2}\right\} - \mathsf{E}^2\left\{t\cos\phi e^{-|r-ate^{j\phi}|^2}\right\}\right) \tag{4.88}$$

By Cauchy-Schwarz inequality, the expression in the parenthesis of (4.88) is non-negative. And we also have

$$q_1(r) = \mathsf{E}\left\{t\cos\phi e^{-|r-ate^{j\phi}|^2}\right\} \tag{4.89}$$

$$= \frac{1}{2\pi}\mathsf{E}\left\{t\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos\phi e^{-|r-ate^{j\phi}|^2}\mathrm{d}\phi + \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}}\cos\phi e^{-|r-ate^{j\phi}|^2}\mathrm{d}\phi\right)\right\} \tag{4.90}$$

$$= \frac{1}{2\pi}\mathsf{E}\left\{t\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos\phi\left(e^{-|r-ate^{j\phi}|^2} - e^{-|r-ate^{j(\phi+\pi)}|^2}\right)\mathrm{d}\phi\right\} \tag{4.91}$$

$$\geq 0 \tag{4.92}$$

where the inequality in (4.92) is due to the fact that $|r-ate^{j\phi}|^2 \leq |r-ate^{j(\phi+\pi)}|^2$ for all $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Therefore, (4.88) is non-negative, which completes our proof. □

At the final iteration of the message passing algorithm, the message from each neighbor is decoded by comparing the complex conditional mean for all corresponding symbols and then choosing the one with the largest magnitude. According to (4.77), it can be

shown that the conditional expectation $\mathsf{E}\{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}\}$ converges in probability to $\mathsf{E}\{X|Z\}$ whose magnitude is a non-decreasing function of $|Z|$ by Lemma 4.1. Therefore, in the large-system limit, we only need to compare the outputs from 2^l independent Gaussian channels to decode the message from each neighbor.

If the input X in (4.80) is zero, $|Z|$ is Rayleigh distributed with parameter $1/\sqrt{2}$. Otherwise denote $a=\sqrt{\gamma_s\eta_f}|U|$, the pdf of $Z=v+\iota w$ in (4.80) can be represented by

$$\frac{1}{\pi}e^{-(v-a\cos\vartheta)^2-(w-a\sin\vartheta)^2}\,. \tag{4.93}$$

Now for a specific neighbor with given channel gain $|U|$, the probability that its message is successfully decoded at X_0 can be calculated as

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}^2}\left(1-e^{-v^2-w^2}\right)^{2^l-1}\cdot\frac{1}{\pi}e^{-(v-a\cos\vartheta)^2-(w-a\sin\vartheta)^2}\mathrm{d}v\mathrm{d}w\mathrm{d}\vartheta\\
&=\frac{1}{2\pi^2}\int_{\mathbb{R}^2}\left(1-e^{-v^2-w^2}\right)^{2^l-1}e^{-v^2-w^2-a^2}\int_0^{2\pi}e^{2a\sqrt{v^2+w^2}\cos\vartheta}\mathrm{d}\vartheta\mathrm{d}v\mathrm{d}w &(4.94)\\
&=\frac{1}{\pi}\int_{\mathbb{R}^2}\left(1-e^{-v^2-w^2}\right)^{2^l-1}e^{-v^2-w^2-a^2}I_0\left(2a\sqrt{v^2+w^2}\right)\mathrm{d}v\mathrm{d}w &(4.95)\\
&=2\int_0^{\infty}\left(1-e^{-r^2}\right)^{2^l-1}e^{-r^2-a^2}I_0(2ar)r\mathrm{d}r &(4.96)
\end{aligned}$$

where (4.95) holds due to the fact that $\int_0^{\pi}e^{\pm\beta\cos x}=\pi I_0(\beta)$. Therefore, the desired error probability $\mathsf{P}_{\mathrm{s}}^{\mathrm{e}}$ can be obtained by averaging (4.96) over $a=\sqrt{\gamma_s\eta_f}|U|$ with $|U|$ distributed as in (4.8):

$$\begin{aligned}
\mathsf{P}_{\mathrm{s}}^{\mathrm{e}}&=1-\frac{8}{\alpha}\theta^{\frac{2}{\alpha}}\int_{\sqrt{\theta}}^{\infty}\int_0^{\infty}\left(1-e^{-r^2}\right)^{2^l-1}e^{-r^2-\gamma_s\eta_f u^2}I_0(2\sqrt{\gamma_s\eta_f}ur)ru^{-\frac{4}{\alpha}-1}\mathrm{d}r\mathrm{d}u &(4.97)\\
&=1-\frac{8}{\alpha}(\gamma_s\eta_f\theta)^{\frac{2}{\alpha}}\int_{\sqrt{\gamma_s\eta_f\theta}}^{\infty}\int_0^{\infty}\left(1-e^{-r^2}\right)^{2^l-1}e^{-r^2-x^2}I_0(2xr)rx^{-\frac{4}{\alpha}-1}\mathrm{d}r\mathrm{d}x\,, &(4.98)
\end{aligned}$$

which completes the proof of Proposition 4.3. □

Proposition 4.3 is basically a single-letter characterization of the performance in the large-system limit. That is, as far as the error probability of an individual neighbor is concerned, the mutual broadcast system with message-passing decoding is asymptotically equivalent to a scalar Gaussian channel with some degradation in the SNR, where the degradation factor is determined from fixed-point equation (4.75).

4.5. Numerical Results

In order for a fair comparison, we assume the same power constraint for both the sparse recovery scheme and random-access schemes, i.e., the average transmit power in each active slot (in which the node transmits energy) is the same. Recall that, in the proposed scheme, where the frame length is M_s, the SNR in the model (4.52) is $\gamma_s = \gamma M_s q(1-q)$. We choose the same transmission probability in each slot for sparse recovery scheme and slotted ALOHA, i.e., $q = p = 1/(c+1)$. Also, the transmission probability in each slot for CSMA is $(1-e^{-c})/c$ (see Section 4.2.2), which is close to $1/(c+1)$ when c is large. The three schemes consume approximately the same amount of average power over any period of time.

Without loss of generality, let one unit of distance be 1 meter. Consider a wireless network of 1000 nodes uniformly distributed in a square with side length 500 meters. The nodes form a Poisson point process in the square conditioned on the node population. Suppose the path-loss exponent $\alpha = 4$. The threshold of channel gain to define neighborhood is set to $\theta = 10^{-6}$. It means that if the transmit power for a node one meter away is 60 dB, then the SNR attenuates to 0 dB at $= 10^{6/\alpha} \approx 31$ meters in the absence of fading,

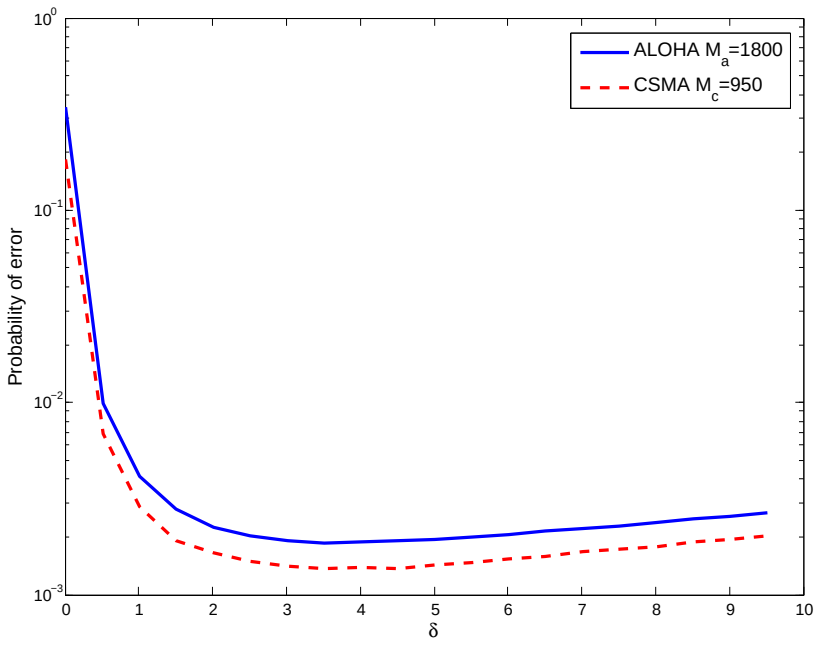

Figure 4.2. Low bounds for error probability in slotted-ALOHA and CSMA for different threshold δ in the case of $l = 10$.

i.e., the coverage of the neighborhood of a node is typically a circle of radius 31 meters. According to (4.12), a node near the center of the square (without boundary effect) has on average $c \approx 11$ neighbors.

We consider two cases for the length of broadcast message $l = 5$ and 10 bits. In random access schemes, a packet of L bits consists of the l-bit message and $\lceil \log_2 c \rceil$ additional bits to identify the sender. Fig. 4.2 shows that $T = 3.5$ minimizes the lower bounds for $\mathsf{P}_{\mathrm{a}}^{\mathrm{e}}$ in (4.16) and $\mathsf{P}_{\mathrm{c}}^{\mathrm{e}}$ in (4.33) in the case of $l = 10$.

The metric for performance comparison is the probability for one node to miss one specific neighbor, averaged over all pairs of neighboring nodes in the network. Suppose the transmit power of each node is $\gamma = 60$ dB. First consider one realization of the network where each node has $c \approx 11$ neighbors on average and $l = 5$ bits to broadcast,

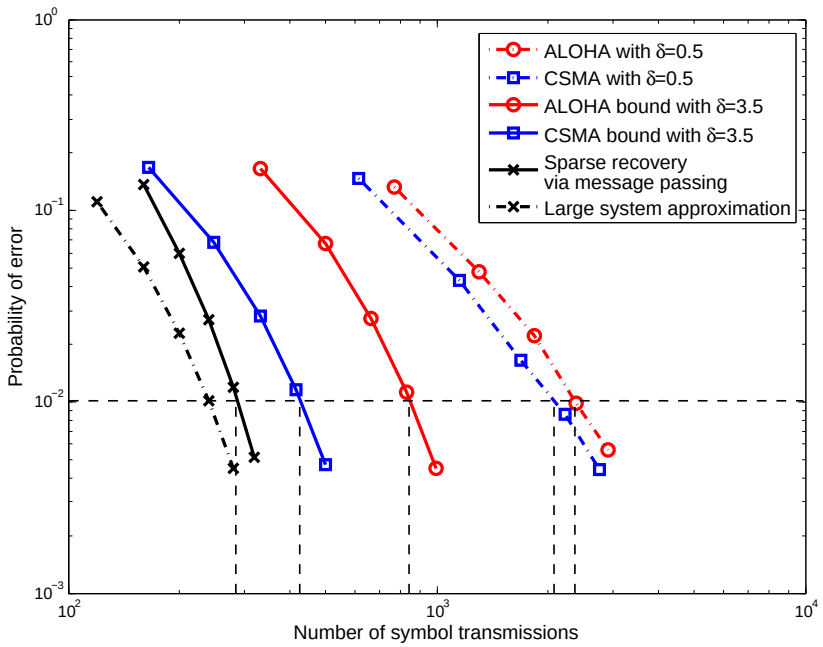

Figure 4.3. Performance comparison between sparse recovery and random access. Each node transmits a 5-bit message.

so that on average $cl \approx 55$ bits are to be collected by each node. In random-access schemes, at least 4 additional bits are needed to identify a sender out of 1000 nodes, so we let $L = 9$. In Fig. 4.3, the error performance of slotted ALOHA and CSMA for $\delta = 0.5$ is compared with that of the sparse recovery scheme with the message-passing algorithm. The simulation result shows the sparse recovery scheme significantly outperforms slotted ALOHA and CSMA, even compared with the minimum of the lower bounds computed from (4.16) and (4.33) for $\delta = 3.5$. For example, to achieve 1% error rate, the sparse recovery scheme takes fewer than 300 symbols. Slotted ALOHA and CSMA take no less than 800 and 400 symbols according to the bounds in (4.16) and (4.33), respectively. In fact, slotted ALOHA and CSMA with threshold $\delta = 0.5$ take more than 2000 symbols. Similar comparison is observed for several other SINR thresholds δ around

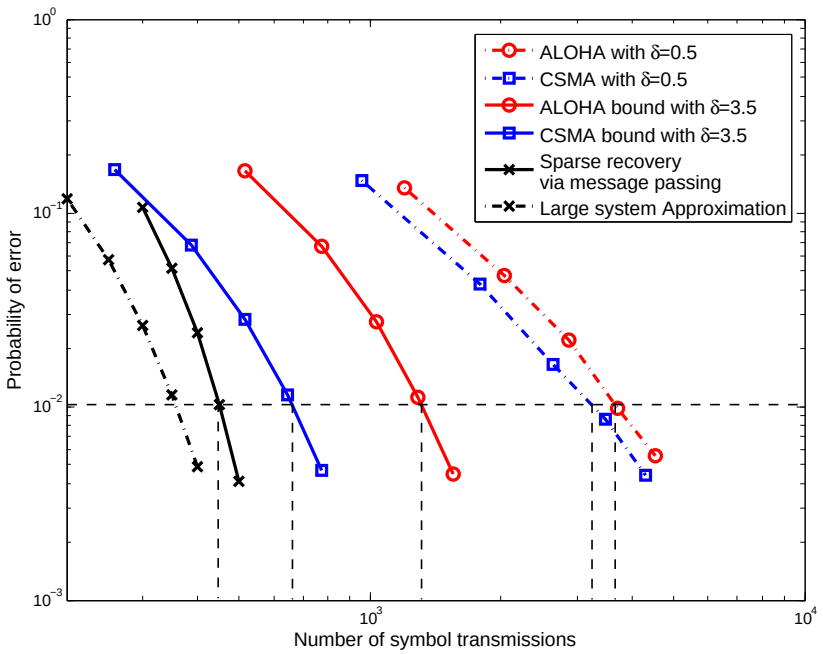

Figure 4.4. Performance comparison between sparse recovery and random access. Each node transmits a 10-bit message.

0.5 and the performance of ALOHA and CSMA are not good for $\delta \geq 1$ because the messages from weaker neighbors may never be successfully delivered. Some additional supporting numerical evidence is, however, omitted due to space limitations. We also make a comparison between the simulation result and the large system approximation from (4.74) for the message-passing algorithm. The approximation is seen to provide a reasonably good characterization of the performance of the message-passing algorithm.

Fig. 4.4 repeats the experiment of Fig. 4.3 with 10-bit messages. The sparse recovery scheme has significant gain compared with slotted ALOHA and CSMA. For example, to achieve the error rate of 1%, the sparse recovery scheme takes about 450 symbols, whereas

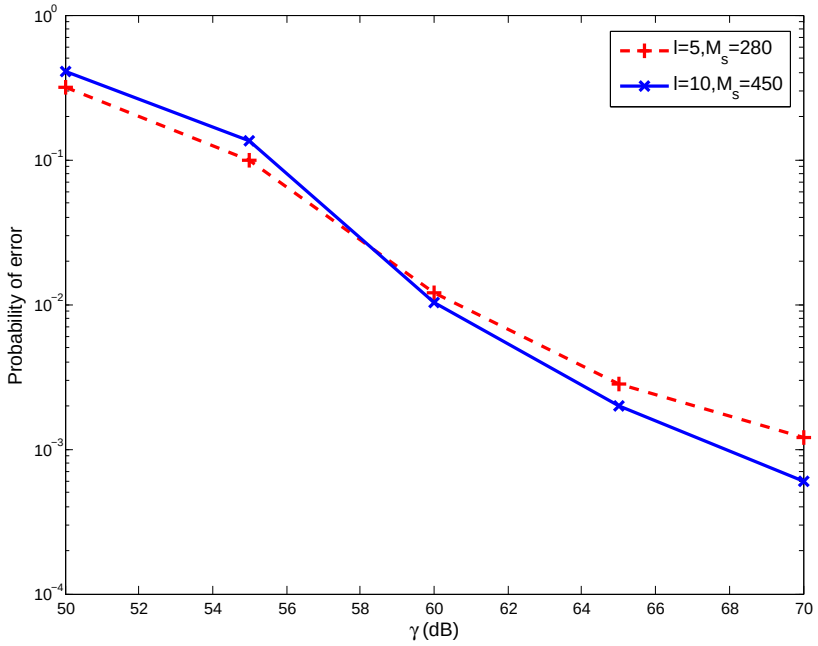

Figure 4.5. Performance of sparse recovery scheme in different nominal SNR (γ).

slotted ALOHA and CSMA take at least 1000 and 650 symbols, respectively. The large-system approximation also gives the trend of the performance of the message-passing algorithm.

In Fig. 4.5, we simulate the same network with different nominal SNRs, i.e., γ varies from 50 dB to 70 dB. In the case that each node transmits a 5-bit message, the frame length is chosen to be 280 symbols. It can be seen from the figure that the probability of error decreases with the increase of SNR. The performance is similar when each node transmits a 10-bit message and the frame consists of 450 symbols.

4.6. Summary

In this chapter, a novel solution based on RODD signaling is proposed to achieve mutual broadcast in wireless networks. Decoding can be viewed as a problem of compressed sensing. In the case that each message consists of a small number of bits, we have developed an iterative message-passing algorithm based on belief propagation, and characterized its performance using a state evolution formula in the limit where each node has a large number of peers. In a network consisting of Poisson distributed nodes with the same transmit power, numerical results demonstrate that, to achieve the same reliability for all nodes to broadcast their messages, the proposed sparse recovery scheme achieves several times the rate of slotted ALOHA and CSMA. The results in this chapter have been published in part in [**93**].

CHAPTER 5

Virtual Full-Duplex Neighbor Discovery

In many wireless networks, each node has direct radio link to only a small number of other nodes, called its neighbors (or peers). Before efficient routing or other network-level activities are possible, nodes have to discover and identify the network interface addresses of their neighbors. This is called *neighbor discovery (or peer discovery)*. The problem is crucial in mobile ad hoc networks (MANETs), which are self-organizing networks without pre-existing infrastructure. The problem is becoming important in increasingly more heterogeneous cellular networks with the deployment of unsupervised picocells and femtocells.

A node interested in its neighborhood, which is henceforth referred to as the *query node*, listens to the wireless channel during the discovery period, and then decodes the NIAs of its neighbors. Neighbors transmit signals which contain their identity information. It is fair to assume that non-neighbors either do not transmit, or their signals are weak enough to be regarded as noise. We make two important observations: 1) the physical channel is a multiaccess channel, where the observation made by the query node is a (linear) superposition of transmissions from its neighbors, corrupted by noise; 2) the goal of neighbor discovery is to identify, out of all valid NIAs, which ones are used by its neighbors.

State-of-the-art neighbor discovery protocols, such as that of the IETF MANET working group [**1**] and the ad hoc mode of IEEE 802.11 standards, can be described as follows:

The query node broadcasts a probe request. Its neighbors then reply with probe response frames containing their respective NIAs. If a response frame does not collide with any other frame, the corresponding NIA is correctly received. Due to lack of coordination, each neighbor has to retransmit its NIA enough times with random delays, so that it can be successfully received by the query node with high probability despite collisions. We refer to such a scheme as *random-access neighbor discovery*. Several such algorithms which operate in or on top of MAC layer have been proposed [**11**, **23**, **45**, **57**, **83**, **84**].

Random access assumes a specific signaling format, namely, a node's response over the discovery period basically consists of repetitions of its NIA interleaved with periods of silence. This signaling format allows the NIA to be directly read out from a successfully received frame. Every node can discover its neighborhood and also be discovered by neighbors given long-enough discovery period. However, such signaling is far from optimal. To design the optimal signaling, we should remove all unnecessary structural restrictions on the responses. Given the duration of the discovery period, the problem is in general to assign each node a distinct response, or *signature* over that period, and to design a decoding algorithm for a query node to identify the constituent signatures (or corresponding NIAs) based on the observed superposition. It would be ideal if all the signatures were orthogonal to each other, but this is impossible in the case that the number of signatures far exceeds the signature length. A good design should make the correlation between any pair of signatures as small as possible.

A crucial observation is that the number of actual neighbors is typically orders of magnitude smaller than the node population, or more precisely, the size of the NIA space, so that neighbor discovery is by nature a compressed sensing (or sparse recovery)

problem [**13**, **19**]. By the wisdom from the compressed sensing literature, the required number of measurements (the signature length) is dramatically smaller than the size of the NIA space.

Based on the preceding observations, this chapter provides a novel solution using RODD signaling, referred to as *compressed neighbor discovery*, which attains highly desirable trade-off between reliability and the length of the discovery period, thus minimizing the neighbor discovery overhead in wireless networks. The defining feature is to let nodes simultaneously transmit their on-off signatures within a single frame interval. Therefore, within the discovery frame a node can make observations during its off-slots and also transmit during its on-slots, i.e., virtual full-duplex neighbor discovery is achieved.

It is proposed in [**52**, **53**] that each node is assigned a pseudo-random on-off signature and uses a simple decoding procedure via eliminating non-neighbors for discovering the neighborhood. The complexity of the decoding algorithm is linear in the size of the address space, which is only feasible for networks with moderately large but not too large NIA spaces. In this chapter, we generate a set of deterministic on-off signatures based on a second-order Reed-Muller (RM) code. First- and second-order RM codes date back to 1950s and are fundamental in the study of error-control codes and algorithms [**79**]. More recently, RM codes have been shown to be excellent for sparse recovery [**35**]. The choice of modified RM codes for neighbor discovery is not incidental: The algebraic structure allows unusually low decoding complexity (sublinear in the number of codewords), so that the scheme is in principle scalable to 2^{48} or more nodes or NIAs in the network.

The organization of the remaining sections of this chapter is as follows. The system model is presented in Section 5.1. In Section 5.2, the generation of the deterministic

signatures based on a second-order RM code is described. The original RM code consists of quadrature phase-shift keying (QPSK) symbols, with no off-slots. In order to achieve full-duplex discovery, we introduce off-slots by replacing roughly a half of the QPSK symbols by zeros. The chirp decoding algorithm of [**35**] is modified to perform despite the erasures. In Section 5.3, compressed neighbor discovery is compared with random-access schemes and shown to require much fewer transmissions to achieve the same error performance. In addition, the new scheme entails much less transmission overhead (such as preambles and parity checks), because it takes a single frame of transmission, as opposed to many frame transmissions in random access. Section 5.4 summarizes the chapter.

5.1. The Channel and Network Models

5.1.1. The Linear Channel

Consider a wireless network where each node is assigned a unique network interface address. Let the address space be $\{0, 1, \ldots, N\}$ (e.g., $N = 2^{48} - 1$ if the space consists of all IEEE 802.11 MAC addresses). The actual number of nodes present in the network can be much smaller than N, but as far as neighbor discovery is concerned, we shall assume that there are exactly $N + 1$ nodes.

We will later discuss the problem of having all nodes simultaneously discover their respective neighborhoods, but for now let us assume that node 0 is the only query node and sends a probe signal to prompt a neighbor discovery period of M symbol intervals. Each node n in the neighborhood responds by sending a signal $\boldsymbol{S}_n = [S_{1n}, \ldots, S_{Mn}]^\top$. The signal identifies node n and is also referred to as the signature of node n. In case a node only transmits over selected time instances, those symbols S_{mn} corresponding to

non-transmissions are regarded as zeros. For the time being let us ignore the variation of the small propagation delays between the query node and its neighbors, and assume symbol-synchronous transmissions from all nodes. We also assume that the discovery period is shorter than the channel coherence time. The received signal of node 0 can thus be expressed as

$$\boldsymbol{Y} = \sqrt{\gamma} \sum_{n \in \mathcal{N}_0} U_n \boldsymbol{S}_n + \boldsymbol{W} \tag{5.1}$$

where $\mathcal{N}_0$ denotes the set of NIAs in the neighborhood of node 0, U_n denotes the complex-valued coefficient of the wireless link from node n to node 0, γ denotes the average channel gain in the SNR, and $\boldsymbol{W}$ consists of M independent unit circularly symmetric complex Gaussian random variables, with each entry $W_m \sim \mathcal{CN}(0, 1)$. For simplicity, transmissions from non-neighbors, if any, are accounted for as part of the additive Gaussian noise.

The goal is to recover the set $\mathcal{N}_0$, given the observation $\boldsymbol{Y}$, the SNR γ, and knowledge of the signatures $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N$. The random coefficients U_n are unknown except for its statistics. For convenience, we introduce binary variables B_n, which is set to 1 if node n is a neighbor of node 0, and set to 0 otherwise. Let $\boldsymbol{X} = [B_1 U_1, \ldots, B_N U_N]^\top$ and $\underline{\boldsymbol{S}} = [\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N]$. Then model (5.1) can be rewritten as

$$\boldsymbol{Y} = \sqrt{\gamma} \underline{\boldsymbol{S}} \boldsymbol{X} + \boldsymbol{W} \tag{5.2}$$

where we wish to determine which entries of $\boldsymbol{X}$ are nonzero, i.e., to recover the support of $\boldsymbol{X}$.

Model (5.2) represents a familiar noisy linear measurement system. We shall refer to $\boldsymbol{Y} = [Y_1, \ldots, Y_M]^\top$ as the measurements, and $\underline{\boldsymbol{S}}_{M \times N}$ as the known signature matrix.

It is reasonable to assume that $B_1, \ldots, B_N$ are i.i.d. Bernoulli random variables with $\mathsf{P}\{B_1 = 1\} = c/N$, where c denotes the average number of neighbors of node 0. Let us further assume that $U_1, \ldots, U_N$ are i.i.d. with known distribution, and are independent of $B_1, \ldots, B_N$ and noise. To recover the support of $\boldsymbol{X}$ is then a well-defined, familiar statistical inference problem.

The node population $N+1$ is typically much larger than the number of symbol epochs in one discovery period M, so that the linear system (5.2) is under-determined even in the absence of noise. An important observation is that the vector variable $\boldsymbol{X}$ is very sparse, so that neighbor discovery is fundamentally a sparse recovery problem, which implies that very few measurements, which can be orders of magnitude smaller than N, are sufficient for reconstructing the N-vector $\boldsymbol{X}$ or its support [**88**].

5.1.2. Propagation Delay and Synchronicity

In general, a receiver has to resolve the timing uncertainty of its neighbors in order to recover their identities. By including sufficient synchronization flags, random-access schemes are robust with respect to abitrary delays. Since it is costly to add enough redundancy to allow accurate estimation of the delays in a multiuser environment, it can be beneficial to let nodes transmit their signatures simultaneously and synchronously. Some common clock, such as access to the GPS can provide the timing needed. In our scheme, it suffices to have all communicating peers be approximately symbol-synchronized, as long as the timing difference (including the propagation delay) is much smaller than the symbol interval. This can be achieved by using distributed algorithms for reaching average consensus [**76**].

By definition neighbors should be physically close to the query node, so that the radio propagation delay is much smaller compared to a symbol epoch. For instance, if neighbors are within 300 meters, the propagation delay is at most 1 microsecond, which is much smaller than the bit or pulse interval of a typical MANET. More pronounced propagation delays can also be explicitly addressed in the physical model, but this is out of the scope of this chapter.

5.1.3. Propagation Loss and Near-Far Problem

Previous work [**53**] considered a single query node and neighbors of the same distance, and simply assumed the channel gains U_n to be Rayleigh fading random variables. Here, we incorporate the effect of network topology and propagation loss in the channel model. Suppose all nodes transmit at the same power, large-scale attenuation follows power law with path loss exponent α, and small-scale attenuation follows i.i.d. fading. Due to reciprocity, the gains of the two directional links between any pair of nodes are identical.

From the viewpoint of a query node, it suffices to describe the statistics of U_n of neighboring nodes in model (5.1) as follows. Suppose all nodes are distributed in a plane according to a homogeneous Poisson point process with intensity λ. Consider a uniformly and randomly selected pair of nodes. The channel power gain between them is $GR^{-\alpha}$, where G denotes small-scale fading and R stands for the distance between them. The nodes are called neighbors of each other if the channel gain between them exceeds a certain threshold, i.e., $GR^{-\alpha} > \theta$ for some fixed threshold θ. We choose not to define the neighborhood purely based on the geometrical closeness because: 1) connectivity between

a pair of nodes is determined by the channel gain; and 2) a receiver cannot separate the attenuations due to path loss and Rayleigh fading in one discovery period.

Consider an arbitrary neighbor, n, of the query node. It is assumed that the phase of U_n is uniform on $[0, 2\pi)$. By following the same steps as in Section 4.1.3, the pdf of $|U_n|$ is expressed as

$$p(u) = \begin{cases} \frac{4}{\alpha}\frac{\theta^{2/\alpha}}{u^{4/\alpha+1}}, & u \geq \sqrt{\theta}; \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

Moreover, the average number of neighbors the query node has is the same as in (4.12), i.e., $c = \frac{2}{\alpha}\pi\lambda\theta^{-2/\alpha}\Gamma(\frac{2}{\alpha})$.

The near-far situation, namely that some neighbors can be much stronger than others is inherently modeled in (5.1)-(5.3). The proposed sparse recovery algorithm is highly resilient to the near-far problem. To be specific, the gain of strong neighbors can be estimated quite accurately so that their interference to weaker neighbors can be removed.

5.1.4. Network-wide Discovery

Unlike in previous work [**52**, **53**], we also considers the problem that many or all nodes in the network need to discover their respective neighborhoods at the same time. A major challenge is posed by the half-duplex constraint.

A random-access scheme naturally supports network-wide discovery. This is because each node transmits its NIA intermittently, so that it can listen to the channel to collect neighbors' NIAs during its own epochs of non-transmission. Collision is inevitable, but

if each node repeats its NIA a sufficient number of times with enough (random) spacing, then with high probability it can be received by every neighbor once without collision.

As we shall see in Section 5.2, the proposed compressed neighbor discovery scheme employs on-off signatures, so that a node can make observations during its own off-slots. All nodes broadcast their signatures and discover their respective neighbors at the same time. Thus network-wide discovery is achieved within a single frame interval.

5.2. On-off Reed-Muller Signatures and Chirp Decoding

In this section, we propose to use deterministic signatures obtained from second-order Reed-Muller codes with erasures, where the complexity of the corresponding chirp decoding algorithm is sub-linear in N. We first discuss the original RM code without erasure. Such a code is sufficient for a single silent query node to acquire its neighborhood. The construction of the RM code is described in detail in [**12**]. We provide a sketch of the construction in Section 5.2.1. The signatures consist of QPSK entries, which prevent a transmitting node from simultaneously discovering its neighborhood. In Section 5.2.2, zero entries are introduced by erasing about 50% of the symbols in each signature, so that full-duplex neighbor discovery is enabled. The chirp decoding algorithm is discussed in Section 5.2.3. As we shall see in Section 5.2.4, using the Reed-Muller code enables more reliable and efficient discovery in networks which are many orders of magnitude larger than allowed by using random on-off signatures as in [**52**, **53**].

For the reader's convenience, the signature generation and chirp decoding procedures are summarized as Algorithms 3 and 4. Examples in the case of very small systems are given to illustrate the encoding and decoding procedures.

5.2.1. The Reed-Muller Code (without Erasure)

RM codes are a family of linear error-control codes. A formal description of RM codes requires a substantial amount of preparation in finite fields. In a general form, RM codes are based on evaluating certain primitive polynomials in finite fields. Due to space limitations, we briefly describe the second-order RM codes used in this chapter using the minimum amount of formalisms. The reader is referred to [**12**] for a more detailed discussion.

Given a positive integer m, we show how to generate up to $2^{m(m+3)/2}$ distinct codewords, each of length 2^m. For example, in the case of $m = 10$, there are up to 2^{65} codewords of length 1,024.

Let $e_l^i = [0, \ldots, 0, 1, 0, \ldots, 0]$ be a row vector of length l in which the i-th entry is equal to 1 whereas all other entries are zeros. Let $\boldsymbol{P}(e_l^i)$ be the $l \times l$ symmetric matrix in which the top row is e_l^i and each of the remaining reverse diagonals (a diagonal from upper right to lower left) is computed from a fixed linear combination of the entries in the top row. The reader is referred to [**12**] for a detailed description of the construction, which is based on evaluating some primitive polynomials in GF(2^m). For example, $\boldsymbol{P}(e_1^1) = 1$ and

$$\boldsymbol{P}(e_2^1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \boldsymbol{P}(e_2^2) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}. \tag{5.4}$$

Given m, we form a linear space of $m \times m$ symmetric matrices with a set $\boldsymbol{B}$ of $m(m+1)/2$ bases constructed using $\{\boldsymbol{P}(e_l^i), i \leq l, l = 1, \ldots, m\}$, where for $l < m$, $\boldsymbol{P}(e_l^i)$ is padded to an $m \times m$ matrix, where the lower right $l \times l$ submatrix is $\boldsymbol{P}(e_l^i)$ and all remaining entries are zeros. In the simple case of $m = 2$, $\boldsymbol{B}$ consists of $m(m+1)/2 = 3$

bases, which are $\boldsymbol{P}(e_2^1)$, $\boldsymbol{P}(e_2^2)$ given by (5.4) and an additional matrix obtained from $\boldsymbol{P}(e_1^1) = 1$ by padding zeros:

$$\boldsymbol{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}. \tag{5.5}$$

Let $\boldsymbol{B}(i)$ denote the i-th basis in $\boldsymbol{B}$ ordered as $\boldsymbol{P}(e_m^1), \ldots, \boldsymbol{P}(e_m^m)$ and then those obtained from $\boldsymbol{P}(e_{m-1}^1), \ldots, \boldsymbol{P}(e_{m-1}^{m-1})$ and so on.

Let the NIA consist of $n = n_1 + n_2$ bits, where $n_1 \leq m$ and $n_2 \leq m(m+1)/2$. Each n-bit NIA is divided into two binary vectors: $\boldsymbol{b}' \in \mathbb{Z}_2^{n_1}$ and $\boldsymbol{c} \in \mathbb{Z}_2^{n_2}$, where $\mathbb{Z}_2 = \{0, 1\}$. Let $\boldsymbol{b} \in \mathbb{Z}_2^m$ be formed by appending $m - n_1$ zeros after $\boldsymbol{b}'$ ($\boldsymbol{b} = \boldsymbol{b}'$ if $n_1 = m$). We map $\boldsymbol{c}$ to an $m \times m$ symmetric matrix according to

$$\boldsymbol{P}(\boldsymbol{c}) = \sum_{i=1}^{n_2} c_i \boldsymbol{B}(i) \mod 2 \tag{5.6}$$

where c_i denotes the i-th bit of $\boldsymbol{c}$. The corresponding codeword is of 2^m symbols, whose entry indexed by $\boldsymbol{a} \in \mathbb{Z}_2^m$ is given by

$$\phi_{\boldsymbol{b},\boldsymbol{c}}(\boldsymbol{a}) = \exp\left[j\pi \left(\frac{1}{2} \boldsymbol{a}^\mathsf{T} \boldsymbol{P}(\boldsymbol{c}) \boldsymbol{a} + \boldsymbol{b}^\mathsf{T} \boldsymbol{a} \right) \right]. \tag{5.7}$$

For example, in the case $m = 2$, there are up to $2^{m(m+3)/2} = 32$ codewords of length $2^m = 4$. Moreover, if the number of nodes is 16, i.e., $n = 4$ only 16 codewords are generated as functions of $(\boldsymbol{b}, \boldsymbol{c})$ and given as column vectors in Table 5.1, where only the first two bases in (5.5) are used as $n_1 = n_2 = 2$.

Table 5.1. 16 Reed-Muller codewords.

$\boldsymbol{b}$	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
$\boldsymbol{c}$	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
$\phi_{\boldsymbol{b},\boldsymbol{c}}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	-1	1	-1	j	$-j$	j	$-j$	j	$-j$	j	$-j$	1	-1	1	-1
	1	1	-1	-1	1	1	-1	-1	j	j	$-j$	$-j$	j	j	$-j$	$-j$
	1	-1	-1	1	$-j$	j	j	$-j$	-1	1	1	-1	$-j$	j	j	$-j$

5.2.2. Generation of On-Off Signatures

The drawback of using the original RM code is that the codewords defined by (5.7) consist of QPSK symbols, so that a node cannot simultaneously receive useful signals while transmitting its own codeword. In order to achieve full-duplex neighbor discovery, we propose to erase about 50% of the entries of each codeword to obtain an on-off signature, so that nodes can listen during their own off-slots. The signature of each node consists of roughly as many off-slots as on-slots, thus two nodes can receive pulses from each other over about 25% of the slots.

For reasons to be explained shortly in conjunction with the chirp decoding algorithm, we apply random erasures to the signatures in the following simple manner: Suppose n_2 is chosen such that the $m \times m$ symmetric matrix generated by each node is determined by its first $m_0 \leq m/2$ rows. For node k, the erasure pattern $\boldsymbol{r}_k$ of length 2^m is constructed as follows: Divide $\boldsymbol{r}_k$ into 2^{m_0} segments with equal length 2^{m-m_0}, let the first segment consist of i.i.d. Bernoulli random variables with parameter $1/2$ and all remaining segments be identical copies of the first segment. It is easy to see that after introducing erasures in the signatures, the network can still accommodate $2^{m(3m+10)/8}$ nodes. For example, if $m = 10$, we have up to 2^{50} signatures of length 1,024.

The procedure for generating the on-off signatures based on the RM code is summarized as Algorithm 2.

Algorithm 2 Signature Generation Algorithm

1: *Input:* n-bit NIA
2: Choose m such that $n = n_1 + n_2$ with $n_1 \leq m$ and $n_2 \leq \frac{m_0}{2}(2m - m_0 + 1)$ where $m_0 \leq m/2$.
3: Divide n-bit NIA into two vectors $\boldsymbol{b}' \in \mathbb{Z}_2^{n_1}$ and $\boldsymbol{c} \in \mathbb{Z}_2^{n_2}$. Form $\boldsymbol{b} \in \mathbb{Z}_2^m$ by appending $m - n_1$ zeros after $\boldsymbol{b}'$.
4: Generate the original RM code $\phi_{\boldsymbol{b},\boldsymbol{c}}$ of length 2^m according to (5.7).
5: Generate the erasure pattern $\boldsymbol{r}$ of length 2^m as follows: Let the first segment of 2^{m-m_0} bits be i.i.d. Bernoulli random variables with parameter $1/2$ and repeat the segment 2^{m_0} times to form the 2^m bits of $\boldsymbol{r}$.
6: *Output:* The on-off signature of length 2^m is the element-wise product of $\phi_{\boldsymbol{b},\boldsymbol{c}}$ and $\boldsymbol{r}$.

5.2.3. The Chirp Decoding Algorithm

We recall that each node makes observations via the multiaccess channel (5.1), which is a superposition of its neighbors' signatures subject to fading and noise. An iterative chirp decoding algorithm has been developed in [**35**] to identify the codewords of the RM code based on their noisy superposition. The general idea is to take the Hadamard transform of the auto-correlation of the signal in each iteration to expose the coefficient of the digital chirps and then cancel the discovered signatures from the signal.

In the case of full-duplex discovery, the original chirp decoding algorithm with some modifications can be applied here for any node (say, node 0) to recover its neighborhood based on the observations through its own off-slots (denoted as $\tilde{\boldsymbol{Y}}$). The details are provided in Algorithm 3.

In the following, we provide a simple example to illustrate the key steps of Algorithm 3. Consider a network of $N = 2^n = 1{,}024$ nodes. Let the parameters in Algorithm 2 be

Algorithm 3 The chirp decoding algorithm

Input: received signal $\boldsymbol{Y}$ in (5.2), signatures of all other nodes $\underline{\boldsymbol{S}}$ and its own erasure pattern $\boldsymbol{r}$.
Choose the maximum iteration number $T_{\max}$, the threshold η and the maximum number n_0 of weak nodes discovered till termination.
Initialize the residual signal $\boldsymbol{Y}_r$ to the pointwise product of $\boldsymbol{Y}$ and $1 - \boldsymbol{r}$.
Initialize the iteration number t to 0, the neighbor set $\boldsymbol{N} = \emptyset$ and the coefficient vector $\boldsymbol{C} = \emptyset$.
Main iterations:
while $t \leq T_{\max}$ **do**
  for $i = 1, 2, \ldots, m_0$ **do**
    Compute the pointwise multiplication of the conjugate of $\boldsymbol{Y}_r$ and the shift of $\boldsymbol{Y}_r$ in the amount of 2^{m-i}.
    Compute the fast Walsh-Hadamard transform of the computed auto-correlation.
    Find the position of the highest peak in the frequency domain and decode the i-th row of an $m \times m$ matrix $\boldsymbol{P}(\boldsymbol{c}_k)$, which corresponds to a certain node k.
  end for
  Use the first m_0 rows of the preceding $\boldsymbol{P}(\boldsymbol{c}_k)$ to determine its remaining rows.
  Compute for all $\boldsymbol{a} \in \mathbb{Z}_2^m$ and apply Hadamard transform to the pointwise product of $\boldsymbol{Y}_r$ and the conjugate of $\boldsymbol{S}_k^0$;
  Recover $\boldsymbol{b}_k$ by finding the highest peak in the frequency domain.
  Compute $\phi_{\boldsymbol{b}_k, \boldsymbol{c}_k}$ according to (5.7) and recover $\boldsymbol{S}_k$ by pointwise product of $\phi_{\boldsymbol{b}_k, \boldsymbol{c}_k}$ and $\boldsymbol{r}_k$.
  Add node k to the neighbor set $\boldsymbol{N}$ and add a corresponding 0 to the coefficient vector $\boldsymbol{C}$.
  Put together all signatures of nodes in $\boldsymbol{N}$ to form a matrix $\underline{\boldsymbol{S}}_{\boldsymbol{N}}$. Construct $\tilde{\underline{\boldsymbol{S}}}$ by pointwise multiplying each column in $\underline{\boldsymbol{S}}_{\boldsymbol{N}}$ with $1 - \boldsymbol{r}$.
  Determine the value of vector $\boldsymbol{X}$ which minimizes $\|\boldsymbol{Y}_r - \tilde{\underline{\boldsymbol{S}}}\boldsymbol{X}\|_2$. Update the coefficient vector $\boldsymbol{C}$ by $\boldsymbol{C} + \boldsymbol{X}$.
  Update the residual signal $\boldsymbol{Y}_r$ by $\boldsymbol{Y}_r - \tilde{\underline{\boldsymbol{S}}}\boldsymbol{X}$.
  if $\boldsymbol{N}$ contains more than n_0 nodes with coefficients less than η **then**
    Stop the main iteration.
  end if
end while
Output: All elements in $\boldsymbol{N}$ whose corresponding coefficients in $\boldsymbol{C}$ are no less than η.

$n_1 = 5$, $n_2 = 5$, $m = 5$, $m_0 = 1$, so that we have 1,024 signatures of length $2^m = 32$.

Suppose for simplicity node 0 has only two neighbors, whose on-off signatures are $\boldsymbol{S}_1$ and

$\boldsymbol{S}_2$, respectively:

$$\begin{aligned}\boldsymbol{S}_1 = [&1,0,0,1,1,1,0,1,0,j,0,j,0,j,-j,j,\\ &-j,0,0,-j,j,j,0,j,0,1,0,1,0,-1,1,-1]\end{aligned} \tag{5.8}$$

$$\begin{aligned}\boldsymbol{S}_2 = [&0,0,j,0,0,-j,1,0,0,j,-1,0,0,0,0,0,\\ &0,0,-j,0,0,-j,1,0,0,-j,1,0,0,0,0,0]\end{aligned} \tag{5.9}$$

where the zeros in the signatures are due to erasures. Suppose the channel gains are $U_1 = 3$ and $U_2 = 2j$. In the absence of noise, node 0 observes the signal $U_1\boldsymbol{S}_1 + U_2\boldsymbol{S}_2$ through its own off-slots as

$$\begin{aligned}\tilde{\boldsymbol{Y}} = [&3,0,0,3,0,0,0,0,0,0,-2j,3j,0,3j,-3j,3j,\\ &-3j,0,0,-3j,0,0,0,0,0,0,2j,3,0,-3,3,-3].\end{aligned} \tag{5.10}$$

Given that $\boldsymbol{Y}_r$ is initialized to $\tilde{\boldsymbol{Y}}$, the key steps of Algorithm 3 leading to the discovery of the first neighbor is described as follows:

(1) Steps 7 to 12:

Note that $m_0 = 1$ in this case. Take the Hadamard transform of the auto-correlation function of $\boldsymbol{Y}_r$ and its shift by 2^{m-1} to expose the chirps in the frequency domain, so that the first row of $\boldsymbol{P}(\boldsymbol{c})$ can be recovered, and then the entire matrix can be determined. Using $\tilde{\boldsymbol{Y}}$ given by (5.10), the index of the highest peak is the 21st. Therefore, the first row of $\boldsymbol{P}(\boldsymbol{c})$ is the binary

representation of 20, i.e., the binary string of 10100. The matrix $\boldsymbol{P}(\boldsymbol{c})$ can then be uniquely determined.

(2) Steps 13 and 14:

Compute $\boldsymbol{S}^0(\boldsymbol{a}) = \exp\left[j\pi\left(\frac{1}{2}\boldsymbol{a}^\top \boldsymbol{P}(\boldsymbol{c})\boldsymbol{a}\right)\right]$ for all $\boldsymbol{a} \in \mathbb{Z}_2^m$ and apply Hadamard transform to the pointwise product of $\boldsymbol{Y}_r$ and the conjugate of $\boldsymbol{S}^0$ to recover $\boldsymbol{b}$. In the example, the index of the highest peak is the 19-th in the first iteration, hence $\boldsymbol{b} = 10010$.

(3) Steps 15 to 18:

Recover the erased signature $\boldsymbol{S}$ by pointwise product of $\phi_{\boldsymbol{b},\boldsymbol{c}}$ and $\boldsymbol{r}$, then put together all signatures already recovered to form a matrix $\underline{\tilde{\boldsymbol{S}}}$, where all rows of $\underline{\tilde{\boldsymbol{S}}}$ corresponding to the on-slots of node 0 are set to zero. Determine the value of $\boldsymbol{X}$ which minimizes $\|\boldsymbol{Y}_r - \underline{\tilde{\boldsymbol{S}}}\boldsymbol{X}\|_2$. In the example, the reconstructed signature in the first iteration corresponds to the signature of the first neighbor ($\boldsymbol{S}_1$) and the corresponding coefficient X_1 is estimated to be 3, which is equal to U_1.

The preceding steps are repeated to discover more nodes. The algorithm terminates if either the total number of iterations reaches a given number as one desires, or among the discovered nodes, enough of them correspond to very weak coefficients, which implies that the algorithm starts to produce non-neighbors.

We now justify the special scheme for generating the erasures in Algorithm 2. In order to recover the i-th row of the $m \times m$ symmetric matrix corresponding to the largest energy component in the residual signal, the auto-correlation is computed between the residual signal of length 2^m and its shift by 2^{m-i}. It is advisable to guarantee that the

positions of erasures in the received signal and its shift are perfectly aligned as designed in Algorithm 2.

5.2.4. A Numerical Example

We illustrate the performance of discovery using RM codes through the following example. Suppose there are $N = 10{,}000$ valid NIAs which belong to nodes uniformly distributed in a square centered at the origin. Let the path loss exponent be $\alpha = 3$. Assume Rayleigh fading and that a node is regarded as a neighbor if the channel gain exceeds $\theta = 0.05$. In each network realization, we consider the average neighbor discovery performance of the 100 nearest nodes to the origin.

Two types of errors are possible: If an actual neighbor is eliminated by the algorithm, it is called a *miss*. On the other hand, if a non-neighbor survives the algorithm and is thus declared a neighbor, it is called a *false alarm*. The *rate of miss* (resp. *rate of false alarm*) is defined as the average number of misses (resp. false alarms) in one node's neighborhood divided by the average number of neighbors a node has.

First let the density of the nodes be such that each node has on average $c = 10$ neighbors. Choose $m = n_1 = n_2 = 10$, then the signature length is $2^m = 1{,}024$. Averaged over 10 network realizations of the large network, the rate of miss and the rate of false alarm of a total $10 \times 100 = 1{,}000$ nodes (with approximately 10,000 neighbors in total) are plotted in Fig. 5.1 against the SNR. Note that there are no false alarms registered during the simulation when SNR is larger than 12 dB. We find that the total error rate can be lower than 0.2% at 13 dB SNR.

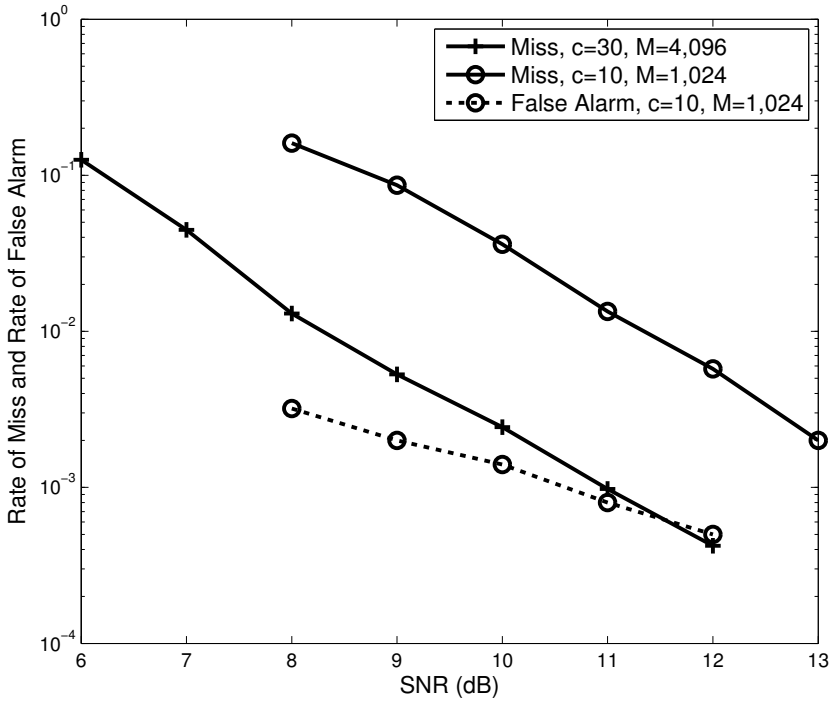

Figure 5.1. The rates of miss and the rate of false alarm versus SNR.

We repeat the simulation with the number of average neighbors changed to $c = 30$ and the parameters changed to $m = n_2 = 12$, and $n_1 = 8$. In this case, the signature length is $2^m = 4{,}096$. During all 10 network realizations, there are no false alarms and the total error rate can be lower than 0.2% at 11 dB SNR.

In order to show that the chirp decoding algorithm is highly resilient to the near-far problem, we demonstrate in Fig. 5.2 that strong neighbors will be detected with very high probability so that their interference to weaker neighbors can be removed. In the case of average $c = 10$ neighbors, when the signature length is 1,024 and SNR is 10 dB, we can see that the rate of miss decreases as the neighbors become stronger, and the rate of miss is below 0.1% at -6 dB attenuation. The simulation is repeated with the number of average neighbors changed to $c = 30$, the length of signature changed to 4,096 and SNR

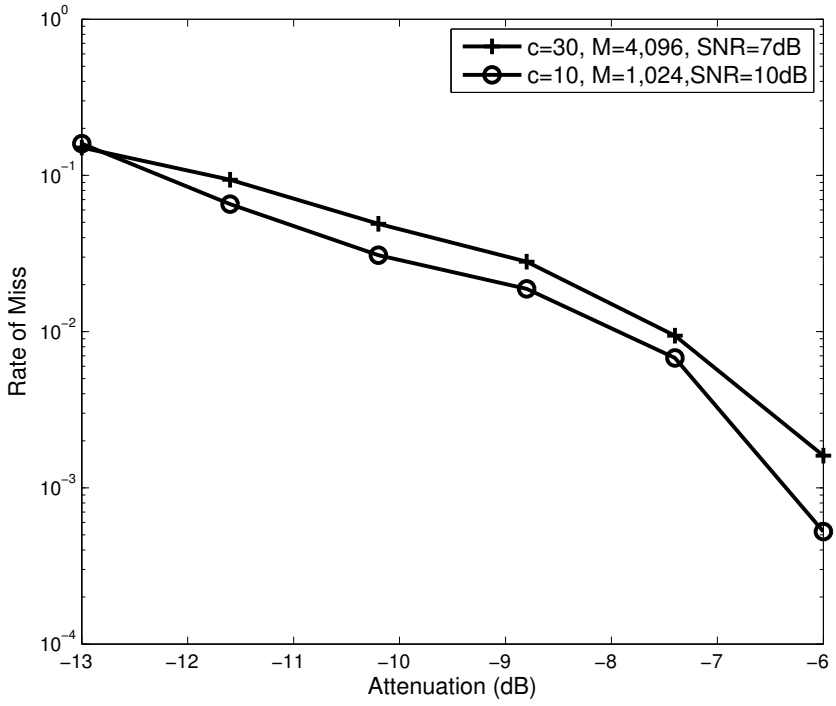

Figure 5.2. The rate of miss versus attenuation.

changed to 7 dB. We can see that all neighbors with attenuation less than -6 dB are successfully discovered with miss rate less than 0.2%.

5.3. Comparison with Random Access

We compare the performance of the compressed neighbor discovery scheme described in Section 5.2 with that of conventional random-access discovery scheme. Only one frame interval is needed by compressed neighbor discovery, as opposed to many frames (often in the hundreds) in the case of random access. Thus compressed neighbor discovery also offers significant reduction of synchronization and error-control overhead embedded in every frame.

5.3.1. Comparison with Generic Random-Access Discovery

Suppose a random-access discovery scheme is used, such as the "birthday" algorithm in [**57**]. Nodes contend to announce their NIAs over a sequence of k contention periods. In each period, each neighbor independently chooses to either transmit (with probability p) or listen (with probability $1-p$). Let $\rho = c/N$. The error rate is equal to the probability of one given neighbor being missed, which is given by

$$\sum_{z=1}^{N} \binom{N}{z} \rho^z \left(1-\rho\right)^{N-z} \left[1-p\left(1-p\right)^{z-1}\right]^k \ . \tag{5.11}$$

Consider a network with 2^{20} NIAs, so in each contention period, the number of bits transmitted is at least $\log_2(2^{20}) = 20$ just to carry the NIA. For a fair comparison with the compressed neighbor discovery scheme, we assume time is slotted and QPSK modulation is used. Table 5.2 lists the amount of transmissions needed by random access discovery according to (5.11) and by compressed discovery based on RM codes with chirp decoding (see Fig. 5.1) in order to achieve the target error rate of 0.002 in the cases of 10 or 30 neighbors.

Evidently, random-access discovery requires hundreds of 20-bit frame transmissions to guarantee the same performance achieved by compressed discovery using a single frame transmission. The latter scheme uses much longer frames. Still, the total number of symbols required by compressed discovery is substantially smaller.

The efficiency of compressed neighbor discovery can be significantly higher than that of random access if all overhead is accounted for. This is because that sending a 20-bit NIA reliably over a fading channel may require up to a hundred symbol transmissions or

Table 5.2. Comparison between random-access discovery and compressed discovery based on RM codes.

	random access	RM codes
$c = 10$	194 frames 1,940 symbols	1 frame 1,024 symbols
$c = 30$	534 frames 5,340 symbols	1 frame 4,096 symbols

more. We believe using compressed discovery can reduce the amount of total discovery overhead by an order of magnitude.

5.3.2. Comparison with IEEE 802.11g

It is also instructive to compare compressed neighbor discovery with the popular IEEE 802.11g technology. Consider the ad hoc mode of 802.11g with active scan, which is basically a random-access discovery scheme. The signaling rate is 4 μs per orthogonal frequency division multiplexing (OFDM) symbol. One probe response frame takes about 850 μs. (The response frame includes additional bits but is dominated by the NIA.) Thus it takes at least 850 $\mu s \times 194 \approx 165$ ms for a query node to discovery 10 neighbors with error rate 0.002 or lower. If compressed neighbor discovery with on-off signature is used, 1,024 symbol transmissions suffice to achieve the same error rate. Using 802.11g symbol interval (4 μs), reliable discovery takes merely 4.1 ms. A highly conservative choice of the symbol interval is 30 μs, which includes carrier (on-off) ramp period (say 10 μs) and the propagation time (less than 1 microsecond for 802.11 range). Compressed neighbor discovery then takes a total of 30 ms, less than 1/5 of that required by 802.11g.

5.4. Summary

In this chapter, we have developed the compressed neighbor discovery scheme using on-off signatures based on a deterministic second-order Reed-Muller code. In order to identify its neighborhood, each node solves a compressed sensing problem using a chirp decoding algorithm. The computational complexity is sub-linear in the address space, so that it is in principle scalable to networks with billions of nodes with 48-bit IEEE 802.11 MAC addresses. The proposed solution has shown to be much more efficient than conventional random-access discovery scheme. The results in this chapter have been published in [**92**, **94**].

CHAPTER 6

Concluding Remarks

In this thesis, a novel paradigm for a clean-state design of the physical and MAC layers of wireless networks has been proposed. The defining feature of the new scheme, which is called rapid on-off-division duplex, enables virtual full-duplex communications by using half-duplex radios. It fully takes advantage of the superposition and broadcast nature of the wireless medium.

As a first step toward quantifying the advantage of on-off signaling in RODD, we have studied in Chapter 2 the capacity of scalar Gaussian channels subject to duty cycle constraint as well as average transmit power constraint. Numerically optimized on-off signaling can achieve much higher rate than Gaussian signaling over a deterministic transmission schedule. It suggests that, compared to intermittently transmitting frames, it is more efficient to disperse nontransmission symbols within each frame to form on-off signaling as in RODD. To further explore the advantage of RODD, in Chapter 3, we have presented the capacity results of RODD network under two simple channel models, namely, deterministic OR-channel and Gaussian channel. The traffic is assumed to be mutual broadcast from each node to all other nodes. The throughput of RODD is shown to be higher than the capacity of ALOHA by a large margin.

As an important application, we have studied in Chapter 4 the mutual broadcast problem by applying RODD signaling. The proposed sparse recovery scheme with random signatures is suitable for the situation where the broadcast messages consist of a relatively

small number of bits. The proposed scheme can also serve as a highly desirable sub-layer of any network protocol stack to provide the important function of simultaneous peer-to-peer mutual broadcast. This sub-layer provides the missing link in many advanced resource allocation schemes, where it is often *assumed* that nodes are provided the state and/or demand of their peers. It is also interesting to note that the proposed scheme departs from the usual solution where a highly-reliable, capacity-achieving, point-to-point physical-layer code is paired with a rather unreliable MAC layer. By treating the physical and MAC layers as a whole, the proposed scheme achieves better overall reliability at much higher efficiency.

In Chapter 5, we have developed the compressed neighbor discovery scheme based on RM codes, which is efficient, scalable, and easy to implement. Using on-off signatures allows half-duplex nodes to achieve network-wide full-duplex discovery. A brief discussion of how neighbor discovery is triggered is in order. If a single node (e.g., a new comer) is interested in its neighborhood, it may send a query message, so that only the neighbors which can hear the message will respond immediately. To implement network-wide discovery, nodes can be programmed to simultaneously transmit their on-off signatures at regular, pre-determined epochs, so that all nodes discover their respective neighbors. This also prevents neighbor discovery from interfering with data transmission.

We conclude by highlighting some further research directions, and indicate which results presented in this thesis can be extended.

In Chapter 2, we only considered scalar Gaussian channels. One step forward would be considering a multiaccess Gaussian channel consisting of more than 2 transmitting nodes with the same duty cycle constraint as well as the average transmit power constraint.

It is interesting to study the maximal achievable sum rate as well as the corresponding optimal input distribution for each node. A conjecture based on the results in Chapter 2 is that the optimal distribution is still discrete. It would also be interesting to find out the optimal signature design (e.g. sparsity, signaling) and the corresponding network capacity in RODD networks.

RODD signaling need not be limited in the time domain. It is in principle compatible with all existing schemes utilizing degrees of freedom in the frequency domain, such as OFDM. Recently, Qualcomm has developed the FlashLinQ technology based on OFDM, which carries out neighbor discovery over a large number of orthogonal time-frequency slots [**60**]. Over each slot, however, the scheme is still based on random access. One possible research direction is to extend the RM-based compressed neighbor discovery scheme in Chapter 5 to multicarrier systems and make a comparison with FlashLinQ.

From an individual receiver's viewpoint, the channel in a RODD network is a multi-access channel with erasure at known positions. Although all good codes for multiaccess channels are in principle good for RODD, no practical codes in the literature can be directly applied in our setting. In Chapter 4, a simple channel code is proposed, which is efficient only if the messages consist of a relatively small number of bits. In the case that each node has many bits to send, one future research direction is to design a structured code with low decoding complexity to make the scheme practical. In Chapter 5, Reed-Muller codes are chosen because its salability and effectiveness in compressed sensing. To find out more efficient codes for neighbor discovery in large networks (i.e., 2^{48} or more NIAs) is also of interest [**44**, **56**].

References

[1] RFC 3684: Topology dissemination based on reverse-path forwarding (TBRPF), 2004. MANET Working Group, The Internet Engineering Task Force (IETF).

[2] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inform. Theory*, 47(4):1290–1301, May 2001.

[3] J. Andrews, N. Jindal, M. Haenggi, R. A. Berry, S. Jafar, D. Guo, S. Shakkottai, R. Heath Jr, M. Neely, S. Weber, A. Yener, and P. Stone. Rethinking information theory for mobile ad hoc networks. *IEEE Communication Magazine*, 46:94–101, Dec. 2008.

[4] D. Angelosante, E. Biglieri, and M. Lops. Neighbor discovery in wireless networks: A multiuser-detection approach. In *Proc. Inform. Theory Appl. Workshop*, pages 46–53, Jan. 29-Feb. 2 2007.

[5] D. Angelosante, E. Biglieri, and M. Lops. A simple algorithm for neighbor discovery in wireless networks. In *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, volume 3, pages 169–172, Apr. 2007.

[6] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank. Asynchronous code-division random access using convex optimization. *Physical Communication*, 5(2):129–147, 2012.

[7] H. Asada, K. Satou, T. Yamazato, M. Katayama, and A. Ogawa. A study on code division duplex (CDD) for distributed CDMA networks. *Technical Report of IEICE*, pages 89–94, 1996.

[8] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse. Wireless network information flow: A deterministic approach. *To appear in IEEE Trans. Inform. Theory.* http://arxiv.org/abs/0906.5394.

[9] F. Baccelli and B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks: Volume I Theory and Volume II Applications*, volume 4. NoW Publishers, 2009.

[10] D. Baron, S. Sarvotham, and R. G. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE Trans. Signal Process.*, 58:269–280, 2010.

[11] S. A. Borbash, A. Ephremides, and M. J. McGlynn. An asynchronous neighbor discovery algorithm for wireless sensor networks. *Ad Hoc Networks*, 5:998–1016, Sep. 2007.

[12] A. R. Calderbank, A. C. Gilbert, and M. J. Strauss. List decoding of noisy reed-muller-like codes. *CoRR*, abs/cs/0607098, 2006.

[13] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, Dec. 2006.

[14] T. H. Chan, S. Hranilovic, and F. R. Kschischang. Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE Trans. Inform. Theory*, 51(6):2073–2088, June 2005.

[15] S. Chen, M.A. Beach, and J.P. McGeehan. Division-free duplex for wireless applications. *IEEE Electronics Letters*, 34:147–148, 1998.

[16] J. Il Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In *Proc. ACM Mobicom*. Chicago, IL, USA, 2010.

[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

[18] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inform. Theory*, 55:2230–2249, 2009.

[19] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, Apr. 2006.

[20] D. L. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing: I. motivation and construction and II. analysis and validation. In *Proc. IEEE Inform. Theory Workshop*. Cairo, Egypt, Jan. 2010.

[21] M. Duarte and A. Sabharwal. Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results. In *Proc. Asilomar Conf. Signals, Systems, & Computers*, 2010.

[22] A. El Gamal and M. H. M. Costa. The capacity region of a class of deterministic interference channels (corresp.). *IEEE Trans. Inform. Theory*, 28(2):343–346, Mar. 1982.

[23] E. Felemban, R. Murawski, E. Ekici, Sangjoon Park, Kangwoo Lee, J. Park, and Z. Hameed. SAND: Sectored-antenna neighbor discovery protocol for wireless networks. In *Proc. IEEE Conf. Sensor Mesh and Ad Hoc Communications and Networks*, pages 1–9, June 2010.

[24] R. K. Ganti, Z. Gong, M. Haenggi, C. Lee, S. Srinivasa, D. Tisza, S. Vanka, and P. Vizi. Implementation and experimental results of superposition coding on software radio. In *Proc. IEEE Int. Conf. Commun.* Cape Town, South Africa, 2010.

[25] S. Gollakota and D. Katabi. Zigzag decoding: combating hidden terminals in wireless networks. In *Proc. ACM SIGCOMM*, pages 159–170, Aug. 2008.

[26] S. Gollakota, S. D. Perli, and D. Katabi. Interference alignment and cancellation. In *Proc. ACM SIGCOMM*, pages 159–170, Aug. 2009.

[27] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products, Fifth Edition*. Academic Press, 5th edition, Jan. 1994.

[28] D. Guo, D. Baron, and S. Shamai (Shitz). A single-letter characterization of optimal noisy compressed sensing. In *Proc. Allerton Conf. Commun., Control, & Computing*. Monticello, IL, USA, Oct. 2009.

[29] D. Guo and S. Verdú. Randomly spread CDMA: Asymptotics via statistical physics. *IEEE Trans. Inform. Theory*, 51(6):1982–2010, June 2005.

[30] D. Guo and C.-C. Wang. Multiuser detection of sparsely spread CDMA. *IEEE J. Select. Areas Commun.*, 26(3):421–431, Apr. 2008.

[31] D. Guo and L. Zhang. Virtual full-duplex wireless communication via rapid on-off-division duplex. In *Proc. Allerton Conf. Commun., Control, & Computing*, pages 412–419, Sep. 29-Oct. 1 2010.

[32] M. C. Gursoy, H. V. Poor, and S. Verdú. The noncoherent Rician fading channel-part I: Structure of the capacity-achieving input. *IEEE Trans. Wireless Commun.*, 4:2193–2206, Sep. 2005.

[33] D. Halperin, T. Anderson, and D. Wetherall. Taking the sting out of carrier sense: interference cancellation for wireless lans. In *Proc. ACM Mobicom*, pages 339–350, Sep. 2008.

[34] M. L. Honig, editor. *Advances in Multiuser Detection*. Wiley-IEEE Press, 2009.

[35] S. D. Howard, A. R. Calderbank, and S. J. Searle. A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes. In *Proc. Conf. Inform. Sciences & Systems*, pages 11–15, Mar. 2008.

[36] J. Huang and S.P. Meyn. Characterization and computation of optimal distributions for channel coding. 51(7):2336 –2351, July 2005.

[37] K. H. Hui, D. Guo, and R. A. Berry. Medium access control via nearest neighbor interactions for regular wireless networks. In *Proc. IEEE Int. Symp. Inform. Theory*, 2010.

[38] M. Jain, J. Il Choi, T. M. Kim, D. Bharadia, S. Set, K. Srinivasan, P. Levis, S. Katti, and P. Sinha. Practical, real-time, full duplex wireless. In *MobiCom*, 2011.

[39] D. Julian and S. Majumdar. Low power personal area communication. In *Proc. Inform. Theory Appl. Workshop*. La Jolla, CA, USA, 2011.

[40] Y. Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 26(43):11111–11121, Oct. 2003.

[41] S. Katti, S. Gollakota, and D. Katabi. Embracing wireless interference: analog network coding. In *Proc. ACM SIGCOMM*, pages 397–408, Aug. 2007.

[42] M. Katz and S. Shamai. On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels. *IEEE Trans. Inform. Theory*, 50(10):2257–2270, Oct. 2004.

[43] G. R. Kenworthy. Self-cancelling full-duplex RF communication system, U.S. Patent 5,691,978, 1997.

[44] M. A. Khajehnejad, J. Yoo, A. Anandkumar, and B. Hassibi. Summary based structures with improved sublinear recovery for compressed sensing. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1427–1431, July-Aug. 2011.

[45] R. Khalili, D. Goeckel, D. Towsley, and A. Swami. Neighbor discovery with reception status feedback to transmitters. In *Proc. IEEE INFOCOM*. San Diego, CA, USA, 2010.

[46] G. Kramer. Communication strategies and coding for relaying. *Wireless Networks*, 143 of The IMA Volumes in Mathematics and its Applications:163–175, 2007.

[47] S. Kumar, V. S. Raghavan, and J. Deng. Medium access control protocols for ad hoc wireless networks: A survey. *Ad Hoc Networks*, 4(3):326–358, May 2006.

[48] S. Lang. *Complex Analysis*. New York: Springer-Verlag, 1999.

[49] W. C. Y. Lee. The most spectrum-efficient duplexing system: CDD. *IEEE Communication Magazine*, pages 163–166, 2002.

[50] D. D. Lin and T. J. Lim. Subspace-based active user identification for a collision-free slotted ad hoc network. *IEEE Trans. Commun.*, 52:612–621, Apr. 2004.

[51] D. G. Luenberger. *Optimization by Vector Sapce Methods*. New York: Wiley, 1969.

[52] J. Luo and D. Guo. Neighbor discovery in wireless ad hoc networks based on group testing. In *Proc. Allerton Conf. Commun., Control, & Computing*. Monticello, IL, USA, 2008.

[53] J. Luo and D. Guo. Compressed neighbor discovery for wireless ad hoc networks: the Rayleigh fading case. In *Proc. Allerton Conf. Commun., Control, & Computing*. Monticello, IL, USA, Oct. 2009.

[54] T. Lutz, C. Hausl, and R. Kötter. Coding strategies for noise-free relay cascades with half-duplex constraint. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 2385–2389. Toronto, ON, Canada, July 2008.

[55] T. Lutz, G. Kramer, and C. Hausl. Capacity for half-duplex line networks with two sources. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 2393–2397. Austin, TX, USA, June 2010.

[56] M. Bakshi, S. Jaggi, S. Cai, and M. Chen SHO-FA: Robust compressive sensing with order-optimal complexity, measurements, and bits. *arXiv:1207.2335v1 [cs.IT]*, 2012.

[57] M. J. McGlynn and S. A. Borbash. Birthday protocols for low energy deployment and flexible neighbor discovery in ad hoc wireless networks. *Proceedings of the 2nd ACM International Symposium on Mobile Ad hoc Networking & Computing*, pages 137–145, Oct. 2001.

[58] P. Mohapatra and S. Krishnamurthy. *AD HOC NETWORKS: technologies and protocols*. Springer Verlag, 2005.

[59] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26:301–321, 2009.

[60] J. Ni, R. Srikant, and X. Wu. Coloring spatial point processes with applications to peer discovery in large wireless networks. In *SIGMETRICS*, pages 167–178, June 2010.

[61] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & systems (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[62] Yu. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.*, 1:157–214, 1956.

[63] B. Radunovic, D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. Dejean. Rethinking indoor wireless: Low power, low frequency, full duplex. Technical report, Microsoft Research, 2009.

[64] A. Raghavan, E. Gebara, E. M. Tentzeris, and J. Laskar. Analysis and design of an interference canceller for collocated radios. *IEEE Transactions on Microwave Theory and Techniques*, 53:3498–3508, 2005.

[65] T. S. Rappaport. *Wireless Communications*. Prentice-Hall, 2nd edition, 2002.

[66] M. G. Rubinstein, I. M. Moraes, M. Campista, L. Costa, and O. Duarte. *A Survey on Wireless Ad Hoc Networks*, volume 211 of *IFIP International Federation for Information Processing*. Springer Boston, Nov 2006.

[67] W. Rudin. *Real and Complex Analysis*. McGraw-Hill Science Engineering, 1986.

[68] A. Sahai, G. Patel, and A. Sabharwal. Pushing the limits of full-duplex: Design and real-time implementation. *arXiv:1107.0607v1*, 2011.

[69] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro. Consensus in ad hoc WSNs with noisy links-part II: Distributed estimation and smoothing of random signals. *IEEE Trans. Signal Process.*, 56(4):1650–1666, 2008.

[70] I. D. Schizas, A. Ribeiro, and G. B. Giannakis. Consensus in ad hoc WSNs with noisy links-part I: Distributed estimation of deterministic signals. *IEEE Trans. Signal Process.*, 56(1):350–364, 2008.

[71] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick. Pricing algorithms for power control and beamformer design in interference networks. *IEEE Signal Processing Mag.*, 26:53–63, 2009.

[72] R. A. Scholtz. Multiple access with time-hopping impulse modulation. In *Proc. IEEE MILCOM*. Bedford, MA, USA, 1993.

[73] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson. Cross-layer design for wireless networks. *IEEE Communication Magazine*, 41(10):74–80, Oct. 2003.

[74] S. Shamai. Capacity of a pulse amplitude modulated direct detection photon channel. *Proc. IEE Communications, Speech and Vision*, 137(6):424–430, Dec. 1990.

[75] S. Shamai and I. Bar-David. The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Trans. Inform. Theory*, 41(4):1060–1071, July 1995.

[76] O. Simeone, U. Spagnolini, Y. Bar-Ness, and S. Strogatz. Distributed synchronization in wireless networks. *IEEE Signal Processing Mag.*, 25(5):81–97, Sep. 2008.

[77] J. G. Smith. The information capacity of amplitude and variance-constrained scalar Gaussian channels. *Inf. Contr.*, 18:203–219, 1971.

[78] D. W. Stroock. *Probability Theory, an Analytic View*. New York: Cambridge Univ. Press, 1993.

[79] M. Sudan. Coding theory: Tutorial & survey. In *Proc. of the 42nd Annual Symposium on Foundations of Computer Science.*, 2001.

[80] P. H. Tan and L. K. Rasmussen. Belief propagation for coded multiuser detection. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1919–1923, 2006.

[81] T. Tanaka and M. Okada. Approximate belief propagation, density evolution, and statistical neurodynamics for CDMA multiuser detection. *IEEE Trans. Inform. Theory*, 51(2):700–706, Feb. 2005.

[82] D. N. C. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[83] S. Vasudevan, J. Kurose, and D. Towsley. On neighbor discovery in wireless networks with directional antennas. *Proc. IEEE INFOCOM*, 4:2502–2512, 2005.

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

[85] M. Z. Win and R. A. Scholtz. Impulse radio: How it works. 2:36–38, 1998.

[86] M. Z. Win and R. A. Scholtz. Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications. *IEEE Trans. Commun.*, 48(4):679 –689, Apr. 2000.

[87] Y. Wu and S. Verdú. MMSE dimension. In *Proc. IEEE Int. Symp. Inform. Theory*. Austin, TX, USA, June 2010.

[88] Y. Wu and S. Verdú. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Trans. Inform. Theory*, 56:3721–3748, 2010.

[89] S. Xu and T. Saadawi. Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? *IEEE Communication Magazine*, 39(6):130–137, June 2001.

[90] R. Zamir. A gaussian input is not too bad. *IEEE Trans. Inform. Theory*, 50(6):1362–1367, June 2004.

[91] L. Zhang and D. Guo. Capacity of Gaussian channels with duty cycle and power constraints. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 424–428, St Petersburg, Russia, July-Aug. 2011.

[92] L. Zhang and D. Guo. Neighbor discovery in wireless networks using compressed sensing with Reed-Muller codes. In *Proc. WiOpt*, May 2011.

[93] L. Zhang and D. Guo. Wireless peer-to-peer mutual broadcast via sparse recovery. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1901–1905, St Petersburg, Russia, July-Aug. 2011.

[94] L. Zhang, J. Luo, and D. Guo. Neighbor discovery for wireless networks via compressed sensing. *Performance Evaluation*, 2012.