Random Wireless Network

An Information Theoretic Perspective

Rahul Vaze

Random Wireless Networks

An Information Theoretic Perspective

Rahul Vaze

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Cambridge House, 4381/4 Ansari Road, Daryaganj, Delhi 110002, India

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107102323

First published 2015

Printed in India

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing-in-Publication data
Vaze, Rahul.
Random wireless networks : an information theoretic perspective / Rahul Vaze.
pages cm
Includes bibliographical references and index.
Summary: "Provides detailed discussion on single hop and multi hop model, feedback constraints and modern communication techniques such as multiple antenna nodes and cognitive radios"– Provided by publisher.
ISBN 978-1-107-10232-3 (hardback)
1. Wireless communication systems. I. Title.
TK5103.2.V39 2015
621.382'1–dc23
2014044738

ISBN 978-1-107-10232-3 Hardback

To my son Niraad;
this book was written while babysitting him.

Contents

List of Figures

Preface

In addition to the traditional cellular wireless networks, in recent past, many other wireless networks have gained widespread popularity, such as sensor networks, military networks, and vehicular networks. In a sensor network, a large number of sensors are deployed in a geographical area for monitoring physical parameters (temperature, rainfall), intrusion detection, animal census, etc., while in a military network, heterogenous military hardware interconnects to form a network in a battlefield, and vehicular networks are being deployed today for traffic management, emergency evacuations, and efficient routing. For efficient scalability, these new wireless networks are envisaged to be self-configurable with no centralized control, sometimes referred to as *ad hoc* networks.

The decentralized mode of operation makes it easier to deploy these networks, however, that also presents with several challenges, such as creating large amount of interference, large overheads for finding optimal routes, complicated protocols for cooperation and coordination. Because of these challenges, finding the performance limits, both in terms of the amount of information that can be carried across the network and ensuring connectivity in the wireless network, is a very hard problem and has remained unsolved in its full generality.

From an information-theoretic point of view, where we are interested in finding the maximum amount of information that can be carried across the network, one of the major bottlenecks in wireless network is the characterization of interference. To make use of the spatial separation between nodes of the wireless network, multiple transmitters communicate at the same time, creating interference at other receivers. The arbitrary topology of the network further compounds the problem by directly affecting the signal interaction or interference profile. Thus, one of the several trade-offs in wireless networks is the extent of spatial reuse viz-a-viz the interference tolerance. Another important trade-off is the relation between the radio range (distance to which each node can transmit) of sensor nodes and the connectivity of the wireless network. Small radio range leads to isolated nodes, while larger radio ranges result in significant interference at the neighboring receivers affecting connectivity.

Over the last decade and a half, these trade-offs have been addressed in a variety of ways, with exact answers derived for *random* wireless networks, where nodes of the wireless network are located uniformly at random in a given area of interest. The primary reason for assuming random location for nodes is the applicability of rich mathematical tools from stochastic geometry, percolation theory, etc. that provide significant mathematical foundation and allow derivation of concrete results. This book ties up the different ideas introduced for understanding the performance limits of random wireless networks and presents a complete overview on the advances made from an information-theoretic (capacity limits) point of view.

In this book, we focus on two capacity metrics for random wireless networks, namely, the transmission and the throughput capacity, that have been defined to capture the successful number

of bits that can be transported across the network. We present a comprehensive analysis of transmission capacity and throughput capacity of random wireless networks. In addition, using the tools from percolation theory, we also discuss the connectivity and percolation properties of random wireless networks, which impact the routing and large-scale connectivity in wireless networks. The book is presented in a cohesive and easy to follow manner, however, without losing the mathematical rigor. Sufficient background and critical details are provided for the advanced mathematical concepts required for solving these problems.

The book is targeted at graduate students looking for an easy and rigorous introduction to the area of information/communication theory of random wireless networks. The book also quantifies the effects of network layer protocols (e.g., automatic repeat requests (ARQs)), physical layer technologies such as multiple antennas (MIMO), successive interference cancelation, information-theoretic security, on the performance of wireless networks. The book is accessible to anyone with a background in basic calculus, probability theory, and matrix theory.

The book starts with an introduction to the signal processing, information theory, and communication theory fundamentals of a point-to-point wireless communication channel. Specifically, a quick overview of the concept of Shannon capacity, outage formulation, basic information-theoretic channels, basics of multiple antenna communication, etc. is provided that lays sufficient background for the rest of the book.

The book is divided into two parts, the first part exclusively deals with single-hop wireless networks, where each source–destination communicates directly with each other, while in the second part, we focus only on the multi-hop wireless networks, where source–destination pairs are out of each others' communication range and use multiple other nodes (called relays) for communication.

For the first part, we begin by deriving analytical expressions for the transmission capacity for a single-hop model with various scheduling protocols such as ALOHA, CSMA, guard-zone based, etc. Next, we discuss in detail the effect of using multiple antennas on the transmission capacity of a random wireless network and derive the optimal role of multiple antennas. We then extend our setup and present performance analysis of random wireless networks under a two-way communication model that allows for bidirectional communication between two nodes. We close the first part of the book by applying stochastic geometry tools to derive a tractable performance analysis of a cellular wireless network in terms of critical measures such as connection probability, average rate, etc. which is extremely useful for practicing engineers.

The second part of the book starts by extending the transmission capacity framework to a multi-hop wireless network, where we derive the transmission capacity expression and find the optimal value of several key parameters relevant to the multi-hop communication model. Then, we give a brief introduction to percolation theory results for both the discrete and the continuum case. The background on percolation theory sets up the platform for deriving several important results for random wireless networks, such as finding the optimal radio range for connectivity, formation of large connected clusters under different connection models, and most importantly for finding tight scaling bounds on the throughput capacity.

We then present the seminal result of Gupta and Kumar which shows that the throughput capacity of a random wireless network scales as square root of the number of nodes. Finally, we discuss the concept of hierarchical cooperation in a wireless network which is used to show that the throughput capacity can scale linearly with the number of nodes.

This book is an effort to present the several disparate ideas developed for deriving capacity of a random wireless network in a unified framework. For effective understanding, extensive effort is made to explain the physical interpretation of all results. As an attempt to reach out to a wider

audience, effects of practical communication models, such as cellular networks, two-way communication (downlink/uplink) feedback constraints, modern communication techniques (such as multiple antenna nodes, interference cancelation and avoidance, cognitive radios), are also analyzed and discussed in sufficient detail.

Most of the ideas/results presented in this book are not more than a decade old and have not yet found a consolidated treatment. The presentation is kept short and lucid with sufficient detail and rigor. For clarity, at instances, places simplified proofs of the original results are provided.

Acknowledgments

I would like to thank everyone who have made this book possible. Starting with Srikanth Iyer, D. Yogeshwaran, Aditya Gopalan, Chandra Murthy, Radhakrishna Ganti, Sibiraj Pillai, Siddharth Banerjee, all have made detailed comments on my various drafts, which undoubtedly made the book more readable. Their critical comments have also shaped the structure and content. I would also like to thank Kaibin Huang, who urged me to write this as a textbook that he could use in his course. I do not know yet whether it will serve his purpose. Help from undergraduate internship students Vivek Bagaria, Dheeraj Narasimha, Jainam Doshi, Rushil Nagda, Siddharth Satpathi, Ajay Krishnan, and Navya Prem in proofreading is gratefully acknowledged. Their comments were valuable in making the book accessible to readers with little or no prior background. The review comments by the two referees helped immensely in changing the structure of the book. Their comments especially helped in the reorganization and pruning of the book to keep a clear and sharp focus.

Notation

$\mathbf{A}$	Matrix A
$\mathbf{A}(i,j)$	(i,j)th entry of matrix A
$\mathbf{a}$	vector a
$\mathbf{a}(i)$	ith element of vector $\mathbf{a}$
$\mathbf{a}^\dagger$	Conjugate transpose of vector $\mathbf{a}$
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}^d$	Set of real numbers in d dimensions
$\mathbb{C}^{m\times n}$	Matrices with m rows and n columns over the set of complex numbers
$\mathbb{P}(A)$	Probability of event A
$\mathbb{E}$	Expectation operator
$\#(A)$	Number of elements in set A
$\nu(A)$	Lebesgue measure of set A
o	Origin in $\mathbb{R}^d$
d_{xy}	Distance between nodes x and y
α	Path-loss exponent for wireless propagation
$\mathbf{B}(x,r)$	Disc of radius r centered at x
ϕ	Null set
$\lvert a\rvert$	Absolute value of a
Φ	A Poisson point process
p	ALOHA access probability
λ	Density of nodes of the network
μ	Density of blockages/obstacles in the network
ϵ	Outage probability constraint
Λ	Density measure of nodes of the network
γ	Interference suppression parameter
ρ	Random variable representing the random radius in the Gilbert's random disc model
$\Gamma(t)$	$\int_0^\infty x^{t-1}\exp^{-x}\,dx$
J	$\sqrt{-1}$
$A\propto B$	$A=cB$, where c is a constant
SNR	Signal-to-noise-ratio
SIR	Signal-to-interference-ratio
SINR	Signal-to-interference-plus-noise-ratio

β	Signal-to-interference-plus-ratio threshold for successful packet reception
B	Rate of transmission corresponding to SINR threshold β, $\beta = 2^B - 1$
M_n	Number of retransmissions required on hop n
N_h	Number of hops
M	Number of end-to-end retransmissions required $\sum_{n=1}^{N_h} M_n = M$
$\mathsf{t}(n)$	Per-node throughput capacity
$\mathsf{T}(n)$	Network wide throughput capacity
C	Transmission capacity
C_{tw}	Two-way transmission capacity
C_d	Delay normalized transmission capacity
C_s	Spatial progress capacity
AWGN	Additive white Gaussian noise
N_0	Variance of the AWGN
$\mathcal{N}(\mathsf{m}, \mathsf{var})$	Normal distribution with mean m and variance var
$\mathcal{CN}(\mathsf{m}, \mathsf{var})$	Complex normal distribution with mean m and variance var
$\mathbf{1}_n$	Indicator variable for node n
$\chi^2(2m)$	Chi-square distribution with m degrees of freedom
$X \sim Y$	Random variable X has distribution Y
$f(n) = \Omega(g(n))$	If $\exists\ k > 0,\ n_0,\ \forall\ n > n_0, \lvert g(n)\rvert k \leq \lvert f(n)\rvert$
$f(n) = \mathcal{O}(g(n))$	If $\exists\ k > 0,\ n_0,\ \forall\ n > n_0, \lvert f(n)\rvert \leq \lvert g(n)\rvert k$
$f(n) = \Theta(g(n))$	If $\exists\ k_1,\ k_2 > 0,\ n_0,\ \forall\ n > n_0, \lvert g(n)\rvert k_1 \leq \lvert f(n)\rvert \leq \lvert g(n)\rvert k_2$
$f(n) = \mathrm{o}(g(n))$	If $\lim_{n\to\infty} \frac{f(n)}{g(n} = 0$

Chapter 1

Introduction

1.1 Introduction

Wireless networks can be broadly classified into two categories: centralized and de-centralized. A canonical example of a centralized network is a cellular network, where all operations are controlled by basestations, for example, when should each user transmit or receive, thereby avoiding simultaneous transmission (interference) by closely located nodes. Prominent examples of de-centralized or *ad hoc* networks include sensor or military networks. Sensor network is deployed in a large physical area to either monitor physical parameters, such as temperature, rainfall, and animal census, or intrusion detection. In a military network, a large number of disparate military equipment, e.g., battle tanks, helicopters, ground forces, is connected in a decentralized manner to form a robust and high throughput network. Ad hoc networks are attractive because of their scalability, self-configurability, robustness, etc.

Vehicular network is a more modern example of an ad hoc wireless network, where a large number of sensors are deployed on the highways as well as mounted on vehicles that are used for traffic management, congestion control, and quick accident information exchange. Many other applications of ad hoc wireless networks are also envisaged such as deploying large number of sensors in large building for helping fire fighters in case of fire emergency and in case of earthquakes.

The key feature that distinguishes centralized and ad hoc wireless networks is interference. With centralized control, interference can be avoided in contrast to ad hoc networks, where there is no mechanism of inhibiting multiple transmitters from being active simultaneously. Thus, ad hoc networks give rise to complicated signal interaction at all receiver nodes. As compared to additive noise, interference is structured, and treating interference as noise is known to be sub-optimal. Thus, performance analysis of ad hoc wireless networks is far more complicated than centralized wireless networks.

In this book, we are interested in studying the physical layer issues of ad hoc wireless networks, such as finding the limits on the reliable rate of information transfer and ensuring connectivity among all nodes of the network. Traditionally, the Shannon capacity has been used to characterize the reliable rate of information transfer in communication systems. In a wireless network, however, finding the Shannon capacity is challenging and has remained unsolved. The main impediment in finding the Shannon capacity of wireless networks is the complicated nature of interference created by multiple simultaneously active transmitters at each other's receivers and network topology that directly influences the signal interaction.

To get some meaningful insights to the fundamental limits of throughput in wireless networks, alternate notions of capacity have been introduced and analyzed, such as transmission [1] and throughput/transport [2] capacity, which are defined by relaxing the reliability constraints compared to the Shannon capacity.

One key relaxation/assumption made for the purposes of analyzing these new capacity metrics is that the nodes of the network are assumed to be distributed uniformly at random in the area of interest, called the *random* wireless networks. The random node location assumption allows the use of tools from stochastic geometry and percolation theory for theoretical capacity analysis. In Chapter 2, we argue that random node location assumption is not very limiting for a practical ad hoc network.

Major focus of this book is on finding the transmission and the throughput capacity of random wireless networks. Through the transmission capacity formulation, we also quantify the effects of using multiple antennas at each node, using two-way communication between source and destination, effect of ARQ protocol, and using "smart" scheduling protocols in the random wireless networks. From here on in this book, when we say wireless network, we mean a random wireless network unless specified differently.

A necessary condition for finding the maximum rate of transmission or throughput between a pair of nodes in a wireless network is to ensure that they are connected to each other or have a connected path between each other, under a suitable definition of connection. Since any source can have an arbitrary choice for its destination, essentially, we need network wide connectivity, that is, each node pair should be reachable from every other node via connected paths. This condition is simply called as *connectivity* of the wireless network. Connectivity in a wireless network depends on the density of nodes, radio (transmission) range of any node, topology of the network, connection model between nodes, etc. In this book, we present relevant results from the percolation theory and then describe their application in finding the network parameters that ensure connectivity in wireless networks. Using percolation theory, we also study the size of the largest connected component in wireless networks and find conditions when the size of the largest connected component is a non vanishing fraction of the total number of nodes, which implies approximate connectivity.

The book is divided into two parts, first part exclusively deals with a single-hop model for wireless networks, where each source has a destination at a fixed distance from it and transmits its information directly to its destination without the help of any other node in the network. We define the notion of transmission capacity for the single-hop model and derive it for single antenna nodes, multiple antenna nodes, with scheduling protocols, and under two-way communication scenarios. The first part of the book also includes the performance analysis of cellular wireless network techniques using tools from stochastic geometry that are developed in the earlier chapters of the first part.

In the second part, we deal with the more relevant model of multi-hop communication for a wireless network and define two notions of capacity, namely the delay normalized transmission capacity and the throughput capacity and present their analysis. In addition, in the second part, we also study the connectivity and percolation properties of a multi-hop wireless network under the signal-to-noise-plus-interfence ratio (SINR) model.

This chapter sets up the background for studying wireless networks from a physical layer point of view. We begin by describing the basics of point-to-point communication, where a single transmitter is interested in communicating with a single receiver. To keep the discussion general, we consider the case when each node is equipped with multiple antennas. We first discuss the role of multiple antennas in improving the error-probability performance as a function of number of transmit and receive antennas with the optimal maximum likelihood (ML) decoder. We then state some difficulties in using the optimal maximum likelihood decoder, such as an exponential

complexity, and present the more popular sub-optimal decoders such as zero-forcing decoder that have linear decoding complexity. We also discuss the error rate performance degradation while using the sub-optimal zero-forcing decoder.

Next, we define the notion of Shannon capacity and present results on the Shannon capacity of the point-to-point communication channel with multi-antenna equipped nodes. We show that the Shannon capacity scales linearly with the minimum of the number of transmit and receive antennas. We next present the outage formulation for characterizing capacity (called outage capacity) in non-ergodic channels, for which the Shannon capacity is zero. The non-ergodic channel is of interest since the popular slow-fading channel model of wireless signal propagation, where channel coefficients remain constant for sufficient amount of time, falls in the class of non-ergodic channels. The outage formulation also helps in defining the transmission capacity of wireless networks.

Next, we describe the received signal model at any node of a wireless network, where multiple transmitters are active at the same time. Using examples of some basic building blocks of a wireless network, we discuss some of the difficulties in finding the Shannon capacity of a wireless network. We then motivate the definitions of alternate capacity metrics, such as transmission capacity and throughput capacity, which are defined under a relaxed reliability constraint compared to the Shannon capacity.

We end this chapter by presenting some details on studying connectivity in wireless networks under various link connection models.

1.2 Point-to-Point Wireless Signal Propagation Model

Consider a wireless communication channel between a single transmitter T_0 equipped with N_t antennas and a single receiver R_0 with N_r antennas. Let the distance between T_0 and R_0 be d, then the received signal at R_0 at time t is given by

$$\mathbf{y}[t] = d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\sum_{m=0}^{M-1}\mathbf{H}_m\mathbf{x}[t-m] + \mathbf{w}[t], \tag{1.1}$$

where M is the number of distinct multiple fading paths between the transmitter and the receiver, $d^{-\alpha/2}$ is the distance-based path-loss function, α is the path-loss exponent that is typically in the range $(2,4)$, $\mathbf{H}_t \in \mathbb{C}^{N_r \times N_t}$ is the channel coefficient matrix at time t between the transmitter and the receiver, where $\mathbf{H}_t(i,j)$ is the channel coefficient between the ith receive and jth transmit antenna. The $N_t \times 1$ transmit signal vector at time t is $\mathbf{x}[t]$ with unit power constraint, $\mathbb{E}\{\mathbf{x}[t]^\dagger\mathbf{x}[t]\} = 1$, P is the average transmitted power, and $\mathbf{w}[t]$ is additive white Gaussian noise vector with entries that are independent and $\mathcal{CN}(0,1)$ distributed.

Assumption 1.2.1 *Throughout this book, we will use the simple distance-based path-loss function of $d^{-\alpha/2}$ that is valid in far-field, however, has a singularity in the near-field at $d = 0$.*

Assumption 1.2.2 *We will also always assume a flat fading channel, that is, no multi-path $\mathbf{H}_t = \mathbf{0}$ for $t > 0$, for which the signal model* (1.1) *simplifies to*

$$\mathbf{y} = d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1.2}$$

where the entries of $\mathbf{H}$ *are assumed to be independent and* $\mathcal{CN}(0,1)$ *distributed to model a rich scattering channel (Rayleigh fading). We also assume throughout this book that matrix* $\mathbf{H}$ *is perfectly known at the receiver.*

To decode the transmit signal $\mathbf{x}$, the optimal decoder is the maximum a posteriori (MAP) decoder that declares that signal to be transmitted which is the most likely signal $\mathbf{x}$ given the knowledge of $\mathbf{y}$. Assuming an uniform distribution over the input signals, MAP decoding is equivalent to ML decoding, where the decoded codeword maximizes the likelihood of $\mathbf{y}$ given $\mathbf{x}$. Mathematically, ML decoding solves the following optimization problem.

$$\max_{\mathbf{x}} \mathbb{P}(\mathbf{y}|\mathbf{x},\mathbf{H}).$$

For the signal model (1.2), since each entry of $\mathbf{w}$ is independent and $\mathcal{CN}(0,1)$ distributed,

$$\mathbb{P}(\mathbf{y}|\mathbf{x},\mathbf{H}) = \frac{1}{\pi}\exp^{-\left(\mathbf{y}-d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{x}\right)\left(\mathbf{y}-d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{x}\right)^{\dagger}},$$

which can be simplified to conclude that the ML decoder decodes vector $\mathbf{x}$ that solves

$$\max_{\mathbf{x}} \mathbb{P}(\mathbf{y}|\mathbf{x},\mathbf{H}) = \min_{\mathbf{x}} ||\mathbf{y} - d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\mathbf{H}\mathbf{x}||^2. \tag{1.3}$$

Thus, the ML decoder decodes $\mathbf{x}$, which is the closest codeword to the received signal $\mathbf{y}$ in terms of the Euclidean distance. With ML decoding, all the components of vector $\mathbf{x}$ are decoded jointly, thereby making the complexity exponential in the size of $\mathbf{x}$ which is N_t.

Assuming that the channel matrix $\mathbf{H}$ remains constant for $T \geq N_t$ time slots, and if the transmitter codes across T time slots to send codeword $\mathbf{X}_i = [\mathbf{x}_i[1] \ldots \mathbf{x}_i[T]]$, the probability of decoding the codeword matrix $\mathbf{X}_j = [\mathbf{x}_j[1] \ldots \mathbf{x}_j[T]]$ instead of $\mathbf{X}_i$ with an ML decoder is [3]

$$\mathbb{P}(\mathbf{X}_i \to \mathbf{X}_j) \leq \left(\prod_{k=1}^{\mathsf{div}} \sigma_k\left(\mathbf{X}_i - \mathbf{X}_j\right)\right)^{-N_r} P^{-\mathsf{div}N_r}, \tag{1.4}$$

where

$$\mathsf{div} = \min_{\mathbf{X}_i \neq \mathbf{X}_j} \{\text{rank}(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^{\dagger}\} \tag{1.5}$$

and $\sigma_k\left(\mathbf{X}_i - \mathbf{X}_j\right)$ are the non-zero eigenvalues of $(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^{\dagger}$. Thus, to minimize the pairwise error probability (1.4), one has to maximize the minimum of the rank of the difference of any two codeword matrices $\mathbf{X}_i$ and $\mathbf{X}_j$ (1.5). Clearly, with $T \geq N_t$, the maximum value of div is N_t (since $\mathbf{X}_i \in \mathbb{C}^{N_t \times T}$, $\forall\ i$) and for achieving $\mathsf{div} = N_t$, the codewords $\mathbf{X}_i$'s should be coded in space and time; hence the codebook consisting of codewords $\mathbf{X}_i$'s is called a space–time block code (STBC). STBCs with $\mathsf{div} = N_t$ are called full-diversity achieving STBCs, and their error probability is proportional to $P^{-N_t N_r}$. Thus, with multiple transmit and receive antennas, the reliability of a wireless channel can be improved exponentially with the increasing transmission power.

Even though ML decoding provides with the best error probability performance, its decoding complexity is very high because of the joint decoding of all elements of transmitted vector $\mathbf{x}$. Several simple decoders with reduced decoding complexity are also known in literature, for example, minimum mean square error (MMSE) decoder and zero forcing (ZF) decoder. ZF decoder is specially attractive for its simple decoding rule and incurs linear decoding complexity in N_t (the

number of elements of $\mathbf{x}$). We describe the ZF decoder in brief and present its error probability performance. We will use the ZF decoder in Chapter 3 to analyze the effects of using multiple antennas in a wireless network.

With ZF decoder, to decode stream $\mathbf{x}(\ell)$ of the transmitted vector

$$\mathbf{x} = [\mathbf{x}(1), \ldots, \mathbf{x}(N_t)]^T,$$

the received signal (1.2) is multiplied with a vector $\mathbf{q}_\ell^\dagger \in \mathbb{C}^{N_r \times 1}$, which belongs to the null space of the columns $\mathbf{H}(j), j = 1, \ldots, \ell-1, \ell+1, \ldots, N_t$ of the channel matrix $\mathbf{H}$, to cancel the inter-stream interference from all other streams $\mathbf{x}(j), j = 1, \ldots, \ell-1, \ell+1, \ldots, N_t$. With this operation, from (1.2), the resulting signal can be written as

$$\mathbf{y}(\ell) = d^{-\alpha/2}\sqrt{\frac{P}{N_t}}\mathbf{q}_\ell^\dagger \mathbf{H}(\ell)\mathbf{x}(\ell) + \mathbf{q}_\ell^\dagger \mathbf{w}, \tag{1.6}$$

$\forall \quad \ell = 1, \ldots, N_t$, where there is no inter-stream interference from $\mathbf{x}(j), j = 1, \ldots, \ell-1, \ell+1, \ldots, N_t$. Thus, with a ZF decoder, each of the N_t data streams of $\mathbf{x}$ can be decoded independently of each other using (1.6), thereby incurring linear decoding complexity compared to the exponential decoding complexity of the ML decoder. This sub-optimal receiver, however, has poor error probability performance because of correlating the noise components in y_ℓ for different $\ell = 1, \ldots, N_t$, and the error probability is proportional to $P^{N_r - N_t + 1}$ [4], instead of $P^{-N_t N_r}$ with the ML decoder.

We next discuss the alternative use of multiple antennas in improving the capacity of the point-to-point communication channel. We first define the concept of Shannon capacity, a measure of reliable throughput and show that Shannon capacity increases linearly with the minimum of the number of transmit and receive antennas.

1.3 Shannon Capacity

Definition 1.3.1 *The Shannon capacity C for a communication channel is defined as the largest quantity such that for any rate $R < C$, reliable communication is possible. By reliable communication, we mean that the probability of error can be driven down to zero with increasing block length. Conversely, if the rate of transmission $R \geq C$, the probability of error is lower bounded by a constant.*

Definition 1.3.2 *Let $x[n]$ and $y[n]$ be the input and output of a channel at time n, respectively, then a channel is called a discrete memoryless channel (DMC), if given the most recent input, the output is independent of all previous inputs and outputs, that is,*

$$\mathbb{P}(y[n] \mid x[1], \ldots, x[n], y[1], \ldots, y[n-1]) = \mathbb{P}(y[n] \mid x[n])$$

for $n = 1, 2, 3, \ldots$. Thus, in a DMC, given the input at time n, the output at time n is independent of all the past inputs and outputs.

C. E. Shannon, in his 1948 seminal paper [5], proved that the capacity of a DMC defined by $\mathbb{P}(\mathbf{y}|\mathbf{x})$, with input $\mathbf{x} = [x[1], \ldots, x[n]]$ and output $\mathbf{y} = [y[1], \ldots, y[n]]$ is given by

$$C = \max_{\mathbb{P}(\mathbf{x})} \mathsf{I}(\mathbf{x}; \mathbf{y}), \tag{1.7}$$

where $\mathsf{I}(\mathbf{x}; \mathbf{y})$ is the mutual information between $\mathbf{x}$ and $\mathbf{y}$ [17]. This result is popularly known as Shannon's channel coding theorem.

Specializing this result for the multiple antenna channel (1.2), when $\mathbf{H}$ is known at the receiver, we have that $\mathsf{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \mathbb{E}_{\mathbf{H}}\left\{\log\det\left(\mathbf{I}_{N_r} + \frac{P}{N_t}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right)\right\}$, and the Shannon capacity of the multiple antenna channel is

$$C = \max_{tr(\mathbf{Q})\leq N_t} \mathbb{E}_{\mathbf{H}}\left\{\log\det\left(\mathbf{I}_{N_r} + \frac{P}{N_t}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right)\right\}, \tag{1.8}$$

where $\mathbf{Q} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}$ is the covariance matrix of the input signal $\mathbf{x}$. The optimization in (1.8) depends on whether the channel coefficient matrix $\mathbf{H}$ is known at the transmitter (referred to as CSIT) or not (called CSIR). With CSIT, the Shannon capacity [7] is

$$C = \mathbb{E}_{\mathbf{H}}\left\{\sum_{k=1}^{\min\{N_t,N_r\}} \log\left(\xi\sigma_k\left(\mathbf{H}\right)\right)^+\right\},$$

where ξ is the Lagrange multiplier satisfying the power constraint

$$\sum_k (\xi - \sigma_k\left(\mathbf{X}_i - \mathbf{X}_j\right)^{-1})^+ = P,$$

and $\sigma_k\left(\mathbf{H}\right)$ is the kth eigenvalue of $\mathbf{H}\mathbf{H}^\dagger$ indexed in the decreasing order.

On the other hand, with CSIR, when transmitter has no information about $\mathbf{H}$, the Shannon capacity [7] is

$$C = \mathbb{E}_{\mathbf{H}}\left\{\log\det\left(\mathbf{I}_{N_r} + \left(\frac{P}{N_t}\right)\mathbf{H}\mathbf{H}^\dagger\right)\right\}.$$

Thus for large signal power P, with CSIT or CSIR, by using multiple antennas at both the transmitter and the receiver, the channel capacity grows linearly with $\min\{N_t, N_r\}$. The $\min\{N_t, N_r\}$ factor is generally referred to as *spatial degrees of freedom*.

Next, we look at an alternate notion of capacity that is useful for non-ergodic channels for which Shannon capacity is zero.

1.4 Outage Capacity

The Shannon capacity formulation is useful for an ergodic multiple antenna fading channel, where in either each time slot or after a block of T time slots, an independent channel realization of $\mathbf{H}$ is drawn from a given distribution. T is generally referred to as the coherence time of the wireless channel. An ergodic model is valid for fast-fading case, where the fading channel coefficients change fast and the communication duration is long enough to get averaging over multiple independent blocks. Another model of interest is the non-ergodic or the slow-fading channel model, where the channel coefficients vary very slowly. To be specific, with the slow-fading model, it is assumed that at the start of the transmission, an independent realization of the channel matrix is drawn from any given distribution, but then is held fixed for the total communication duration. This model is well suited for low mobility wireless channels requiring short duration communication, where the coherence time is large enough compared to the total transmission time.

It is easy to see that the Shannon capacity of any non-ergodic channel is zero, because with increasing block length no averaging is available, and the error probability is lower bounded by a constant for any non-zero rate of transmission. To have a meaningful definition of capacity for

the non-ergodic channel, concept of outage capacity was introduced in [7], which is described as follows. Let B bits/sec/Hz be the desired rate of communication. Then channel outage at rate B is defined to be the event that the mutual information is less than B, and the outage probability is defined as

$$P_{\text{out}}(B) = \mathbb{P}(\mathsf{I}(\mathbf{x};\mathbf{y}) < B)).$$

The outage capacity $C_{\text{out}}(\epsilon)$ is defined to be the maximum rate of transmission B for which the outage probability is below a certain threshold ϵ, that is,

$$C_{\text{out}}(\epsilon) := \max_{P_{\text{out}}(B)\leq\epsilon} B.$$

The outage capacity can be interpreted as the maximum possible rate for which there exists a code whose probability of error can be made arbitrarily small for all but a set of $\mathbf{H}$, whose total probability is less than ϵ. Thus, in essence, outage capacity is the maximum rate which is guaranteed with success probability of at least $(1-\epsilon)$.

The outage capacity formulation naturally extends to a wireless network and will be used to define a throughput metric for a wireless network called the transmission capacity in Chapter 2.

For the multiple antenna channel (1.2), with an ML decoder, the outage probability can be simplified to obtain

$$P_{\text{out}}(B) = \inf_{\mathbf{Q},\mathbf{Q}\geq 0, tr(\mathbf{Q})\leq N_t} \mathbb{P}\left(\log\det\left(\mathbf{I}_{N_r} + \frac{P}{N_t}\mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right) < B\right),$$

where $\mathbf{Q}$ is the covariance matrix of the transmitted vector $\mathbf{x}$.

For the most popular Rayleigh channel fading model, where each entry of the channel matrix $\mathbf{H}$ is i.i.d. $\mathcal{CN}(0,1)$ distributed, the distribution of the maximum mutual information expression $\log\det(\mathbf{I}+\mathbf{H}\mathbf{H}^\dagger)$ is unknown. Consequently, finding the outage capacity of the multiple antenna channel has remained unsolved. The mutual information expression can be significantly simplified if instead of an ML decoder, we use a ZF decoder, where different data streams sent by the transmitter are decoupled before decoding. From [4] for (1.6), with N_t independent data streams, and assuming that each data stream is required to have rate B, and outage probability constraint ϵ, the outage capacity of a $N_t \times N_r,\ N_r \geq N_t$ multiple antenna channel with ZF decoder is

$$C_{\text{out}}^{ZF}(\epsilon) = \max_{\mathbb{P}(\log(1+|g|^2)<B)\leq\epsilon} N_t B, \tag{1.9}$$

where $|g|^2$ is the signal power after zero forcing other $N_t - 1$ signal components and hence $|g|^2 \sim \chi^2(2(N_r - N_t + 1))$. Thus, the outage capacity of the multiple antenna channel with ZF decoder can be found by using the CDF of a χ^2 distributed random variable with $N_r - N_t + 1$ degrees of freedom.

After discussing the point-to-point communication scenario, we next look at the signal interactions in a wireless network, which is of primary interest in this book.

1.5 Wireless Network Signal Model

Consider a wireless network with K nodes, where the nth node's location is denoted by T_n. We assume that each node has N antennas for transmission and reception. The received signal at the

mth node is

$$\mathbf{y}_m = d_{mm}^{-\alpha}\mathbf{H}_{mm}\mathbf{x}_m + \sum_{k=1,k\neq m}^{K} \mathbf{1}_k d_{km}^{-\alpha}\mathbf{H}_{km}\mathbf{x}_k + \mathbf{w}_m, \tag{1.10}$$

where $\mathbf{x}_m$ is the transmitted signal from transmitter T_m, d_{km} and $\mathbf{H}_{km} \in \mathbb{C}^{N\times N}$ are the distance and channel coefficient matrix between the kth transmitter and the mth receiver node, respectively, $\mathbf{1}_k$ is the indicator function that represents whether kth node is active/transmitting, and $\mathbf{w}_m$ is the AWGN vector. We will assume throughout this book that each entry of $\mathbf{H}_{km}$ is i.i.d. and $\mathcal{CN}(0,1)$ distributed to model a Rayleigh fading channel. Scheduling policy of transmitter k defines the indicator function $\mathbf{1}_k$ that critically determines the network performance, since it controls the amount of interference seen at any receiver.

Remark 1.5.1 *In a wireless network, the signal transmitted by the mth node ($\mathbf{x}_m$) could be its own signal or a signal that is being forwarded by it to facilitate communication between some other source–destination pair, in which case $\mathbf{x}_m$ is function of the received signal in previous time slots.*

In a wireless network, there are various source–destination configurations possible, for example, a single node might be interested in communicating with a single node (unicast), few nodes (broadcast), or all nodes (multicast), or two different nodes might be interested in communicating with the same node, or a relay might be helping a single source–destination pair communicate. A wireless network can essentially be broken down into four building block channels that are listed as follows:

- Interference channel: A canonical example of an interference channel is where there are two source–destination pairs that are interested in receiving their own information and do not care about the other pair's data.
- Relay channel: In its simplest form, in a relay channel, a single node (designated relay) helps a single source–destination pair communicate. In more complex form, multiple relays can help multiple source–destination pairs to communicate.
- Broadcast channel: The simplest broadcast channel is where a single source wants to communicate with two destinations, where the information content for the two destinations has both common and private components. Extensions to multiple destinations are also possible.
- Multiple access channel: A multiple access channel is where multiple sources want to communicate with a single destination at the same time.

1.5.1 Information Theoretic Limits of Wireless Networks

From an information theoretic point of view, one of the basic questions is to find the limit on reliable rate of information transfer (Shannon capacity) in a wireless network. In comparison to point-to-point communication, where the Shannon capacity is a scalar quantity, the Shannon capacity of a wireless network is a region spanned by rate tuples corresponding to various source–destination pairs that can be simultaneously supported, such that the error probability can be made arbitrarily small for large block lengths.

Clearly, finding the Shannon capacity of the four basic building block channels discussed above is a prerequisite for finding the Shannon capacity of a wireless network. Unfortunately, the Shannon

capacity of the relay channel and the interference channel is unknown, precluding the possibility of finding the Shannon capacity of a wireless network. Today, one of the biggest challenges (some people call it the "holy grail") in information theory is to find the Shannon capacity of a wireless network. A simple upper bound on the Shannon capacity of the wireless network can be found using the Fano's inequality, called the cut-set bound; however, there is no known strategy with achievable rates close to the upper bound.

Primary reason for the intractability of the Shannon capacity of the wireless network is the strict reliability constraint that requires the error probability to be arbitrarily small for large block lengths and complicated signal interaction resulting in interference, which is hard to charcterize. In practice, however, if the SINR seen at the receiver is above a threshold, communication can be deemed successful with sufficient reliability. This SINR model of successful transmission gives rise to the concept of *transmission capacity* [1] and *throughput capacity* [2] that were introduced to understand the fundamental limits on the overall throughout of the wireless network as a function of the number of nodes.

Transmission capacity definition uses the concept of outage probability as a reliability metric. Assuming that all the source–destination pairs are at a fixed distance from each other, the transmission capacity is defined to be the maximum density of nodes per unit area such that the outage probability at each node is below a threshold for a fixed rate of transmission by each node. In essence, given a quality of service QoS constraint (rate of transmission and outage probability), transmission capacity counts the maximum number of concurrently allowed transmissions in a given area. In Chapter 2, we discuss the concept of transmission capacity in detail and derive it using tools from stochastic geometry. We also quantify the effects of multiple antenna nodes, interference cancelation, spectrum sharing on the transmission capacity and bi-directional communication in Chapters 3 and 4. We also highlight the use of stochastic geometric tools to analyze some important performance measures in cellular wireless network in Chapter 5, which are hard to find otherwise.

The alternate notion of capacity, called the per-node throughput capacity for a random wireless network with density n is defined to be $\mathtt{t}(n)$ bits/sec/Hz, if there is a spatial and temporal scheduling strategy, such that each node can send $\mathtt{t}(n)$ bits/sec/Hz on average to its randomly chosen destination with high probability. The network wide throughput capacity is obtained by multiplying the density of nodes n with $\mathtt{t}(n)$. In Chapter 9, we will discuss the concept of throughput capacity and derive the seminal result of [2], which showed that the network-wide throughput capacity of a random wireless network with density n, scales as order $\sqrt{n}$ under the SINR model. The order $\sqrt{n}$ scaling is specific to the SINR model and is not an information theoretic limit. We next derive an information theoretic upper bound of order $n \log n$ on the throughput capacity and then show that a hierarchical cooperation strategy can achieve a throughput capacity of order n by using multi-antenna transmission using distributed antennas of different nodes in Chapter 9.

Both the throughput and transmission capacity yield the same scaling with respect to the number of nodes of the wireless network. Because of the use of the outage probability framework, however, quantifying the effects of advanced physical layer techniques, such as equipping nodes with multiple antennas, using successive interference cancelation, and ARQ, on the transmission capacity is easier than on throughput capacity.

An important distinction between the transmission and throughput capacity is in the averaging of successful transmission event. The transmission capacity uses an outage probability constraint

$\mathbb{P}(\mathsf{SINR} < \beta) \leq \epsilon$ and counts the number of successful transmissions satisfying the outage probability constraint. In contrast, in the throughput capacity definition, we are counting the number of nodes that can simultaneously transmit so that the SINR (realization and not the probability) for each pair of transmissions is above a threshold.

1.6 Connectivity in Wireless Networks

Global network connectivity is more or less a prerequisite for ensuring efficient operation of a wireless network. For example, in a military network, where for obvious reasons, it is imperative to have an active communication link between any pair of nodes. Similarly, in a sensor network, fusion node needs to have a path from each of nodes for data collection and processing. Even though connectivity is desirable, ensuring it is quite challenging, since even a single isolated node breaks the connectivity of the network.

In addition to connectivity, efficient routing protocols that are robust to node failures/outages, intelligent network management tools for transmission scheduling, smart application layer protocols are equally important for a smooth operation of a wireless network. In this book, however, we will restrict ourselves to studying the physical layer properties of the wireless network such as connectivity and refer the reader to [8] for the discussion on higher layer issues such as routing, link management, and scheduling etc.

Connectivity in a wireless network is defined for a variety of link connection models, for example, the disc model, and the SINR model. With the disc connection model, two nodes within a fixed distance are assumed to be connected. The motivation behind this model comes from the radio range of each node—the distance to which each node's signal can be received with sufficient strength. The disc connection model, however, assumes that simultaneously active transmitters do not interfere with each which is an idealization. A more realistic connection model is the SINR model that allows multiple nodes to transmit at the same, where a link between two nodes exists if the SINR between them is above a threshold. SINR model is far more complicated than the disc model, since it gives rise to directed links in contrast to the disc model, and the existence of a link between any pair of nodes depends on the formation of all other links.

Under any connection model, a wireless network can be naturally thought of a graph, where an edge in the graph corresponds to a link in the wireless network. Because of this association, graph theoretic tools, namely, percolation theory is used to study the connectivity properties of the graph. In particular, questions like: what is the minimum radio range required to ensure connectivity in a large wireless network and when does a connected component of unbounded size exists as a function of the density of the nodes are answered using percolation theory. The event of formation of an unbounded component is generally referred to as *percolation* and from a wireless network perspective, percolation guarantees long-range communication possibility. Percolation theory is not only useful for studying connectivity properties, but as we will see in Chapter 9, it is also useful in deriving the throughput capacity of a wireless network.

In Chapter 7, we give a brief introduction to the basics of percolation theory that are required for deriving the results presented in this book. In particular, we describe the basic ideas behind main results in discrete percolation theory over square lattice and hexagonal face lattice, and study some properties of the continuum percolation. In Chapters 7 and 8, we discuss in detail the connectivity properties of the wireless network under the disc model and the SINR connection models, respectively. For the disc model, we derive the critical radio range required for connectivity as a function of the number of nodes. For the SINR model, we show that if nodes use multiple

non-overlapping time slots/frequency bands that are of the order of the logarithm of the number of nodes, connectivity of a wireless network is ensured.

As discussed above, percolation is a slightly weaker condition compared to connectivity and only requires the formation of an unbounded connected cluster size. In general, for several network parameters, percolation can be shown to exist with high probability even when connectivity cannot be guaranteed. In Chapter 8, we study the percolation properties of the wireless network under the SINR connection model, and information theoretically secure connection model in the presence of eavesdroppers.

Bibliography

[1] S. Weber, X. Yang, J. Andrews, and G. de Veciana. 2005. "Transmission capacity of wireless ad hoc networks with outage constraints." *IEEE Trans. Inf. Theory* 51 (12): 4091–102.

[2] P. Gupta and P. Kumar. 2000. "The capacity of wireless networks." *IEEE Trans. Inf. Theory* 46 (2): 388–04.

[3] V. Tarokh, H. Jafarkhani, and A. Calderbank. 1999. "Space-time block coding for wireless communications: Performance results." *IEEE J. Sel. Areas Commun.* 17 (3): 451–60.

[4] Robert W. Heath. 2008. *Introduction to Wireless Digital Communication: A Signal Processing Perspective*.

[5] C. E. Shannon. 2001. "A mathematical theory of communication." *ACM SIGMOBILE Mobile Computing and Communications Review* 5 (1): 3–55.

[6] T. Cover and J. Thomas. 2004. *Elements of Information Theory*. John Wiley and Sons.

[7] E. Telatar. 1999. "Capacity of multi-antenna gaussian channels." *European Trans. on Telecommunications* 10 (6): 585–95.

[8] V. Garg. 2010. *Wireless Communications & Networking*. Morgan Kaufmann.

Chapter 2

Transmission Capacity of ad hoc Networks

2.1 Introduction

In the first part of this book, starting with this chapter, we focus on the single-hop model for wireless networks, where each source has a destination at a fixed distance from it and each source transmits information directly to its destination without the help of any other node in the network.

For the single-hop model, in this chapter, we begin by introducing the concept of transmission capacity of wireless networks, which measures the largest number of simultaneously allowed transmissions across space, satisfying a per-user outage probability constraint. To facilitate the analysis of transmission capacity, we assume that the nodes of the wireless network are distributed 'uniformly randomly' in space as a Poisson point process (PPP). We first describe some tools/results from stochastic geometry that are necessary for the transmission capacity analysis with PPP-distributed nodes.

Using these tools/results from stochastic geometry, we then present an exact characterization of the transmission capacity of a wireless network under the Rayleigh fading model, when each node has a single antenna and uses an ALOHA protocol. The derived results reveal the dependence of several important parameters on the transmission capacity, such as outage probability constraint, path-loss exponent, rate of transmission per node, and ALOHA transmission probability. Using the exact expression for the transmission capacity, we find the optimal transmission probability for the ALOHA protocol that maximizes the transmission capacity. For the path-loss model, where the effects of multi-path fading are ignored, we also present a lower and upper bound on the transmission capacity that are tight in the parameters of interest. The technique used in finding the lower and upper bound has wider ramifications for cases where exact expressions on the transmission capacity cannot be found.

We then present a surprising result that even when each node transmits independently using an ALOHA protocol in a wireless network with PPP-distributed node locations, the interference received from all nodes at any point in space has spatial as well as temporal correlation. The inherent-shared randomness arising out of PPP distribution on the location of nodes gives rise to this correlation. Temporal correlation impacts the performance of ARQ-type protocols, and the

results presented in this chapter help in analyzing the performance of ARQ-type protocols in wireless networks in Chapter 6.

Finally, we consider guard-zone-based and collision sense multiple access (CSMA) scheduling protocols that are *smarter* than ALOHA protocol in choosing which transmitters should be active simultaneously to maximize the transmission capacity. With the guard-zone-based strategy, only those nodes that are at a distance greater than a threshold from any receiver are allowed to transmit, thus restricting the interference seen by any receiver. With the CSMA protocol, each node measures the channel and transmits depending on a function of the channel measurement. Guard-zone-based strategy or CSMA protocol, however, does not lend itself to exact analysis because of correlations across different node transmissions, and we present only an approximate characterization of its performance that is known to be accurate by extensive simulations.

2.2 Transmission Capacity Formulation

We begin by defining a homogenous PPP as follows, which will be used to model the location of the nodes of the wireless network.

Definition 2.2.1 *For compact sets $A \subset \mathbb{R}^2, B \subset \mathbb{R}^2$, with finite area $\nu(A) < \infty, \nu(B) < \infty$, a homogenous PPP Φ with density λ is a random point process with*

$$\mathbb{P}_\Phi(\#(A) = k) = \frac{(\lambda\nu(A))^k}{k!}\mathrm{e}^{-\lambda\nu(A)} \tag{2.1}$$

and

$$\mathbb{P}_\Phi(\#(A) = k, \#(B) = m) = \mathbb{P}_\Phi(\#(A) = k)\mathbb{P}_\Phi(\#(B) = m) \tag{2.2}$$

for $A \cap B = \phi$. The specific case of A containing no nodes $k = 0$ is called the void probability of the PPP,

$$\mathbb{P}_\Phi(\#(A) = 0) = \mathrm{e}^{-\lambda\nu(A)}. \tag{2.3}$$

The first condition (2.1) requires that the number of points of a PPP lying in a finite region is a Poisson random variable with mean λ times the area of the region, where λ is the density or the expected number of points of PPP in an unit area. The second condition (2.2) states that the number of points in non-overlapping regions should be independent.

Consider the following example that shows the correspondence between PPP-distributed node locations and nodes that are located uniformly randomly in a given area of interest.

Example 2.2.2 *Let n nodes be distributed uniformly in region A with area $\nu(A)$. Then the probability that there are k nodes in region $B \subseteq A$ is binomial with parameters $\left(n, \frac{\nu(B)}{\nu(A)}\right)$. Taking the limit in the number of nodes $n \to \infty$, while fixing thedensity of nodes $\frac{n}{\nu(A)} = \lambda$, the probability that there are k nodes in region B is Poisson distributed with $\lambda\nu(B)$.*

To understand the fundamental throughput limits (capacity) of a wireless network, we model the location of nodes of a wireless network to be distributed according to a PPP. The PPP assumption corresponds to having nodes distributed uniformly random in a given area of interest. Even though this is not entirely accurate, assuming that nodes are distributed uniformly randomly is reasonable for some of the main examples of wireless networks such as sensor networks, where sensors are thrown randomly in a given area, and military or vehicular networks, where nodes are

mobile and their locations are close to being symmetrically distributed across the network. The uniformly distributed node locations assumption serves as a reasonable approximation to a realistic wireless network scenario as well as allows analytical tractability.

To be specific, we consider a wireless network located in $\mathbb{R}^2$ with the location of nodes distributed as a homogenous PPP with density λ, that is, the expected number of nodes per unit area is λ. A PPP is characterized by two properties: the probability of the number of nodes in any area A with Lebesgue measure $\nu(A)$ is Poisson distributed with parameter $\lambda\nu(A)$ and the number of nodes lying in non-overlapping areas are independent.

To avoid the overload of notation, whenever required, we let T_n and R_n denote both the node and the location of the nth transmitter and receiver, respectively. Let $\Phi_T = \{T_n\}$ be the transmitter location PPP process and $\Phi_R = \{R_n\}$ be the receiver location process. We assume that the receiver R_n corresponding to the transmitter T_n is at a fixed distance d away in a random orientation. This assumption is not really binding but made for purposes of simple exposition. Results could be extended to random distances between transmitters and receivers by taking an expectation with respect to the distance distribution.

In this chapter, we assume each transmitter and receiver to have a single antenna. Extension to multiple antennas is the subject matter of Chapter 3. We start by considering the ALOHA protocol, where each transmitter is active with probability p independently of all other transmitters. Since an independent thinning of a PPP is also a PPP (Proposition 2.3.2), the active set of nodes of a wireless network with the ALOHA protocol is also a PPP, allowing analytical tractability. More sophisticated scheduling policies provide better performance; however, they entail correlation among different transmitters breaking the PPP assumption on the transmitter locations leading to a more complicated analysis. We consider two such protocols in Sections 2.8.1 and 2.8.2.

Under this ALOHA protocol model, the received signal at receiver R_n is

$$y_n = \sqrt{P} d^{-\alpha/2} h_{nn} x_n + \sum_{m:T_m \in \Phi_T \setminus \{T_n\}} \sqrt{P} d_{mn}^{-\alpha/2} \mathbf{1}_m h_{mn} x_m + w, \tag{2.4}$$

where x_n is the signal transmitted from transmitter T_n with power P, d_{mn} and h_{mn} are the distance and the channel coefficient between T_m and receiver R_n, respectively, $\mathbf{1}_m$ is an indicator function that represents if T_m is active or not, and w is the additive white Gaussian noise (AWGN) with zero mean and unit variance. With the ALOHA protocol, $\mathbf{1}_m = 1$ with probability p and 0 with probability $1-p$. In (2.4), $P d_{mn}^{-\alpha} |h_{mn}|^2$ is the interference power of transmitter T_m at receiver R_n, and

$$I = \sum_{m:T_m \in \Phi_T \setminus \{T_n\}} P d_{mn}^{-\alpha} |h_{mn}|^2 \tag{2.5}$$

is the total interference power seen at R_0.

As stated in Assumption 1.2.1, throughout this book, we consider the path-loss function to be $d^{-\alpha}$ for ease of exposition. Most of the analysis presented in this book is applicable for more general path-loss functions such as $\frac{1}{1+d^{\alpha}}$, which compared to $d^{-\alpha}$ model does not have a singularity at 0. The $d^{-\alpha}$ path-loss function models the far field communication fairly well but results in unbounded signal power at extremely close distances.

From (2.4), the signal-to-interference-plus-noise ratio (SINR) at receiver R_n is

$$\mathsf{SINR}_n = \frac{P d^{-\alpha} |h_{nn}|^2}{\sum_{m:T_m \in \Phi_T \setminus \{T_n\}} P d_{mn}^{-\alpha} \mathbf{1}_m |h_{mn}|^2 + 1}, \tag{2.6}$$

and the outage probability $P_{\text{out}}^n(B)$ at receiver R_n is defined to be the event that the SINR is below a threshold $\beta(B)$ that is a function of rate of transmission B bits/sec/Hz,

$$P_{\text{out}}^n(B) = \mathbb{P}(\mathsf{SINR}_n \leq \beta). \tag{2.7}$$

Remark 2.2.3 *With PPP-distributed transmitter locations, the SINR is identically distributed for all receivers; hence, we drop the index n from SINR definition* (2.6) *and outage probability definition* (2.7) *and represent it as* SINR *and* $P_{\text{out}}(B)$*, respectively.*

With mutual information to be equal to $\log(1 + \mathsf{SINR})$, for B bits/sec/Hz transmission rate, $\beta(B) = 2^B - 1$. For ease of notation, we just write β in place of $\beta(B)$. We consider the quality of service requirement as the constraint on the outage probability. In particular, we assume an outage probability constraint of ϵ with transmission rate B bits/sec/Hz. The outage probability constraint of ϵ allows on average $(1-\epsilon)B$ bits/sec/Hz of successful transmission rate between any transmitter receiver pair. Since the transmitter density is λ nodes/m^2, on average, $(1-\epsilon)B\lambda$ bits/sec/Hz/m^2 can be transmitted in the wireless network. This is essentially the idea behind the concept of transmission capacity defined in [8] that captures the spatial capacity of the network or the number of simultaneously allowed transmissions under an outage probability constraint. The formal definition of transmission capacity is as follows.

Definition 2.2.4 *Assuming B bits/sec/Hz of transmission rate between any transmitter receiver pair, and an outage probability constraint of ϵ at each receiver, let*

$$\lambda^\star = \sup\{\lambda : P_{\text{out}}(B) \leq \epsilon,\ \forall n\}.$$

The transmission capacity of a wireless network with PPP-distributed nodes with density $\lambda^\star$ is defined as

$$C := \lambda^\star(1-\epsilon)B \text{ bits/sec/Hz/m}^2.$$

Assumption 2.2.5 *To keep the problem non-trivial, we assume that the power transmitted P by each transmitter is sufficient to satisfy the outage probability constraint of ϵ in the absence of interference. In the absence of interference,*

$$\mathsf{SINR} = \mathsf{SNR} = Pd^{-\alpha}|h_{nn}|^2.$$

Thus, if h_{nn}'s are Rayleigh distributed, that is, $|h_{nn}|^2 \sim \exp(1)$, then the outage probability without interference ($I = 0$ in (2.5)*) is*

$$P_{\text{out}}(B) = \mathbb{P}(\mathsf{SNR} \leq \beta) = \mathbb{P}\left(d^{-\alpha}|h_{nn}|^2 \leq \beta\right) = 1 - \exp\left(-\frac{\beta d^\alpha}{P}\right).$$

Thus, we assume that power P is such that $1 - \exp\left(-\frac{\beta d^\alpha}{P}\right) \leq \epsilon$.

To find the transmission capacity, we first need to derive an expression for the outage probability $P_{\text{out}}(B)$ in terms of λ and B. Then optimizing over the constraint $P_{\text{out}}(B) \leq \epsilon$, we can obtain $\lambda^\star$. To find the outage probability expression, we need tools from stochastic geometry, which are detailed as follows.

2.3 Basics of Stochastic Geometry

Proposition 2.3.1 *A homogenous PPP is stationary, that is, if $\Phi = \{x_1, x_2, \dots\}$ is any homogenous PPP with density λ, then $\Phi_x = \{x_1 + x, x_2 + x, \dots\}$ is also a homogenous PPP with density λ.*

Proof: Follows easily from Definition 2.2.1, since for a PPP, the probability for any number of nodes to lie in any region only depends on its Lebesgue measure/area in $\mathbb{R}^2$. □

Proposition 2.3.2 *If the points of a homogenous PPP are independently retained with probability p and removed with probability $1-p$, the resulting process is also a homogenous PPP with density λp. This is called the random thinning property of a PPP.*

Proof: Left as an excercise. □

Theorem 2.3.3 *Slivnyak's theorem: Let Φ be any homogenous PPP. Then conditioned on $x \in \Phi$, $\mathbb{P}(f(\Phi)|x \in \Phi) = \mathbb{P}(f(\Phi \cup \{x\}))$ for any function f.*

Slivnyak's theorem is an important result that allows us to compute probabilities of events conditioned on the event that there is a point of Φ at location x, which is a zero probability event. It essentially says that conditioned on the event that there is a point of Φ at location x, the distribution is equivalent to a new process that is a union of Φ and an extra point at x.

The utility of Slivnyak's theorem is illustrated by the following example:

Example 2.3.4 *Let $B \subseteq \mathbb{R}^2$ be a compact set containing the origin o. Let Φ be a PPP, and let $o \in \Phi$, that is, there is a point of the PPP at the origin. Then*

$$\mathbb{P}(\#_\Phi(B) = k | o \in \Phi) = \mathbb{P}(\#_{\Phi \cup \{o\}}(B) = k) = \mathbb{P}(\#_\Phi(B) = k-1).$$

Note that without Slivnyak's theorem, the same answer can be derived by conditioning on the event that a point of Φ is in $\mathbf{B}(o,r)$ and letting $r \to 0$.

Definition 2.3.5 *Let $\mathcal{G}$ be the family of all non-negative, bounded measurable functions $g : \mathbb{R}^d \to \mathbb{R}$ on $\mathbb{R}^d$ whose support $\{x \in \mathbb{R}^d : g(x) > 0\}$ is bounded. Let $\mathcal{F}$ be the family of all functions $f = 1 - g$, for $g \in \mathcal{G},\ 0 \le g \le 1$. Then the probability-generating functional (PGF) for a point process $\Phi = \{x_n\}$ is defined as*

$$PGF(f) = \mathbb{E}\left\{ \prod_{x_n \in \Phi} f(x_n) \right\}.$$

Theorem 2.3.6 *For a homogenous PPP Φ with density λ, the PGF is given by*

$$PGF(f) = \exp^{-\int (1-f(x))\lambda \mathrm{d}x}.$$

Theorem 2.3.6 is very useful for deriving the outage probability in a PPP-distributed wireless network by allowing us to compute the expectation of a product of functions over the entire PPP. We will make use of Theorem 2.3.6 quite frequently in the book.

Theorem 2.3.7 *Campbell's theorem: For any measurable function $f : \mathbb{R}^d \to \mathbb{R}$ and for a homogenous PPP Φ with density λ,*

$$\mathbb{E}\left\{ \sum_{n: x_n \in \Phi} f(x_n) \right\} = \int_{\mathbb{R}^d} f(x) \lambda \mathrm{d}x.$$

Using Campbell's Theorem, one can show that the interference received at the origin (or any other point using stationarity) from all points of the PPP with the path-loss model of $x^{-\alpha}$ is unbounded for any α as follows.

Example 2.3.8 *Let $I = \sum_{n:x_n\in\Phi} x_n^{-\alpha}|h_n|^2$ be the interference received at the origin from all points of the PPP, where the fading gains $|h_n|^2$ are i.i.d. with mean 1. Then from Campbell's theorem*

$$\mathbb{E}\{I\} = \mathbb{E}\left\{\sum_{n:x_n\in\Phi} x_n^{-\alpha}|h_n|^2\right\} = \mathbb{E}\left\{\sum_{n:x_n\in\Phi} x_n^{-\alpha}\right\} = \int_{\mathbb{R}^2} x^{-\alpha}\lambda \mathrm{d}x, \tag{2.8}$$

since $\mathbb{E}\{|h_n|^2\} = 1$. Since the path-loss function $x^{-\alpha}$ has a singularity at 0, the expected interference $\mathbb{E}\{I\}$ is unbounded for any value of α. However, if we can avoid interference coming from a small disc of radius ϵ around the origin, $\mathbf{B}(0,\epsilon)$, that is, somehow inhibit all points of the PPP within $\mathbf{B}(0,\epsilon)$, then $\mathbb{E}\{I\}$ will be finite. We will use this technique of inhibition to compute the expected interference for finding meaningful bounds on the transmission capacity.

Definition 2.3.9 *Marked Poisson Process: A marked PPP Φ_M is obtained by attaching a mark m_n to each point of $x_n \in \Phi$, where Φ is a PPP and $\Phi_M = \{(x_n, m_n) : x_n \in \Phi\}$. Marks could represent the power transmitted by point x_n, its colour, and shape of any other characteristic. The marks m_n belong to set M with distribution $\mathcal{M}$, such that for any bounded set $A \subset \mathbb{R}^2$, $\#(\Phi_M(A\times\mathsf{M})) < \infty$ almost surely, that is, the number of points of Φ_M lying in a bounded region is finite.*

The next theorem allows us to make a correspondence between a marked PPP Φ_M and a PPP defined over $\mathbb{R}^2 \times \mathsf{M}$.

Theorem 2.3.10 *Marking theorem: The following statements are equivalent*

- *a marked PPP Φ_M, where if conditioned on the points $\{x_n\}$ of the PPP Φ, the marks m_n of the marked PPP Φ_M with marks in M, are independent, with probability distribution $\mathcal{M}$ on M,*
- *a PPP on $\mathbb{R}^2 \times \mathsf{M}$ with measure $\Lambda(A\times B) = \lambda\nu(A)\times\mathcal{M}(B)$, where $\nu(A)$ is the Lebesgue measure of A.*

The proofs of Theorems 2.3.3, 2.3.6, 2.3.7, and 2.3.10 can be found in [5].

We next present three examples to illustrate how Marking theorem is useful for analyzing PPP-distributed *ad hoc* networks.

Example 2.3.11 *Finding the distribution of the L^{th} strongest interferer in a PPP. Let Φ be a PPP and let $I_n = x_n^{-\alpha}|h_n|^2$ be the interference power of the n^{th} (unordered) transmitter $x_n \in \Phi$ at the origin, where $|h_n|^2$ are i.i.d. We want to find the distribution of the L^{th} largest interference power.*

Let us define I_n to be a mark corresponding to $x_n \in \Phi$ and consider the marked PPP as

$$\Phi_M = \{(x_n, I_n) : x_n \in \Phi\}.$$

Note that $I_n \in \mathbb{R}^+$ is independent given x_n, since $|h_n|^2$ is independent. Thus, from the Marking theorem, Φ_M is equivalent to a PPP Ψ on $\mathbb{R}^2\times\mathbb{R}^+$ with an appropriate density measure $\Lambda(B)$ on subsets B of $\mathbb{R}^2\times\mathbb{R}^+$. Let us define

$$\mathcal{B}(g) = \{(x_n, I_n) : I_n > g\}$$

to be the set of points x_n of Φ such that the interference they cause at the origin is more than some threshold g. Note that $\mathcal{B}(g) \subset \Psi$ is the set of points lying in a subset of $\mathbb{R}^2\times\mathbb{R}^+$, and hence the

number of points in $\mathcal{B}(g)$ is Poisson distributed with mean equal to the density measure of the subset corresponding to $\mathcal{B}(g)$, that is,

$$\begin{aligned}\Lambda(\mathcal{B}(g)) &= \lambda\mathbb{P}(x^{-\alpha}|h|^2 > g),\\ &= \lambda\mathbb{E}_{|h|^2}\left\{\int_{0,x\in\mathbb{R}^2}^{(|h|^2/g)^{1/\alpha}} \mathrm{d}x\right\},\\ &\overset{(a)}{=} \lambda\mathbb{E}_{|h|^2}\left\{\int_{0,x\in\mathbb{R}}^{(|h|^2/g)^{1/\alpha}} 2\pi x\ \mathrm{d}x\right\},\\ &= \lambda\int_0^\infty \int_{0,x\in\mathbb{R}}^{(|h|^2/g)^{1/\alpha}} 2\pi x\,\mathrm{d}x f_h(h)dh,\end{aligned}$$

where we get the term $2\pi x$ in (a) by changing the integration from $\mathbb{R}^2$ to $\mathbb{R}^1$. The cumulative distribution function $F_L(g)$ of the L^{th} strongest interferer is equal to the probability that there are less than or equal to $L-1$ points in $\mathcal{B}(g)$.

Marking theorem can also be used to obtain many other interesting results as described in the next two examples.

Example 2.3.12 *Consider a random disc process on $\mathbb{R}^2$ whose centers form a PPP Φ with density λ, and where for each point $x_n \in \Phi$ an independent random radius r_n (mark) is selected with distribution $F_{\mathsf{R}}(\mathsf{r}) = \mathbb{P}(\mathsf{R} \le \mathsf{r})$. From the Marking theorem, this process is equivalent to a PPP on $\mathbb{R}^2 \times \mathbb{R}^+$ with density measure $\Lambda(A \times [0,\ \mathsf{r}]) = \lambda\nu(A)F_{\mathsf{R}}(\mathsf{r})$ for compact $A \subset \mathbb{R}^2$ and $\mathsf{r} > 0$.*

A disc $\mathbf{B}(x, \mathsf{r})$ contains the origin o if and only if (x, r) belongs to set $S = \{(x, \mathsf{r}) : x \in \mathbf{B}(o, \mathsf{r})\}$, that is, if x lies in the circle with center as origin and radius r. Since S is a subset of $\mathbb{R}^2 \times \mathbb{R}^+$, the number of points lying in S (also the number of discs containing the origin) are Poisson distributed with mean that is equal to the density measure of S. From the Marking theorem the density measure of S is

$$\Lambda(S) = \lambda\int_S F_{\mathsf{R}}(d\mathsf{r})\mathrm{d}x = \lambda\int_0^\infty \nu(\mathbf{B}(o,\mathsf{r}))F_{\mathsf{R}}(d\mathsf{r})\mathrm{d}x = \lambda\mathbb{E}(\pi\mathsf{R}^2).$$

[1]Next, we present an application of the Marking theorem to prove the random thinning property of the PPP (Proposition 2.3.2).

Example 2.3.13 *Random Thinning: With each point x_n of a PPP Φ on $\mathbb{R}^2$ associate a mark $m_n \in \{0, 1\}$ with $\mathbb{P}(m_n = 1) = p$ independently of all other points $x_m, m \ne n$. Then from the Marking theorem, this marked PPP is equivalent to a PPP on $\mathbb{R}^2 \times \{0, 1\}$, with density measure*

$$\Lambda(A \times \{y\}) = \lambda\nu(A)\left(yp + (1-y)(1-p)\right)$$

for compact $A \subset \mathbb{R}^2$ and $y \in \{0, 1\}$. Define a subset

$$S = \{(A, m_n) : m_n = 1\} \subset \mathbb{R}^2 \times \{0, 1\}$$

that corresponds to the thinned version of the original PPP Φ. The number of points in S is Poisson distributed with density measure $\Lambda(S) = \lambda\nu(A)p$.

[1]Example 2.3.12 is taken from the lecture notes of Gustavo De Veciana, ECE Department, the University of Texas at Austin.

Now with relevant background of stochastic geometry tools at hand, we proceed toward analyzing the outage probability (2.7) and consequently the transmission capacity. Note that the outage probability (2.7) is invariant to the choice of any transmitter receiver pair because of the stationarity of the PPP. Thus, without any loss of generality, we consider the pair (T_0, R_0) for the transmission capacity analysis. From (2.6), the SINR for the (T_0, R_0) pair is

$$\mathsf{SINR} = \frac{Pd^{-\alpha}|h_{00}|^2}{\sum_{m:T_m \in \Phi_T \setminus \{T_0\}} Pd_{m0}^{-\alpha} \mathbf{1}_m |h_{m0}|^2 + 1}.$$

Remark 2.3.14 *Since we have considered a typical transmitter receiver pair (T_0, R_0), to derive the outage probability $\mathbb{P}(\mathsf{SINR} \leq \beta)$, we need to know the distribution of interference I received by R_0, where*

$$I := \sum_{m:T_m \in \Phi_T \setminus \{T_0\}} Pd_{m0}^{-\alpha} \mathbf{1}_m |h_{m0}|^2$$

conditioned on a transmitter being located at T_0.

From Slivnyak's theorem (Theorem 2.3.3), we know that conditioned on the event that there is a transmitter at T_0, the equivalent point process is $\Phi_T^0 = \Phi_T \cup \{T_0\}$. Thus, now working with Φ_T^0, for communication between transmitter T_0 and receiver R_0, all transmitters belonging to $\Phi_T^0 \setminus \{T_0\} = \Phi_T$ are interferers. Thus, the conditional interference seen at the receiver R_0 is $I^0 := \sum_{m:T_m \in \Phi_T} Pd_{m0}^{-\alpha} \mathbf{1}_m |h_{m0}|^2$.

Therefore, the conditional distribution of interference I^0 seen at receiver R_0 is the sum of interferences from all points of the homogenous PPP with density λ, which is also called as the shot-noise process [11]. Thus, the conditional outage probability $P_{\mathrm{out}}(B)$ is equal to

$$\begin{aligned} \mathbb{P}(\mathsf{SINR} \leq \beta) &= \mathbb{P}\left(\frac{Pd^{-\alpha}|h_{00}|^2}{\sum_{m:T_m \in \Phi_T \setminus \{T_0\}} Pd_{m0}^{-\alpha} \mathbf{1}_m |h_{m0}|^2 + 1} \,\middle|\, T_0 \text{ is a transmitter} \right), \\ &= \mathbb{P}\left(\frac{Pd^{-\alpha}|h_{00}|^2}{\sum_{m:T_m \in \Phi} Pd_{m0}^{-\alpha} \mathbf{1}_m |h_{m0}|^2 + 1} \right), \end{aligned} \tag{2.9}$$

where in the last step, we have replaced Φ_T with Φ that is also a PPP with density λ to avoid confusion whether it contains T_0 or not.

Now we are ready to derive a closed form expression for the outage probability, and consequently the transmission capacity, as described in the next section.

2.4 Rayleigh Fading Model

In this section, we consider the received signal model of (2.4), when the fading channel magnitudes h_{nm} are i.i.d. exponentially distributed (Rayleigh fading) across different users n, m. Rayleigh fading is the most popular fading model for analyzing wireless communication systems and represents the scenario of richly scattered fading environment. In Section 2.5, we will consider just the path-loss model with no fading, that is, $h_{nm} = 1$, which models the line-of-sight communication, and find tight bounds on the transmission capacity.

2.4.1 Derivation of Transmission Capacity

Theorem 2.4.1 *The outage probability at a typical receiver R_0 with PPP-distributed transmitter locations with density λ is*

$$P_{\text{out}}(B) = 1 - \exp\left(-\frac{d^\alpha \beta}{P}\right) \exp\left(-\lambda p c \beta^{\frac{2}{\alpha}} d^2\right),$$

where $c = \frac{2\pi\Gamma(\frac{2}{\alpha})\Gamma(1-\frac{2}{\alpha})}{\alpha}$ and $\Gamma(.)$ is the Gamma function. Hence, under an outage probability constraint of ϵ, the transmission capacity is

$$C = \frac{-\ln(1-\epsilon) - \frac{d^\alpha \beta}{P}}{c\beta^{\frac{2}{\alpha}} d^2}(1-\epsilon)B \ \text{ bits/sec/Hz/m}^2$$

Remark 2.4.2 *Recall from Assumption 2.2.5, $-\ln(1-\epsilon) > \frac{d^\alpha \beta}{P}$. Thus, the transmission capacity is always non-negative.*

Proof: From the outage probability definition (2.7) and conditional SINR distribution with a transmitter located at T_0 (2.9),

$$\begin{aligned} P_{\text{out}}(B) &= \mathbb{P}\left(\frac{Pd^{-\alpha}|h_{00}|^2}{\sum_{m:\ T_m\in\Phi} P\mathbf{1}_m d_{m0}^{-\alpha}|h_{m0}|^2 + 1} \le \beta\right), \\ &\overset{(a)}{=} 1 - \mathbb{E}\left\{\exp\left(-\frac{d^\alpha\beta}{P}\left(\sum_{m:\ T_m\in\Phi} P\mathbf{1}_m d_{m0}^{-\alpha}|h_{m0}|^2 + 1\right)\right)\right\}, \\ &= 1 - \exp\left(-\frac{d^\alpha\beta}{P}\right)\mathbb{E}\left\{\exp\left(-d^\alpha\beta \sum_{m:\ T_m\in\Phi} \mathbf{1}_m d_{m0}^{-\alpha}|h_{m0}|^2\right)\right\}, \end{aligned}$$

where (a) follows by taking the expectation with respect to $|h_{00}|^2$, where $|h_{00}|^2 \sim \exp(1)$. It remains to take expectation with respect to $|h_{m0}|^2 \sim \exp(1)$, and ALOHA protocol's indicator variable $\mathbf{1}_m$, that is 1 with probability p and 0 otherwise. We first take the expectation with respect to $|h_{m0}|^2$ that are i.i.d. $\forall\ m$, and obtain

$$P_{\text{out}}(B) = 1 - \exp\left(-\frac{d^\alpha\beta}{P}\right)\mathbb{E}\left\{\prod_{m:\ T_m\in\Phi}\left(\frac{1}{1+\mathbf{1}_m d^\alpha\beta d_{m0}^{-\alpha}}\right)\right\}$$

Now we take the expectation with respect to the ALOHA protocol indicator function $\mathbf{1}_m$. Note that

$$\frac{1}{1+\mathbf{1}_m d^\alpha\beta d_{m0}^{-\alpha}} = \frac{\mathbf{1}_m}{1+d^\alpha\beta d_{m0}^{-\alpha}} + 1 - \mathbf{1}_m.$$

Thus,

$$\mathbb{E}\left\{\frac{1}{1+\mathbf{1}_m d^\alpha\beta x^{-\alpha}}\right\} = \frac{p}{1+d^\alpha\beta d_{m0}^{-\alpha}} + 1 - p. \tag{2.10}$$

Hence,

$$\begin{aligned}
P_{\text{out}}(B) &\overset{(b)}{=} 1-\exp\left(-\frac{d^{\alpha}\beta}{P}\right)\mathbb{E}\left\{\prod_{x\in\Phi}\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)\right\},\\
&\overset{(c)}{=} 1-\exp\left(-\frac{d^{\alpha}\beta}{P}\right)\exp\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)\mathrm{d}x\right),\\
&= 1-\exp\left(-\frac{d^{\alpha}\beta}{P}\right)\exp\left(-\lambda\int_{\mathbb{R}^2}\frac{pd^{\alpha}\beta x^{-\alpha}}{1+d^{\alpha}\beta x^{-\alpha}}\mathrm{d}x\right),\\
&= 1-\exp\left(-\frac{d^{\alpha}\beta}{P}\right)\exp\left(-2\pi\lambda\int_{0}^{\infty}\frac{pd^{\alpha}\beta x^{-\alpha}}{1+d^{\alpha}\beta x^{-\alpha}}x\mathrm{d}x\right),\\
&\overset{(d)}{=} 1-\exp\left(-\frac{d^{\alpha}\beta}{P}\right)\exp\left(-\lambda pc\beta^{\frac{2}{\alpha}}d^{2}\right),
\end{aligned}$$

where (b) follows by replacing the distance of the mth interferer d_{m0} by x, (c) follows by using the probability-generating function of the PPP (Theorem 2.3.6), and finally in (d) $c=\frac{2\pi\Gamma(\frac{2}{\alpha})\Gamma(1-\frac{2}{\alpha})}{\beta}$ and Γ is the Gamma function. The transmission capacity expression follows easily by using the outage probability constraint of $P_{\text{out}}(B)\leq\epsilon$. □

In Theorem 2.4.1, we derived the exact expression for the transmission capacity using the probability-generating functional of the PPP. The probability-generating functional allowed us to compute the expectation of a product of functions over the entire PPP. Also, the fact that the channel coefficients are Rayleigh distribution was instrumental in allowing us to derive the closed form distribution using the probability-generating functional. Instead of using the probability-generating functional, an alternate way to compute an exact expression for the transmission capacity is via the use of the Laplace transform of the shot-noise process as described in [11].

The derived expression reveals the exact dependence of critical parameters such as the outage probability constraint ϵ, rate B, and distance d between transmitter and receiver on the transmission capacity. To see how does the transmission capacity scales with the outage probability constraint ϵ, it is useful to look at the regime of small values of ϵ that corresponds to an extremely strict outage probability constraint. For small values of ϵ, using the Taylor series expansion of $\log$, the transmission capacity is seen to be directly proportional to the outage probability constraint of ϵ. So tightening the outage probability constraint leads to a linear fall in the transmission capacity.

To reveal the interplay between distance d between any transmitter receiver pair and the transmission capacity, we look at the interference limited regime, where the interference power at any receiver is much larger than the AWGN power that is,

$$I=\sum_{T_m\in\phi}Pd_{m0}^{-\alpha}|h_{m0}|^2>>1,$$

and we can ignore the AWGN contribution safely without losing accuracy. Ignoring the AWGN term that gives rise to $-d^{\alpha}\beta$ term in the numerator of the transmission capacity expression in Theorem 2.4.1, the transmission capacity scales as $\Theta(d^{-2})$. Thus, Theorem 2.4.1 reveals an interesting spatial packing relationship, where the transmission capacity can be interpreted as the packing of as many simultaneous spatial transmissions where each transmission occupies an area

of $\Theta(d^2)$. This packing relationship is similar to the $\sqrt{n}$ scaling of throughput capacity result of [2] (discussed in Chapter 9), where n nodes are distributed uniformly in an unit area. Throughput capacity measures the sum of the rate of successful transmissions achievable between all pairs of nodes simultaneously with high probability, where source destination pairs are randomly chosen. In this section's model, n corresponds to the density λ and each communication happens over a fixed distance d, while in the throughput capacity model, the expected distance between each source destination pair is a constant. Thus, the transport capacity, that is, the capacity multiplied with the distance is same for the transmission capacity and throughput capacity metric, since from Theorem 2.4.1, $\lambda \propto \frac{1}{d^2}$, and transport capacity is $\lambda d = \sqrt{n}$, similar to the transport capacity of order $\sqrt{n}$.

After deriving the transmission capacity with the Rayleigh fading model, next, we consider the path-loss model, where the multi-path fading component is ignored. At first, this might appear futile, and to make it appear even more ridiculous, we will only obtain bounds on the transmission in contrast to an exact expression. The real advantage of the presented bounding techniques is that they are extremely useful in analyzing the transmission capacity of many advanced signal processing techniques, where exact expressions for transmission capacity cannot be found.

2.5 Path-Loss Model

In this section, we consider a slightly restrictive path-loss model, that does not account for multi-path fading and where the received signal at a receiver R_n is given by

$$y_n = \sqrt{P} d^{-\alpha/2} x_n + \sum_{m:T_m \in \Phi_T \setminus \{T_n\}} \sqrt{P} d_{mn}^{-\alpha/2} x_m + w, \tag{2.11}$$

where compared to (2.4), $h_{nm} = 1, \ \forall, \ n, m$, and we have absorbed the ALOHA parameter $\mathbf{1}_m$ into the density of the PPP Φ_T which is now equal to $p\lambda$ (Proposition 2.3.2). Generally, with a simplified model, the analysis becomes easier. Finding an exact expression for the transmission capacity is one exception, where it is known only for the Rayleigh fading model, and not for the path-loss model. Tight lower and upper bounds on the transmission capacity [1] are, however, known with the path-loss model, and described in this section.

From (2.11), the SINR at receiver R_n is

$$\mathsf{SINR} = \frac{P d^{-\alpha}}{\sum_{m:T_m \in \Phi_T \setminus \{T_n\}} P d_{mn}^{-\alpha} \mathbf{1}_m + 1}. \tag{2.12}$$

As before, we consider a typical transmitter receiver pair (T_0, R_0) and similar to (2.9), the outage probability conditioned on the event that there is a transmitter at T_0 is given by

$$\begin{aligned} P_{\text{out}}(B) &= \mathbb{P}\left(\frac{P d^{-\alpha}}{\sum_{m:T_m \in \Phi_T \setminus \{T_n\}} P d_{mn}^{-\alpha} + 1} \leq \beta \,\middle|\, T_0 \in \Phi_T \right), \\ &\overset{(a)}{=} \mathbb{P}\left(\frac{P d^{-\alpha}}{\sum_{m:T_m \in \Phi_T} P d_{mn}^{-\alpha} + 1} \leq \beta \right), \\ &\overset{(b)}{=} \mathbb{P}\left(I \geq \frac{d^{-\alpha}}{\beta} - \frac{1}{P} \right), \end{aligned} \tag{2.13}$$

where (a) follows from the Slivnyak's Theorem, and (b) follows by defining the total interference from Φ_T, as $I := \sum_{m:T_m \in \Phi_T} d_{mn}^{-\alpha}$.

2.5.1 Upper Bound on the Transmission Capacity

Theorem 2.5.1 *The transmission capacity with the path-loss model is upper bounded by*

$$C_{\rm ub} = \frac{\epsilon}{\pi}\kappa^{\frac{2}{\alpha}}(1-\epsilon)B + \Theta\left(\epsilon^2\right) \text{ bits/sec/Hz/m}^2,$$

as $\epsilon \to 0$*, where* $\kappa = \frac{d^{-\alpha}}{\beta} - \frac{1}{P}$.

Proof: Let $\kappa = \frac{d^{-\alpha}}{\beta} - \frac{1}{P}$. Then the interference power received at R_0 from a transmitter T_m located in $\mathbf{B}(R_0, \kappa^{-\frac{1}{\alpha}})$ is $\geq \kappa$. Thus, by using the definition of κ in (2.13), it is clear that $\#\mathbf{B}(R_0, \kappa^{-\frac{1}{\alpha}}) > 0$ is sufficient to cause outage at R_0, since $I > \kappa$.

To lower bound the outage probability (upper bound the transmission capacity), we assume that there is at least one transmitter from Φ_T in the disc with radius $\kappa^{-\frac{1}{\alpha}}$ centered at R_0. Consequently, the total interference $I > \kappa$, and the outage is guaranteed. Thus, $P_{\rm out}(B) \geq \mathbb{P}\left(\#(\mathbf{B}(R_0, \kappa^{-\frac{1}{\alpha}})) > 0\right)$. From the void probability of the PPP, we know that

$$\mathbb{P}\left(\#(\mathbf{B}(R_0, \kappa^{-\frac{1}{\alpha}})) > 0\right) = 1 - \exp^{-\lambda\pi\kappa^{\frac{2}{\alpha}}}.$$

With the given outage probability constraint of $P_{\rm out}(B) \leq \epsilon$, we find an upper bound on the largest permissible λ to be $\lambda_{\rm ub} = -\ln(1-\epsilon)\frac{1}{\pi}\kappa^{-\frac{2}{\alpha}}$. Expanding, we get

$$\lambda_{\rm ub} = \frac{\epsilon}{\pi}\kappa^{\frac{2}{\alpha}} + \Theta\left(\epsilon^2\right). \tag{2.14}$$

Corresponding upper bound on the transmission capacity follows immediately by noting that $C_{\rm ub} = \lambda_{\rm ub}(1-\epsilon)B$. □

Keeping ALOHA probability p separately in (2.13), we get $\lambda_{\rm ub} = \frac{\epsilon}{\pi p}\kappa^{\frac{2}{\alpha}} + \Theta\left(\epsilon^2\right)$ and the upper bound on the transmission capacity identical to (2.14), since $C_{\rm ub} = p\lambda_{\rm ub}(1-\epsilon)B$.

2.5.2 Lower Bound on the Transmission Capacity

Theorem 2.5.2 *The transmission capacity with the path-loss model is lower bounded by*

$$C_{\rm lb} = \left(1 - \frac{1}{\alpha}\right)\frac{\epsilon}{\pi}\kappa^{\frac{2}{\alpha}}(1-\epsilon)B + \Theta\left(\epsilon^2\right) \text{ bits/sec/Hz/m}^2.$$

Proof: Finding this lower bound is slightly more involved than the upper bound. Consider the typical receiver R_0, for which we divide $\mathbb{R}^2$ into two regions, near field $\mathbf{B}(R_0, s)$ and far field $\mathbb{R}^2 \backslash \mathbf{B}(R_0, s)$ for some s that will be chosen later. Let us define two events

$$\mathcal{E}_N = \{\#(\mathbf{B}(R_0, s) > 0)\}$$

which corresponds to having at least one transmitter in disc $\mathbf{B}(R_0, s)$, and

$$\mathcal{E}_F = \left\{ \sum_{m: T_m \in \Phi_T, T_m \in \mathbb{R}^2 \backslash \mathbf{B}(R_0, s)} d_{mn}^{-\alpha} > \kappa \right\}$$

which corresponds to the case that the interference received at R_0 from transmitters lying outside of $\mathbf{B}(R_0, s)$ is more than κ, and hence sufficient to cause outage.

Let $\mathcal{E} = \mathcal{E}_N \cup \mathcal{E}_F$. Then note that for $s \leq \kappa^{\frac{-1}{\alpha}}$, if outage happens (2.13), that is, $I > \kappa$, then either there is a transmitter in $\mathbf{B}(R_0, s)$ or the interference from transmitters lying outside $\mathbf{B}(R_0, s)$ is more than κ. Thus, outage implies either of the two events $\mathcal{E}_N$ or $\mathcal{E}_F$ is true. Hence, we have

$$P_{\text{out}}(B) \leq \mathbb{P}(\mathcal{E}).$$

Moreover, because of the PPP property, events $\mathcal{E}_N$ and $\mathcal{E}_F$ are independent since they are defined over non-overlapping regions. Thus, we bound $\mathbb{P}(\mathcal{E})$ to get a lower bound on the transmission capacity.

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &= \mathbb{P}(\mathcal{E}_N \cup \mathcal{E}_F), \\ &= \mathbb{P}(\mathcal{E}_N) + \mathbb{P}(\mathcal{E}_F) - \mathbb{P}(\mathcal{E}_N)\mathbb{P}(\mathcal{E}_F),\end{aligned}$$

where we have used the independence of $\mathcal{E}_N$ and $\mathcal{E}_F$.

Now we try to find the largest density λ for which $\mathbb{P}(\mathcal{E}) \leq \epsilon$. It is sufficient for our purpose since $\mathbb{P}(\mathcal{E}) \leq \epsilon$ implies $P_{\text{out}}(B) \leq \epsilon$. We can write $\mathbb{P}(\mathcal{E}) \leq \epsilon$ equivalently as $\mathbb{P}(\mathcal{E}_N) \leq \epsilon_1$ and $\mathbb{P}(\mathcal{E}_F) \leq \epsilon_2$ such that $\epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2 \leq \epsilon$. Defining

$$\lambda_N = \sup\{\lambda | \mathbb{P}(\mathcal{E}_N) \leq \epsilon_1\}$$

and

$$\lambda_F = \sup\{\lambda | \mathbb{P}(\mathcal{E}_F) \leq \epsilon_2\},$$

we get the lower bound on the optimal density $\lambda^\star$ to be

$$\lambda^\star \geq \sup_{\epsilon_1 \geq 0, \epsilon_2 \geq 0, \epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2 \leq \epsilon} \{\inf\{\lambda_N, \lambda_F\}\}. \tag{2.15}$$

From (2.14), we know that

$$\lambda_N = \frac{\epsilon}{\pi} s^{-\frac{2}{\alpha}} + \Theta\left(\epsilon^2\right). \tag{2.16}$$

For computing λ_F, we make use of the Chebyeshev's inequality and get

$$\mathbb{P}(\mathcal{E}_F) = \mathbb{P}\left(\sum_{m:T_m \in \Phi_T, T_m \in \mathbb{R}^2 \setminus \mathbf{B}(R_0,s)} d_{mn}^{-\alpha} > \kappa\right) \leq \frac{\mathsf{var}}{(\kappa - \mathsf{m})^2}, \tag{2.17}$$

where

$$\mathsf{var} = \text{Var}\left(\sum_{m:T_m \in \Phi_T, T_m \in \mathbb{R}^2 \setminus \mathbf{B}(R_0,s)} P d_{mn}^{-\alpha}\right) = \frac{\pi}{\alpha - 1} s^{2(1-\alpha)} \lambda,$$

and

$$\mathsf{m} = \mathbb{E}\left\{\sum_{m:T_m \in \Phi_T, T_m \in \mathbb{R}^2 \setminus \mathbf{B}(R_0,s)} P d_{mn}^{-\alpha}\right\} = \frac{\pi}{\alpha - 2} s^{2-\alpha} \lambda,$$

computed directly using the Campbell's theorem (Theorem 2.3.7).

By equating the bound on $\mathbb{P}(\mathcal{E}_F)$ (2.17) to ϵ_2, and keeping the dominant terms, we get

$$\lambda_F = \frac{(\alpha-1)\kappa^2}{\pi^2} s^2(\alpha-1)\epsilon_2 + \Theta(\epsilon_2^2). \tag{2.18}$$

From (2.15), we know that for a given ϵ_1, ϵ_2 pair, the optimal solution satisfies $\lambda_N = \lambda_F$. Equating $\lambda_N = \lambda_F$ from (2.16) and (2.18), we get

$$s = \left(\frac{\epsilon_1}{\epsilon_2}\right)^{\frac{1}{2\alpha}} ((\alpha-1)\kappa)^{\frac{-1}{2\alpha}}.$$

Thus, for a given ϵ_1, ϵ_2 pair, by substituting for s, we get

$$\lambda_N = \lambda_F = (\alpha-1)^{\frac{1}{\alpha}} \kappa^{\frac{2}{\alpha}} \frac{1}{\pi} \epsilon_1^{1-\frac{1}{\alpha}} \epsilon_2^{\frac{1}{\alpha}}. \tag{2.19}$$

Moreover, for small outage probability constraint ϵ, $\epsilon_2 = \epsilon - \epsilon_1 + \Theta(\epsilon^2)$ and to get the lower bound, we need to solve,

$$\max_{\epsilon=\epsilon_1+\epsilon_2} \lambda_N.$$

Using (2.19), the optimum is $\epsilon_1^\star = \left(1-\frac{1}{\alpha}\right)\epsilon$ and $\epsilon_2^\star = \frac{\epsilon}{\alpha}$, and we get the required lower bound (2.15) on $\lambda^\star$ as

$$\lambda_N = \lambda_{\text{lb}} = \left(1-\frac{1}{\alpha}\right) \kappa^{\frac{2}{\alpha}} \frac{1}{\pi} \epsilon + \Theta(\epsilon^2). \tag{2.20}$$

Required lower bound on the transmission capacity $C_{\text{lb}} = \lambda_{\text{lb}}(1-\epsilon)B$. □

The derived upper and lower bounds on the transmission capacity for the path-loss model have the exact same scaling in terms of the parameters κ (that depends on d, and β) and ϵ (the outage probability constraint), and only differ in constants. Thus, this technique of dividing the overall interference into two regions and using simple void probability expressions and Chebyeshev's inequality is powerful enough to derive meaningful expressions for the transmission capacity. These bounding techniques will come in handy when we analyze more complicated protocols and techniques such as multiple antennas, and ARQ protocols, and so on.

Similar to the Rayleigh fading model, comparing the upper and lower bounds derived in Theorems 2.5.1 and 2.5.2, using the definition of κ, it is clear that the transmission capacity is inversely proportional to d^2, where d is the distance between each transmitter and receiver. So operationally, the transmission capacity exhibits the same spatial packing behaviour with or without taking fading into account.

In Fig. 2.1, we plot the transmission capacity with respect to the outage probability constraint ϵ for both the Rayleigh fading and path-loss models. For the path-loss model, we plot both the derived upper and lower bounds in addition to the simulated performance. We see that the upper bound is very close to the simulated performance. We can also see that there is some performance degradation in the transmission capacity while including the multi-path fading that is Rayleigh distributed.

For both the Rayleigh fading and path-loss models, the transmission capacity expressions (in Theorems 2.4.1, 2.5.1, and 2.5.2) are independent of the ALOHA probability of access p. This happens since we have constrained the outage probability to be below a threshold ϵ, which in turn gives an upper bound on the effective density of the PPP λp, and the transmission capacity loses its dependence on p. If we define the transmission capacity as the product of the density of the PPP and the success probability of any node, which we call as *goodput*, then we can unravel the dependence of p on network performance which is presented in the next section.

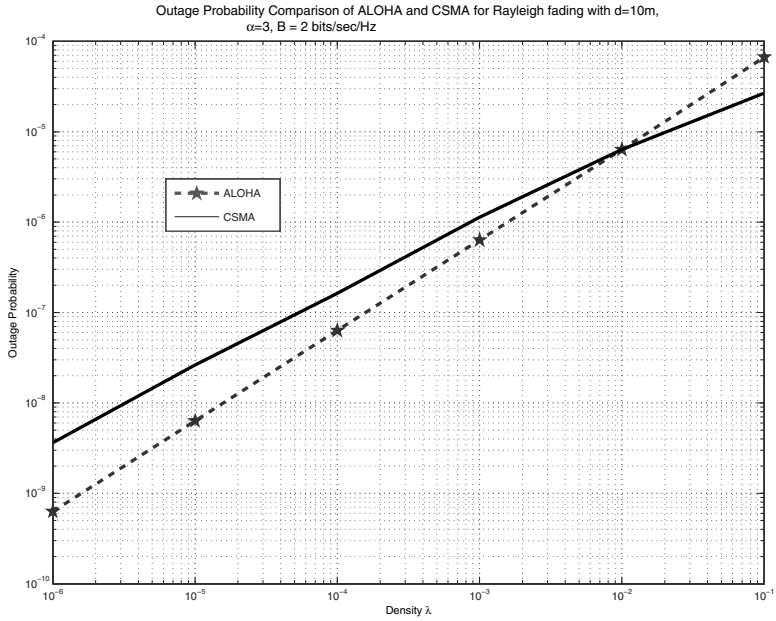

Figure 2.1: Transmission capacity with Rayleigh fading and path-loss model with ALOHA protocol.

2.6 Optimal ALOHA Transmission Probability

Let the network goodput of a wireless network be defined as

$$\mathsf{G} = \lambda(1 - P_{\text{out}}(B))B \text{ bits/sec/Hz/m}^2,$$

by accounting for λ concurrent transmissions per meters square at rate B bits/sec/Hz with outage probability $P_{\text{out}}(B)$.

Then from Theorem 2.4.1 using the expression for $P_{\text{out}}(B)$, we have

$$\mathsf{G} = \lambda p \exp\left(-\frac{d^\alpha \beta}{P}\right) \exp\left(-\lambda p c \beta^{\frac{2}{\alpha}} d^2\right), \tag{2.21}$$

where $c = \frac{2\pi\Gamma(\frac{2}{\alpha})\Gamma(1-\frac{2}{\alpha})}{\beta}$. Ignoring the AWGN contribution $\exp\left(-\frac{d^\alpha\beta}{P}\right)$, we get

$$\mathsf{G} = \lambda p \exp\left(-\lambda p c \beta^{\frac{2}{\alpha}} d^2\right). \tag{2.22}$$

Differentiating G with respect to p and equating it to 0, the optimal ALOHA access probability is $p^\star = \min\left\{1, \frac{1}{\lambda c \beta^{\frac{2}{\alpha}} d^2}\right\}$, and

$$\mathsf{G}^* = \begin{cases} \frac{1}{\exp(1) c \beta^{\frac{2}{\alpha}} d^2} & \text{if } \lambda c \beta^{\frac{2}{\alpha}} d^2 > 1, \\ \lambda \exp\left(-\lambda c \beta^{\frac{2}{\alpha}} d^2\right) & \text{o.w.} \end{cases} \tag{2.23}$$

Thus, even without an outage probability constraint, we see that the goodput expression (2.23) is independent of both λ and p, similar to Theorem 2.4.1 whenever $\lambda c \beta^{\frac{2}{\alpha}} d^2 > 1$, since the product of

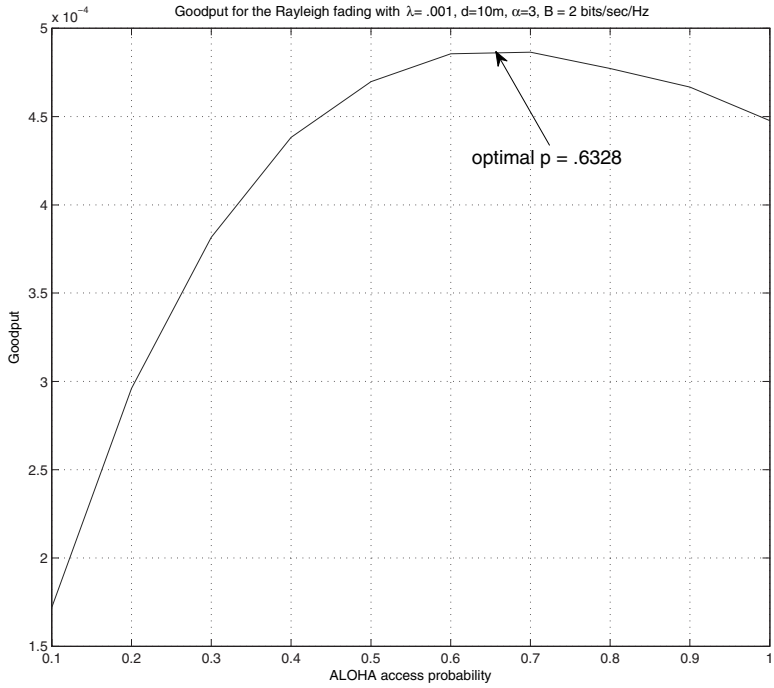

Figure 2.2: Network goodput G with Rayleigh fading as a function of ALOHA access probability p.

the density λ and the optimal ALOHA probability p is a constant. This is a result of an underlying fundamental limit on the maximum density of successful transmissions in a wireless network, which in case of the ALOHA protocol is equal to λp.

In Fig. 2.2, we plot the network goodput as a function of the ALOHA access probability p. As derived, we can see that the optimal $p = .63$ for $\lambda = 10^{-3}$ with $d = 10\,\mathrm{m}$, $\alpha = 3$ for $B = 2$ bits/sec/Hz transmission rate. After analyzing the transmission capacity of wireless network with ALOHA protocol in detail, we next highlight a surprising feature of the ALOHA protocol of having both spatial and temporal correlations in interference received at any point in space. With ALOHA protocol, at each time slot, each node transmits independently of all other nodes, but the shared randomness between node locations due to PPP assumption gives rise to counterintuitive correlations across time and space. We capture this critical phenomenon in the next section, which will be useful for analyzing the performance of ARQ-type protocols in Chapter 6.

2.7 Correlations with ALOHA Protocol

In this section, we show a counterintuitive result from [4] that shows the interference received at any point in space in a PPP network is both temporally and spatially correlated while using the ALOHA protocol. One would assume that given that the locations of nodes in a PPP network are uniformly random in any given area, and with ALOHA protocol, each node transmits independently across space and time, the interference received at different locations or time instants would be independent; however, that is shown to be incorrect as follows.

To be concrete, in a PPP-distributed wireless network with the ALOHA protocol, we fix the transmitter locations drawn from a single realization of PPP, while at each time slot, each node decides to transmit with probability p independently of all other nodes. Finally, the averaging is taken with respect to the PPP. Thus, the shared randomness of transmitter locations gives rise to spatial and temporal correlations.

Let the interference seen at location $u \in \mathbb{R}^2$ at time k be

$$I_u(k) := \sum_{x \in \Phi_T} Pg(x-u)\mathbf{1}_p(x,k)|h_{xu}(k)|^2,$$

where $g(.)$ be the path-loss function that only depends on the distance. Throughout this book, we use $g(x-u) = |x-u|^{-\alpha}$. The indicator function $\mathbf{1}_p(x,k)$ means that the transmitter at $x \in \Phi_T$ transmits with probability p using the ALOHA protocol at time k.

The spatio-temporal correlation coefficient between $I_u(k)$ and $I_v(\ell)$ is

$$\begin{aligned} \mathsf{corr}_{uv}(k,l) &= \frac{\mathbb{E}\left\{\left(I_u(k) - \mathbb{E}\{I_u(k)\}\right)\left(I_v(\ell) - \mathbb{E}\{I_v(\ell)\}\right)\right\}}{Var(I_u(k))^{\frac{1}{2}} Var(I_v(\ell))^{\frac{1}{2}}} \\ &= \frac{\mathbb{E}\{I_u(k)I_v(\ell)\} - (\mathbb{E}\{I_u(k)\})^2}{\mathbb{E}\{I_u(k)^2\} - (\mathbb{E}\{I_u(k)\})^2}, \end{aligned} \tag{2.24}$$

since $I_u(k)$ and $I_v(\ell)$ are identically distributed. We now compute the three expectations in (2.24) using the Campbell's theorem and second-order product density (correlation) of the PPP as follows.

First the expected value of $I_u(k)$, which is

$$\begin{aligned} \mathbb{E}\{I_u(k)\} &\stackrel{(a)}{=} \mathbb{E}\{I_o(k)\}, \\ &= \mathbb{E}\left\{\sum_{x \in \Phi_T} Pg(x)\mathbf{1}_p(x,k)|h_{xo}(k)|^2\right\}, \\ &\stackrel{(b)}{=} \mathbb{E}\left\{\sum_{x \in \Phi(k)} Pg(x)\mathbf{1}_p(x,k)\right\}, \\ &\stackrel{(c)}{=} p\lambda \int_{\mathbb{R}^2} g(x)\mathrm{d}x, \end{aligned} \tag{2.25}$$

where (a) follows since the distribution of $I_u(k)$ is invariant to location of u and o is the origin, (b) follows since the $|h_{xo}|^2$ is Rayleigh distributed with $\mathbb{E}\{|h_{xo}|^2\} = 1$, and, finally, (c) follows from the Campbell's theorem for PPP.

Next, we derive the expression for second moment of the interference as follows.

$$\begin{aligned} \mathbb{E}\{I_u(k)^2\} &= \mathbb{E}\{I_o(k)^2\}, \\ &= \mathbb{E}\left\{\left(\sum_{x \in \Phi_T} Pg(x)\mathbf{1}_p(x,k)|h_{xo}(k)|^2\right)^2\right\}, \\ &= \mathbb{E}\left\{\sum_{x \in \Phi_T} Pg^2(x)\mathbf{1}_p(x,k)|h_{xo}(k)|^4\right\} \end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left\{\sum_{x,y\in\Phi_T,x\neq y} Pg(x)g(y)\mathbf{1}_p(x,k)\mathbf{1}_p(y,k)|h_{xo}(k)|^2|h_{yo}(k)|^2\right\},\\
= \;& p\mathbb{E}\{h^4\}\lambda\int_{\mathbb{R}^2} g^2(x)\mathrm{d}x,\\
&+p^2\left(\mathbb{E}\{h^2\}\right)^2\lambda^2\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} g(x)g(y)\mathrm{d}x\,\mathrm{d}y, \qquad (2.26)
\end{aligned}
$$

where the first term in the last equality follows from Campbell's theorem (Theorem 2.3.7) for PPP, and the second term from the fact that $|h_{xo}|^2$ and $|h_{y0}|^2$ are independent, and second-order product density of the PPP [5] which states that

$$
\mathbb{E}\left\{\sum_{x,y\in\Phi(k),x\neq y} f(x)f(y)\right\} = \lambda^2\int\int f(x)f(y)\,\mathrm{d}x\,\mathrm{d}y.
$$

Finally, by exactly following the same procedure as above, the cross-correlation of the interference is given by

$$
\begin{aligned}
\mathbb{E}\{I_u(k)I_v(\ell)\} \;&=\; p^2\lambda\int_{\mathbb{R}^2} g(x-u)g(x-v)\mathrm{d}x\\
&\quad +p^2\lambda^2\left(\int_{\mathbb{R}^2} g(x)\mathrm{d}x\right)^2, \qquad (2.27)
\end{aligned}
$$

where we have used $\mathbb{E}\{|h|^2\}=1$.

Thus, using (2.25), (2.26), and (2.27), the spatio-temporal correlation of the interference from (2.24) is

$$
\mathsf{corr}_{u,v}(k,\ell) = \frac{p\int_{\mathbb{R}^2} g(x-u)g(x-v)\mathrm{d}x}{\mathbb{E}\{|h|^4\}\int_{\mathbb{R}^2} g^2(x)\mathrm{d}x}. \qquad (2.28)
$$

For the $g(x-u)=|x-u|^{-\alpha}$ path-loss function, which we use throughout this book, we next show that the spatial correlation coefficient is zero.

Example 2.7.1 *For the special case of $g(x)=x^{-\alpha}$, the spatial correlation coefficient is zero. For the purposes of analysis, we let $g_\psi(x)=\frac{1}{\psi+x^\alpha}$ and let $\psi\to 0$, since otherwise $\int x^{-\alpha}\mathrm{d}x=\infty$. From* (2.28)*, the spatial correlation coefficient with path-loss function $g_\psi(x)$ is*

$$
\begin{aligned}
\mathsf{corr}_{u,v}(k,\ell) \;&=\; \frac{p\int_{\mathbb{R}^2} g_\psi(x-u)g_\psi(x-v)\mathrm{d}x}{\mathbb{E}\{|h|^4\}\int_{\mathbb{R}^2} g_\psi^2(x)\mathrm{d}x},\\
&=\; \frac{p\int_{\mathbb{R}^2}\frac{1}{\psi+(x-u)^\alpha}\frac{1}{\psi+(x-v)^\alpha}\mathrm{d}x}{\mathbb{E}\{|h|^4\}\int_{\mathbb{R}^2}\frac{1}{\psi+x^\alpha}\mathrm{d}x}.
\end{aligned}
$$

Taking the limit as $\psi\to 0$, it follows that $\lim_{\psi\to 0}\mathsf{corr}_{u,v}(k,\ell)=0$. This result is an artifact of the path-loss model of $x^{-\alpha}$, where the nearest interferers are the dominant interferers. Thus, for two distinct receivers, most of the interference comes from small discs around them that are non-overlapping, and since the number of points of PPP lying in non-overlapping discs are independent, the result follows.

Thus, even though the interferences seen at different receivers are not independent, however, they are uncorrelated. Consequently, assuming independence of interference for ease of analysis is not too limiting.

Example 2.7.2 *From (2.28), we can get the temporal correlation coefficient as a special case of* $\mathsf{corr}_{u,v}(k,\ell)$ *by specializing* $v = u$ *as* $\mathsf{corr}(k,\ell) = \frac{p}{\mathbb{E}\{|h|^4\}}$.

Clearly from (2.28), the temporal correlation coefficient is non-zero, and hence repeated transmissions (SINR at different times) between a transmitter receiver pair are not independent. Direct impact of this observation is encountered in the analysis of ARQ protocols, where repeated transmission attempts are made till the packet is successfully received. Typically, for the ease of exposition, SINRs at repeated attempts are assumed independent, which is inaccurate as shown in the next example. We next present a simple example to derive the joint success probability at a receiver from [4]. A more general exact derivation for ARQ protocols will be described in Chapter 6 under a maximum retransmission/delay constraint.

Example 2.7.3 *We consider a receiver located at origin* o *and for simplicity drop the AWGN contribution and define the success to be the event that*

$$\mathsf{SINR}(k) = \frac{d^{-\alpha}|h(k)|^2}{I_o(k)} > \beta.$$

Then

$$\begin{aligned}
\mathbb{P}\left(\mathsf{SINR}(k) > \beta, \mathsf{SINR}(\ell) > \beta\right) &= \mathbb{P}\left(|h(k)|^2 > d^{\alpha} I_o(k)\beta, |h(\ell)|^2 > d^{\alpha} I_o(\ell)\beta\right), \\
&\overset{(a)}{=} \mathbb{E}\left\{\exp^{-d^{\alpha} I_o(k)\beta} \exp^{-d^{\alpha} I_o(\ell)\beta}\right\}, \\
&= \mathbb{E}\left\{\exp^{-d^{\alpha}\beta \sum_{x\in\Phi} Pg(x)\mathbf{1}_p(x,k)|h_{xo}(k)|^2}\right. \\
&\quad \left.\exp^{-d^{\alpha}\beta \sum_{x\in\Phi} Pg(x)\mathbf{1}_p(x,\ell)|h_{xo}(\ell)|^2}\right\} \\
&\overset{(b)}{=} \mathbb{E}\left\{\prod_{x\in\Phi}\left(\frac{p}{1+d^{\alpha}x^{-\alpha}\beta} + 1 - p\right)^2\right\}, \\
&\overset{(c)}{=} \exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \left(\frac{p}{1+d^{\alpha}x^{-\alpha}\beta} + 1 - p\right)^2 \mathrm{d}x\right),
\end{aligned}$$

where (a) *follows from the fact that* $|h(\ell)|^2, |h(k)|^2 \sim \exp(1)$ *and are independent,* (b) *follows by taking the expectation with respect to ALOHA (similar to* (6.9)*) and* $|h_{xo}(k)|^2, |h_{xo}(\ell)|^2$ *that are i.i.d.* $\sim \exp(1)$*, and finally* (c) *follows from the probability-generating functional of the PPP (Theorem 2.3.6).*

Solving for this integral we get

$$\frac{\mathbb{P}\left(\mathsf{SINR}(k) > \beta, \mathsf{SINR}(\ell) > \beta\right)}{\mathbb{P}\left(\mathsf{SINR}(k) > \beta\right)^2} = \exp^{\left(2\lambda p^2 \beta^{2/\alpha} d^2 \pi^2 \frac{\alpha-2}{\alpha^2} csc\left(\frac{2\pi}{\alpha}\right)\right)} > 1.$$

Thus, link success probabilities are positively correlated. Hence, if the transmission between a transmitter and a receiver is successful at a given instant, it is more likely to be successful again. Thus, the analysis of ARQ-type strategies, where a packet is repeatedly sent until it is successfully

received is complicated, since the probability of success or failure in successive slots is not independent. Most works assume the independence and make an inaccurate prediction on the performance of ARQ protocols. In Chapter 6, we will illustrate the exact performance of a ARQ protocol in a wireless network.

Until now in this chapter we have concentrated on analyzing the performance of wireless network when each transmitter uses an ALOHA protocol. However, clearly, one can choose the set of simultaneously active transmitters better than the ALOHA protocol by considering received SINRs, neighbor distance, and so on, to reduce interference and improve the transmission capacity. In the next section, we present two such protocols and analyze their performance.

2.8 Transmission Capacity with Scheduling in Wireless Networks

In a PPP network, the most significant contributor of interference at each receiver is its nearest interferer. The expected interference from the nearest interferer is of the same order as the expectation of the sum of the interference from all other interferers. With the ALOHA protocol, each node transmits independently and there is no restriction on the set of simultaneously transmitting nodes. Thus, the nearest interferer is active with a fixed probability and limits the performance of the ALOHA protocol. To improve the outage probability at any receiver, thus there is a case for inhibiting transmission from some of the nearest interferers. This in turn, however, reduces the number of simultaneous transmissions (spatial capacity) in the network. Thus, an efficient scheduling strategy(transmitter inhibition strategy) has to find which transmitters to turn off such that the improvement in the outage probability compensates for the spatial capacity loss.

One such strategy is based on *guard zones*, where any transmitter within a distance of d_{gz} from a receiver is not allowed to transmit [6]. Other class of strategies includes various versions of CSMA, where each trasnmitter measures the channel and decides to transmit depending on a function of its measurement. Intuitively, both these strategies, should improve the transmission capacity with respect to ALOHA; however, exactly quantifying the improvement is complicated analytically. The challenge is that with the guard zone strategy or CSMA, the set of active transmitters is correlated (not PPP anymore) and the distribution of interference seen at any receiver does not have a closed form expression or known Laplace transform or probability-generating functionals. To overcome this difficulty, typically, approximations are made on the interference distributions to get some analytical tractability. In this section, we first discuss the guard-zone-based strategy, and then follow it up with analyzing two variants of CSMA.

2.8.1 Guard Zone Strategy

Consider a PPP network Φ with density λ nodes per unit area as described in Section 2.2, where each transmitter has a corresponding receiver at distance d. With a guard zone, only those transmitters that are not within a distance of d_{gz} from any receiver are allowed to transmit, see Fig. 2.3. For the typical transmitter–receiver pair (T_0, R_0), where R_0 is at the origin, let the active set of interferers be denoted by $\Phi_{\rm gz} = \{T_m : T_m \in \Phi \backslash \mathbf{B}(0, d_{\rm gz})\}$. Then, the outage probability at receiver R_0 is

$$P_{\rm out}^{\rm gz}(B) = \mathbb{P}\left(\frac{Pd^{-\alpha}|h_{00}|^2}{\sum_{m:T_m \in \Phi_{gz}} Pd_{m0}^{-\alpha}|h_{m0}|^2 + 1} \leq \beta\right). \tag{2.29}$$

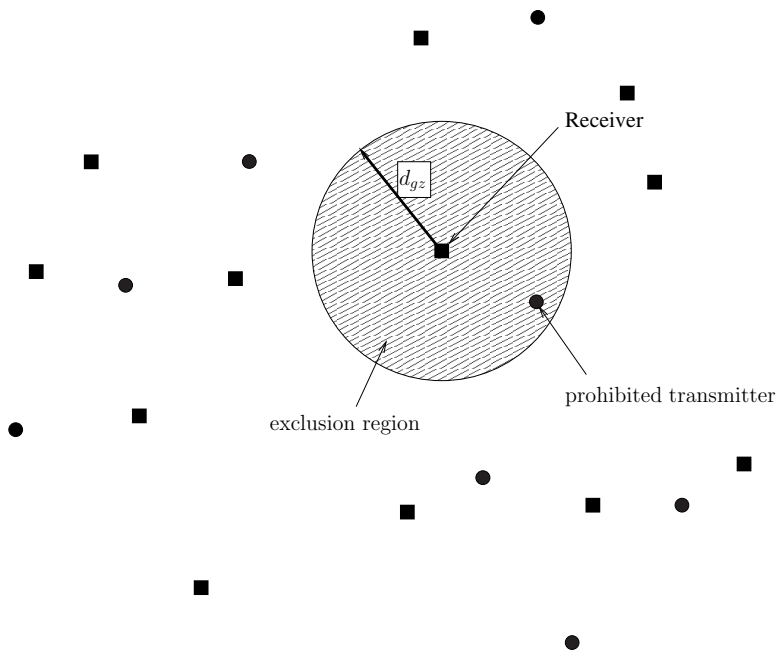

Figure 2.3: Dots represent transmitters and squares represent receivers. Only those transmitters (squares) are allowed to transmit that lie outside the discs of radius $d_{\rm gz}$ centered at all the receivers.

Then with an outage probability constraint of ϵ, the maximum density of successful transmissions is

$$\lambda_{\rm gz}^{1} = \sup\{\lambda : P_{\rm out}^{\rm gz}(B) \leq \epsilon\}.$$

Since the distribution of interference $I_{gz} = \sum_{m:T_m \in \Phi_{gz}} P d_{m0}^{-\alpha} |h_{m0}|^2$ received at any receiver is not known, deriving exact expression for the outage probability is not possible. To facilitate the analysis, I_{gz} is assumed to follow a Gaussian distribution. Similar to the computation of mean and variance in (2.17), the exact mean m_{gz} and variance var_{gz} of I_{gz} can be computed using the Campbell's theorem (Theorem 2.3.7) (proof is left as an excercise) as

$$\mathsf{m}_{gz} = \frac{4\pi d^{\alpha} d_{gz}^{2-\alpha}}{\alpha^2 - 4}\lambda, \tag{2.30}$$

$$\mathsf{var}_{gz} = \frac{\pi d^{2\alpha} d_{gz}^{2(1-\alpha)}}{\alpha^2 - 1}\lambda. \tag{2.31}$$

Thus, using the Gaussian distribution approximation on I_{gz} with mean m_{gz} and variance var_{gz}, we have that

$$\begin{aligned} P_{\rm out}^{gz}(B) &= \mathbb{P}\left(\frac{P d^{-\alpha}|h_{00}|^2}{\sum_{m:T_m \in \Phi_{gz}} P d_{m0}^{-\alpha}|h_{m0}|^2 + 1} \leq \beta\right), \\ &\overset{(a)}{=} \exp^{\left\{-\frac{d^{\alpha}\beta}{P}\right\}} \exp^{\{-d^{\alpha}\beta I_{gz}\}}, \end{aligned}$$

$$\stackrel{(b)}{=} \quad \exp^{\left\{-\frac{d^{\alpha}\beta}{P}\right\}} \exp^{\left\{-d^{\alpha}\beta \mathsf{m}_{gz}+\frac{d^{2\alpha}\beta^{2}\mathsf{var}_{gz}}{2}\right\}}, \tag{2.32}$$

where (a) follows since $|h_{00}|^2 \sim \exp(1)$ and (b) follows from using the moment generating function of the Gaussian distribution. Thus, (2.32) reveals how the outage probability decreases with the increasing guard zone distance d_{gz}, and we can get λ_g^1 by equating it to the outage probability constraint ϵ.

The probability for any receiver to not have any transmitter in a disc of radius of d_{gz} around it is equal to the void probability of transmitter PPP Φ in a disc of radius d_{gz}, which is given by $\exp^{-\lambda\pi d_{gz}^2}$. Thus, with the inhibition criterion of the guard-zone-based policy, the density of the active transmitter process is $p_a\lambda$, where $p_a = \exp^{-\lambda\pi d_{gz}^2}$. Hence, the operational density of transmitters is $\lambda_{gz}^2 = \lambda \exp^{-\lambda\pi d_{gz}^2}$. Note that this separate analysis of λ_{gz}^1 and λ_{gz}^2 is not completely accurate since they depend on each other; however, it acts as a reasonable approximation.

There is inherent tension between inhibition radius d_{gz} and the transmission capacity; increasing d_{gz} decreases the interference and the outage probability (increases λ_g^1), but at the same time decreases the number of active transmitters (decreases λ_g^2). Note λ_g^1 is the maximum density that can be supported given an outage probability constraint of ϵ, and considering only the interference coming from outside the disc of radius d_{gz}, under the guard-zone-based strategy. Hence, we want to find the best d_{gz} and λ such that λ_g^2 (density corresponding to inhibition) is equal to λ_g^1 (corresponding to the outage probability constraint). Equivalently, one can also write the optimization problem as

$$\lambda^{\star}(d_{gz}) = \max_{d_{gz},\lambda} \left\{\min\{\lambda_{gz}^1, \lambda_{gz}^2\}\right\}. \tag{2.33}$$

Corresponding transmission capacity is $\lambda^{\star}(d_{gz})(1-\epsilon)B$ bits/sec/Hz/m^2.

Problem (2.33) is a non-linear optimization problem that can be solved using numerical computations. In Fig. 2.4, we plot the transmission capacity as a function of d_{gz} for outage probability constraint of 10 percent ($\epsilon = .1$). We can see that the transmission capacity increases with d_{gz} for $d_{gz} \leq d_{gz}^{\star}$ and decreases thereafter, since the decrease in the number of concurrent transmissions for $d_{gz} > d_{gz}^{\star}$ overtakes the improvement in the outage probability. We can also see that there is significant improvement by employing a guard-zone-based inhibition policy over uninhibited transmissions ($d_{gz} = 0$).

An alternative to guard-zone-based scheduling strategy, is the CSMA protocol, where each node monitors the channels and follows a contention resolution strategy. We discuss two versions of CSMA protocols in the next section for wireless networks and analyze their performance.

2.8.2 CSMA

Channel Gain-Based

We consider the same signal model as in Section 2.2, where the location of transmitters T_n is assumed to follow a PPP Φ with density λ. Each transmitter T_n is defined to be *qualified* to transmit if its channel gain with its associated receiver R_n, $|h_{nn}|^2$ exceeds a threshold τ_h. Thus, this protocol requires channel feedback from each R_n to T_n. Let

$$\Phi_Q = \{T_n \in \Phi : |h_{nn}|^2 \geq \tau_h\}$$

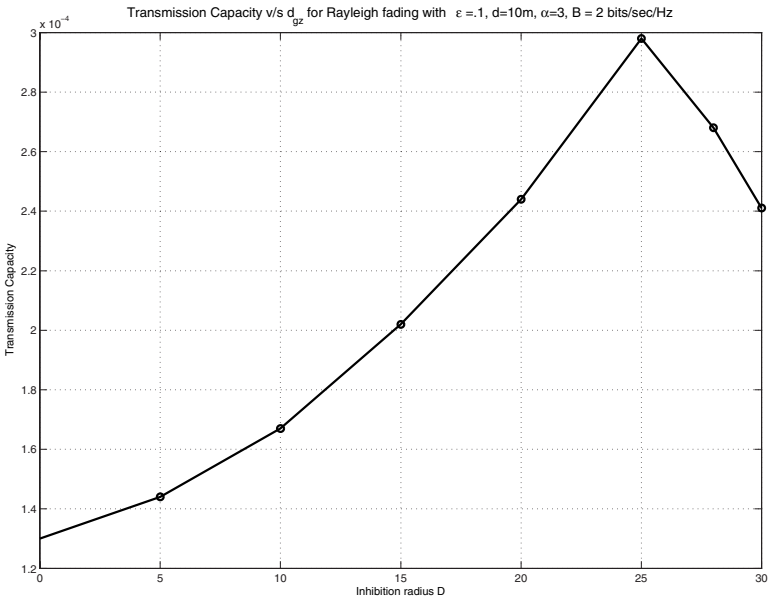

Figure 2.4: Transmission capacity with Rayleigh fading as a function of guard zone distance $d_{\rm gz}$.

denote the set of qualified nodes or contenders. Note that Φ_Q is a randomly thinned version of Φ_T, since $|h_{nn}|^2$ are i.i.d., and therefore Φ_Q is also a homogenous PPP with density λp_{τ_h}, where $p_{\tau_h} = \mathbb{P}(|h_{nn}|^2 \geq \tau_h)$.

We define that two transmitters T_n and T_m contend with each other if the received interference power they see from each other, $|h_{nm}|^2 d^{-\alpha}$, is greater than the CSMA threshold τ_c, $|h_{nm}|^2 d^{-\alpha} > \tau_c$. For a transmitter T_n, its contention neighborhood is the set of nodes that contend with it,

$$\Phi_{\rm CN}(n) = \{T_m \in \Phi_Q : |h_{mn}|^2 |T_m - T_n|^{-\alpha} \geq \tau_c\}.$$

The *inhibition* module of the CSMA protocol allows only one of the transmitters from $\Phi_{\rm CN}(n)$ to transmit at any time to suppress interference.

To decide which node of $\Phi_{\rm CN}(n)$ gets to transmit in a decentralized manner, each node of $\Phi_{\rm CN}(n)$ is equipped with a timer value clk_n, which is a uniformly distributed random variable between $[0, 1]$. Thus, the node with the minimum timer value transmits in each slot, and if any node in Φ_N hears a transmission from node $T_{n^\star}$, it does not transmit in that entire slot and resets its timer value. For each slot, the node $n^\star \in \Phi_{\rm CN}(n)$ transmits at time $\mathsf{clk}_{n^\star}$, where

$$n^\star = \arg\min_{n:T_n \in \Phi_{\rm CN}(n)} \mathsf{clk}_n.$$

Remark 2.8.1 *There are two modules in this CSMA protocol, the first that finds qualified nodes that have sufficient channel gains to their respective receivers, and the second that chooses one node among the set of qualified nodes to minimize interference. Allowing only qualified nodes to contend increases the chance of success; however, it limits the number of simultaneously spatial transmissions, thus the choice of τ_h, (parameter controlling the qualification) is critical. Similar tradeoff exists as a function of τ_c that controls the size of the neighborhood, since only one node in each neighborhood is allowed to transmit.*

Let $\mathcal{E}_n = \mathbf{1}_{\{|h_{00}|^2 \geq \tau_h, \mathsf{clk}_n \leq \min_{T_m \in \Phi_{\mathrm{CN}}(n)} \mathsf{clk}_n\}}$ represent the event that transmitter T_n is qualified and has the least timer in its neighborhood and gets to transmit. Then a typical transmitter T_0 located at the origin gets to transmit with the above CSMA protocol if $\mathcal{E}_0 = 1$. Thus, the probability that a typical transmitter T_0 transmits is $p_{\mathrm{CSMA}} = \mathbb{E}\{\mathcal{E}_0\}$. Note that events $\{|h_{00}|^2 \geq \tau_h\}$ and $\{\mathsf{clk}_n \leq \min_{T_m \in \Phi_{\mathrm{CN}}(n)} \mathsf{clk}_n\}$ are independent.

Next, we find the distribution of cardinality of $\Phi_{\mathrm{CN}}(0)$ (the contention neighborhood set of T_0 located at origin). By definition, the set $\Phi_{\mathrm{CN}}(0)$ consists of all nodes of Φ_Q that lie in a disc $\mathbf{B}\left(o, \left(\frac{\tau_c}{|h_{m0}|^2}\right)^{\frac{1}{\alpha}}\right)$. Since Φ_Q is a PPP with density λp_{τ_h}, the number of nodes of Φ_Q lying in $\mathbf{B}\left(o, \left(\frac{\tau_c}{|h_{m0}|^2}\right)^{\frac{1}{\alpha}}\right)$ is Poisson distributed with mean $\lambda p_{\tau_h} \pi \mathbb{E}\left\{\left(\frac{\tau_c}{|h_{m0}|^2}\right)^{\frac{2}{\alpha}}\right\}$.

Thus,

$$\begin{aligned} p_{\mathrm{CSMA}} &= p_{\tau_h} \mathbb{E}\{\mathsf{clk}_0 \leq \min_{T_m \in \Phi_N^0} \mathsf{clk}_n\}\}, \\ &\overset{(a)}{=} p_{\tau_h} \mathbb{E}\left\{\frac{1}{1 + \#(\Phi_N^0)}\right\}, \\ &\overset{(b)}{=} p_{\tau_h} \frac{1 - \exp^{-p_{\tau_h}\bar{N}}}{p_{\tau_h}\bar{N}}, \\ &= \frac{1 - \exp^{-p_{\tau_h}\bar{N}}}{\bar{N}}, \end{aligned}$$

where (a) follows since T_0 has the least timer among its $\#(\Phi_{\mathrm{CN}}(0))$ neighbors, and (b) follows since $\#(\Phi_{\mathrm{CN}}(n))$ is Poisson distributed with mean $p_{\tau_h}\bar{N}$.

Next, we compute the outage probability for the typical transmitter receiver pair (T_0, R_0). The set of active users $\Phi_a = \{T_m \in \Phi_Q : \mathcal{E}_m = 1\}$ and the interference received at R_0 from active users is

$$I_0^a = \sum_{T_m \in \Phi_a \setminus \{T_0\}} d_{m0}^{-\alpha} |h_{m0}|^2.$$

Hence, the outage probability is given by

$$P_{\mathrm{out}}(B) = \mathbb{P}\left(|h_{00}|^2 < \beta d^\alpha I_0^a \;\middle|\; |h_{00}|^2 > \tau_h\right).$$

Deriving the outage probability with this CSMA protocol is challenging since the set of active transmitters is no longer a homogenous PPP, and there are correlations among the active node locations.

To facilitate the analysis, following [7], we will approximate I_0^a by the interference from a non-homogenous PPP Φ_{h} with density $p_{\tau_h}\lambda \mathsf{h}(x, p_{\tau_h}\lambda)$, which is a function of $x > 0$. The function $\mathsf{h}(x, p_{\tau_h}\lambda)$ is the conditional probability that T_0 at origin is active and in addition, there is another active transmitter T_m with $\mathcal{E}_m = 1$ at a distance x from the origin, and where the density of qualified nodes Φ_Q is $p_{\tau_h}\lambda$. The difference between a homogenous and a non-homogenous PPP is that in the non-homogenous case, the density is not constant and depends on the location $x \in \mathbb{R}^2$.

By using the non-homogenous PPP Φ_{h}, we are trying to model the inhibition induced by the CSMA among transmitters of the PPP Φ. Note that $\mathsf{h}(x, p_{\tau_h}\lambda) \to 0$ as $x \to 0$. Thus, the density of PPP Φ_h goes to zero for short distance x, and correspondingly there is large scale inhibition

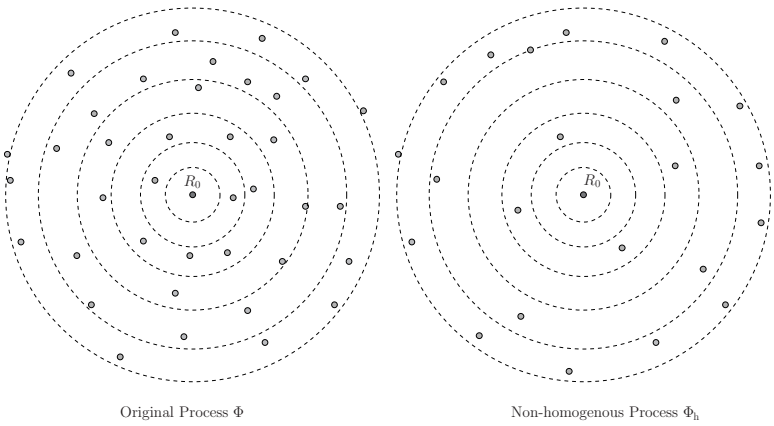

Figure 2.5: A pictorial description of PPP Φ_{h} in comparison to original process Φ, where the density increases with increasing distance from the origin.

induced by the CSMA protocol at distances close to origin where T_0 is located, allowing only very few nodes to transmit at the same time as T_0. On the other hand, as $x \to \infty$, $\mathsf{h}(x, p_{\tau_h}\lambda) \to \mathbb{P}\left(\mathsf{clk}_m \leq \min_{T_n \in \Phi_N^m} \mathsf{clk}_n\right)$ for some $T_m \in \Phi_Q$, since as x increases, the effect of conditioning (having an active transmitter T_0 at the origin) over the event that there is an active transmitter at distance x from the origin diminishes. Eventually at $x = \infty$, having an active transmitter T_0 at the origin has no effect on having an active transmitter at distance x from the origin, leading to $\mathsf{h}(x, p_{\tau_h}\lambda)$ being equal to the unconditional probability of having an active node among the qualified nodes, which is equal to $\mathbb{P}\left(\mathsf{clk}_m \leq \min_{T_n \in \Phi_N^m} \mathsf{clk}_n\right)$ for some $T_m \in \Phi_Q$. Thus, at large distances x, the density of process Φ_{h} is equal to $\lambda p_{\mathrm{CSMA}}$ having no inhibition effect from the transmitter located at the origin.

Thus, as a function of distance x from the origin, PPP Φ_{h} essentially models the inhibiting nature of the CSMA with respect to the typical node at the origin, having increasing number of active transmitters with increasing x. We illustrate the PPP Φ_{h} pictorially in Fig. 2.5.

The utility of this new process Φ_{h} is that its Laplace transform is known to be

$$\mathcal{L}_{\Phi_{\mathsf{h}}\setminus\{T_0\}}(s) = \exp^{\left(-p_{\tau_h}\lambda \int_0^\infty \int_0^{2\pi} \frac{\mathsf{h}(x, p_{\tau_h}\lambda) x d\theta \mathrm{d}x}{1+f(x,d,\theta)/s}\right)}, \tag{2.34}$$

where $f(x, d, \theta) = (x^2 + d^2 - 2xr\cos(\theta))^{-\alpha/2}$ and d is the distance between each transmitter receiver pair [7]. This can also be derived from first principles similar to Laplace transform of a homogenous PPP.

Using (2.34), by approximating I_0^a with interference from nodes in Φ_{h}, we can write

$$P_{\mathrm{out}}(B) = 1 - \mathcal{L}_{\Phi_{\mathsf{h}}\setminus\{T_0\}}(2i\pi r^{-\alpha} ts) \frac{\frac{1}{1-2i\pi s} \exp^{2i\pi s\tau_h} - 1}{2i\pi s}, \tag{2.35}$$

using the Plancherel Parseval theorem [8]. The details of this derivation are intentionally deleted because of being laborious and too technical.

Thus, using (2.35), we can numerically compute the outage probability and consequently the transmission capacity, which is the density of successful transmissions multiplied

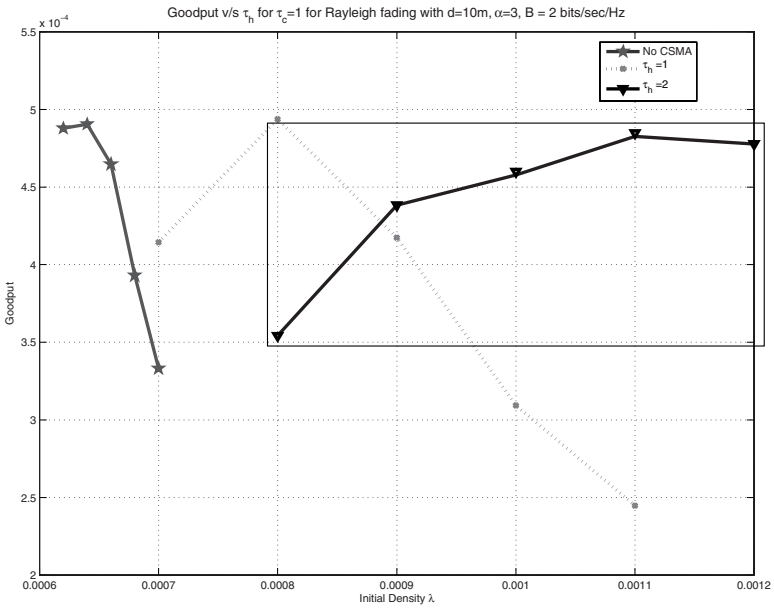

Figure 2.6: Network goodput with Rayleigh fading as a function of CSMA transmitter channel access threshold of τ_h for neighbourhood contention threshold of $\tau_c = 1$.

with rate of transmission $2^\beta - 1$. In this inhibition-based CSMA protocol, the two key parameters are τ_h and τ_c, which control the number of qualified transmitters, and the size of the neighborhood.

In Fig. 2.6, we plot the goodput $(1 - P_{\text{out}}(B))\lambda B$ as a function of density λ for fixed neighborhood contention threshold $\tau_c = 1$, and different values of transmitter channel access threshold τ_h. We see that for small values of λ, no CSMA (ALOHA with access probability 1) is better than CSMA, since interference is very low and inhibition provided by CSMA is unnecessary. On the other hand, as we increase the density, the role of CSMA transmitter channel access threshold τ_h becomes more prominent, since with sufficient interference it is important to control how many transmitters are allowed to transmit. We see a similar performance comparison in Fig. 2.7, where we plot the goodput as a function of density λ for fixed transmitter channel access threshold τ_h and different values of neighbourhood contention threshold $\tau_c = 1$.

Remark 2.8.2 *Recently, a more detailed analytical analysis of CSMA protocol with just the neighbourhood contention model, that is, with $\tau_c = 0$ (no qualification criteria) has been done in [9] for small densities (λ) regime, to show that the transmission capacity scales as $\Theta\left(\epsilon^{\frac{2}{\alpha\psi}}\right)$, for $\epsilon \to 0$, where $\psi \geq 1$ depends on the fading coefficient distribution. For Rayleigh fading, $\psi = 1$.*

Next, we present an alternate SINR-based CSMA protocol, where each node monitors the SINR to its corresponding receiver and transmits only if the SINR is larger than a threshold. In all prior sections in this chapter, we have assumed that each node's data queue is backlogged, that is, it always has a packet to transmit. This is only an abstraction, and a more realistic data arrival process model is considered with the SINR-based CSMA protocol in the next section.

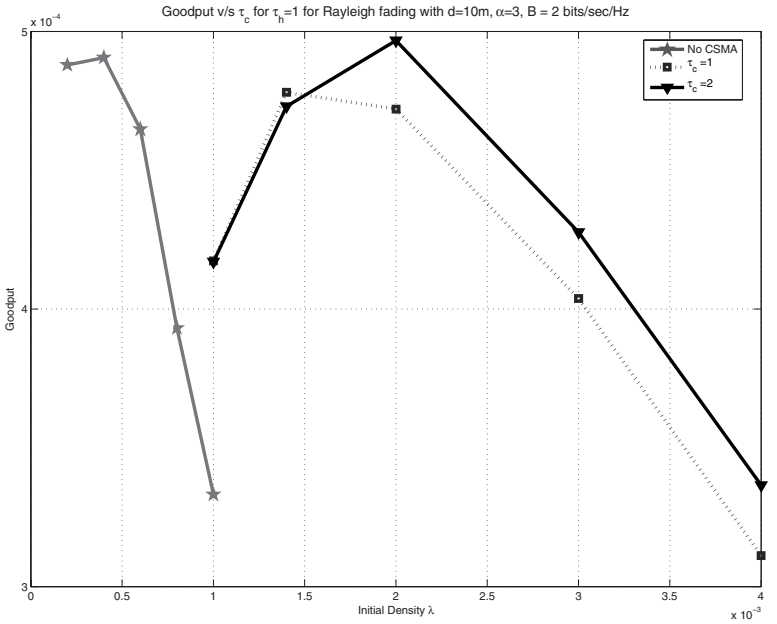

Figure 2.7: Network goodput with Rayleigh fading as a function of CSMA neighbourhood contention threshold τ_c with channel access threshold of $\tau_h = 1$.

SINR-based

In this section, we consider an SINR-based CSMA protocol and consider that packets arrive at each node according to a 1-dimensional PPP. In all earlier sections, we have assumed that each node always has a packet to transmit, which is only an abstraction. To analyze the SINR-based CSMA protocol with random packet arrivals, we consider a slightly different network model compared to Section 2.2, which has been introduced in [10]. We consider an area A and model the packet arrival process as a 1-dimensional PPP with arrival rate $(A/L)\lambda$, where L is the fixed packet duration. Each packet after arrival is assigned to a transmitter location that is uniformly distributed in area A, and the receiver corresponding to a particular transmitter is located at a fixed distance d away with a random orientation, as shown in Fig. 2.8. For $A \to \infty$, this process corresponds to a 2-dimensional PPP of transmitter locations with density λ (Section 2.2), where each transmitter has a packet arrival rate of $\frac{1}{L}$. Note that the performance of ALOHA protocol with data packet arrival process follows similar to what follows next and omitted for brevity and can be found in [10].

Remark 2.8.3 *If we use the model of Section 2.2, we would first fix the transmitter locations according to a 2-dimensional PPP, and then packet traffic is generated, and each transmitting node receives a packet according to 1-dimensional PPP over time. The packets are then transmitted to the respective receivers that are located at a fixed distance d. Thus, with the model of Section 2.2, one has to average over the spatial (to fix locations) and temporal (packet arrivals) statistics, which is rather challenging. Instead, by slightly altering the model, as described above, there is a single process that defines both the spatial location and temporal packet arrival process.*

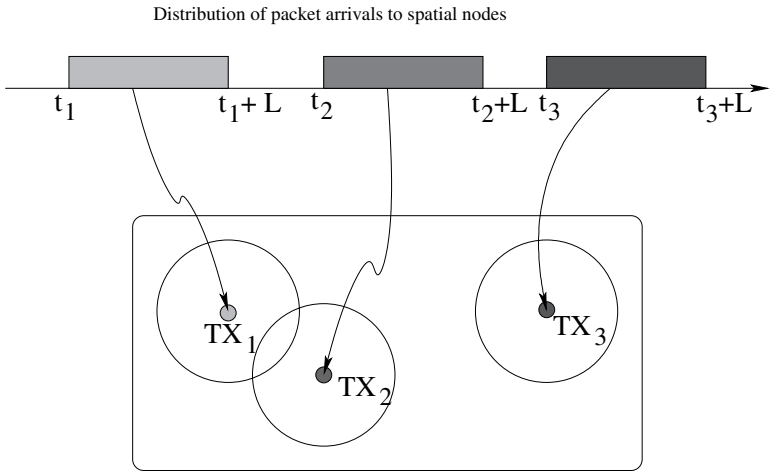

Figure 2.8: Spatial model for CSMA with packet arrivals.

For simplicity, we ignore the AWGN contribution. Hence, the SIR between transmitter T_n and its receiver R_n at time t is given by

$$\mathsf{SIR}_n(t) := \frac{d^{-\alpha}|h_{nn}|^2}{\sum_{T_s \in \Phi \setminus \{T_0\}} \mathbf{1}_{T_s}(t) d_{mn}^{-\alpha}|h_{mn}|^2}, \tag{2.36}$$

where $\mathbf{1}_{T_m}(t) = 1$, if the transmitter T_s is not in back-off, and 0 otherwise, and Φ is a PPP with density λ. With CSMA, transmitter T_s sends its packet at time t if the channel is sensed *idle* at time t, which corresponds to $\mathsf{SIR}_n(t) > \beta$. Otherwise, the transmitter backs off and makes a retransmission attempt after a random amount of time. If T_n transmits the packet, the packet transmission can still fail if SIR_n falls below β during the packet transmission time L. Thus, the outage probability

$$P_{\text{out}} = P_b + (1 - P_b) P_{\text{fail|no back-off}},$$

where P_b is the back-off probability, and $P_{\text{fail|no back-off}}$ is the probability that the transmission fails during transmission. Hence, the transmission capacity with CSMA is defined as

$$C = \lambda(1 - P_{\text{out}})B \text{ bits/sec/Hz/m}^2.$$

Remark 2.8.4 *CSMA introduces correlation among different transmitter's back-off events, and hence the number of simultaneously active transmitters no longer follow a PPP. Nevertheless, for analytical tractability, as an approximation, we assume that the transmitter back-off events are independent, and simultaneously active transmitter locations are still PPP distributed. The simulation results show that this assumption is reasonable [10].*

In the following theorem, we derive the back-off probability for any transmitter with the SINR-based CSMA.

Theorem 2.8.5 *The back-off probability follows a recursive relationship*

$$P_b = 1 - \exp\left(-\lambda(1 - P_b) c \beta^{\frac{2}{\alpha}} d^2\right),$$

which can be solved using Lambert's function $W_0(.)$.

Proof: Under the independent back-off assumption, the set of active transmitters at time 0 is a PPP with density $\sum_{i=-L}^{0} \frac{\lambda}{L}(1-P_b)$ by counting for all active transmitters during the packet length of L time slots. Thus, the density of active transmitters at time 0 is

$$\lambda_a = \lambda(1-P_b).$$

Hence, from Theorem 2.4.1, we get the recursive relation $P_b = \mathbb{P}(\mathsf{SIR}_n(0) < \beta) = 1 - \exp\left(-\lambda(1-P_b)c\beta^{\frac{2}{\alpha}}d^2\right)$. □

Next, we derive an explicit expression for the packet failure probability $P_{\text{out}}(B)$ with the CSMA protocol.

Theorem 2.8.6 $P_{\text{fail|no back-off}} = 1 - \frac{\sum_{\ell=0}^{L+1}(-1)^{\ell}\binom{L+1}{\ell} e^{-\frac{\lambda}{T}\left(\int_{\mathbb{R}^2} 1-\left(\frac{(1-P_b)}{1+d^{\alpha}\beta x^{-\alpha}}+1-(1-P_b)\right)^{\ell} \mathrm{d}x\right)}}{1-P_b}$.

Proof: Note that $P_{\text{fail|no back-off}}$ is the probability that at any time t, $\mathsf{SIR}_n(t) < \beta$ for $0 < t \leq L$ given that $\mathsf{SIR}_n(0) > \beta$. Hence,

$$\begin{aligned} 1 - \mathbb{P}_{\text{fail|no back-off}} &= \mathbb{P}(\mathsf{SIR}_n(1) > \beta, \ldots, \mathsf{SIR}_n(L) > \beta | \mathsf{SIR}_0 > \beta), \\ &= \frac{\mathbb{P}(\mathsf{SIR}_0 > \beta, \mathsf{SIR}_n(1) > \beta, \ldots, \mathsf{SIR}_n(L) > \beta)}{\mathbb{P}(\mathsf{SIR}_0 > \beta)}, \end{aligned} \tag{2.37}$$

and the desired expression for the joint probability in the numerator follows from Proposition 6.2.2, using the probability-generating functional of the PPP (Theorem 2.3.6), similar to Example 2.7.3. The transmitters that become active at any time t between time 0 and L is a PPP with density $\frac{\lambda}{L}(1-P_b)$.

□

Hence using $P_{\text{out}} = P_b + (1-P_b)P_{\text{fail|no back-off}}$, we get the transmission capacity $C = \lambda(1-P_{\text{out}})B$ for CSMA by combining Theorems 2.8.5 and 2.8.6. Finding the closed form expression for $P_{\text{fail|no back-off}}$ derived in Theorem 2.8.6 is quite challenging. An upper bound on the $P_{\text{fail|no back-off}}$, however, can be found using the Fortuin-Kasteleyn-Ginibre inequality as follows.

Definition 2.8.7 *Let $(\Upsilon, \mathcal{F}, \mathcal{P})$ be the probability space. Let $A \in \mathcal{F}$, and $\mathbf{1}_A$ be the indicator function of A. Event $A \in \mathcal{F}$ is called increasing if $\mathbf{1}_A(\omega) \leq \mathbf{1}_A(\omega')$, whenever $\omega \leq \omega', \omega, \omega' \in \Upsilon$ for some partial ordering on ω. The event A is called decreasing if its complement A^c is increasing.*

Lemma 2.8.8 *[FKG Inequality [1]] If both $A, B \in \mathcal{F}$ are increasing or decreasing events then $\mathbb{P}(AB) \geq \mathbb{P}(A)\mathbb{P}(B)$.*

Lemma 2.8.9 *For SINR-based CSMA, $P_{\text{out}} \leq 1 - (1-P_b)^{L+1}$.*

Proof: Clearly, $\mathsf{SIR}_n(t)$ is a decreasing function of the number of interferers, since larger the number of interferers, lesser the SIR. Therefore, the success event $\{\mathsf{SIR}_n(t) > \beta\}$ is a decreasing event. Hence, from the FKG inequality,

$$\mathbb{P}(\mathsf{SIR}_n(0) > \beta, \mathsf{SIR}_n(1) > \beta, \ldots, \mathsf{SIR}_n(L) > \beta) \geq \mathbb{P}(\mathsf{SIR}_0 > \beta)^{L+1},$$

since $\mathsf{SIR}_n(t)$ is identically distributed for any t. Hence, from (2.37), $\mathbb{P}_{\text{fail|no back-off}} \leq 1-(1-P_b)^L$, and $P_{\text{out}} \leq 1-(1-P_b)^{L+1}$.

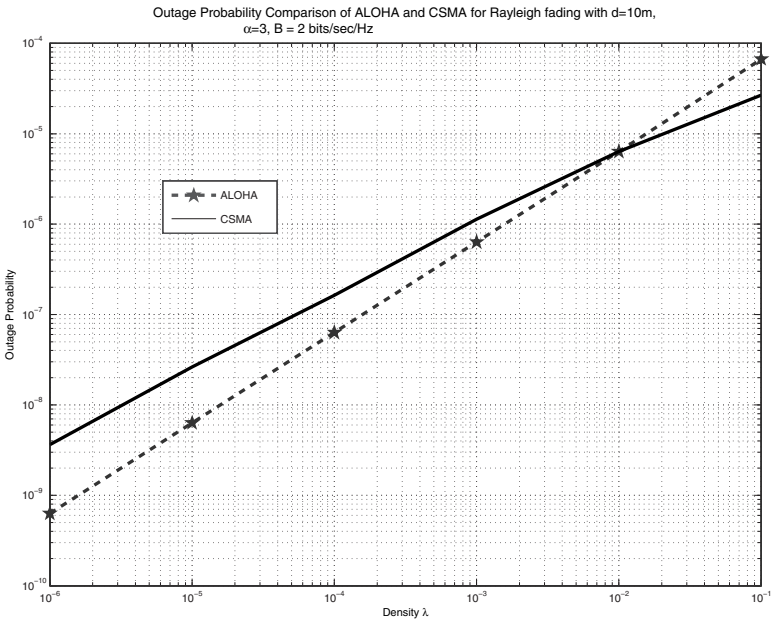

Figure 2.9: Outage probability comparison of ALOHA and SINR-based CSMA with Rayleigh fading.

Consequently, we get a lower bound on the transmission capacity with CSMA as

$$C \geq \lambda(1 - P_b)^{L+1} B.$$

□

Even though we have obtained closed form expression for the outage probability and consequently the transmission capacity, it is not easy to directly compare the SINR-based CSMA protocol and the ALOHA protocol. We hence turn to numerical simulation for comparing their performance.

In Fig. 2.9, we plot the outage probabilities of ALOHA and SINR-based CSMA. For low densities λ, we see that the performance of ALOHA is better than CSMA, because of unnecessary back-offs initiated by CSMA that are not required. However, as the density λ increases, the back-off mechanism of CSMA kicks in and reduces the interference and consequently outperforms the ALOHA protocol.

2.9 Reference Notes

The notion of transmission capacity was introduced in [1], where upper and lower bounds for the path-loss model were presented. The exact transmission capacity expression presented in Section 2.4.1 for the Rayleigh fading model, and the optimal ALOHA probability that maximizes the goodput is derived from [11]. The spatial and temporal correlations with the ALOHA model presented in Section 2.7 can be found in [4]. Transmission capacity analysis with scheduling using guard zone is derived from [6], while the case of scheduling with CSMA can be found in [7, 10].

Bibliography

[1] S. Weber, X. Yang, J. Andrews, and G. de Veciana. 2005. "Transmission capacity of wireless ad hoc networks with outage constraints." *IEEE Trans. Inf. Theory* 51 (12): 4091–02.

[2] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler. 2006. "An ALOHA protocol for multihop mobile wireless networks." *IEEE Trans. Inf. Theory* 52 (2): 421–36.

[3] P. Gupta and P. Kumar. 2000. "The capacity of wireless networks." *IEEE Trans. Inf. Theory* 46 (2): 388–04.

[4] R. Ganti and M. Haenggi. 2009. "Spatial and temporal correlation of the interference in ALOHA ad hoc networks." *IEEE Commun. Lett.* 13 (9): 631–33.

[5] D. Stoyan, W. Kendall, and J. Mecke. 1995. *Stochastic Geometry and Its Applications*. New York, NY: John Wiley and Sons.

[6] A. Hasan and J. G. Andrews. 2007. "The guard zone in wireless ad hoc networks." *IEEE Trans. Wireless Commun.* 6 (3): 897–06.

[7] K. Yuchul, F. Baccelli, and G. De Veciana. 2011. "Spatial reuse and fairness of mobile ad-hoc networks with channel-aware CSMA protocols." In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2011 International Symposium on IEEE*, 360–65.

[8] F. Baccelli and B. Blaszczyszyn. 2009. *Stochastic Geometry and Wireless Networks Part 2*. Boston - Delft: Now Publishers.

[9] R. K. Ganti, J. G. Andrews, and M. Haenggi. 2011. "High-SIR transmission capacity of wireless networks with general fading and node distribution." *IEEE Trans. Inf. Theory* 57 (5): 3100–16.

[10] M. Kaynia, N. Jindal, and G. Oien. 2011. "Improving the performance of wireless ad hoc networks through MAC layer design." *IEEE Trans. Wireless Commun.* 10 (1): 240 –52.

[11] G. Grimmett. 1980. *Percolation*. Berlin-Heidelberg-New York: Springer-Verlag.

Chapter 3

Multiple Antennas

3.1 Introduction

In this chapter, we address an important question on the optimal role of multiple antennas in a wireless network. For a point-to-point channel with no interference, from Chapter 1, we know that employing multiple antennas at both the transmitter and the receiver either linearly increases the capacity or exponentially decreases the error rate with SNR. In contrast, in a wireless network, where interference is the performance limiter, finding how to best use the multiple antennas is a fairly complicated issue.

The problem is challenging because in the presence of interference, multiple antennas have dual roles at both the transmitter and the receiver side. On the transmitter side, multiple antennas can be used to beamform the signal toward the intended receiver or to suppress transmission (construed as interference) toward other receivers. Similarly, on the receiver side, each receiver can use its multiple antennas to improve the SNR from its intended transmitter or cancel the interference coming from other transmitters. To further compound the problem, the roles of multiple antennas at both the transmitter and the receiver side are inter-dependent on each other.

In this chapter, we derive results on the scaling of the transmission capacity with the number of antennas for two cases; i) CSIR case, where only the receivers have channel coefficient/state information (CSI), and ii) CSIT case, where in addition to CSIR, each transmitter also has CSI for its intended receiver. We derive upper and lower bounds on the transmission capacity with multiple antennas that do not match each other exactly but have a negligible gap for path-loss exponent values close to 2.

We show that with linear decoders, for example, zero-forcing (ZF) or minimum mean square error (MMSE), the transmission capacity scales at least linearly with the number of antennas for both the CSIR and the CSIT case, and sending only one data stream from each transmitter achieves the linear scaling of the transmission capacity in both cases. The derived upper and lower bounds are identical for both the CSIR and the CSIT case; thus, we conclude that the value of CSIT is limited in a wireless network. We obtain exact scaling results for transmission capacity with respect to the number of antennas for two important special cases: having only a single antenna at each transmitter/receiver and a simplified receiver with no interference cancelation capability.

We end the chapter by characterizing the effect of the interference suppression capability of multiple antennas at the transmitter. For this end, we consider a cognitive/secondary wireless network that is overlaid over a licensed/primary wireless network, which is allowed to operate

under an outage probability constraint at each receiver node of the primary wireless network. The secondary nodes are equipped with multiple antennas and use them at the transmitter side to suppress the interference they cause to any primary user and at the receiver side to cancel dominant interferers. We obtain explicit results on the scaling of the transmission capacity of the secondary wireless network as a function of the number of transmit and receive antennas available at the secondary nodes.

3.2 Role of Multiple Antennas in ad hoc Networks

In a point-to-point channel with no interference, the only objective with multiple antennas at both the transmitter and the receiver is to improve the received signal strength. In an ad hoc network, however, the role of multiple antennas is more diverse because of the presence of interference. For example, each transmitter could attempt to increase its own data rate by transmitting multiple data streams, or in the presence of CSI, could improve the signal strength by steering its beam toward the direction of the receiver or suppress its interference toward other receivers by nulling its signal toward them. Similarly, each receiver could decode its signal of interest after mitigating interference using its multiple antennas. So inter-dependent questions like how many data streams to transmit and how many interferers to cancel at receiver are critical in finding the optimal role of multiple antennas in a wireless network.

One way to frame these questions more concretely is by defining the spatial transmit (receive) degrees of freedom (STDOF)/(SRDOF). The STDOF refers to the signaling dimensions used at the transmitter for either transmitting to the intended receiver or to suppress interference toward other receivers. For example, with N transmit antennas, there are total N STDOF, out of which possibly k can be used to send k independent data streams, leaving the remaining $N-k$ STDOF for interference suppression in the presence of CSI at the transmitter or not using the $N-k$ STDOF at all to decrease the overall interference at all other receivers. Similarly, the SRDOF refers to the number of spatial dimensions, that through linear processing (linear decoder/receiver), can be used to separate multiple source symbols at the receiver. For example, with N antennas at the receiver, the total SRDOF is equal to N, out of which m can be used for mitigation/cancelation and leaving the remainder of $N-m$ SRDOF for decoding the signal of interest.

When k STDOF are used by each transmitter to send k independent data streams, the number of interferers that can be canceled at any receiver using its m SRDOF is at most $\left\lfloor \frac{m}{k} \right\rfloor$. Larger STDOFs help in improving per-user transmission rates by sending more data streams but limit the interference suppression ability of any receiver. In this chapter, we find the optimal values of STDOF used for transmission k and SRDOF m used for interference cancelation that maximize the transmission capacity with linear decoders under different CSI assumptions. The choice of linear decoders is made for both their analytical tractability and low-complexity implementation.

3.3 Channel State Information Only at Receiver

We consider the fixed distance model of Section 2.2, where each transmitter–receiver pair is at a fixed distance of d from each other. The transmitter locations $\{T_n\}$ are assumed to follow a Poisson point process (PPP) distribution with density λ_0. Each transmitter is assumed to transmit independently with probability p using an ALOHA protocol. Consequently, the active transmitter density is $\lambda = p\lambda_0$. We let $\Phi = \{T_n : T_n \text{ is active}\}$ to represent the active transmitter locations that is a PPP with density λ.

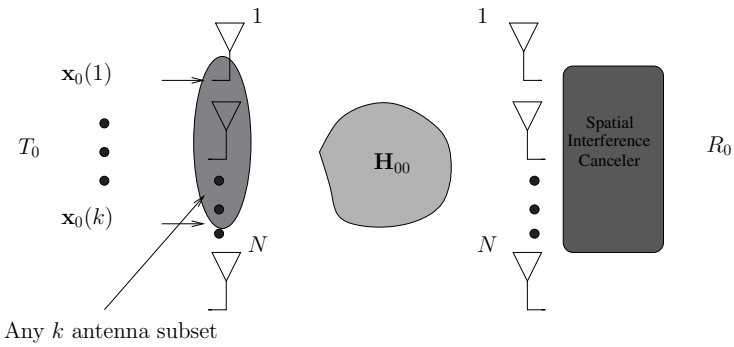

Figure 3.1: Transmit-receive strategy with no CSI at the transmitter.

In this section, we consider the practically efficient model where each receiver has instantaneous CSI, while no transmitter has any instantaneous or delayed CSI. We refer to this scenario as CSIR (CSI at the receiver). The case with CSI at the transmitter (referred to as CSIT) is dealt in Section 3.4.

With no CSI at any transmitter, we assume that each transmitter uses any k, $k = 1, 2, \ldots, N$, of its N antennas to transmit k independent data streams to its receiver with equally distributing the power over all the k antennas. In terms of STDOF, each transmitter uses k STDOF for transmission out of its total N STDOF. Since CSI is not available, the choice of which antennas to use does not impact the performance. Each receiver is assumed to have CSI for the channel from its intended transmitter as well as from all the other interferers that are canceled/suppressed at that receiver.

Let $\mathbf{x}_n = [\mathbf{x}_n(1)\ \mathbf{x}_n(2) \ldots \mathbf{x}_n(k)]^T$ be the $k \times 1$ signal sent from transmitter T_n, where each element $\mathbf{x}_n(\ell),\ \ell = 1, 2, \ldots, k$ is independent and $\mathcal{CN}\left(0, \frac{1}{k}\right)$ distributed, so that the total power transmitted through $\mathbf{x}_n$ is unity. Then, the multiple antennas counterpart of received signal (2.4) at the typical receiver R_0 is

$$\mathbf{y}_0 = d^{-\alpha/2}\mathbf{H}_{00}\mathbf{x}_0 + \sum_{T_n \in \Phi \setminus \{T_0\}} d_n^{-\alpha/2}\mathbf{H}_{0n}\mathbf{x}_n, \tag{3.1}$$

where d_n is the distance between T_n and R_0, $\mathbf{H}_{0n} \in \mathbb{C}^{N \times k}$ is the channel coefficient matrix between T_n and R_0, such that the i, jth entry $\mathbf{H}_{0n}(i, j)$ of $\mathbf{H}_{0n}$ is the channel coefficient between the ith receive antenna of R_0 and jth transmit antennas of T_n. Each entry of $\mathbf{H}_{0n}$ is assumed to be independent and Rayleigh distributed. We consider the interference-limited regime and ignore the AWGN contribution. For analysis, we will consider the typical transmitter–receiver pair (T_0, R_0).

Interference cancelation: To cancel interference, each receiver multiplies its received signal with vector $\mathbf{q}^\dagger$ that lies in the null space of the channel matrices corresponding to the interferers that are chosen for cancelation. Thus, if $\mathcal{C} \in \Phi$ is the subset of interferers to be canceled, then $\mathbf{q} \in \mathsf{O}(\mathcal{H}_\mathcal{C})$, where $\mathsf{O}(\mathcal{H}_\mathcal{C})$ represents the null/orthogonal space of matrix $\mathcal{H}_\mathcal{C} = [\mathbf{H}_{0n}], n \in \mathcal{C}$. The system model is illustrated in Fig. 3.1.

Which interferers to cancel: Each receiver R_n with multiple antennas has to make a judicious choice of which interferers it should cancel before decoding its signal of interest $\mathbf{x}_n$. The most

natural choice is to cancel those interferers that maximize the post-cancelation SIR, that is, to find subset $\mathcal{C}$ that solves

$$\max_{\mathcal{C}} \mathsf{SIR} = \max_{\mathcal{C}} \frac{d^{-\alpha}\mathbf{q}\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}\mathbf{q}^{\dagger}}{\sum_{T_n \in \Phi \setminus \{T_0, \mathcal{C}\}} d_n^{-\alpha}\mathbf{q}\mathbf{H}_{0n}\mathbf{H}_{0n}^{\dagger}\mathbf{q}^{\dagger}}. \tag{3.2}$$

Solving (3.2) is however, complicated and also the performance analysis is difficult. As we have seen in Chapter 2, typically, the closest interferers dominate the total interference seen at any receiver. This motivates the choice of canceling the nearest interferers in terms of distance from each receiver R_n. Canceling the nearest interferers is also efficient in terms of CSI requirement, since CSI from only the nearest interferers to be canceled is required, in comparison to the global CSI requirement for solving (3.2). Since any receiver with N antennas can cancel at most N interferers, CSI is only needed from at most N nearest interferers. Throughout this chapter, for analyzing the transmission capacity with multiple antennas, we assume that each receiver cancels its nearest interferers.

Another choice for interference cancelation is to cancel those interferers that have the largest interference power at the receiver. With this choice, some of the nearby interferers may not be canceled if their channel gains are very low. After multiplication by the cancelation vector $\mathbf{q}^{\dagger}$, however, the situation might change, and the post-cancelation channel gain values of the nearby uncanceled interferers could become moderately high, and they could start dominating the performance. We discuss this choice briefly in Remark 3.3.11 from the transmission capacity point of view.

Choice of decoder: To decode $\mathbf{x}_0$ from (3.1), the optimal decoder is the maximum likelihood (ML) decoder which finds $\mathbf{x}_0$ that maximizes the likelihood $\mathbb{P}(\mathbf{y}_0|\mathbf{x}_0)$. As discussed in Chapter 1, the complexity of the ML decoder is quite high since it finds the jointly optimal vector $\mathbf{x}_0$. Moreover, for the transmission capacity analysis with the ML decoder, we need to be able to analyze the outage probability $\mathbb{P}(\mathsf{I}(\mathbf{x}_0;\mathbf{y}_0) < \beta)$, where $\mathsf{I}(\mathbf{x}_0;\mathbf{y}_0)$ is the mutual information between input $\mathbf{x}_0$ and output $\mathbf{y}_0$. The exponent of outage probability $\mathbb{P}(\mathsf{I}(\mathbf{x}_0;\mathbf{y}_0) < \beta)$ for the multiple input multiple output (MIMO) channel is only known for the high SNR regime [1] and that too in the absence of interference. Thus, in the presence of interference, meaningful analysis of transmission capacity is not possible with the optimal ML decoder. The obvious other choice is to consider linear decoders, such as ZF or MMSE decoder. As discussed in Section 1.2, with linear decoders, each element of the input signal vector $\mathbf{x}_0$ is decoded separately allowing the use of scalar outage probability expressions, while incurring linear decoding complexity in the size of vector $\mathbf{x}_0$. For detailed analysis purposes, we will consider the ZF decoder and point out that identical results can be obtained for MMSE decoder as well in Remark 3.3.7. In particular, throughout this chapter, we consider a general ZF decoder called the *partial ZF decoder* that allows the flexibility of choosing a variable number of SRDOF for interference cancelation and leaving the remaining SRDOF for decoding the signal of interest.

3.3.1 Transmission Capacity With Partial ZF Decoder

With k data streams sent from each transmitter, and each receiver using m SRDOF for interference cancelation, let $N_{\mathsf{canc}} = \left\lfloor \frac{m}{k} \right\rfloor$ be the number of nearest canceled interferers. To cancel the nearest interferers, let the indices of the interferers be sorted in an increasing order in terms of their distance

from the typical receiver R_0, that is, $d_1 \leq d_2 \leq \ldots \leq d_{N_{\mathsf{canc}}} \leq d_{N_{\mathsf{canc}}+1} \leq \ldots$. Then the received signal (3.1) is

$$\begin{aligned}\mathbf{y}_0 &= d^{-\alpha/2}\mathbf{H}_{00}(\ell)\mathbf{x}_0(\ell) + \sum_{j=1, j\neq\ell}^{k} d^{-\alpha/2}\mathbf{H}_{00}(j)\mathbf{x}_0(j) \\ &\quad + \sum_{n=1}^{\infty} d_n^{-\alpha/2} \sum_{j=1}^{k} \mathbf{H}_{0n}(j)\mathbf{x}_n(j), \end{aligned} \tag{3.3}$$

where we have intentionally separated the data stream $\mathbf{x}_0(\ell)$ and the rest of the data streams $\mathbf{x}_0(1), \ldots, \mathbf{x}_0(\ell-1),\ \mathbf{x}_0(\ell+1), \ldots, \mathbf{x}_0(k)$ sent by the typical transmitter T_0.

To decode the $\mathbf{x}_0(\ell)^{\text{th}}$ data stream sent from transmitter T_0, $\ell = 1, 2, \ldots, k$, receiver R_0 uses partial ZF decoder to remove the inter-stream interference from all the other data streams

$$\mathbf{x}_0(1), \ldots, \mathbf{x}_0(\ell-1),\ \mathbf{x}_0(\ell+1), \ldots, \mathbf{x}_0(k)$$

sent by transmitter T_0, and all the k data streams transmitted by the N_{canc} nearest interferers $\mathbf{x}_n(j),\ n = 1, 2, \ldots, N_{\mathsf{canc}},\ j = 1, 2, \ldots, k$.

Let

$$\mathcal{H} = [\mathbf{H}_{00}(1) \ldots \mathbf{H}_{00}(\ell-1)\ \mathbf{H}_{00}(\ell+1) \ldots \mathbf{H}_{00}(k)\ \mathbf{H}_{01}\ \mathbf{H}_{02} \ldots \mathbf{H}_{0N_{\mathsf{canc}}}],$$

where $\mathcal{H} \in \mathbb{C}^{N \times m+k-1}$ be the channel matrix corresponding to the $k-1$ inter-stream interferers, and the N_{canc} nearest interferers in (3.3), where $\mathbf{H}_{0n} \in \mathbb{C}^{N \times k}$ is the channel matrix corresponding to the nth nearest interferer, $N > m + k - 1$. Since each channel coefficient is i.i.d. Rayleigh distributed, the rank of matrix $\mathcal{H}$ is $m + k - 1$ with probability 1. Let S be the orthonormal basis of the null space $\mathsf{O}(\mathcal{H})$ of the matrix $\mathcal{H}$, where S has dimension $N - (m + k - 1)$. To decode stream $\mathbf{x}_0(\ell)$, the receiver R_0 multiplies $\mathbf{q}_\ell^\dagger,\ \mathbf{q}_\ell^\dagger \in \mathsf{O}(\mathcal{H})$ to the received signal (3.3) to get

$$\mathbf{q}_\ell^\dagger \mathbf{y}_0 = d^{-\alpha/2}\mathbf{q}_\ell^\dagger \mathbf{H}_{00}(\ell)\mathbf{x}_0(\ell) + \sum_{n=N_{\mathsf{canc}}+1}^{\infty} d_n^{-\alpha/2} \sum_{j=1}^{k} \mathbf{q}_\ell^\dagger \mathbf{H}_{0n}(j)\mathbf{x}_n(j), \tag{3.4}$$

$\ell = 1, 2, \ldots, k$. Similar to the choice of which interferers to cancel, there is choice for selecting the interference cancelation vector $\mathbf{q}_\ell^\dagger$. The obvious choice is the one that maximizes the SIR, however, that leads to analytic intractability.

So we consider the next best option of choosing $\mathbf{q}_\ell^\dagger \in \mathsf{O}(\mathcal{H})$ that maximizes the signal power $s = |\mathbf{q}_\ell^\dagger \mathbf{H}_{00}(\ell)|^2$. In Lemma 3.3.1, we show that the optimal $\mathbf{q}_\ell$ that maximizes the signal power s is given by

$$\mathbf{q}_\ell = \frac{\mathbf{H}_{00}(\ell)^\dagger \mathsf{S}\mathsf{S}^\dagger}{|\mathbf{H}_{00}(\ell)^\dagger \mathsf{S}\mathsf{S}^\dagger|},$$

and the signal power $s = |\mathbf{q}_\ell^\dagger \mathbf{H}_{00}(\ell)|^2$ is $\chi^2(2(N - m - k + 1))$, since the dimension of S is $N - k - m + 1$. Moreover, since $\mathbf{q}_\ell^\dagger$ is chosen to maximize the signal power $s = |\mathbf{q}_\ell^\dagger \mathbf{H}_{00}(\ell)|^2$, it does not depend on the uncanceled interferer's channels $\mathbf{H}_{0n}$ for $n \geq N_{\mathsf{canc}} + 1$ in (3.3). Hence, the power of the jth stream of the nth interferer, $n \geq N_{\mathsf{canc}} + 1$, $|\mathbf{q}_\ell^\dagger \mathbf{H}_{0n}(j)|^2$ is $\chi^2(2)$, since each entry of $\mathbf{H}_{0n}$ is independent and Rayleigh distributed. Adding the contribution from k independent data streams of each interferer, the total interference power of the nth uncanceled interferer from its k data streams in (3.3) is $\mathsf{pow}_n = \sum_{j=1}^{k} |\mathbf{q}_\ell^\dagger \mathbf{H}_{0n}(j)|^2$ that is $\chi^2(2k)$ distributed.

Lemma 3.3.1 *Let* $\mathbf{Q} \in \mathbb{C}^{N\times\ell}$, $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$. *Then*

$$\arg \max_{v\in\mathbf{Q},|v|^2=1} |\mathbf{v}^\dagger\mathbf{h}_0|^2 = |\mathbf{Q}^\dagger\mathbf{h}_0|^2,$$

and $|\mathbf{Q}^\dagger\mathbf{h}_0|^2$ *is* $\chi^2(2\ell)$ *if* $\mathbf{h}_0 \in \mathbb{C}^{N\times 1}$ *is complex Gaussian distributed with independent entries that have zero mean and unit variance.*

Proof: Without loss of generality, let $\mathbf{v} = \frac{\mathbf{Q}x}{|\mathbf{Q}x|}$. From Cauchy–Schwarz inequality,

$$< \mathbf{h}_0^\dagger, \mathbf{Q}x > = < \mathbf{h}_0^\dagger\mathbf{Q}, x > \leq |\mathbf{h}_0^\dagger\mathbf{Q}|^2$$

and the maximum is achieved by $x = \mathbf{Q}^\dagger\mathbf{h}_0$. Thus, we get that

$$\max_{v\in\mathbf{Q},|\mathbf{v}|^2=1} |\mathbf{v}^\dagger\mathbf{h}_0|^2 = |\mathbf{h}_0^\dagger\mathbf{Q}\mathbf{Q}^\dagger\mathbf{h}_0|,$$

which is the norm of vector $\mathbf{h}_0^\dagger\mathbf{Q}$.

Since the columns of $\mathbf{Q}$ are orthonormal, the covariance matrix of vector $\mathbf{h}_0^\dagger\mathbf{Q}$ of length ℓ is diagonal, where the expectation is with respect to entries of $\mathbf{h}_0$. Thus, the elements of $\mathbf{h}_0^\dagger\mathbf{Q}$ that are Gaussian distributed are uncorrelated and hence are independent. Since the norm of an ℓ-length independent Gaussian vector is $\chi^2(2\ell)$ distributed, the result follows. More details can be found in [2]. □

Lemma 3.3.2 *Let* $\mathsf{x}_1, \ldots, \mathsf{x}_m$ *be a set of* m n*-length Gaussian vectors, whose each element is independent with zero mean and unit variance. If vector* y *of length* n *is independent of* $\mathsf{x}_1 \ldots \mathsf{x}_m$, *then* $|\mathsf{y}^\dagger\mathsf{x}_i|^2 \sim \chi^2(2)$, $\forall\, i$ *and* $\sum_{i=1}^m |\mathsf{y}^\dagger\mathsf{x}_i|^2 \sim \chi^2(2m)$.

Definition 3.3.3 *With partial ZF decoder, from* (3.4)*, the SIR for the* ℓth *stream is given by*

$$\mathsf{SIR}_\ell = \frac{d^{-\alpha}s}{\sum_{n=N_{\mathsf{canc}}+1}^{\infty} d_n^{-\alpha}\mathsf{pow}_n}, \tag{3.5}$$

where from Lemma 3.3.1, signal power $s = |\mathbf{q}_\ell^\dagger\mathbf{H}_{00}(\ell)|^2 \sim \chi^2(2(N-k-m+1))$, *and from Lemma 3.3.2, interference power* $\mathsf{pow}_n = \sum_{j=1}^k |\mathbf{q}_\ell^\dagger\mathbf{H}_{0n}(j)|^2 \sim \chi^2(2k)$.

Note that the same decoding strategy is used for each stream $\ell = 1, 2, \ldots, k$ sent by any transmitter T_n; therefore, the SIR for each stream ℓ is identically distributed. Henceforth we drop the subscript ℓ from SIR_ℓ and represent it as SIR. Thus, for each stream ℓ, $\ell = 1, 2, \ldots, k$, the outage probability at rate B bits/sec/Hz is given by

$$\begin{aligned} P_{\text{out}}(B) &= \mathbb{P}\left(\log(1+\mathsf{SIR}) \leq B\right), \\ &= \mathbb{P}\left(\mathsf{SIR} \leq 2^B - 1\right). \end{aligned} \tag{3.6}$$

Let $2^B - 1 = \beta$. Since k streams are transmitted simultaneously, the transmission capacity is defined as

$$C = k\lambda(1-\epsilon)B \text{ bits/sec/Hz/m}^2, \tag{3.7}$$

where ϵ is the outage probability constraint for each data stream, and λ is the maximum density of nodes such that $P_{\text{out}}(B) \leq \epsilon$ in (3.6). Here, C represents the average successful rate of information transfer across the network when each transmitter sends k independent data streams.

From (3.5, 3.6),

$$P_{\text{out}}(B) \quad = \quad \mathbb{P}\left(\frac{d^{-\alpha}s}{I_{nc}} \leq \beta\right), \tag{3.8}$$

where $I_{nc} = \sum_{n=N_{\mathsf{canc}}+1}^{\infty} d_n^{-\alpha}\mathsf{pow}_n$ is the total interference from uncanceled interferers, $d_n \leq d_m,\ n < m$, where d_n's are ordered in increasing distance from the receiver R_0.

Remark 3.3.4 *For the case of $k = 1$ (single data stream transmission), the outage probability expression* (3.6) *and consequently the transmission capacity expression* (3.7) *is also valid for the ML decoder. Thus, the performance of ML decoder and ZF decoder is identical for $k = 1$. We will show in Theorem 3.3.5 that the optimal $k^\star = 1$ with the ZF decoder and the transmission capacity scales at least linearly with N. Thus, the same conclusion holds true for the ML decoder. What is left open is the fact whether $k^\star = 1$ for ML decoder or not? The simulation results (Fig. 3.3) point out that $k^\star = 1$ even for the ML decoder.*

To find the transmission capacity expression (3.7), and to maximize that with respect to the number of transmitted data streams k, and the number of SRDOF m used to cancel the nearest interferers, we need to find a closed form expression for the outage probability (3.8). Unfortunately, this is hard to find since the distribution of I_{nc} is unknown. We thus rely on deriving upper and lower bounds on the outage probability that allow us to find the optimal values of k and m that maximize the transmission capacity.

The main result of this section is as follows (derived from [3] and [4]):

Theorem 3.3.5 *The transmission capacity with multiple antennas and partial ZF decoder receiver scales as*

$$C = \Omega(N) \text{ and } C = \mathcal{O}(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}}),$$

with the number of antennas N. The optimal lower bound is achieved by sending a single data stream from each transmitter, $k^\star = 1$, and using $m^\star = \left(1-\frac{2}{\alpha}\right) N$ SRDOF for interference cancelation to cancel the $\left(1-\frac{2}{\alpha}\right) N$ nearest interferers.

Theorem 3.3.5 tells us that similar to point-to-point channels without interference, the transmission capacity of wireless network scales at least linearly with the number of antennas N. The derived upper bound does not match with the lower bound; however, the gap is negligible for path-loss exponents α close to 2 and the maximum gap is $N^{1/4}$ at $\alpha = 4$ for any $2 < \alpha \leq 4$. The simulation results, however, suggest that this gap is only a manifest of the proof technique and not the underlying principle and transmission capacity cannot scale faster than order N. Using simulations, we show that the transmission capacity can be at best scale linearly with N by checking for many different combinations of k and m. Thus, a more finer analysis is required for exactly characterising the scaling of the transmission capacity with multiple antennas. For the special case, when each transmitter has a single antenna or chooses $k = 1$, we can obtain the exact results for the scaling of transmission capacity with respect to number of antennas N in Theorem 3.3.6.

A more important conclusion from Theorem 3.3.5 is that to maximize the lower bound, that is, to achieve linear scaling with N, $k^\star = 1$, and $m^\star \propto N$, that is, one should only transmit a single data stream from each transmitter, while linearly scaling the SRDOF dedicated for interference cancelation. To interpret this result, note that the number of transmitted data streams is directly related to the interference power and the number of interferers that can be canceled by each receiver. Thus, in an interference limited network, such as an ad hoc wireless

network, minimizing the number of data streams transmitted by each node keeps the interference power low and at the same time leaves enough room for canceling significant number of nearest interferers at the receiver. Moreover, by scaling the SRDOF used for interference cancelation linearly with N, the number of nearest interferers that can be canceled scales linearly with N, while leaving sufficient SRDOF (that also scales linearly in N) for decoding the signal of interest.

For the special case when each transmitter is equipped with a single transmit antenna, or chooses $k = 1$ irrespective of N, the proof of Theorem 3.3.5 can be used to obtain the exact scaling of transmission capacity with the number of receiver antennas as follows.

Theorem 3.3.6 *With a single antenna at each transmitter or fixed $k = 1$, the transmission capacity with partial ZF decoder scales linearly with the number of receive antennas N, that is,*

$$C = \Theta(N),$$

and the optimal SRDOF for interference cancelation is $m^\star = \left(1 - \frac{2}{\alpha}\right) N$.

Theorem (3.3.6) shows that linear scaling of transmission capacity is possible even if each transmitter has a single antenna. Thus, in a wireless network, the role of multiple antennas at the receiver side is more important than the transmit side. This result is in contrast to a point-to-point channel where the capacity scales linearly with the minimum of the transmit and the receive antennas.

Remark 3.3.7 *An alternate choice of linear decoder is the MMSE decoder, where to decode the data stream $\mathbf{x}_0(\ell)$ from received signal* (3.1)*, $\Sigma_\ell^{-1}\mathbf{H}_{00}(\ell)$ is multiplied to the received signal, where $\mathbf{H}_{00}(\ell)$ is the ℓth column of $\mathbf{H}_{00}$, and*

$$\Sigma_\ell = \sum_{i=1, i\neq\ell}^{k} d^{-\alpha}\mathbf{H}_{00}(i)\mathbf{H}_{00}(i)^\dagger + \sum_{T_n\in\Phi\setminus\{T_0\}} d_n^{-\alpha/2}\mathbf{H}_{0n}\mathbf{H}_{0n}^\dagger$$

is the spatial covariance matrix of the inter-stream interference and interference caused by other transmitters. The MMSE decoder is known to maximize the received SINR, and hence the lower bound derived for the transmission capacity in Theorem 3.3.5 with partial ZF decoder also holds for the MMSE case as well.

To upper bound the transmission capacity with MMSE decoder, we can let the signal power with the MMSE decoder to be distributed as $\chi^2(2N)$, which is clearly an idealization since after canceling interferers by multiplying $\Sigma_\ell^{-1}\mathbf{H}_{00}(\ell)$, the signal power is $\mathbf{H}_{00}^\dagger(\ell)\Sigma_\ell^{-1}\mathbf{H}_{00}(\ell)$ which is less than the norm of vector $\mathbf{H}_{00}(\ell)$ that is distributed as $\chi^2(2N)$. Moreover, by selecting $r = N$ (where r is the number of uncanceled nearest interferers used for lower bounding the outage probability in Theorem 3.3.5) an upper bound identical to Theorem 3.3.5 on the transmission capacity can also be found for the MMSE decoder as well [5]. Thus, results obtained for the ZF decoder also hold for the MMSE decoder as well.

Towards proving Theorem 3.3.5, we first upper and lower bound the outage probability (3.8) as follows:

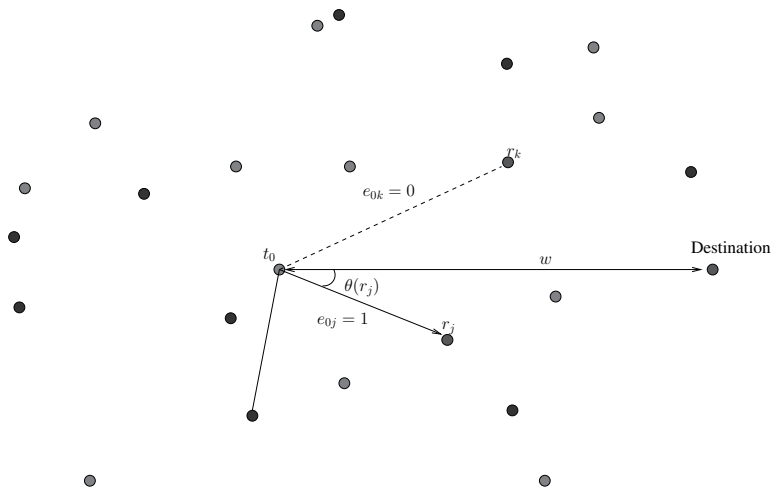

Figure 3.2: Squares represent the N_{canc} nearest canceled interferers with dashed lines, solid circles represent the r nearest uncanceled interferers whose interference contribution will be used to derive the lower bound on the outage probability, and unfilled circles are all the other uncanceled interferers.

Theorem 3.3.8 *The outage probability* (3.8) *when the transmitter sends k independent data streams and the receiver cancels the N_{canc} nearest interferers using the partial ZF decoder is lower bounded by*

$$P_{\text{out}}(B) \quad \geq \quad 1 - \frac{N-m-k+1}{(kr-1)d^{\alpha}\beta\,(\pi\lambda)^{\frac{\alpha}{2}}} \left(N_{\mathsf{canc}} + r + \frac{\alpha}{2}\right)^{\frac{\alpha}{2}},$$

for any $r \in \mathbb{N}^{+}$ such that $kr > 1$.

To derive this lower bound, we consider the interference contribution from only the r nearest uncanceled interferer (the $N_{\mathsf{canc}} + 1^{st}$ interferer to $N_{\mathsf{canc}} + r\text{th}$ interferer) and consider their aggregate interference

$$I_{nc}^{r} = \sum_{j=1}^{r} d_{N_{\mathsf{canc}}+j}^{-\alpha}\, \mathsf{pow}_{N_{\mathsf{canc}}+j}. \tag{3.9}$$

Since $I_{nc}^{r} < I_{nc}$, from (3.8), we have

$$P_{\text{out}}(B) = \mathbb{P}\left(\frac{d^{-\alpha}s}{I_{nc}} \leq \beta\right) \geq \mathbb{P}\left(\frac{d^{-\alpha}s}{I_{nc}^{r}} \leq \beta\right).$$

For any r, we can efficiently bound the outage probability $\mathbb{P}\left(\frac{d^{-\alpha}s}{I_{nc}^{r}} \leq \beta\right)$ using the Markov's inequality as follows.

Proof: Consider the interference contribution from only the r nearest uncanceled interferers, I_{nc}^{r}. Fig. 3.2 illustrates this scenario, where the N_{canc} (squares) have been canceled, and only the interference coming from the r nearest uncanceled neighbors of receiver R_0 are considered toward computing the outage probability. To derive the lower bound, we use the Markov's inequality with

s (signal power) as the random variable and compute the expectation with respect to the interference power I_{nc}^r. From (3.8),

$$\begin{aligned} 1 - P_{\text{out}}(B) &= \mathbb{P}\left(s > d^{\alpha}\beta I_{nc}\right) \\ &\leq \mathbb{P}\left(s > d^{\alpha}\beta I_{nc}^r\right), \text{ since } I_{nc}^r \leq I_{nc} \text{ from (3.9)}, \\ &\leq \mathbb{E}\left\{\frac{\mathbb{E}\{s\}}{d^{\alpha}\beta I_{nc}^r}\right\}, \quad \text{from Markov's inequality.} \end{aligned} \tag{3.10}$$

Since interferers are ordered in increasing distance from the receiver R_0, $d_{N_{\mathsf{canc}}+j} \leq d_{N_{\mathsf{canc}}+r},\ j = 1, \ldots, r-1$, we have $I_{nc}^f \geq I_{nc}^r$, where

$$I_{nc}^f = d_{N_{\mathsf{canc}}+r}^{-\alpha} \sum_{j=1}^{r} \mathsf{pow}_{N_{\mathsf{canc}}+j} \tag{3.11}$$

is obtained by substituting for each of the path-loss term $d_{N_{\mathsf{canc}}+j}^{-\alpha}$ in I_{nc}^r by the path-loss term of the farthest interferer $d_{N_{\mathsf{canc}}+r}^{-\alpha}$. Hence, from (3.10),

$$1 - P_{\text{out}}(B) \leq \mathbb{E}\left\{\frac{\mathbb{E}\{s\}}{d^{\alpha}\beta I_{nc}^f}\right\}. \tag{3.12}$$

There are three random variables involved in the analysis: signal power s, channel power of uncanceled interferers $\mathsf{pow}_{N_{\mathsf{canc}}+j}$, and the distance of the farthest uncanceled interferers $d_{N_{\mathsf{canc}}+r}$ considered for deriving the bound. From Definition 3.3.3, $s \sim \chi^2(2(N-m-k+1))$ and $\mathsf{pow}_{N_{\mathsf{canc}}+j}$ are independent and $\chi^2(2k)$ distributed, hence, their sum $\mathsf{pow} = \sum_{j=1}^{r} \mathsf{pow}_{N_{\mathsf{canc}}+j} \sim \chi^2(2kr)$. Moreover, from Lemma 3.3.9, we have $\pi\lambda d_{N_{\mathsf{canc}}+r}^2 \sim \chi^2(2(N_{\mathsf{canc}}+r))$. □

Lemma 3.3.9 *Let d_n be the distance of the nth nearest node of a PPP Φ with density λ from the origin. Then $\pi\lambda d_n^2 \sim \chi^2(2n)$.*

Proof: The result follows from the direct computation of the distribution of $\pi\lambda d_n^2$ by finding the distribution $\mathbb{P}(d_n > r)$ using the void probability of PPP Φ with density λ. □

Hence, from (3.12), we have $1 - P_{\text{out}}(B)$

$$\begin{aligned} &\overset{(a)}{\leq} \frac{\mathbb{E}\{s\}}{d^{\alpha}\beta}\mathbb{E}\left\{\frac{1}{d_{N_{\mathsf{canc}}+r}^{-\alpha}}\right\}\mathbb{E}\left\{\frac{1}{\mathsf{pow}}\right\}, \\ &\overset{(b)}{=} \frac{N-k-m+1}{d^{\alpha}\beta\left(\pi\lambda\right)^{\frac{\alpha}{2}}}\int_0^{\infty}\frac{x^{N_{\mathsf{canc}}+r+\frac{\alpha}{2}}e^{-x}}{N_{\mathsf{canc}}+r!}\mathrm{d}x\int_0^{\infty}\frac{\mathsf{pow}^{kr-2}e^{-\mathsf{pow}}}{kr-1!}\mathrm{d}\mathsf{pow}, \\ &\leq \frac{N-k-m+1}{d^{\alpha}\beta\left(\pi\lambda\right)^{\frac{\alpha}{2}}}\frac{\Gamma\left(N_{\mathsf{canc}}+r+1+\frac{\alpha}{2}\right)}{\Gamma\left(N_{\mathsf{canc}}+r+1\right)}\left(\frac{1}{kr-1}\right), \end{aligned} \tag{3.13}$$

where in (a) we have substituted for I_{nc}^f from (3.11) and (b) follows since $s \sim \chi^2(2(N-m-k+1))$, $\mathsf{pow} = \sum_{j=1}^{r} \mathsf{pow}_{N_{\mathsf{canc}}+j} \sim \chi^2(2kr)$, and $\pi\lambda d_{N_{\mathsf{canc}}+r}^2 \sim \chi^2(2r)$.

From Kershaw's inequality [6] that states that $\frac{\Gamma(x+1)}{\Gamma(x+s)} \leq \left(x - \frac{1}{2} + \sqrt{s+\frac{1}{4}}\right)^{1-s}$ for $x > 0, 0 < s < 1$, since $\Gamma(x+1) = x\Gamma(x)$ we have

$$\frac{\Gamma\left(N_{\mathsf{canc}} + r + 1 + \frac{\alpha}{2}\right)}{\Gamma\left(N_{\mathsf{canc}} + r + 1\right)} \leq \left(N_{\mathsf{canc}} + r + 1 + \frac{\alpha}{2}\right)^{\frac{\alpha}{2}}.$$

Hence, from (3.13), we have,

$$1 - P_{\text{out}}(B) \quad \leq \quad \frac{N-k-m+1}{(kr-1)d^{\alpha}\beta\left(\pi\lambda\right)^{\frac{\alpha}{2}}}\left(N_{\mathsf{canc}} + r + \frac{\alpha}{2}\right)^{\frac{\alpha}{2}}.$$

Next, we derive an upper bound on the outage probability (3.8) using the Markov's inequality with I_{nc} as the random variable. For that purpose, we compute an upper bound on $\mathbb{E}\{I_{nc}\}$ using the Campbell's theorem (Theorem 2.3.7).

Theorem 3.3.10 *When each transmitter sends k independent data streams, and the receiver uses m SRDOF for canceling the N_{canc} nearest interferers using the partial ZF decoder, the outage probability* (3.8) *is upper bounded by*

$$P_{\text{out}}(B) \quad \leq \quad \begin{cases} \frac{2k\beta(\pi\lambda)^{\frac{\alpha}{2}}}{d^{-\alpha}(N-k-m)}\left(\frac{\alpha}{2}-1\right)^{-1}\left(N_{\mathsf{canc}} - \lceil\frac{\alpha}{2}\rceil\right)^{1-\frac{\alpha}{2}}, & k+m<N, \\ 1-\exp\left(\frac{-d^{\alpha}\beta 2k(\pi\lambda)^{\frac{2}{\alpha}}\left(N_{\mathsf{canc}}-\lceil\frac{\alpha}{2}\rceil\right)^{1-\frac{\alpha}{2}}}{\left(\frac{\alpha}{2}-1\right)}\right), & k+m=N. \end{cases}$$

Proof: To upper bound the outage probability $\mathbb{P}\left(\frac{d^{-\alpha}s}{I_{nc}} \leq \beta\right)$ (3.8), we use Markov's inequality with $I_{nc} = \sum_{n=N_{\mathsf{canc}}+1}^{\infty} d_n^{-\alpha}\mathsf{pow}_n$, as the random variable. To apply Markov's inequality on $\mathbb{P}(I_{nc} \geq \frac{d^{-\alpha}s}{R})$, we need to bound the expected interference from uncanceled interferers $\mathbb{E}\{I_{nc}\}$ as follows. From (3.8),

$$\begin{aligned} \mathbb{E}\{I_{nc}\} &= \mathbb{E}\left\{\sum_{j\geq N_{\mathsf{canc}}+1} d_j^{-\alpha}\mathsf{pow}_j\right\}, \\ &\overset{(a)}{=} 2k\mathbb{E}\left\{\sum_{j\geq N_{\mathsf{canc}}+1} d_j^{-\alpha}\right\}, \\ &\overset{(b)}{=} 2k\sum_{j\geq N_{\mathsf{canc}}+1}(\pi\lambda)^{\frac{\alpha}{2}}\int_0^{\infty} x^{-\alpha/2}\frac{x^{j-1}\exp(-x)}{(j-1)!}\mathrm{d}x, \\ &= 2k(\pi\lambda)^{\frac{\alpha}{2}}\sum_{j\geq N_{\mathsf{canc}}+1}\frac{\Gamma\left(j-\frac{\alpha}{2}\right)}{\Gamma(j)}, \end{aligned} \tag{3.14}$$

where (a) follows since the power of jth interferer $\mathsf{pow}_j \sim \chi^2(2k)$ and (b) follows from Lemma 3.3.9 where $\pi\lambda d_j^2 \sim \chi^2(2j)$. Note that $\Gamma\left(j-\frac{\alpha}{2}\right)$ is finite only for $j > \frac{\alpha}{2}$; thus, we at least need to cancel at least $\frac{\alpha}{2}$ nearest interferers. Since, typically, $2 < \alpha < 4$, this is not much of a restriction. Using the Kershaw's inequality [6], $\frac{\Gamma(j-\frac{\alpha}{2})}{\Gamma(j)} \leq \left(j - \lceil\frac{\alpha}{2}\rceil\right)^{-\frac{\alpha}{2}}$, from (3.14), we get

$$\mathbb{E}\{I_{nc}\} \quad \leq \quad 2k(\pi\lambda)^{\frac{\alpha}{2}}\sum_{j\geq N_{\mathsf{canc}}+1}\left(j - \left\lceil\frac{\alpha}{2}\right\rceil\right)^{-\frac{\alpha}{2}},$$

$$\begin{aligned} &\overset{(d)}{\leq} && 2k(\pi\lambda)^{\frac{\alpha}{2}} \int_{N_{\mathsf{canc}}} \left(x - \left\lceil \frac{\alpha}{2} \right\rceil\right)^{-\frac{\alpha}{2}} \mathrm{d}x, \\ &= && 2k(\pi\lambda)^{\frac{\alpha}{2}} \left(\frac{\alpha}{2} - 1\right)^{-1} \left(N_{\mathsf{canc}} - \left\lceil \frac{\alpha}{2} \right\rceil\right)^{1-\frac{\alpha}{2}}, \end{aligned} \tag{3.15}$$

where (d) follows since $x^{-\alpha/2}$ is a decreasing function.

Using (3.15), we now derive the required upper bound. From (3.8), $P_{\text{out}}(B) =$

$$\begin{aligned} \mathbb{P}\left(I_{nc} \geq \frac{d^{-\alpha}s}{\beta}\right) &= \mathbb{E}_s\left\{\mathbb{P}\left(I_{nc} \geq \frac{d^{-\alpha}s}{\beta}\right)\right\}, \\ &\leq \frac{\mathbb{E}\{I_{nc}\}\beta}{d^{-\alpha}} \mathbb{E}\left\{\frac{1}{s}\right\}, \text{ from Markov's inequality.} \end{aligned}$$

From Definition 3.3.3, signal power $s \sim \chi^2(2(N-k-m+1))$. Thus, for $N > k+m$, we have $\mathbb{E}\left\{\frac{1}{s}\right\} = \frac{1}{N-k-m}$. Thus, substituting for the upper bound on the expected interference from (3.15),

$$\mathbb{P}\left(I_{nc} \geq \frac{d^{-\alpha}s}{\beta}\right) \leq \frac{2\beta k(\pi\lambda)^{\frac{\alpha}{2}}}{d^{-\alpha}(N-k-m)} \left(\frac{\alpha}{2} - 1\right)^{-1} \left(N_{\mathsf{canc}} - \left\lceil \frac{\alpha}{2} \right\rceil\right)^{1-\frac{\alpha}{2}}.$$

Since $s \sim \chi^2(2(N-k-m+1))$, with $N = k+m$, s is an exponential random variable with parameter 1, and hence

$$\begin{aligned} P_{\text{out}}(B) &= \mathbb{P}(s \leq d^{\alpha}\beta I_{nc}), \\ &= \mathbb{E}\left\{1 - \exp\left(-d^{\alpha}\beta I_{nc}\right)\right\}, \\ &= 1 - \exp\left(-d^{\alpha}\beta \mathbb{E}\left\{I_{nc}\right\}\right), \end{aligned} \tag{3.16}$$

since $\exp$ is a convex function. Thus, we get the required bound by plugging in the upper bound on $\mathbb{E}\left\{I_{nc}\right\}$ from (3.15).

□

Using Theorems 3.3.8 and 3.3.10, we prove Theorem 3.3.5 as follows.

Proof: (Theorem 3.3.5) Recall that the number of nearest canceled interferers are $N_{\mathsf{canc}} = \left\lfloor \frac{m}{k} \right\rfloor$. Using the definition of transmission capacity $C = (1-\epsilon)\lambda B$, and fixing $P_{\text{out}}(B) = \epsilon$, from the lower bound derived on outage probability in Theorem 3.3.8, for any $r \in \mathbb{N}$ such that $kr > 1$,

$$C \leq \frac{kB(1-\epsilon)^{1-\frac{2}{\alpha}}}{\pi} \left(\frac{N-m-k+1}{(kr-1)d^{\alpha}\beta}\right)^{\frac{2}{\alpha}} \left(\left\lfloor \frac{m}{k} \right\rfloor + r + \frac{\alpha}{2}\right). \tag{3.17}$$

In terms of scaling with N, the upper bound is increasing in m, the SRDOF used for interference cancelation, as long as the total SRDOF do not exceed N, that is, $k + m < N$. Thus, we fix $m = \Theta(N)$, the largest scaling factor with respect to N. Let the number of data streams to transmit $k = \Theta(N^{\kappa})$, $\kappa \leq 1$. We will find the tightest upper bound as a function of κ for all values of α, and not just $2 < \alpha \leq 4$. Recall that we can choose the parameter r, the number of uncanceled interferers whose aggregate interference we accounted for while deriving the upper bound in Theorem 3.3.10.

We let $r = N^{2/\alpha}$, one can take ceil or floor function if $N^{2/\alpha}$ is not an integer. Then, for $1-\kappa < \frac{\alpha}{2}$, the upper bound (3.17) is $\mathcal{O}\left(N^{\kappa\left(1-\frac{2}{\alpha}\right)+\frac{4}{\alpha}-\frac{4}{\alpha^2}}\right)$, for which the optimal $\kappa = 1$, and yields

$$C = \mathcal{O}\left(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}}\right).$$

For the other case of $1-\kappa \geq \frac{\alpha}{2}$, the upper bound (3.17) is $\mathcal{O}\left(N^{1-\frac{2}{\alpha}\left(1-\kappa-\frac{2}{\alpha}\right)}\right)$ and the optimal $\kappa = 0$ and upper bound is

$$C = \mathcal{O}\left(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}}\right).$$

Thus, for any κ, we get

$$C = \mathcal{O}\left(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}}\right).$$

Moving on to the lower bound, from Theorem 3.3.10, we consider the case of $k+m < N$ which provides a better lower bound than $k+m = N$. Equating $P_{\text{out}}(B) = \epsilon$, we have

$$C \geq \frac{kB(1-\epsilon)}{\pi}\left(\frac{\epsilon(N-k-m)}{k\beta d^{\alpha}}\right)^{\frac{2}{\alpha}}\left(\left\lfloor\frac{m}{k}\right\rfloor - \left\lceil\frac{\alpha}{2}\right\rceil\right)^{1-\frac{2}{\alpha}}. \tag{3.18}$$

Clearly, $k = 1$, $m = \theta N, \theta \in (0,1]$, yields $C = \Omega(N)$. Finding the best constant θ that maximizes the lower bound (3.18) is equivalent to solving,

$$\max_{\theta}(1-\theta)^{\frac{2}{\alpha}}\theta^{1-\frac{2}{\alpha}}.$$

By setting the derivative to zero, the optimal value of $\theta^{\star} = 1-\frac{2}{\alpha}$. Note that the lower bound on the transmission capacity is concave in m. Thus, to enforce the integer constraint on m, m should be chosen as $\left\lfloor\left(1-\frac{2}{\alpha}\right)N\right\rfloor$ or $\left\lceil\left(1-\frac{2}{\alpha}\right)N\right\rceil$ depending on whichever value maximizes the lower bound. □

Theorem 3.3.5 shows that with single data stream transmission, $k = 1$, and using a linearly increasing (with N) SRDOF for interference cancelation, $m = \theta N$, transmission capacity scales linearly with the number of antennas N. To interpret the optimality of $k = 1$ and $m = \theta N$, we need to look at (3.18), where the term $\left(\frac{(N-k-m)}{k}\right)^{\frac{2}{\alpha}}$ corresponds to the gain obtained by coherently combining the signal of interest using $N-k-m$ SRDOF, while the term $\left(\left\lfloor\frac{m}{k}\right\rfloor - \left\lceil\frac{\alpha}{2}\right\rceil\right)^{1-\frac{2}{\alpha}}$ is attributed to the gain obtained by canceling the m nearest interferers. Thus, using $k = 1$ and $m = \theta N$ allows the two terms to balance out each other and allows a linear increase of the transmission capacity with number of antennas N.

To illustrate the scaling behavior of the transmission capacity with respect to the number of antennas N, we plot the simulated transmission capacity with different transmit–receive strategies, for example, $(k = 1, m = N-1)$, $(k = N/2, N-m = N/2)$ and $(k = 1, m = (1-2/\alpha)N)$ in Fig. 3.3 with increasing N. We plot the transmission capacity with both the partial ZF decoder and the ML decoder. Since for $k = 1$, both the ML and partial ZF decoder are identical, their transmission capacities are also the same. As expected from our derived results, sending a single data stream and using a constant fraction of SRDOF for interference cancelation, $k = 1, m = (1-2/\alpha)N$, achieves a linear increase of transmission capacity with increasing N, in contrast to sublinear increase for the other two cases. More importantly, Fig. 3.3 shows that the upper bound derived in Theorem 3.3.5 where transmission capacity scales super-linearly with N is loose, and

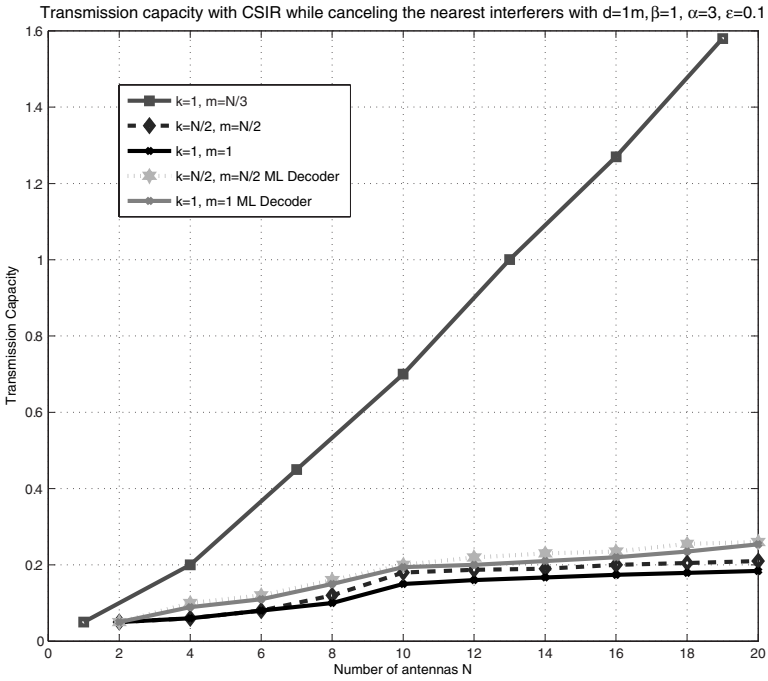

Figure 3.3: Transmission capacity versus N with CSIR while canceling the nearest interferers for $k = 1$, $d = 1$ m, $\beta = 1$ bits, $\alpha = 3$, $\epsilon = 0.1$.

at best only linear increase in N is possible for transmission capacity. Fig. 3.3 also shows that performance of partial ZF decoder is very close to the ML decoder and all the conclusions we draw from Theorem 3.3.5 hold reasonably well for the ML decoder as well.

An important lesson from Theorem 3.3.5 is the utility of using Markov's inequality. Typically, bounds obtained by Markov's inequality are fairly loose, but for transmission capacity purposes this seems to be a handy tool for obtaining tight enough scaling bounds. Next, we indicate how to obtain the exact scaling results of Theorem 3.3.6, when each transmitter has a single antenna or chooses $k = 1$.

Proof: (Theorem 3.3.6) For a single transmit antenna or single stream transmission $k = 1$, we know from Theorem 3.3.5 that $C = \Omega(N)$ by using $m = \theta N$. Moreover, from (3.17), for $k = 1$, choosing $r = N$ with $m = \Theta(N)$, we get $C = \mathcal{O}(N)$, thus finishing the proof. □

Remark 3.3.11 *In this section, we have analyzed the case when each receiver cancels the nearest interferers using its multiple antennas. Another logical choice is to cancel those interferers that have the largest interference power at the receiver. With a single transmit antenna and N receive antennas, the scaling behavior of transmission capacity while canceling the $N - 1$ strongest interferers has been analyzed in [7], and it is shown that the transmission capacity scales as $\epsilon^{1/N}$, where ϵ is the outage probability constraint. Thus, using multiple antennas for canceling the strongest interferers leads to diminished gains compared to canceling the nearest interferers, where the transmission capacity scales linearly with the number of antennas.*

This might appear counter intuitive, however, it can be explained by nothing that while canceling the strongest interferers, some of the nearby interferers may not be canceled if their channel gains are very low. Post-cancelation, that is, after multiplication by the cancelation vector, the situation might change, and the interference power from some of the nearest interferers could become moderate, in which case the nearby interferers dominate the performance.

Next, using the results of this section, we find exact transmission capacity scaling result with respect to the number of antennas when no CSI about other interferers' channels is available at any receiver, thereby precluding the possibility of interference cancelation.

3.3.2 No Interference Cancelation

In this section, we consider the case when no receiver employs any interference cancelation and uses all its SRDOF for decoding the data streams transmitted by its intended transmitter. This scenario is motivated for two important practical reasons. First, interference cancelation requires the knowledge of channel coefficients between the interferer and the receiver which is typically hard to get, especially in a wireless network. Second, the hardware complexity of the receiver is fairly low without the interference cancelation capability. We next show that there is no loss in terms of transmission capacity with or without interference cancelation in terms of scaling with respect to the number of antennas. Thus, the restricted receiver design has no effect on the transmission capacity performance, though the optimal transmit strategy used at each transmitter differs significantly with respect to the interference cancelation case. The advantage of CSI shows up in simplified encoding/decoding, since with CSI, only one data stream needs to be transmitted and decoded to achieve linear scaling of the transmission capacity with multiple antennas, in comparison to a constant fraction of N data streams that are transmitted and decoded without CSI.

Theorem 3.3.12 *The transmission capacity with multiple antennas when receiver uses ZF decoder and does not employ any interference cancelation, scales as $C = \Theta(N)$, and the optimal number of data streams to transmit, k, scales linearly with N.*

Proof: When no interferers are canceled, the SRDOF used for interference cancelation is $m = 0$ or $N_{\mathsf{canc}} = 0$. Then from Theorem 3.3.10, for any r such that $kr > 1$, the transmission capacity is upper bounded by

$$C \leq \frac{kB(1-\epsilon)^{1-\frac{2}{\alpha}}}{\pi}\left(\frac{N-k+1}{(kr-1)d^{\alpha}\beta}\right)^{\frac{2}{\alpha}}\left(r+\frac{\alpha}{2}\right). \tag{3.19}$$

Since we have the freedom to choose r in the upper bound (Theorem 3.3.10), let r be a constant independent of N. Then for number of transmitted data streams $k = N^{\kappa}$, from (3.19), we have

$$C = \mathcal{O}\left(N^{\kappa}N^{\frac{2}{\alpha}(1-\kappa)}\right) = \mathcal{O}\left(N^{\frac{2}{\alpha}+\kappa\left(1-\frac{2}{\alpha}\right)}\right) = \mathcal{O}(N),$$

for the optimal value of $\kappa = 1$.

For the lower bound, similar to (3.18), for $k < N$ with no interference cancelation $m = 0$,

$$C \geq \frac{kB(1-\epsilon)}{\pi}\left(\frac{\epsilon(N-k)}{k\beta d^{\alpha}}\right)^{\frac{2}{\alpha}}. \tag{3.20}$$

Letting $k = \Theta(N)$, immediately from (3.20), we get $C = \Omega(N)$, finishing the proof. □

Remark 3.3.13 *If we fix single data stream transmission $k = 1$, then from* (3.19) *and* (3.20)*, we get that with no interference cancelation, the transmission capacity scales only sublinearly as $\Theta\left(N^{\frac{2}{\alpha}}\right)$. This result was originally found in [8] using a direct outage probability computation. Thus to obtain linear scaling, we have to scale the number of transmitted data streams with N.*

Remark 3.3.14 *Theorem 3.3.12 has been independently derived in [9] by explicitly computing the outage probability, rather than finding tight lower and upper bounds.*

Comparing Theorem 3.3.12 with Theorem 3.3.5, interestingly, we conclude that with or without interference cancelation, the transmission capacity scales linearly with N, and only the transmit–receive strategy changes. When no interference cancelation is employed, the number of data streams sent from each transmitter should scale with the number of antennas, as opposed to the case of interference cancelation where only single data stream should be transmitted. With no interference cancelation, if the number of data streams is not scaled with N, the transmission capacity scales only sublinearly with N (Remark 3.3.13). Thus, with single data stream transmission $k = 1$, the maximal ratio combining gain available at the receiver for decoding the only data stream scales as $N^{2/\alpha}$, and to achieve a linear growth of transmission capacity, we need to linearly scale the number of transmitted data streams with N.

From (3.20), one can easily show that using a $\left(1 - \frac{2}{\alpha}\right)$ fraction of the total transmit antennas N maximizes the lower bound on the transmission capacity with no interference cancelation. Thus, for small path-loss exponents α, that is, when the interference is dominating, only a small number of data streams should be transmitted, while for large path-loss exponents that corresponds to the weak interference regime, almost all transmit antennas should be used to maximize the transmission capacity.

Remark 3.3.15 *For a cellular communication network, when each transmitter sends k independent data streams with equal power allocation, and no interference cancelation is employed at the receiver, using a single transmit antenna is shown to maximize the ergodic Shannon capacity in the presence of small number of strong co-channel interferers in [10–13]. Thus, the results obtained in this section with no interference cancelation for PPP-distributed transmitter locations in a wireless network match with results on cellular networks only for small path-loss exponents α.*

After discussing the case of having no CSI at any of the transmitters in this section, we move on to the more general (but practically challenging) scenario in the next section, when each transmitter is also assumed to have CSI for its corresponding receiver and find the impact of CSI availability at each transmitter on the transmission capacity.

3.4 Channel State Information at Both Transmitter and Receiver

In this section, we consider the case when in addition to each receiver having the CSI for all the channels (Section 3.3), the transmitter also has CSI for the channel between itself and its intended receiver. We refer to this as the CSIT case. From a transmission capacity perspective, the CSIT case is fundamentally different from the CSIR case, since with CSI, each transmitter can increase the signal power at its intended receiver by steering the beam towards it, and possibly the role of

multiple receiver antennas and consequently the transmission capacity scaling is different from the CSIR case.

Remark 3.4.1 *With global CSIT, each transmitter could also use its multiple antennas for interference suppression by nulling out its signal toward unintended receivers. We consider the interference suppression role of multiple transmit antennas in Section 3.5 together with a cognitive radio network.*

We begin with a brief background on using CSIT for a point-to-point multiple antenna channel without interference. Lets consider a point-to-point multiple antenna channel, where $\mathbf{H} \in \mathbb{C}^{N\times N}$ is the channel matrix between the transmitter–receiver pair with N antennas each. If $\tilde{\mathbf{x}}$ is the transmit signal, then the received signal is given by

$$\mathbf{y} = \mathbf{H}\tilde{\mathbf{x}} + \mathbf{w}, \tag{3.21}$$

where $\mathbf{w}$ is the AWGN vector with independent entries that have zero mean and unit variance. Let $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger}$ be the singular value decomposition of $\mathbf{H}$. To maximize the mutual information, the transmitter sends its signal over the strongest singular values of the channel $\mathbf{H}$ [7]. Let $\mathbf{V}^k$ be the matrix consisting of the first k columns of $\mathbf{V}$ corresponding to the k strongest eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$. Thus, if $\mathbf{x} \in \mathbb{C}^{k\times 1}$ is the input signal, then the transmitter sends $\tilde{\mathbf{x}} = \mathbf{V}^k\mathbf{x}$ through its N antennas, and the received signal is

$$\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger}\mathbf{V}^k\mathbf{x} + \mathbf{w}. \tag{3.22}$$

The case of $k = 1$ is referred to as beamforming, while the case of $k > 1$ is called multimode beamforming. With multimode beamforming, if the receiver multiplies the received signal (3.22) with $\mathbf{U}^{\dagger}$, the equivalent received signal is given by

$$\mathbf{y}(\ell) = \mathbf{D}(\ell,\ell)\mathbf{x}(\ell) + \hat{\mathbf{w}}(\ell), \tag{3.23}$$

for $\ell = 1,\ldots,k$, where noise contributions $\hat{\mathbf{w}}(\ell)$ are Gaussian and independent $\forall\ \ell$, since $\mathbf{U}^{\dagger}$ is unitary. Thus, with CSIT, the received signal decouples into k independent signals, where the signal power of the ℓth channel is equal to the ℓth eigenvalue of $\mathbf{H}\mathbf{H}^{\dagger}$. Thus, the knowledge of $\mathbf{H}$ not only helps in increasing the received SNR but also simplifies the decoding since each element of the input signal $\mathbf{x}$ can be decoded independently.

Now we look at our model with interference. We assume that each transmitter uses multimode beamforming even in the presence of interferers. As before, $\mathbf{H}_{nm} \in \mathbb{C}^{N\times N}$ represents the channel matrix between transmitter T_m and receiver R_n. Transmitter T_n is assumed to only know $\mathbf{H}_{nn}$, the channel between itself and its corresponding receiver.

Consider the typical transmitter–receiver pair (T_0, R_0). Let the singular value decomposition of $\mathbf{H}_{00} \in \mathbb{C}^{N\times N}$ (channel between T_0 and R_0) be $\mathbf{U}_{00}\mathbf{D}_{00}\mathbf{V}_{00}^{\dagger}$. Let $k,\ k \in [1,2,\ldots,N]$ denote the number of independent data streams (STDOF) sent by each transmitter to its receiver. Then with multimode beamforming, transmitter T_0 sends $\mathbf{V}_{00}^k\mathbf{x}_0$, where $\mathbf{V}_{00}^k$ be the matrix consisting of first k columns of $\mathbf{V}_{00}$ corresponding to the k strongest eigenvalues of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$, and $\mathbf{x}_0 \in \mathbb{C}^{k\times 1}$ is the data vector consisting of k independent streams, where each stream is $\mathcal{CN}\left(0, \frac{1}{k}\right)$ distributed. Note that in contrast to the CSIR case, with CSIT, the k data streams are transmitted by all the N transmit antennas via processing through $\mathbf{V}_{00}^k$. For keeping the analysis tractable, we consider equal power allocation among the k-transmitted streams.

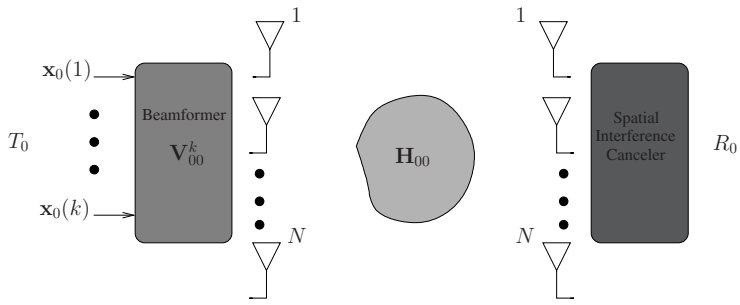

Figure 3.4: Transmit-receive strategy with beamforming at the transmitter.

Similar to the case of point-to-point channel with no interference (3.23), we will show in (3.27) that even in a wireless network, with multimode beamforming, the channel between each transmitter–receiver pair is equivalent to k scalar parallel channels with no inter-stream interference in contrast to the CSIR case. Thus, we assume that each receiver uses k SRDOF to receive the intended signal, while the remaining $N-k$ SRDOF are used for canceling the $c(k) = \left\lfloor \frac{N}{k} \right\rfloor - 1$ nearest interferers.

To cancel the $c(k)$ interferers, the receiver projects the received signal on to the null space of the $c(k)$ interferers. A block diagram depicting the transmit–receive strategy is illustrated in Fig. 3.4.

Using multimode beamforming at each transmitter, the received signal $\mathbf{y}_0 \in \mathbb{C}^{N\times 1}$ at receiver R_0 is

$$\mathbf{y}_0 \;=\; d^{-\alpha/2}\mathbf{H}_{00}\mathbf{V}_{00}^k\,\mathbf{x}_0 + \sum_{T_n:\Phi\setminus\{T_0\}} d_n^{-\alpha/2}\mathbf{H}_{0n}\mathbf{V}_{nn}^k\mathbf{x}_n, \tag{3.24}$$

where $\mathbf{V}_{nn}$ is the matrix of the right singular vectors of the channel between transmitter n and receiver n, $\mathbf{V}_{nn}^k$ are the first k columns of $\mathbf{V}_{nn}$, $\mathbf{H}_{nn} = \mathbf{U}_{nn}\mathbf{D}_{nn}\mathbf{V}_{nn}^\dagger$. Let the indices of the interferers be sorted in an increasing order in terms of their distance from R_0, that is, $d_1 \le d_2 \le \ldots \le d_{c(k)} \le d_{c(k)+1} \le \ldots$. Let S be the basis of the null space of the channel matrices $[\mathbf{H}_{01}\ldots\mathbf{H}_{0c(k)}]$ corresponding to the $c(k)$ nearest interferers to be canceled. Since $N-k$ SRDOF are used for interference cancelation, $\mathsf{S} \in \mathbb{C}^{k\times N}$. Multiplying S to the received signal (3.24),

$$\begin{aligned}\mathsf{S}\mathbf{y}_0 \;&=\; d^{-\alpha/2}\mathsf{S}\mathbf{U}_{00}\mathbf{D}_{00}\mathbf{V}_{00}^\dagger\mathbf{V}_{00}^k\,\mathbf{x}_0 + \sum_{n=c(k)+1}^{\infty} d_n^{-\alpha/2}\mathsf{S}\mathbf{H}_{0n}\mathbf{V}_{nn}^k\mathbf{x}_n,\\ &=\; d^{-\alpha/2}\mathsf{S}\mathbf{U}_{00}^k\mathbf{D}_{00}^k\mathbf{x}_0 + \sum_{n=c(k)+1}^{\infty} d_n^{-\alpha/2}\mathsf{S}\mathbf{H}_{0n}\mathbf{V}_{nn}^k\mathbf{x}_n,\end{aligned} \tag{3.25}$$

where $\mathbf{U}_{00}^k$ is the $N\times k$ matrix consisting of the first k columns of $\mathbf{U}_{00}$, and $\mathbf{D}_{00}^k \in \mathbb{C}^{k\times k}$ is the diagonal matrix consisting of the first k entries of $\mathbf{D}_{00}$. Since S and $\mathbf{U}_{00}^k$ are both of rank k and are independent of each other with each entry drawn from a continuous distribution, $\mathsf{S}\mathbf{U}_{00}^k$ is full rank

with probability 1. Multiplying $\left(\mathsf{S}\mathbf{U}_{00}^{k}\right)^{-1}$ to the received signal (3.25)

$$\hat{\mathbf{y}}_0 \quad = \quad d^{-\alpha/2}\mathbf{D}_{00}^{k}\mathbf{x}_0 + \sum_{n=c(k)+1}^{\infty} d_n^{-\alpha/2}\left(\mathsf{S}\mathbf{U}_{00}^{k}\right)^{-1}\mathsf{S}\mathbf{H}_{0n}\mathbf{V}_{nn}^{k}\mathbf{x}_n. \tag{3.26}$$

Note that $\mathbf{D}_{00}$ is the diagonal matrix of the eigenvalues of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$. Denoting the ℓth eigenvalue of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$ by $\sigma_\ell(\mathbf{H}_{00})$, the received signal (3.26) can be decomposed into k parallel channels as

$$\hat{\mathbf{y}}_0(\ell) = d^{-\alpha/2}\sqrt{\sigma_\ell(\mathbf{H}_{00})}\mathbf{x}_0(\ell) + \sum_{n=c(k)+1}^{\infty} d_n^{-\alpha/2}\sum_{j=1}^{k} g_n(\ell,j)\mathbf{x}_n(j),\ \ell = 1,2,\ldots,k, \tag{3.27}$$

where $\mathbf{x}_n(j)$ is the jth element of the transmitted vector $\mathbf{x}_n$, $g_n(\ell,j)$ is the (ℓ,j)th element of $\left(\mathsf{S}\mathbf{U}_{00}^{k}\right)^{-1}\mathsf{S}\mathbf{H}_{0n}\mathbf{V}_{nn}^{k}$.

Thus, with multimode beamforming, as shown in (3.27), the received signal can be decomposed into k parallel channels, with the ℓth channel corresponding to the data stream $\mathbf{x}_0(\ell)$ having no contribution from data streams $\mathbf{x}_0(1),\ldots,\mathbf{x}_0(\ell-1),\mathbf{x}_0(\ell+1),\ldots,\mathbf{x}_0(k)$. Therefore, with multimode beamforming, there is no inter-stream interference from the other $k-1$ data streams sent by the same transmitter. Thus, using $N-k$ SRDOF for interference cancelation, $c(k) = \left\lfloor \frac{N-k}{k} \right\rfloor$ nearest interferers can be canceled at each receiver.

Since S, $\mathbf{U}_{00}^{k}$, and $\mathbf{V}_{nn}$ are independent of $\mathbf{H}_{0n}$, $g_n(\ell,j)$'s in (3.27) are independent for $j = 1,\ldots,k$, and each $g_n(\ell,j) \sim \chi^2(2)$ from Lemma 3.3.2. Thus, the interference power of the ℓth data stream of the nth interferer

$$\mathsf{pow}_n(\ell) = \mathbb{E}\left\{\left|\sum_{j=1}^{k} g_n(\ell,j)\mathbf{x}_n(j)\right|^2\right\} \sim \chi^2(2k).$$

Let

$$I_n(\ell) = d_n^{-\alpha}\mathsf{pow}_n(\ell)$$

be the interference power of the nth interferer for the ℓth channel in (3.27). Since S and $\mathbf{U}_{00}^{k}$ are independent of $\mathbf{H}_{0n}$, $g_n(\ell,j)$'s and consequently $\mathsf{pow}_n(\ell)$'s are identically distributed for all ℓ, and it follows that $I_n(\ell)$ is identically distributed for all ℓ. Then the total interference power seen at receiver R_0 for the ℓth channel corresponding to signal $\mathbf{x}_0(\ell)$ in (3.27) is $I_{nc}(\ell) = \sum_{n=c(k)+1}^{\infty} I_n(\ell)$. Since $I_n(\ell)$ is identically distributed for all ℓ, it follows that $I_{nc}(\ell)$ is also identically distributed for all $\ell = 1,\ldots,k$ channels.

We assume an uniform data rate of B bits/sec/Hz on each of the k-transmitted data streams.[1] By combining the k streams, the total rate of transmission between a source and destination is kB bits/sec/Hz. To define outage probability, we consider the outage event of the data stream with the worst channel gain, which in this case is the kth data stream, since the eigenvalues of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$ are indexed in the decreasing order. Thus, the outage probability for any channel in (3.27) is at most

$$P_{\text{out}}(B) \quad = \quad \mathbb{P}\left(\log\left(1 + \frac{d^{-\alpha}\sigma_k(\mathbf{H}_{00})}{I_{nc}(k)}\right) \le B\right),$$

[1] In general with multimode beamforming, data rates can be a function of the magnitude of the eigenvalues, however, that requires finding the optimal rate allocation that minimizes the maximum of the outage probability on k different streams, which is an unsolved problem.

$$= \quad \mathbb{P}\left(\frac{d^{-\alpha}\sigma_k(\mathbf{H}_{00})}{I_{nc}(k)} \leq 2^B - 1\right), \tag{3.28}$$

where $\sigma_k(\mathbf{H}_{00})$ is the kth eigenvalue of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$. Since $I_{nc}(\ell)$ is identically distributed for each $\ell = 1, 2, \ldots, k$, from here on we drop the index ℓ and represent $I_{nc}(\ell)$ as I_{nc} for each $\ell = 1, 2, \ldots, k$. Thus, with $2^B - 1 = \beta$

$$P_{\text{out}}(B) = \mathbb{P}\left(\frac{d^{-\alpha}\sigma_k(\mathbf{H}_{00})}{I_{nc}} \leq \beta\right), \tag{3.29}$$

where

$$I_{nc} = \sum_{n=c(k)+1}^{\infty} d_n^{-\alpha}\mathsf{pow}_n, d_i \leq d_j,\ i < j,$$

and pow_n are i.i.d. with $\chi^2(2k)$ distribution.

This definition of outage probability (3.29) implies that if $P_{\text{out}}(B) = \epsilon$, then all the k streams can at least support a data rate of B bits/sec/Hz with probability $1 - \epsilon$, and the transmission capacity is defined as

$$C = k\lambda(1 - \epsilon)B \text{ bits/sec/Hz/m}^2,$$

by combining the contribution from all the k-transmitted data streams. Deriving a closed form expression for the outage probability requires the distribution of I_{nc}, and $\sigma_k(\mathbf{H}_{00})$, the kth maximum eigenvalue of the Wishart matrix $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$. Unfortunately, both these distributions are unknown, and hence finding an exact expression for the outage probability is difficult. To facilitate analysis, we use upper and lower bounds on the outage probability derived in Theorems 3.3.8 and 3.3.10, and then find the optimal number of data streams k that maximize the transmission capacity.

Use of Theorems 3.3.8 and 3.3.10 also allows us to circumvent the problem of requiring a simple closed form expression for the probability density function (PDF) of $\sigma_k(\mathbf{H}_{00})$. For our analysis, it will suffice to know the expected value of the maximum eigenvalue of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$, $\sigma_1(\mathbf{H}_{00})$, and the expected value of the reciprocal of $\sigma_1(\mathbf{H}_{00})$. For large N, the maximum eigenvalue of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$, $\sigma_1(\mathbf{H}_{00})$, converges to $4N$ [14], and $\mathbb{E}\{\sigma_1(\mathbf{H}_{00})\} \approx 4N$. With extensive simulation results (see Fig. 3.5) we observe that $\mathbb{E}\{\frac{1}{\sigma_1(\mathbf{H}_{00})}\} \approx \frac{1}{3.5N}$, however, an analytical proof for this result cannot be found readily in literature. Note that the constant $1/3.5$ is immaterial for us, we are only interested in the scaling of the mean of the reciprocal of the maximum eigenvalue of Wishart matrix with N and our simulations show that mean of the reciprocal of the maximum eigenvalue of Wishart matrix does not decrease faster than N^{-1}. We will use both these large N approximations on $\mathbb{E}\{\sigma_1(\mathbf{H}_{00})\}$ and $\mathbb{E}\{\frac{1}{\sigma_1(\mathbf{H}_{00})}\}$ for our analysis.

The main result of this section is as follows that characterizes the scaling of transmission capacity with multiple antennas using multimode beamforming.

Theorem 3.4.2 *With multimode beamforming and ZF decoder, the transmission capacity scales as*

$$C = \Omega(N), \text{ and } C = \mathcal{O}(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}})$$

with the number of antennas N. The optimal lower bound is achieved by $k^{\star} = 1$ and $c(k) = N - 1$, that is, sending only one data stream on the strongest eigenvector, and canceling the maximum number of interferers $N - 1$, is optimal.

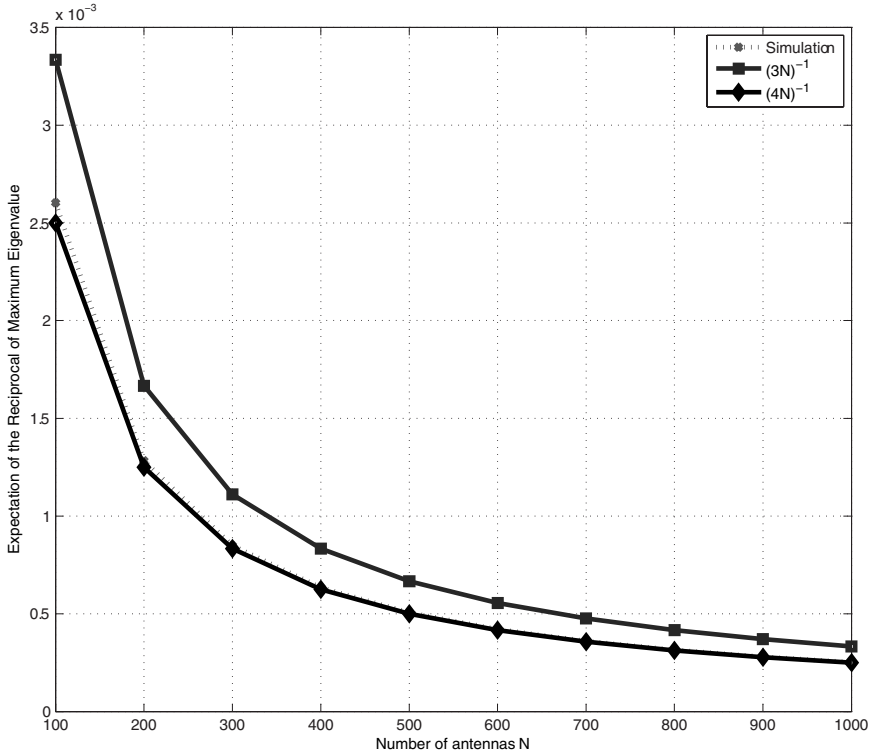

Figure 3.5: Empirical expected value of the reciprocal of the largest eigenvalue of $\mathbf{H}_{00}\mathbf{H}_{00}^{\dagger}$.

Proof: With the number of nearest canceled interferers to be $c(k) = \left\lfloor \frac{N}{k} \right\rfloor - 1$, from Theorem 3.3.10, for any r such that $kr > 1$,

$$C \leq \frac{kB(1-\epsilon)^{1-\frac{2}{\alpha}}}{\pi} \left(\frac{\mathbb{E}\{\sigma_k(\mathbf{H}_{00})\}}{(kr-1)d^{\alpha}\beta} \right)^{\frac{2}{\alpha}} \left(\left\lfloor \frac{N}{k} \right\rfloor - 1 + r + \frac{\alpha}{2} \right), \tag{3.30}$$

where we have replaced the expected signal power $\mathbb{E}\{s\} = N - m - k + 1$ of the CSIR case with $\mathbb{E}\{\sigma_k(\mathbf{H}_{00})\}$, the expected signal power of the kth data stream with multimode beamforming (3.26). Recall that we have ordered the eigenvalues $\sigma_k(\mathbf{H}_{00})$ in decreasing order, $\sigma_k(\mathbf{H}_{00}) \geq \sigma_m$ if $k > m$. From [14], $\mathbb{E}\{\sigma_1(\mathbf{H}_{00})\} = 4N$, hence $\mathbb{E}\{\sigma_k(\mathbf{H}_{00})\} < 4N$ for $k > 1$. Hence, from (3.30), similar to the proof of Theorem 3.3.5, we can show that

$$C = \mathcal{O}\left(N^{1+\frac{2}{\alpha}-\frac{4}{\alpha^2}} \right)$$

with $r = N^{2/\alpha}$ by parametrizing $k = N^{\kappa}$ and finding the best κ.

For the lower bound, by substituting $\mathbb{E}\left\{\frac{1}{s}\right\} = \mathbb{E}\left\{\frac{1}{\sigma_k(\mathbf{H}_{00})}\right\}$ in Theorem 3.3.8 for $k < N$, we have

$$C \geq \frac{kB(1-\epsilon)}{\pi} \left(\frac{\epsilon}{k\mathbb{E}\left\{\frac{1}{\sigma_k(\mathbf{H}_{00})}\right\}\beta d^{\alpha}} \right)^{\frac{2}{\alpha}} \left(\left\lfloor \frac{N}{k} \right\rfloor - 1 - \left\lceil \frac{\alpha}{2} \right\rceil \right)^{1-\frac{2}{\alpha}}. \tag{3.31}$$

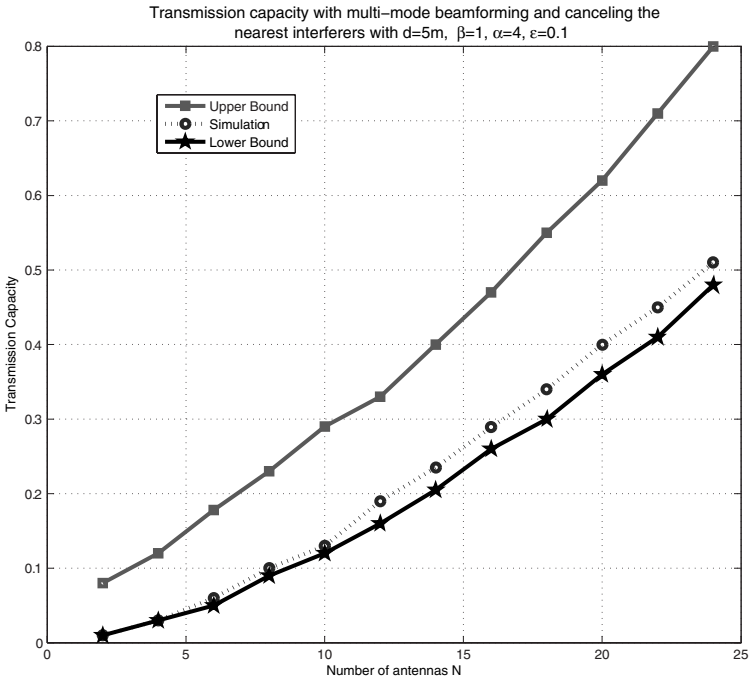

Figure 3.6: Transmission capacity versus the number of antennas N with multimode beamforming and canceling the nearest interferers with single stream data transmission $k = 1$, $d = 5$ m, $\beta = 1$ ($B = 1$ bits/sec/Hz), $\alpha = 4$, $\epsilon = 0.1$.

As pointed out earlier, $\mathbb{E}\{\frac{1}{\sigma_1(\mathbf{H}_{00})}\} = \frac{1}{3.5N}$, $\mathbb{E}\{\frac{1}{\sigma_k(\mathbf{H}_{00})}\} > \frac{1}{3.5N}$ for $k > 1$. Thus, evaluating the lower bound (3.31) at $k = 1$, we get $C = \Omega(N)$. □

With multimode beamforming, the lower bound on the transmission capacity is maximized by using single stream beamforming ($k = 1$) together with canceling the $N - 1$ nearest interferers, and the lower bound scales linearly with N. Thus, comparing the CSIT and CSIR cases, the transmission strategy remains identical, but the reception strategy is completely different (with the CSIR case $m = \Theta(N)$ nearest interferers are canceled). This difference is because in the CSIT case, the average signal power (strongest eigenvalue) scales linearly with N without any processing at the receiver, while in the CSIR case, it is independent of N if signals received at multiple receive antennas are not combined at the receiver. Thus, in the CSIR case, to boost the signal power, so that it scales with N, $\Theta(N)$ SRDOF are required for decoding the signal of interest allowing only $m = \Theta(N)$ nearest interferers to be canceled.

The derived bounds on the transmission capacity in both the CSIR and CSIT cases are identical, implying that the value of channel feedback (which is generally costly) is fairly limited. There is, however, a constant multiplicative gain of 4 in terms of signal power with CSIT, since with the optimal mode of $k = 1$, the signal power $\mathbb{E}\{\sigma_1(\mathbf{H}_{00})\} = 4N$ in comparison to order N for the CSIR case. The real advantage of CSIT is the simplified encoding and decoding, since with CSIT, the multiple data streams sent by the transmitter can be resolved as parallel channels at the receiver resulting in independent decoding.

To illustrate the scaling behavior of the transmission capacity with CSIT as a function of the number of antennas N, we plot the derived lower and upper bound, and the simulated transmission capacity in Fig. 3.6, for $k = 1$, $d = 5$ m, path-loss exponent $\alpha = 4$ and outage probability constraint

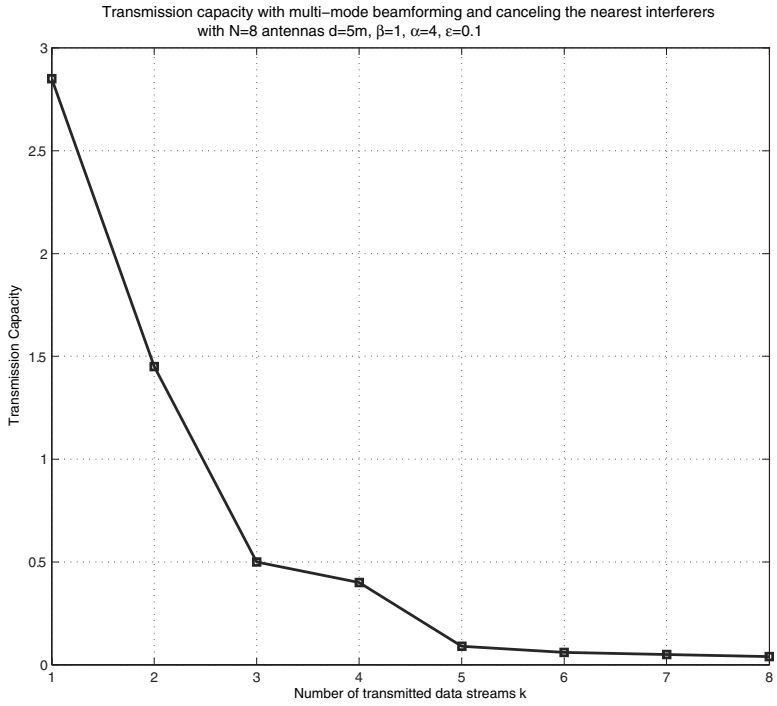

Figure 3.7: Transmission capacity versus the number of transmitted data streams k with multimode beamforming and canceling the nearest interferers with $d = 5$ m, $\beta = 1$, $\alpha = 4$, $\epsilon = 0.1$, total number of antennas $N = 8$.

of $\epsilon = 0.1$ with increasing N. We see that the transmission capacity grows linearly even for $\alpha = 4$, for which the upper bound suggests super-linear scaling of $N^{1+1/4}$. Thus, the derived lower bound that scales linearly N is tight; however, some more analytical work is required to tighten the upper bound to make it scale linearly with N. To show the optimality of using a single data stream from each transmitter k, in Fig. 3.7, we plot the transmission capacity as a function of k for total $N = 8$ antennas. Fig. 3.7 clearly shows that the transmission capacity with multimode beamforming is a decreasing function of k and sending a single data stream is optimal in a wireless network. Similar to the CSIR case, for the CSIT case, we can get the exact result for the special case when each receiver employs no interference cancelation. We show that with no interference cancelation, the transmission capacity scales as $\Theta(N)$ and the optimal number of data streams to transmit is $k = \Theta(N)$. Further, if the number of receive antennas is 1, then we show that the transmission capacity is $\Theta(N^{\frac{2}{\alpha}})$, that is, scales sublinearly with N.

Theorem 3.4.3 *Without interference cancelation at any receiver, with multimode beamforming, the transmission capacity is $C = \Theta(N)$, and the optimal number of data streams to transmit is $k = \theta N, \theta \in (0, 1]$. If the number of receive antennas is 1, then the transmission capacity is $C = \Theta(N^{\frac{2}{\alpha}})$, where N is the number of transmit antennas.*

Proof: Follows similarly to the proof of Theorem 3.4.2 □

Remark 3.4.4 *In this section, even though we have assumed the availability of CSI at each transmitter, we have not accounted for resources required for feeding back CSI from each receiver. In general, it is a hard problem to quantify the effects of feedback. In Chapter 4, we present some results in that direction.*

3.5 Spectrum-Sharing/Cognitive Radios

After considering the dual role of multiple antennas in previous sections, sending multiple data streams from the transmitter and canceling interference at the receiver, in this section, we look at the third possible role of multiple antennas in a wireless network: using them at the transmitter for suppressing interference toward other receivers. To highlight this feature, we consider a cognitive/secondary wireless network that is overlaid over a pre-existing/primary wireless network that consists of licensed/primary nodes.

In particular, we consider two co-existing networks, one primary and other secondary, where primary network is oblivious to the presence of the secondary network, while the secondary network is aware of the primary network. For the primary network, we assume the same model as in Section 3.3, where each primary transmitter has a primary receiver associated with it at a fixed of fixed distance d_p, with SIR threshold β_p, and under an outage probability constraint of ϵ_p at each receiver, except that each transmitter and receiver has a single antenna. Thus, from Theorem 2.4.1, the maximum density of primary network is

$$\lambda_p^\star = \frac{\ln(1-\epsilon_p)}{c\beta_p^{\frac{2}{\alpha}} d_p^2},$$

for a constant c.

The secondary network is overlaid on top of the primary network, where each secondary transmitter has a secondary receiver associated with it at a fixed of fixed distance d_s, with density λ_s and SIR threshold β_s, under an outage probability constraint of ϵ_s at each secondary receiver. Clearly, the presence of secondary transmitters increases the interference seen at any primary receiver; thus, if $\lambda_s \neq 0$, the primary outage probability constraint of ϵ_p cannot be met if the primary network is operating with density $\lambda_p^\star$. Therefore, $\lambda_s = 0$ if primary network density is $\lambda_p^\star$ and primary outage probability constraint is ϵ_p.

To make the problem non-degenerate, we relax the primary outage probability constraint of ϵ_p to $\epsilon_p + \Delta_p$ while keeping the primary network density to be $\lambda_p^\star$, and find the maximum value of λ_s such that the relaxed primary outage probability constraint of $\epsilon_p + \Delta_p$, and the secondary outage probability constraint of ϵ_s is satisfied simultaneously.

We assume that the secondary nodes are equipped with multiple transmit and receive antennas. Multiple antennas at each secondary transmitter node are used for interference suppression toward primary receivers, while multiple receive antennas are used for interference cancelation at each secondary receiver. Since the secondary network has to operate under an outage probability constraint at each primary node, it is important to control the interference that each secondary node creates, and this is where the interference suppression feature of multiple transmit antennas comes to the fore.

We let the locations of primary and secondary transmitters to be distributed as two independent homogenous PPPs with density λ_1 and λ_2, respectively. We consider an ALOHA random access protocol for both the primary and secondary transmitters, with access probability p. Consequently,

the active primary and secondary transmitter processes are also homogenous PPPs on a two-dimensional plane with density $\lambda_p = p\lambda_1$ and $\lambda_s = p\lambda_2$, respectively.

Let the location of the nth active primary transmitter be T_{pn} and the nth active secondary transmitter be T_{sn}. The set of all active primary and secondary transmitters is denoted by $\Phi_p = \{T_{pn},\ n \in \mathbb{N}\}$ and $\Phi_s = \{T_{sn},\ n \in \mathbb{N}\}$, respectively.

We assume that each secondary transmitter has N_t antennas, while each secondary receiver has N_r antennas. We also assume that each secondary transmitter has CSI for its corresponding receiver as well as for its N_t nearest primary receivers that is used to suppress interference towards them. Each secondary receiver is assumed to have CSI for its intended transmitter as well as for its N_r nearest interferers (from the union $\Phi_s \cup \Phi_p$). The system model of overlaid wireless networks under consideration is illustrated in Fig. 3.8, where the squares represent the primary transmitters and receivers, while the dots represent the secondary transmitters and receivers, and a dashed line indicates a suppressed interferer. We restrict ourselves to the case when each secondary transmitter sends only one data stream using its multiple antennas to its intended secondary receiver.

Let the beamformer used by the nth secondary transmitter for interference suppression toward primary receivers is denoted by $\mathbf{b}_n \in \mathbb{C}^{N\times 1}$. Then, the received signal at the primary receiver R_{p0} is

$$\begin{aligned} y_0 &= \sqrt{P_p} d_p^{-\alpha/2} h_{00} x_{p0} + \sum_{n:T_{pn}\in\Phi_p\setminus\{T_{p0}\}} \sqrt{P_p} d_{pp,n}^{-\alpha/2} h_{0n} x_{pn} \\ &\quad + \sum_{n:T_{sn}\in\Phi_s} \sqrt{\frac{P_s}{N}} d_{sp,n}^{-\alpha/2} \mathbf{g}_{0n} \mathbf{b}_n x_{sn}, \end{aligned} \tag{3.32}$$

where P_p and P_s is the transmit power of each primary and secondary transmitter, respectively, $h_{0n} \in \mathbb{C}$ is the channel between the nth primary transmitter T_{pn} and a primary receiver R_{p0}, $\mathbf{g}_{0n} \in \mathbb{C}^{1\times N}$ is the channel vector between the nth secondary transmitter T_{sn} with N_t antennas and R_{p0}, $d_{pp,n}$ and $d_{sp,n}$ is the distance between T_{pn} and R_{p0}, and T_{sn} and R_{p0}, respectively, x_{pn} and x_{sn} are data signals transmitted from T_{pn} and T_{sn}, respectively, with $x_{pn}, x_{sn} \sim\ \mathcal{C}N(0,1)$.

The second term of (3.32) corresponds to the interference received from primary transmitters at the primary receiver R_{p0}, while the third term corresponds to the interference received from secondary transmitters at the primary receiver R_{p0}.

We consider the interference limited regime, that is, noise power is negligible compared to the interference power and drop the AWGN contribution. We assume that each h_{0n}, and each entry of $\mathbf{g}_{0n}$ is i.i.d. Rayleigh distributed.

Similarly, the $\mathbb{C}^{N_r\times 1}$ received signal $\mathbf{v}_0$ at the secondary receiver R_{s0} is

$$\begin{aligned} \mathbf{v}_0 &= \sqrt{\frac{P_s}{N}} d_s^{-\alpha/2} \mathbf{Q}_{00} \mathbf{b}_0 x_{s0} + \sum_{n:T_{sn}\in\Phi_s\setminus\{T_{s0}\}} \sqrt{\frac{P_s}{N}} d_{ss,n}^{-\alpha/2} \mathbf{Q}_{0n} \mathbf{b}_n x_{sn} \\ &\quad + \sum_{n:T_{pn}\in\Phi_p} \sqrt{P_p} d_{ps,n}^{-\alpha/2} \mathbf{f}_{0n} x_{pn}, \end{aligned} \tag{3.33}$$

where $d_{ss,n}$ and $d_{ps,n}$ is the distance between T_{sn} and R_{s0}, and T_{pn} and R_{s0}, respectively, $\mathbf{Q}_{0n} \in \mathbb{C}^{N_r\times N}$ is the multiple antenna channel between the secondary transmitter T_{sn} and the secondary receiver R_{s0}, $\mathbf{f}_{0n} \in \mathbb{C}^{N_r\times 1}$ is the channel vector between T_{pn} and R_{s0}. Each of the channel coefficients are assumed to be Rayleigh distributed.

Secondary Transmitter Interference Suppression: To minimize the interference caused at primary receivers, the N_t transmit antennas at each secondary transmitter are used to suppress

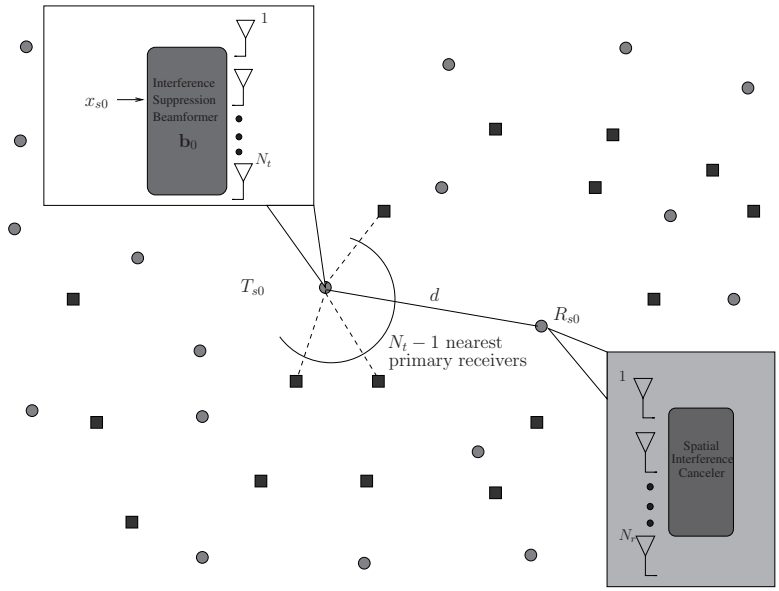

Figure 3.8: Transmit-receive strategy of secondary transmitters and receivers (dots) and primary transmitters and receivers (squares), where each secondary transmitter suppresses its interference toward its $N_t - 1$ nearest primary receivers.

interference towards its $N_t - 1$ nearest primary receivers. Thus, the beamformer (suppressing vector) employed by the nth secondary transmitter $\mathbf{b}_n$ lies in the null space of the channel vectors of the $N_t - 1$ nearest primary receivers, that is, $[\mathbf{g}_{1n}^\dagger \dots \mathbf{g}_{N_t-1n}^\dagger]$, to suppress the interference toward its $N_t - 1$ nearest primary receivers.

Remark 3.5.1 *Note that each secondary transmitter nulls/suppresses its signal toward its $N_t - 1$ nearest primary receivers. However, from a primary receiver's perspective this does not translate into not receiving any interference from its $N_t - 1$ nearest secondary transmitters.*

Let N_{supp} be the random variable denoting the number of consecutive nearest secondary interferers that appear suppressed at the typical primary receiver R_{p0}. For example, as shown in Fig. 3.9, each secondary transmitter tries to suppress interference toward its 3 nearest primary receivers. A dashed line indicates suppressed interferer while a solid line indicates non-suppressed interferer. In Fig. 3.9, we can see that the primary receiver R_{p0} can still receive interference from its second nearest secondary transmitter T_{s1}, in which case $N_{\mathsf{supp}} = 1$.

With $N_{\mathsf{supp}} = \mathsf{c}$ nearest secondary interferers suppressed at primary receiver R_{p0}, the interference received from both the primary and secondary transmitters at the primary receiver R_{p0} in (3.32) is

$$I_{\mathrm{MIMO}}(\mathsf{c}) = \underbrace{\sum_{n:T_{pn}\in\Phi_p\setminus\{T_{p0}\}} P_p d_{pp,n}^{-\alpha}|h_{0n}|^2}_{I_{pp}} + \underbrace{\sum_{n:n>\mathsf{c},\ T_{sn}\in\Phi_s} P_s d_{sp,n}^{-\alpha}|\mathbf{g}_{0n}\mathbf{b}_n|^2}_{I_{sp}^{\mathsf{c}}}. \tag{3.34}$$

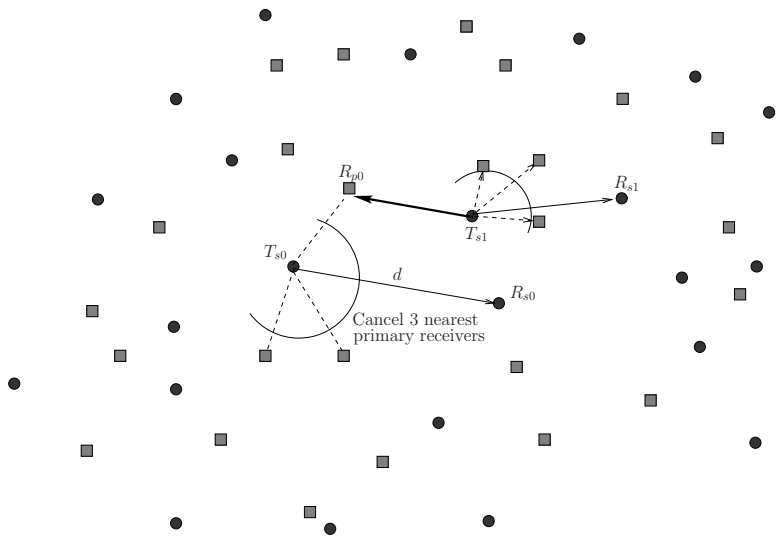

Figure 3.9: Each dot (secondary transmitter) suppresses its interference toward its 3 nearest squares (primary receivers) denoted by dashed lines, but still a square can receive interference from one of its 3 nearest dots.

Thus, with signal model (3.32), the SIR at R_{p0} is

$$\mathsf{SIR}_p = \frac{P_p d_p^{-\alpha} |h_{00}|^2}{I_{\mathrm{MIMO}}(\mathsf{c})}. \tag{3.35}$$

Secondary Receiver Interference Cancelation: Similar to Section 3.3, we consider the use of a partial ZF decoder at each secondary receiver that uses its m SRDOF for canceling the nearest interferers from the union of the primary and the secondary interferers, and the remaining $N - m$ SRDOF are used for decoding the signal of interest. Since each primary and secondary transmitter sends a single data stream, the number of interferers that can be canceled at each secondary receiver using m SRDOF is m.

For interference cancelation, let the nth secondary receiver multiply $\mathbf{t}_n^\dagger$ to the received signal (3.33). Then $\mathbf{t}_n^\dagger$ lies in the null space of channel vectors corresponding to its m nearest interferers from $\{\Phi_p \cup \Phi_s\} \backslash \{T_{sn}\}$ chosen such that it maximizes the signal power $|\mathbf{t}_n^\dagger \mathbf{Q}_{nn} \mathbf{b}_n|^2$ in (3.33).

Thus, from (3.33), the SIR at the secondary receiver R_{s0} is

$$\mathsf{SIR}_s = \frac{P_s d_s^{-\alpha} |\mathbf{t}_0^\dagger \mathbf{Q}_{00} \mathbf{b}_0|^2}{\sum_{n:T_{sn} \in \Phi_s \backslash \{T_{s0}\}} P_s d_{ss,n}^{-\alpha} |\mathbf{t}_n^\dagger \mathbf{Q}_{0n} \mathbf{b}_n|^2 + \sum_{n:T_{pn} \in \Phi_p} P_p d_{ps,n}^{-\alpha} |\mathbf{t}_0^\dagger \mathbf{f}_{0n}|^2}, \tag{3.36}$$

where the beamforming vector $\mathbf{b}_n$ used by secondary transmitter T_{sn} lies in the null space of $[\mathbf{g}_{1n}^\dagger \dots \mathbf{g}_{N-1n}^\dagger]$ to suppress the interference from T_{sn} toward its $N_t - 1$ nearest primary receivers, and $\mathbf{t}_n$ lies in the null space of channel vectors corresponding to the m nearest interferers of R_{s0} from $\{\Phi_p \cup \Phi_s\} \backslash \{T_{sn}\}$ chosen such that it maximizes the signal power $|\mathbf{t}^\dagger \mathbf{Q}_{nn} \mathbf{b}_n|^2$. From Lemma 3.5.2, optimal

$$\mathbf{t}_n = \frac{(\mathbf{Q}_{nn}\mathbf{u})^\dagger \mathsf{S}\mathsf{S}^\dagger}{|(\mathbf{Q}_{nn}\mathbf{u})^\dagger \mathsf{S}\mathsf{S}^\dagger|},$$

where $\mathsf{S} \in \mathbb{C}^{N_r \times N_r - m}$ is the orthonormal basis of the null space of channel vectors corresponding to the m nearest interferers of R_{s0} from $\Phi_p \cup \Phi_s \backslash \{T_{sn}\}$.

Lemma 3.5.2 *The signal power* $s = |\mathbf{t}_0^\dagger \mathbf{Q}_{00} \mathbf{b}_0|^2$ *in* (3.36) *at the secondary receiver with* $\mathbf{t}_n = \frac{(\mathbf{Q}_{nn}\mathbf{u})^\dagger \mathsf{S}\mathsf{S}^\dagger}{|(\mathbf{Q}_{nn}\mathbf{u})^\dagger \mathsf{S}\mathsf{S}^\dagger|}$ *is* $\sim \chi^2(2(N_r - m))$. *The interference power at secondary receiver from the secondary transmitter* n *in* (3.36), $\mathsf{pow}_{ss}^{0n} = |\mathbf{t}_0^\dagger \mathbf{q}_{0n} \mathbf{b}_n|^2$, *and the interference power at secondary receiver from the primary transmitter* n *in* (3.36), $\mathsf{pow}_{ps}^{0n} = |\mathbf{t}_0^\dagger \mathbf{f}_{0n}|^2$ *is* $\sim \chi^2(2)$.

Proof: The first statement follows from Lemma 3.3.1. The second and third statements of the Lemma follow since $\mathbf{t}_0^\dagger$, $\mathbf{b}_n$, and $\mathbf{q}_{0n}$ are independent, and since each entry of $\mathbf{q}_{0n}, \mathbf{f}_{0n} \sim \mathcal{CN}(0,1)$. □

We next present an alternate way of representing the interference term in (3.36) which allows both easy analysis and simpler notation.

Lemma 3.5.3 *The interference term in* (3.36)

$$I_s = \sum_{n: T_{sn} \in \Phi_s \backslash \{T_{s0}\}} P_s d_{ss,n}^{-\alpha} \mathsf{pow}_{ss}^{0n} + \sum_{n: T_{pn} \in \Phi_p} P_p d_{ps,n}^{-\alpha} \mathsf{pow}_{ps}^{0n}$$

received at the typical secondary receiver R_{s0} *can also be written as*

$$\sum_{n: T \in \Phi \backslash \{T_{s0}\}} P_n d_n^{-\alpha} \mathsf{pow}^{0n},$$

where pow^{0n} *is* $\sim \chi^2(2)$, $\Phi = \{\Phi_s \cup \Phi_p\}$, *and* P_n *is a binary random variable which takes value* P_p *with probability* $\frac{\lambda_p}{\lambda_p + \lambda_s}$, *and value* P_s *with probability* $\frac{\lambda_s}{\lambda_p + \lambda_s}$.

Essentially, the Lemma says that the aggregate interference seen at a secondary receiver can be thought of as interference coming for a single PPP Φ that is an union of the primary and the secondary transmitter's PPP, and where each node of PPP Φ transmits with either power P_s or P_p with probability $\frac{\lambda_s}{\lambda_p + \lambda_s}$ and $\frac{\lambda_p}{\lambda_p + \lambda_s}$, respectively.

Proof: Since the superposition of two independent PPPs is a PPP, consider the union of Φ_p and Φ_s that are independent as a single PPP $\Phi = \{\Phi_s \cup \Phi_p\}$. Thus, the interference received at the typical secondary receiver R_{s0} is derived from the transmitters corresponding to Φ with channel gains pow_{ss}^{0n} or pow_{ps}^{0n}, where both pow_{ss}^{0n} and pow_{ps}^{0n} are $\sim \chi^2(2)$ and is denoted as pow. Note that the primary transmitters use power P_p, and the secondary transmitters use power P_s. The probability that any randomly chosen node of Φ belongs to Φ_p is $\frac{\lambda_p}{\lambda_p + \lambda_s}$; hence, the power transmitted by any node of Φ is P_p with probability $\frac{\lambda_p}{\lambda_p + \lambda_s}$, and P_s with probability $\frac{\lambda_s}{\lambda_p + \lambda_s}$. □

Thus, we can write the SIR expression (3.36) at the typical secondary receiver R_{s0} after canceling the m nearest interferers from $\Phi = \Phi_p \cup \Phi_s$ at secondary receiver R_{s0}, more compactly as

$$\mathsf{SIR}_s = \frac{P_s d_s^{-\alpha} |\mathbf{t}_0^\dagger \mathbf{Q}_{00} \mathbf{b}_0|^2}{\sum_{n > m,\ T_n \in \Phi \backslash \{T_{s0}\}} P_n d_n^{-\alpha} \mathsf{pow}^{0n}}. \tag{3.37}$$

We assume that the rate of transmission for each primary (secondary) transmitter is $B_p = \log(1+\beta_p)$ ($B_s = \log(1+\beta_s)$) bits/sec/Hz. Therefore, a packet transmitted by T_{p0} (T_{s0}) can be successfully

decoded at R_{p0} (R_{s0}), if $\mathsf{SIR}_p \geq \beta_p$ ($\mathsf{SIR}_s \geq \beta_s$). Without the presence of secondary network, the SIR at the primary receiver R_{p0} is

$$\mathsf{SIR}_p^{nc} = \frac{P_p d_p^{-\alpha} |h_{00}|^2}{\sum_{n:T\in\Phi_p\setminus\{T_{s0}\}} P_n d_n^{-\alpha} \mathsf{pow}^{0n}}. \tag{3.38}$$

Primary Network Outage Model: For a given rate B_p bits/sec/Hz and primary outage probability constraint ϵ_p, let $\lambda_p^\star$ be the maximum density for which the outage probability of the primary network

$$P_{p,\text{out}}^{nc} = \mathbb{P}\left(\mathsf{SIR}_p^{nc} \leq \beta_p\right) \leq \epsilon_p. \tag{3.39}$$

From Theorem 2.4.1, $\lambda_p^\star = \frac{\ln(1-\epsilon_p)}{c\beta_p^{\frac{2}{\alpha}} d_p^2}$. We assume that the primary network operates at the largest permissible density $\lambda_p^\star$.

Allowing secondary transmissions to co-exist with the primary transmissions, increases the interference received at R_{p0} as quantified in SIR_p (3.35) compared to SIR_p^{nc} (3.38), and thereby increases the outage probability from $P_{p,\text{out}}(\beta_p)^{nc}$ (3.39) to

$$P_{p,\text{out}}(\beta_p) = \mathbb{P}\left(\mathsf{SIR}_p \leq \beta_p\right). \tag{3.40}$$

Thus, if the primary outage probability constraint is fixed at ϵ_p, and the primary network density is $\lambda_p^\star$, the density of the secondary transmitters cannot be non-zero.

To make the problem non-trivial, we consider an increased outage probability tolerance at the primary receivers of $\epsilon_p + \Delta_p$.

Secondary Network Outage Model: For the secondary network, we consider the usual outage probability constraint of $P_{s,\text{out}}(\beta_s) = \mathbb{P}\left(\mathsf{SIR}_s \leq \beta_s\right) \leq \epsilon_s$. Thus, we want to find the maximum density of secondary transmitters λ_s satisfying both the outage constraint at primary receivers $P_{p,\text{out}}(\beta_p) \leq \epsilon_p + \Delta_p$, and secondary receivers $P_{s,\text{out}}(\beta_s) = \mathbb{P}\left(\mathsf{SIR}_s \leq \beta_s\right) \leq \epsilon_s$ for primary nodes' density $\lambda_p^\star$. Thus, the maximum density of the secondary network is

$$\lambda_s^\star = \max_{P_{p,\text{out}}(\beta_p)\leq\epsilon_p+\Delta_p,\ P_{s,\text{out}}(\beta_s)\leq\epsilon_s} \lambda.$$

Consequently, the transmission capacity of the secondary network is defined as

$$C_s = \lambda_s^\star(1-\epsilon_s)B_s \text{ bits/sec/Hz/m}^2.$$

In the following, we derive $\lambda_s^\star$ as a function of secondary transmit (N_t) and receive (N_r) antennas via computing the outage probabilities. To compute the outage probability $P_{p,\text{out}}(\beta_p)$ and $P_{s,\text{out}}(\beta_s)$, we once again consider a typical transmitter–receiver pair (T_{p0}, R_{p0}) and (T_{s0}, R_{s0}), respectively.

We next state the main theorem of this section, on the scaling of transmission capacity of secondary nodes with multiple antennas under a primary and secondary outage probability constraint.

Theorem 3.5.4 *When each secondary transmitter uses $N_t - 1$ STDOF for suppressing interference toward its $N_t - 1$ nearest primary receivers, and each secondary receiver uses m SRDOF for canceling the m nearest interferers from $\{\Phi_s \cup \Phi_p\}\backslash\{T_{s0}\}$, then*

$$C_s = \Omega\left(\min\{N_r, {N_t}^{1-\frac{2}{\alpha}}\}\right), \text{ and } C_s = \mathcal{O}\left(\min\{N_t, N_r\}\right),$$

and $m = \theta N_r,\ \theta \in (0,1]$ maximizes the lower bound on the transmission capacity of the secondary wireless network.

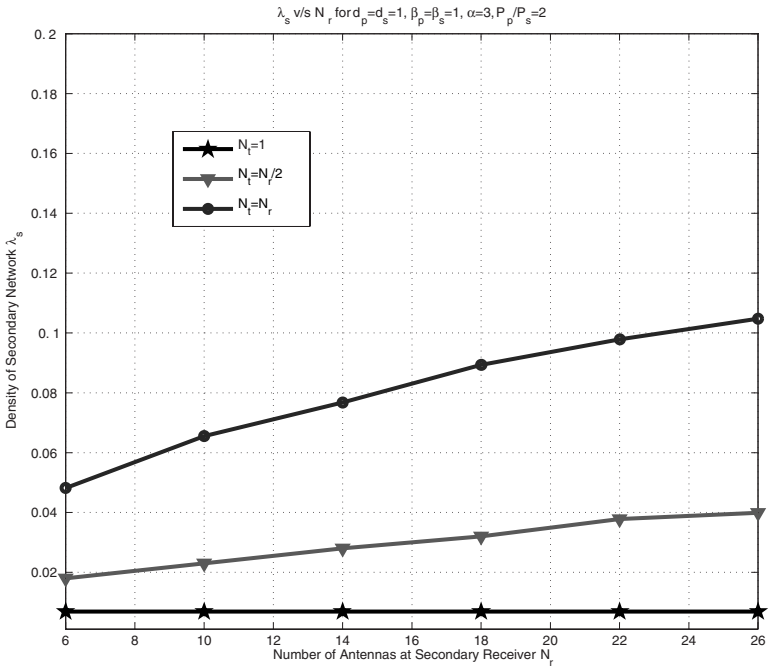

Figure 3.10: Density of the secondary network with respect to number of transmit and receive antennas N_t, N_r at the secondary nodes.

Theorem 3.5.4 highlights the dependence of transmission capacity of the secondary network on the number of transmit and receive antennas, when multiple antennas are allowed to exploit their full capability; perform interference suppression at the transmit side and interference cancelation at the receive side. It also identifies that increasing only the transmit or receiver antennas is futile and to get non-vanishing gain, both the transmit and receive antennas have to be increased simultaneously, which is expected since there are two outage probability constraints.

Theorem 3.5.4 shows that if the number of transmit antennas is much larger than the number of receive antennas $N_t >> N_r$, then the transmission capacity increases linearly with N_r, the number of receive antennas. With large number of transmit antennas at the secondary nodes $N_t >> N_r$, each secondary transmitter can suppress its interference toward a very large number of primary receivers and hence the outage probability constraint at each primary receiver is always met. Thus, with $N_t >> N_r$, only the outage probability constraint at the secondary nodes is active, and the situation is identical to that of Theorem 3.3.5, where each transmitter has a single antenna and each receiver has N_r antennas with a single outage probability constraint, and hence the result is identical to that of Theorem 3.3.5.

When $N_r >> N_t$, the transmission capacity is limited by the interference suppression capability of secondary transmitters, and Theorem 3.5.4 shows that the transmission capacity scales at least as ${N_t}^{1-2/\alpha}$. This result is intuitive since larger the path-loss exponent α, less is the interference caused by each secondary transmitter at any primary receiver.

In Fig. 3.10, we plot the density of the secondary network with respect to the number of secondary transmit and receive antennas N_t and N_r for outage probabilities $\epsilon_p = \epsilon_s = .1$. We see that for $N_t = N_r$, the density of the secondary network scales sublinearly with N_t, however, for

$N_t = 1$ the density of the secondary network is constant as expected.

Proof: Similar to the proof of Theorem 3.3.5, we will first find an upper and a lower bound on the outage probability, but in this case, we have two outage probabilities to bound, one at the primary receiver and the other at the secondary receiver. The outage probability bounds for the secondary receiver follow from Theorems 3.3.8 and 3.3.10, since the secondary receiver employs partial ZF decoder for interference cancelation, similar to Section 3.3. Thus, we only need to derive the bounds for the outage probability expression at the primary receiver in Theorem 3.5.5, when each secondary transmitter uses its $N_t - 1$ STDOF for interference suppression.

Considering the relaxed outage probability constraint of $\epsilon_p + \Delta_p$ at any primary receiver when each secondary transmitter uses $k = N_t - 1$ STDOF for interference suppression, from Theorem 3.5.5,

$$\lambda_s^\star = \Omega\left({N_t}^{1-\frac{2}{\alpha}}\right) \text{ and } \lambda_s^\star = \mathcal{O}\left(N_t\right). \tag{3.41}$$

Next, we consider the outage probability constraint of ϵ_s on each secondary receiver. From (3.37), the outage probability at the secondary receiver R_{s0} is

$$P_{s,\text{out}}(\beta_s) \quad = \quad \mathbb{P}\left(\frac{P_s d_s^{-\alpha}|\mathbf{t}_0^\dagger \mathbf{Q}_{00}\mathbf{b}_0|^2}{\sum_{n>m,\ T_n\in\Phi\setminus\{T_{s0}\}} P_n d_n^{-\alpha}\mathsf{pow}^{0n}} \leq \beta_s\right),$$

where $\Phi = \{\Phi_s \cup \Phi_p\}$, and interference power $\mathsf{pow}^{0n} \sim \chi^2(2)$ (Lemma 3.5.2) and signal power $|\mathbf{t}_0^\dagger \mathbf{Q}_{00}\mathbf{b}_0|^2 \sim \chi^2(2(N_r - m))$ (Lemma 3.5.2). Thus, with m SRDOF used for interference cancelation at each secondary receiver, from Theorems 3.3.8 and 3.3.10, for $\Phi = \Phi_p \cup \Phi_s$ with density $\lambda_p + \lambda_s$, we get for any $r > 1$,[2] with a single data stream transmission $k = 1$,

$$P_{s,\text{out}}(\beta_s) \geq 1 - \frac{(N_r - m)(m + r + \frac{\alpha}{2})^{\frac{2}{\alpha}}}{(r-1)\frac{d_p^\alpha \beta_s}{P_s}(\pi(\lambda_s + \lambda_p))^{\frac{\alpha}{2}}}\left(\frac{\lambda_p}{P_p(\lambda_p + \lambda_s)} + \frac{\lambda_s}{P_s(\lambda_p + \lambda_s)}\right), \tag{3.42}$$

and $P_{s,\text{out}}(\beta_s)$

$$\leq \frac{(\pi(\lambda_p + \lambda_s))^{\frac{\alpha}{2}}\beta_p d_p^\alpha \left(\frac{\alpha}{2} - 1\right)^{-1}\left(m - \left\lceil\frac{\alpha}{2}\right\rceil\right)^{1-\frac{\alpha}{2}}}{N_r - m - 1}\left(\frac{\lambda_p P_p}{\lambda_p + \lambda_s} + \frac{\lambda_s P_s}{\lambda_p + \lambda_s}\right), \tag{3.43}$$

respectively, where we have taken the expectation with respect to power transmitted P_n by any node of $\Phi = \Phi_p \cup \Phi_s$, which is a binary random variable taking values P_p and P_s, with probability $\frac{\lambda_p}{\lambda_p + \lambda_s}$ and $\frac{\lambda_s}{\lambda_p + \lambda_s}$, respectively.

From the lower bound on the outage probability (3.42), we get that

$$\lambda_s^\star = \mathcal{O}(N_r), \tag{3.44}$$

by choosing $r = {N_r}^{2/\alpha}$, similar to Theorem 3.3.5, by fixing outage probability $P_{s,\text{out}}(\beta_s) = \epsilon_s$. Moreover, with $m = \theta N_r,\ \theta \in (0, 1]$, using the upper bound on the outage probability (3.43), we get

$$\lambda_s^\star = \Omega(N_r). \tag{3.45}$$

Hence, considering both the outage probability constraints together, from (3.41), (3.44) and (3.45), we get

$$\lambda_s^\star = \Omega\left(\min\{N_r, {N_t}^{1-\frac{2}{\alpha}}\}\right), \text{ and } \lambda_s^\star = \mathcal{O}\left(\min\{N_r, N_t\}\right).$$

□

[2] r represents the number of nearest uncanceled interferers considered for bounding the outage probability.

Finally, we prove a more general result than required by Theorem 3.5.4, where we derive bounds on the outage probability at any primary receiver when each secondary transmitter uses k STDOF for interference suppression towards its k nearest primary receivers. The proof of Theorem 3.5.4 is slightly long and complicated. For Theorem 3.5.4, we fix $k = N_t - 1$ to get (3.41).

Theorem 3.5.5 *If k STDOF are used for interference suppression at each secondary transmitter, then*

$$\lambda_s^\star = \Omega\left(k^{1-\frac{2}{\alpha}}\right), \text{ and } \lambda_s^\star = \mathcal{O}\left(k\right).$$

Proof: Since we are interested in establishing the scaling behavior of the density of the secondary network with respect to N_t, we consider the case when both N_t and k, the number of STDOF used for interference suppression, are large enough. We bound the outage probability (3.40) at a typical primary receiver and find the density of secondary network λ_s that satisfies the primary outage constraint of $P_{p,\text{out}} = \epsilon_p + \Delta_p$.

Lower Bound: Recall that N_{supp} is the random variable representing the number of consecutive nearest secondary interferers suppressed at the typical primary receiver R_{p0}. Let $N_{\mathsf{supp}} = \mathsf{c}$, and recall the definition of interference received at R_{p0}, $I_{\text{MIMO}}(\mathsf{c}) = I_{pp} + I_{sp}^{\mathsf{c}}$ from (3.34), where I_{pp} is the interference contribution from primary transmitters other than T_{p0}, and I_{sp}^{c} is the secondary transmitters other than the c consecutive nearest secondary interferers, at the primary receiver R_{p0}.

From (3.40) and (3.35), the outage probability at primary receiver R_{p0} is $P_{p,\text{out}}(\beta_p)$

$$\begin{aligned}
&= \mathbb{E}_{N_{\mathsf{supp}}}\left\{\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(\mathsf{c})} \le \beta_p\right)\right\},\\
&\overset{(a)}{=} \mathbb{E}_{N_{\mathsf{supp}}}\left\{\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(\mathsf{c})} \le \beta_p\right)\middle| N_{\mathsf{supp}} < \lfloor k/\eta \rfloor\right\} \times \mathbb{P}\left(N_{\mathsf{supp}} < \lfloor k/\eta \rfloor\right)\\
&\quad + \mathbb{E}_{N_{\mathsf{supp}}}\left\{\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(\mathsf{c})} \le \beta_p\right)\middle| N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right\} \times \mathbb{P}\left(N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right),\\
&\overset{(b)}{\le} \delta + \mathbb{E}_{N_{\mathsf{supp}}}\left\{\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(\mathsf{c})} \le \beta_p\right)\middle| N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right\} \times \mathbb{P}\left(N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right),\\
&\overset{(c)}{\le} \delta + \mathbb{E}_{N_{\mathsf{supp}}}\left\{1 - \mathbb{E}\left\{\exp\left(-\frac{\beta_p I_{\text{MIMO}}(\mathsf{c}) d_p^{\alpha}}{P_p}\right)\right\}\middle| N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right\},\\
&\overset{(d)}{=} \delta + \mathbb{E}_{N_{\mathsf{supp}}}\left\{1 - \int_0^\infty \exp\left(-\beta_p (I_{pp}) d_p^{\alpha}\right) f_{I_{pp}}(s)\mathrm{d}s\right.\\
&\quad \left.\int_0^\infty \exp\left(-\frac{\beta_p (I_{sp}^{\mathsf{c}}) P_s d^{\alpha}}{P_p}\right) f_{I_{sp}^{\mathsf{c}}}(t)dt\middle| N_{\mathsf{supp}} \ge \lfloor k/\eta \rfloor\right\},\\
&\overset{(e)}{=} \delta + \mathbb{E}_{N_{\mathsf{supp}}}\left\{1 - \mathcal{L}_{I_{pp}}\left(\beta_p d_p^{\alpha}\right)\right.\\
&\quad \left.\left(1 - \mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{\sum_{n:\ n>\mathsf{c}, T_{sn}\in\Phi_s} P_s d_{sp,n}^{-\alpha}|g_{0n}|^2} \le \beta_p\right)\right)\middle| C \ge \lfloor k/\eta \rfloor\right\},
\end{aligned} \tag{3.46}$$

where (a) follows by splitting the expectation over conditioning the event $N_{\mathsf{supp}} < \lfloor k/\eta \rfloor$ where η is a constant (b) follows by letting $\eta \in \mathbb{N}$ such that $\mathbb{P}\left(N_{\mathsf{supp}} < \lfloor k/\eta \rfloor\right) \le \delta$, $\delta \le \Delta_p$, where

Δ_p is the additional tolerance of outage probability at the primary receivers, and η is independent of k. Existence of $\eta \in \mathbb{N}$ such that $\mathbb{P}\left(N_{\mathsf{supp}} < \lfloor k/\eta \rfloor\right) \leq \Delta_p$ is guaranteed, since for large values of k, canceling only a few nearest secondary interferers has a very small probability. Inequality (c) follows by taking expectation with respect to $|h_{00}|^2 \sim \exp(1)$ and using $\mathbb{P}\left(N_{\mathsf{supp}} \geq \lfloor k/\eta \rfloor\right) \leq 1$. Equality (d) follows since $I_{\mathrm{MIMO}}(\mathsf{c}) = I_{pp} + I_{sp}^{\mathsf{c}}$, and I_{pp} and I_{sp}^{c} are independent. Equality (e) follows by defining $\mathcal{L}_I(.)$ as the Laplace transform of I, and noting that

$$\int_0^\infty \exp\left(-\frac{\beta_p(I_{sp}^{\mathsf{c}})P_s d^\alpha}{P_p}\right) f_{I_{sp}^{\mathsf{c}}}(t)dt = \mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{sp}^{\mathsf{c}}} > \beta_p\right), \tag{3.47}$$

since $|h_{00}|^2 \sim \exp(1)$, and where the expectation in the R.H.S. is taken only with respect to $|h_{00}|^2$. Thus, from (3.46), $P_{p,\mathrm{out}}(\beta_p)$

$$\begin{aligned}
&\overset{(f)}{\leq} \quad \delta + \mathbb{E}_{N_{\mathsf{supp}}}\left\{1 - \exp\left(-\lambda_p c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right)\right. \\
&\qquad \left.\left(1 - (\pi\lambda_s)^{\frac{\alpha}{2}}\beta_p\left(\frac{P_s}{P_p}\right) d_p^\alpha \left(\frac{\alpha}{2} - 1\right)^{-1}\left(\mathsf{c} - \left\lceil \frac{\alpha}{2}\right\rceil\right)^{1-\frac{\alpha}{2}}\right)\middle| N_{\mathsf{supp}} \geq \lfloor k/\eta \rfloor\right\}, \\
&= \quad \delta + 1 - \exp\left(-\lambda_p c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right) + \exp\left(-\lambda_p c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right)(\pi\lambda_s)^{\frac{\alpha}{2}}\beta_p \\
&\qquad \left(\frac{P_s}{P_p}\right) d_p^\alpha \left(\left(\frac{\alpha}{2} - 1\right)^{-1} \mathbb{E}_{N_{\mathsf{supp}}}\left\{\left(\mathsf{c} - \left\lceil \frac{\alpha}{2}\right\rceil\right)^{1-\frac{\alpha}{2}}\middle| \; N_{\mathsf{supp}} \geq \lfloor k/\eta \rfloor\right\}\right), \\
&\overset{(g)}{\leq} \quad \delta + \epsilon_p + \exp\left(-\lambda_p c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right)(\pi\lambda_s)^{\frac{\alpha}{2}}\beta_p \left(\frac{P_s}{P_p}\right) d_p^\alpha \\
&\qquad \left(\left(\frac{\alpha}{2} - 1\right)^{-1} (\lfloor k/\eta \rfloor + \left\lceil \frac{\alpha}{2}\right\rceil)^{1-\frac{\alpha}{2}}\right),
\end{aligned} \tag{3.48}$$

where (f) follows by using the lower bound on the success probability

$$\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{\sum_{n:\; n>\mathsf{c}, T_{sn}\in\Phi_s} P_s d_{sp,n}^{-\alpha}|g_{0n}|^2} > \beta_p\right)$$

from Theorem 3.3.10, by substituting $k = 1$ data stream, and $N_r - k - m = 1$, since the signal strength $|h_{00}|^2 \sim \chi^2(2)$ in this case. Finally (g) follows since for $N_{\mathsf{supp}} \geq \lfloor k/\eta \rfloor, \mathbb{E}_{N_{\mathsf{supp}}}\{(\mathsf{c} - \lceil \frac{\alpha}{2} \rceil)^{1-\frac{\alpha}{2}} | N_{\mathsf{supp}} \geq \lfloor k/\eta \rfloor\} \leq (\lfloor k/\eta \rfloor - \lceil \frac{\alpha}{2} \rceil)^{1-\frac{\alpha}{2}}$ for $\alpha > 2$.

From the primary outage probability constraint in the absence of a secondary wireless network (Theorem 2.4.1)

$$\epsilon_p = 1 - \exp\left(-\lambda_p^\star c \beta_p^{\frac{2}{\alpha}} d_p^2\right),$$

where $\lambda_p^\star$ is the largest density of primary nodes satisfying the outage constraint of ϵ_p.

Hence, equating (3.48) with the relaxed outage probability of $P_{p,\mathrm{out}}(\beta_p) = \epsilon_p + \Delta_p$ at each primary receiver, and substituting for ϵ_p, we get

$$\lambda_s \geq \frac{1}{\pi}\left(\frac{\Delta_p - \delta}{\exp\left(-\lambda_p^\star c_1 \beta_p^{\frac{2}{\alpha}}\right) d_p^2 \beta_p \left(\frac{P_s}{P_p}\right) d_p^\alpha \left(\left(\frac{\alpha}{2} - 1\right)^{-1}(\lfloor k/\eta \rfloor + 1)^{1-\frac{\alpha}{2}} + c_3\right)^{\frac{2}{\alpha}}}\right),$$

and

$$\lambda_s = \Omega\left(k^{1-\frac{2}{\alpha}}\right). \tag{3.49}$$

Upper bound: To find an upper bound on λ_s, we consider the case when exactly k consecutive nearest secondary interferers are suppressed at each primary receiver. Clearly, when each secondary transmitter uses k STDOF for interference suppression toward the primary receivers, at best k consecutive nearest secondary interferers are suppressed at each primary receiver, thus yielding the upper bound. This can also be seen from Fig. 3.9, where each secondary transmitter tries to suppress interference toward its 3 nearest primary receivers.

Thus, from (3.40) and (3.35),

$$\begin{aligned}
P_{p,\text{out}}(\beta_p) &= \mathbb{E}_{N_{\mathsf{supp}}}\left\{\mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(\mathsf{c})} \le \beta_p\right)\right\}, \\
&\overset{(a)}{\ge} \mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{I_{\text{MIMO}}(k)} \le \beta_p\right), \\
&\overset{(b)}{=} 1 - \mathbb{E}\left\{\exp\left(-\frac{\beta_p(P_p I_{pp} + P_s I_{sp}^k) d_p^\alpha}{P_p}\right)\right\}, \\
&= 1 - \int_0^\infty \exp\left(-\beta_p(I_{pp}) d_p^\alpha\right) f_{I_{pp}}(s)\mathrm{d}s \\
&\quad \int_0^\infty \exp\left(-\frac{\beta_p(I_{sp}^k) P_s d^\alpha}{P_p}\right) f_{I_{sp}^x}(t)\mathrm{d}t, \\
&\overset{(c)}{=} 1 - \mathcal{L}_{I_{pp}}\left(\beta_p d_p^\alpha\right) \int_0^\infty \exp\left(-\frac{\beta_p(I_{sp}^k) P_s d^\alpha}{P_p}\right) f_{I_{sp}^x}(t)\mathrm{d}t, \\
&\overset{(d)}{=} 1 - \mathcal{L}_{I_{pp}}\left(\beta_p d_p^\alpha\right)\left(1 - \mathbb{P}\left(\frac{P_p d_p^{-\alpha}|h_{00}|^2}{\sum_{n>k, T_{sn}\in\Phi_s} P_s d_{sp,n}^{-\alpha}|g_{0n}|^2} \le \beta_p\right)\right), \\
&\overset{(e)}{\ge} 1 - \exp\left(-\lambda_p^\star c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right) \frac{\left(k + r + \frac{\alpha}{2}\right)^{\frac{\alpha}{2}}}{d^\alpha \beta (\pi\lambda_s)^{\frac{2}{\alpha}}},
\end{aligned}$$

where (a) follows from the fact that using k STDOF for interference suppression by each secondary transmitter, at best k consecutive nearest secondary interferers are suppressed at each primary receiver, (b) follows by definition of $I_{\text{MIMO}}(k)$ (3.34). Equality (c) follows since the Laplace transform of I_{pp}, the interference contribution from PPP Φ_p with density $\lambda_p^\star$, evaluated at $\left(\beta_p d_p^\alpha\right)$ is

$$\mathcal{L}_{I_{pp}}\left(\beta_p d_p^\alpha\right) = \exp\left(-\beta_p(I_{pp}) d_p^\alpha\right) f_{I_{pp}}(s)\mathrm{d}s = \exp\left(-\lambda_p^\star c_1 \beta_p^{\frac{2}{\alpha}} d_p^2\right).$$

Equality (d) follows similar to (3.47), since $|h_{00}|^2 \sim \exp(1)$, and finally (e) follows from the lower bound on outage probability (Theorem 3.3.8), since the signal power is $\sim \chi^2(2)$ instead of $N - m - k + 1$ for Theorem 3.3.8 and $N_{\mathsf{canc}} = k$ nearest interferers are canceled at each primary receiver.

Thus, we get the upper bound

$$\lambda_s = \mathcal{O}(k). \tag{3.50}$$

Combining (3.49), and (3.50),

$$\lambda_s^\star = \Omega\left(k^{1-\frac{2}{\alpha}}\right), \text{ and } \lambda_s^\star = \mathcal{O}\left(k\right).$$

□

3.6 Reference Notes

The results presented in Section 3.3 and 3.4 can be found in [3]. The study of transmission capacity with multiple antennas was initiated in [8], followed up in [4, 5, 7, 9], and mostly settled in [3]. Results on multiple antennas in cellular networks can be found in [10–13]. Results on space-division multiple access with multiple antennas can be found in [15], and impact of multiple antennas with scheduling can be found in [16]. The results of Section 3.5 with multiple antennas in overlaid networks are presented from [17]. Transmission capacity result for single antenna equipped secondary nodes can be found in [18–23, 23–26].

Bibliography

[1] L. Zheng and D. Tse. 2003. "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels." *IEEE Trans. Inf. Theory* 49 (5): 1073–96.

[2] R. J. Muirhead. 2009. *Aspects of Multivariate Statistical Theory*. Wiley.com, 197.

[3] R. Vaze and R. Heath. 2012. "Transmission capacity of ad-hoc networks with multiple antennas using transmit stream adaptation and interference cancellation." *IEEE Trans. Inf. Theory* 58 (2): 780–92.

[4] N. Jindal, J. Andrews, and S. Weber. 2009. "Rethinking MIMO for wireless networks: Linear throughput increases with multiple receive antennas." In *IEEE International Conference on Communications, 2009. ICC '09.* 1–6.

[5] ——. 2011. "Multi-antenna communication in ad hoc networks: Achieving MIMO gains with SIMO transmission." *IEEE Trans. Commun.* 59 (2): 529–40.

[6] D. Kershaw. 1983. "Some extensions of W. Gautschis inequalities for the gamma function." *Math. Comput.* 41 (164): 607–11.

[7] K. Huang, J. Andrews, D. Guo, R. Heath, and R. Berry. 2012. "Spatial interference cancellation for multiantenna mobile ad hoc networks." *IEEE Trans. Inf. Theory* 58 (3): 1660–76.

[8] A. M. Hunter, J. G. Andrews, and S. P. Weber. 2008. "Capacity scaling of ad hoc networks with spatial diversity." *IEEE Trans. Wireless Commun.* 7 (12): 5058–71.

[9] R. Louie, M. McKay, and I. Collings. 2009. "Spatial multiplexing with MRC and ZF receivers in ad hoc networks." In *IEEE International Conference on Communications, 2009. ICC '09.* 1–5.

[10] M. Demirkol and M. Ingram. 2001. “Power-controlled capacity for interfering MIMO links.” In *Proceedings of IEEE Vehicular Technology Conference (VTC 2001)*.

[11] R. Blum. 2003. “MIMO capacity with interference.” *IEEE J. Sel. Areas Commun.* 21 (5): 793–801.

[12] A. Lozano and A. Tulino. 2002. “Capacity of multiple-transmit multiple-receive antenna architectures.” *IEEE Trans. Inf. Theory* 48 (12): 3117–28.

[13] A. Moustakas, S. Simon, and A. Sengupta. 2003. “MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large n analysis.” *IEEE Trans. Inf. Theory* 49 (10): 2545–61.

[14] A. Edelman. 1989. *Eigenvalues and Condition Numbers of Random Matrices*. PhD. Dissertation, MIT.

[15] M. Kountouris and J. G. Andrews. 2009. “Transmission capacity scaling of SDMA in wireless ad hoc networks.” In *Information Theory Workshop, 2009. IEEE*, 2009 534–38.

[16] A. M. Hunter, R. K. Ganti, and J. G. Andrews. 2010. “Transmission capacity of multi-antenna ad hoc networks with CSMA.” In *Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2010* 1577–81.

[17] R. Vaze. 2011. “Transmission capacity of spectrum sharing ad hoc networks with multiple antennas.” *IEEE Trans. Wireless Commun.* 10 (7): 2334–40.

[18] R. Zhang and Y.-C. Liang. 2008. “Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks.” *IEEE J. Sel. Areas Commun.* 2 (1): 88–102.

[19] M. Vu and V. Tarokh. 2009. “Scaling laws of single-hop cognitive networks.” *IEEE Trans. Wireless Commun.* 8 (8): 4089–97.

[20] S.-W. Jeon, N. Devroye, M. Vu, S.-Y. Chung, and V. Tarokh. 2009. “Cognitive networks achieve throughput scaling of a homogeneous network.” In *7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, June 2009. WiOPT 2009.* 1–5.

[21] C. Yin, C. Chen, and S. Cui. 2009. “Stable distribution based analysis of transmission capacities for overlaid wireless networks.” In *International Conference on Wireless Communications Signal Processing* 1–5.

[22] C. Yin, C. Chen, T. Liu, and S. Cui. 2009. “Generalized results of transmission capacities for overlaid wireless networks.” In *IEEE International Symposium on Information Theory* 1774–78.

[23] K. Huang, V. Lau, and Y. Chen. 2009. “Spectrum sharing between cellular and mobile ad hoc networks: Transmission-capacity trade-off.” *IEEE J. Sel. Areas Commun.* 27 (7): 1256–67.

[24] C. Yin, L. Gao, T. Liu, and S. Cui. 2009. “Transmission capacities for overlaid wireless ad hoc networks with outage constraints.” In *IEEE International Conference on Communications, 2009* 14–18 1–5.

[25] O. Bakr, M. Johnson, R. Mudumbai, and K. Ramchandran. 2009. “Multi-antenna interference cancellation techniques for cognitive radio applications.” In *IEEE WCNC April 2009.* 1–6.

[26] C. Yin, L. Gao, and S. Cui. 2010. “Scaling laws for overlaid wireless networks: A cognitive radio network versus a primary network.” PP (99): 1–1.

Chapter 4

Two-Way Networks

4.1 Introduction

In this chapter, we consider a two-way transmission model for the wireless network, where all source–destination pairs want to exchange information in both directions over a single hop. This model is more realistic than the one-way single-hop model of Chapter 2, since most often in practice, the communication between nodes is two-way, for example, uplink/downlink in cellular networks, feedback for packet acknowledgments, control and channel state information.

For defining the transmission capacity with the two-way communication model, we define an outage event if there is outage in any one of the two directions. Since the outage events in the two directions in wireless networks are correlated, the transmission capacity analysis with the two-way model is non-trivial and does not follow easily from the one-way communication case. Thus, instead of deriving exact expressions for the transmission capacity as done for the one-way communication in Chapter 2, we resort to deriving tight upper and lower bounds on the transmission capacity for the two-way case, which only differ in constants. The tight bounds derived on the transmission capacity also allow us to find the optimal bandwidth/resource partitioning between the communication in two directions that maximizes the transmission capacity.

An added benefit of the two-way model of transmission model is that it allows us to quantify the loss in the multiple antenna transmission capacity with practical feedback in comparison to the genie-aided feedback. The multiple antenna transmission capacity analysis of the CSIT case in Chapter 3 assumes a genie-aided feedback, that is, the transmitter has error-free access to channel coefficients without accounting for resources used for feedback. In this chapter, we characterize the effect of practical channel feedback on the transmission capacity with CSIT and show that genie-aided feedback is a severely simplifying assumption and there is significant performance loss while accounting for realistic feedback requirements.

4.2 Two-Way Communication

In this chapter, we consider an ad hoc network with two-way communication, where each source–destination pair has data to exchange in both directions and extend the one-way transmission capacity framework of Chapter 2 to allow two-way communication, by defining an outage event if there is outage in any one of the two communication directions. With this new

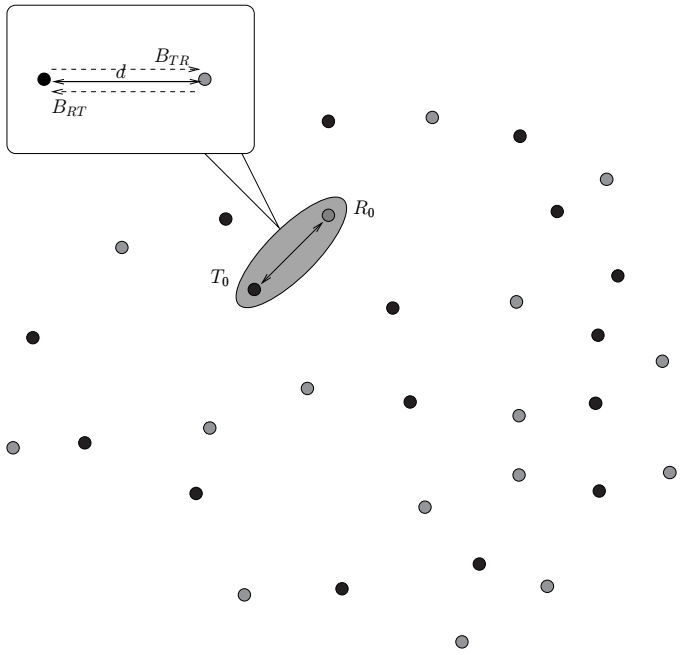

Figure 4.1: Schematic for wireless network with two-way communication, where black dots represent nodes of Φ_T and gray dots represent nodes of Φ_R.

definition, we find the maximum density of nodes with two-way communication that satisfies a per-user data rate and outage probability constraint for transmissions in both directions. This framework also allows us to find the optimal resource allocation (time/bandwidth) in the two directions, given data rate requirements of the two directions.

Consider an *ad hoc* network with two sets of nodes $\Phi_T = \{T_n,\ n \in \mathbb{N}\}$ and $\Phi_R = \{R_n,\ n \in \mathbb{N}\}$, where T_n and R_n want to exchange data between each other for each n. Similar to Chapter 2, we let T_n, R_n denote the nodes as well as their locations. We assume each T_n and R_n to have a single antenna and consider a slotted ALOHA protocol, where at any given time, the pair $(T_n,\ R_n)$ transmits data to each other with an access probability p for each n, independent of all other nodes $\Phi_T \backslash \{T_n\},\ \Phi_R \backslash \{R_n\}$, respectively. We assume that the distance between each T_n and R_n pair is fixed to be d with random orientation. We consider a general system, where the data requirement in both directions can be different, and a frequency division duplex (FDD) communication model, where two separate frequency carriers are used for two directions, thereby forming a full-duplex link. Analysis with time division duplex (TDD) is identical.

As in Chapter 2, for purposes of analysis, the set $\Phi_T = \{T_n\}$ is modeled as a homogenous Poisson point process (PPP) on a two-dimensional plane with density λ_0. Since R_n is at a fixed distance d from the T_n in a random direction, the set $\Phi_R = \{R_n\}$ is also a homogenous PPP on a two-dimensional plane with density λ_0. With ALOHA protocol, at any given time, the active node location processes $\Phi_T^a = \{T_n | T_n \text{ is active}\}$ and $\Phi_R^a = \{R_n | R_n \text{ is active}\}$ are homogenous PPPs on a two-dimensional plane with density $\lambda = p\lambda_0$, where p is the ALOHA access probability. The total available bandwidth is F_{total}, out of which F_{TR} is dedicated for $T_n \rightarrow R_n$ communication to support a rate demand B_{TR} bits/sec/Hz, and the rest $F_{RT} = F_{\text{total}} - F_{TR}$ for the $R_n \rightarrow T_n$

communication to support a rate demand of B_{RT} bits/sec/Hz. Fig. 4.1 illustrates the considered system model.

In a time slot, when a typical pair $(T_0,\ R_0)$ is active, the received signal at R_0 is

$$y_0 = d^{-\alpha/2} h_{00} x_0 + \sum_{n:T_n \in \Phi_T^a} d_{T_n}^{-\alpha/2} h_{0n} x_n, \tag{4.1}$$

and the received signal at T_0 is

$$z_0 = d^{-\alpha/2} g_{00} u_0 + \sum_{n:R_n \in \Phi_R^a} d_{R_n}^{-\alpha/2} g_{0n} u_n, \tag{4.2}$$

where h_{00} and g_{00} are the channel coefficients between T_0 and R_0 and R_0 and T_0, respectively, h_{0n} and g_{0n} are the channel coefficients between T_n and R_0 and R_n and T_0, respectively, d_{T_n} and d_{R_n} are the distances between T_n and R_0 and R_n and T_0, respectively, $\alpha > 2$ is the path loss exponent, and x_n and u_n are the signals transmitted from T_n and R_n, respectively. We consider an interference-limited regime and ignore the effects of AWGN. We assume that h_{00}, g_{00}, h_{0n}, and g_{0n}, are i.i.d. with Rayleigh distribution.

With the received signal models (4.1) and (4.2), SIRs from $T_0 \to R_0$ and from $R_0 \to T_0$ are

$$\mathsf{SIR}_{TR} = \frac{d^{-\alpha}|h_{00}|^2}{\sum_{n:T_n \in \Phi_T^a \setminus \{T_0\}} d_{T_n}^{-\alpha}|h_{0n}|^2}, \tag{4.3}$$

$$\mathsf{SIR}_{RT} = \frac{d^{-\alpha}|g_{00}|^2}{\sum_{n:R_n \in \Phi_R^a \setminus \{R_0\}} d_{R_n}^{-\alpha}|g_{0n}|^2}, \tag{4.4}$$

respectively.

The mutual informations [17] for the $T_0 \to R_0$ channel using bandwidth F_{TR}, and for the $R_0 \to T_0$ channel using bandwidth F_{RT} are

$$\mathsf{I}_{TR} = F_{TR} \log\left(1 + \mathsf{SIR}_{TR}\right) \ \text{ bits/sec/Hz},$$

and

$$\mathsf{I}_{RT} = F_{RT} \log\left(1 + \mathsf{SIR}_{RT}\right) \ \text{bits/sec/Hz},$$

respectively. Recall that the rate requirement for the $T_0 \to R_0$ transmission is B_{TR} bits/sec/Hz, and for the $R_0 \to T_0$ communication is B_{RT} bits/sec/Hz. Thus, to account for the two-way or bidirectional nature of communication, we define the success probability (complement of the outage probability ϵ) as the probability that communication in both directions is successful simultaneously, that is,

$$P_{\text{suc}} = P\left(\mathsf{I}_{TR} > B_{TR},\ \mathsf{I}_{RT} > B_{RT}\right). \tag{4.5}$$

This is a natural definition since it models the fact that successful communication in one direction depends on the success of the communication in the other direction, for example channel feedback, ACK-NACK, signals.

Let λ_{tw} be maximum density of nodes of Φ_T (or Φ_R) per unit area that can support rate B_{TR} from $T_0 \to R_0$ and B_{RT} from $R_0 \to T_0$, with success probability $P_{\text{suc}} = 1 - \epsilon$, using bandwidth F_{total}. Then a natural extension of single-hop transmission capacity defined in Chapter 2 to capture the two-way communication model is as follows.

Definition 4.2.1 *The two-way transmission capacity is defined as*

$$C_{tw} = (1-\epsilon)\left(\frac{B_{TR}+B_{RT}}{F_{\text{Total}}}\right)\lambda_{tw} \text{ bits/sec/Hz/m}^2,$$

by accounting for λ_{tw} two-way transmissions per meter square with communication rate of B_{TR} and B_{RT} in two directions under a joint success probability constraint of $(1-\epsilon)$.

It remains to find λ_{tw}, which will allows us to derive C_{tw} for a given rate $B_{TR},\ B_{RT}$, success probability $1-\epsilon$, and bandwidth F_{total}.

Remark 4.2.2 *Assuming different success probability requirements in two directions, we can define*

$$P^f_{\text{suc}} = P\left(\mathsf{I}_{TR} > B_{TR}\right)$$

for the $T_0 \to R_0$ direction and

$$P^r_{\text{suc}} = P\left(\mathsf{I}_{RT} > B_{RT}\right)$$

for the $R_0 \to T_0$ direction. Let λ_f be the maximum density of nodes per unit area that can support rate B_{TR} from $T_0 \to R_0$ with success probability $P^f_{\text{success}} = 1-\epsilon_f$, and λ_r be the maximum density of nodes per unit area that can support rate B_{RT} bits from $R_0 \to T_0$ with success probability $P^r_{\text{success}} = 1-\epsilon_r$. Then, with the total bandwidth of F_{total}, the two-way transmission capacity can be defined as

$$C_{\epsilon_f,\epsilon_r} = (1-\epsilon_f)\lambda_f\left(\frac{B_{TR}}{F_{TR}}\right) + (1-\epsilon_r)\lambda_r\left(\frac{B_{RT}}{F_{\text{Total}}-F_{TR}}\right) \text{bits/sec/Hz/m}^2.$$

With this definition, the two-way network gets decoupled into two one-way networks, and $C_{\epsilon_f,\epsilon_r}$ can be derived easily by finding λ_f, λ_r similar to the results from transmission capacity of one-way networks (Theorem 2.4.1).

4.2.1 Computing the Two-Way Transmission Capacity

We consider a typical pair (T_0, R_0) to compute the success probability P_{suc} (4.5). From Slivnyak's theorem (2.3.3), we know that conditioned on the event that $T_0 \in \Phi$, the locations of the interferers for R_0 form a homogenous PPP with density λ (Remark 2.3.14). Similarly, for the interference seen at T_0, when R_0 transmits. Thus, from (4.3) and (4.4), the conditional SIRs ($T_0 \in \Phi^a_T, R_0 \in \Phi^a_R$) are

$$\mathsf{SIR}_{TR} = \frac{d^{-\alpha}|h_{00}|^2}{\sum_{n:T_n\in\Phi^a_T} d^{-\alpha}_{T_n}|h_{0n}|^2}, \tag{4.6}$$

$$\mathsf{SIR}_{RT} = \frac{d^{-\alpha}|g_{00}|^2}{\sum_{n:R_n\in\Phi^a_R} d^{-\alpha}_{R_n}|g_{0n}|^2}. \tag{4.7}$$

From (4.5), the success probability is

$$P_{\text{suc}} \quad = \quad P\left(\mathsf{SIR}_{TR} > 2^{\frac{B_{TR}}{F_{TR}}}-1, \mathsf{SIR}_{RT} > 2^{\frac{B_{RT}}{F_{RT}}}-1\right). \tag{4.8}$$

Define $\beta_1 = \left(2^{\frac{B_{TR}}{F_{TR}}}-1\right),\ \ \beta_2 = \left(2^{\frac{B_{RT}}{F_{RT}}}-1\right)$, and aggregate interference seen at R_0 and T_0, respectively: as $I_{TR} = \sum_{n:T_n\in\Phi^a_T} d^{-\alpha}_{T_n}|h_{0n}|^2,\ \ I_{RT} = \sum_{n:R_n\in\Phi^a_R} d^{-\alpha}_{Rn}|g_{0n}|^2$.

Then, from (4.8)

$$\begin{aligned} P_{\text{suc}} &= P\left(\frac{|h_{00}|^2}{I_{TR}} > d^{\alpha}\beta_1,\ \frac{|g_{00}|^2}{I_{RT}} > d^{\alpha}\beta_2\right), \\ &\overset{(a)}{=} \mathbb{E}\left\{\exp\left(-d^{\alpha}\beta_1 I_{TR}\right)\exp\left(-d^{\alpha}\beta_2 I_{RT}\right)\right\}, \\ &\overset{(b)}{=} \mathbb{E}\left\{\exp\left(-d^{\alpha}\beta_1\left(\sum_{n:T_n\in\Phi_T^a} d_{T_n}^{-\alpha}|h_{0n}|^2\right)\right)\right. \\ &\qquad \left.\exp\left(-d^{\alpha}\beta_2\left(\sum_{n:R_n\in\Phi_R^a} d_{R_n}^{-\alpha}|g_{0n}|^2\right)\right)\right\}, \\ &\overset{(c)}{=} \mathbb{E}\left\{\prod_{n:T_n\in\Phi_T^a}\left(\frac{1}{1+d^{\alpha}\beta_1 d_{T_n}^{-\alpha}}\right)\prod_{n:R_n\in\Phi_R^a}\left(\frac{1}{1+d^{\alpha}\beta_2 d_{R_n}^{-\alpha}}\right)\right\}, \end{aligned} \tag{4.9}$$

where (a) follows since $P\left(|h_{00}|^2 > x\right) = P\left(|g_{00}|^2 > x\right) = \exp\left(-x\right)$ as $|h_{00}|^2, |g_{00}|^2 \sim \exp(1)$, and h_{00} and g_{00} are independent, (b) follows by using the definition of I_{TR} and I_{RT}, and (c) follows by taking the expectation with respect to $h_{0n} \sim \exp(1)$, and $g_{0n} \sim \exp(1)$, and noting that h_{0n}, and g_{0n} are independent.

Finding exact expression for (4.9) is challenging since the distances $\{d_{T_n}\}$ and $\{d_{R_n}\}$ are not independent in a two-way wireless network. To visualize their dependence, consider Fig. 4.2, where there are only two active pairs of nodes, (T_0, R_0) and (T_1, R_1). For R_0, the transmission from T_1 at a distance d_{T1} is interference, while for receiver T_0, the transmission from R_1 at distance d_{R1} is interference. For the case when d is very small $d \to 0$, $d_{R1} \approx d_{T1}$, and thus distances d_{R1} and d_{T1} are dependent. Moreover, explicitly computing the correlation between d_{T_n} and d_{R_n} is also a hard problem. Thus, to get a meaningful insight into the two-way transmission capacity, we derive tight lower and upper bounds. The upper bound is derived using the FKG inequality (Lemma 2.8.8), similar to Section 2.8.2, while for deriving a lower bound, we make use of the Cauchy–Schwarz inequality.

4.2.2 Lower Bound on the Success Probability

From Definition 2.8.7, both the terms inside the expectation of (4.9),

$$\prod_{n:T_n\in\Phi_T^a}\left(\frac{1}{1+d^{\alpha}\beta_1 d_{T_n}^{-\alpha}}\right)$$

and

$$\prod_{n:R_n\in\Phi_R^a}\left(\frac{1}{1+d^{\alpha}\beta_2 d_{R_n}^{-\alpha}}\right)$$

are decreasing random variables, since each term in the product is less than 1, and with the increasing number of terms (number of interferers) in the product, the total value of each expression decreases.

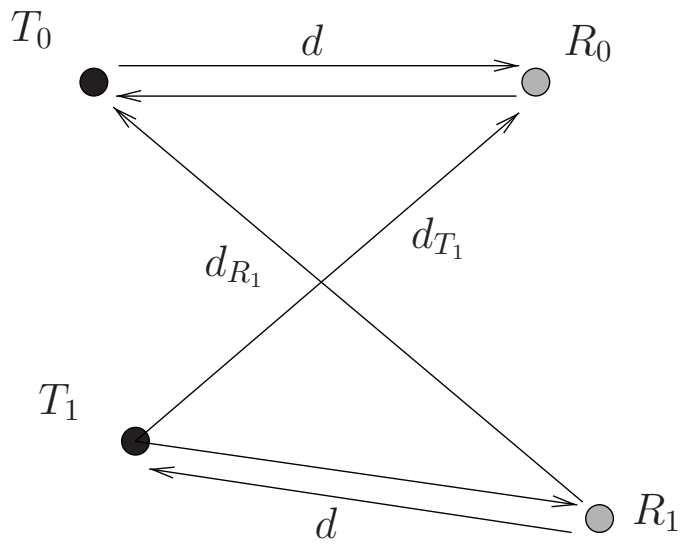

Figure 4.2: Schematic of two-way communication with two pairs of nodes.

Thus, using FKG inequality (Lemma 2.8.8), from (4.9)

$$
\begin{aligned}
P_{\text{suc}} &\geq \mathbb{E}\left\{\prod_{n:T_n\in\Phi_T^a}\left(\frac{1}{1+d^\alpha\beta_1 d_{T_n}^{-\alpha}}\right)\right\}\mathbb{E}\left\{\prod_{n:R_n\in\Phi_R^a}\left(\frac{1}{1+d^\alpha\beta_2 d_{R_n}^{-\alpha}}\right)\right\},\\
&\overset{(a)}{=} \exp\left(\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{1}{1+d^\alpha\beta_1 x^{-\alpha}}\right)\,\mathrm{d}x\right)\right)\\
&\quad \exp\left(\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{1}{1+d^\alpha\beta_2 x^{-\alpha}}\right)\,\mathrm{d}x\right)\right),\\
&\overset{(b)}{=} \exp\left(\left(-2\pi\lambda\int_0^\infty\left(\frac{d^\alpha\beta_1 x^{-\alpha+1}}{1+d^\alpha\beta_1 x^{-\alpha}}\right)\,\mathrm{d}x\right)\right)\\
&\quad \exp\left(\left(-2\pi\lambda\int_0^\infty\left(\frac{d^\alpha\beta_2 x^{-\alpha+1}}{1+d^\alpha\beta_2 x^{-\alpha}}\right)\,\mathrm{d}x\right)\right),\\
&= \exp\left(-\lambda c_1 d^\alpha\beta_1^{\frac{2}{\alpha}}\right)\exp\left(-\lambda c_1 d^\alpha\beta_2^{\frac{2}{\alpha}}\right),\\
&= \exp\left(-\lambda c_1\left(d^\alpha\beta_1^{\frac{2}{\alpha}}+d^\alpha\beta_2^{\frac{2}{\alpha}}\right)\right), \qquad (4.10)
\end{aligned}
$$

where (a) follows from the probability-generating functional of the PPP (Theorem 2.3.6) (b) follows by changing the integration from over $\mathbb{R}^2$ to $\mathbb{R}$, and $c_1 = \frac{2\pi^2}{\alpha}\csc(\frac{2\pi}{\alpha})$ is a constant, where csc is the co-secant.

4.2.3 Upper Bound on the Success Probability

Using the Cauchy–Schwarz inequality $\mathbb{E}\{XY\} \leq \sqrt{\mathbb{E}\{X^2\}\mathbb{E}\{Y^2\}}$, from (4.9)

$$\begin{aligned}
P_{\text{suc}} &\leq \left(\mathbb{E}\left\{ \prod_{n:T_n \in \Phi_T^a} \left(\frac{1}{1+d^\alpha \beta_1 d_{T_n}^{-\alpha}} \right)^2 \right\} \right. \\
&\quad \left. \mathbb{E}\left\{ \prod_{n:R_n \in \Phi_R^a} \left(\frac{1}{1+d^\alpha \beta_2 d_{R_n}^{-\alpha}} \right)^2 \right\} \right)^{\frac{1}{2}}, \\
&\overset{(a)}{=} \left(\exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \left(\frac{1}{1+d^\alpha \beta_1 x^{-\alpha}} \right)^2 \mathrm{d}x \right) \right. \\
&\quad \left. \exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \left(\frac{1}{1+d^\alpha \beta_2 x^{-\alpha}} \right)^2 \mathrm{d}x \right) \right)^{\frac{1}{2}}, \\
&\overset{(b)}{=} \left(\exp\left(-2\pi\lambda \int_0^\infty \left(\frac{d^{2\alpha}\beta_1^2 x^{-2\alpha+1} + 2d^\alpha \beta_1 x^{-\alpha+1}}{\left(1+d^\alpha \beta_1 x^{-\alpha}\right)^2} \right) \mathrm{d}x \right) \right. \\
&\quad \left. \exp\left(-2\pi\lambda \int_0^\infty \left(\frac{d^{2\alpha}\beta_2^2 x^{-2\alpha+1} + 2d^\alpha \beta_2 x^{-\alpha+1}}{\left(1+d^\alpha \beta_2 x^{-\alpha}\right)^2} \right) \mathrm{d}x \right) \right)^{\frac{1}{2}}, \\
&\overset{(c)}{=} \exp\left(-\lambda c_2 d^\alpha \beta_1^{\frac{2}{\alpha}} \right) \exp\left(-\lambda c_2 d^\alpha \beta_2^{\frac{2}{\alpha}} \right) \exp\left(-\lambda c_2 \left(d^\alpha \beta_1^{\frac{2}{\alpha}} + d^\alpha \beta_2^{\frac{2}{\alpha}} \right) \right), \quad (4.11)
\end{aligned}$$

where (a) follows from the probability-generating functional of the PPP (Theorem 2.3.6), (b) follows by changing the integration from over $\mathbb{R}^2$ to $\mathbb{R}$, and (c) follows by evaluating the integral where $c_2 = \frac{\pi^2}{\alpha^2} \csc\left(\frac{2\pi}{\alpha}\right)(\alpha+2)$ is a constant. This leads us to the following theorem that characterizes the two-way transmission capacity upto a constant.

Theorem 4.2.3 *The two-way transmission capacity is lower and upper bounded by*

$$\begin{aligned}
C_{tw} &\geq (1-\epsilon) \frac{-\ln(1-\epsilon)}{c_1 \left(d^\alpha \beta_1^{\frac{2}{\alpha}} + d^\alpha \beta_2^{\frac{2}{\alpha}} \right)} \frac{B_{TR}+B_{RT}}{F_{\text{Total}}}, \\
C_{tw} &\leq (1-\epsilon) \frac{-\ln(1-\epsilon)}{c_2 \left(d^\alpha \beta_1^{\frac{2}{\alpha}} + d^\alpha \beta_2^{\frac{2}{\alpha}} \right)} \frac{B_{TR}+B_{RT}}{F_{\text{Total}}},
\end{aligned}$$

where

$$c_1 = \frac{2\pi^2}{\alpha} \csc\left(\frac{2\pi}{\alpha}\right), \quad c_2 = \frac{\pi^2}{\alpha^2} \csc\left(\frac{2\pi}{\alpha}\right)(\alpha+2),$$

and $c_2/c_1 = \frac{1}{2} + \frac{1}{\alpha}$.

Proof: With $P_{\text{suc}} = 1-\epsilon$, and using the definition of C_{tw} in Definition 4.2.1, the result follows from (4.10) and (4.11). □

The bounds derived in Theorem 4.2.3 differ only in constants c_1 and c_2, where the ratio of c_2 and c_1 depends only on the path-loss exponent α. Thus, Theorem 4.2.3 gives a simple characterization of the exact dependence of transmission capacity on the outage probability constraint ϵ, the rate

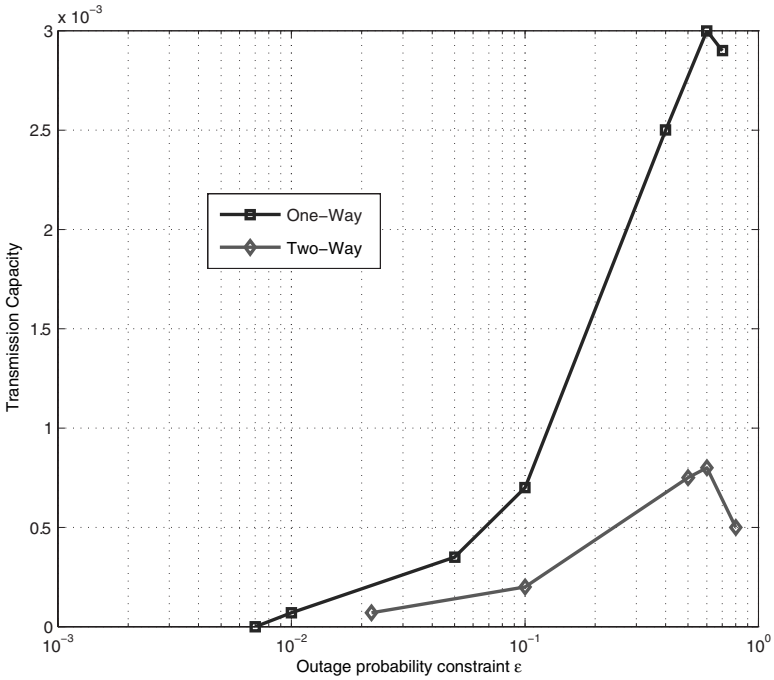

Figure 4.3: Comparison of one-way and two-way transmission capacity with $d = 5\,\text{m}$, $\alpha = 4$, $B_{TR} = 1$ Mbits, $B_{RT} = 0.03$ Mbits, $F = 1.1$ MHz, and $F_{TR} = 1$ MHz.

of information transfer in two directions B_{TR}, B_{RT} and consequently the thresholds β_1, β_2, and density λ.

Note that if we define $P_{\text{suc}} = P\left(\mathsf{I}_{TR} > B_{TR}\right) P\left(\mathsf{I}_{RT} > B_{RT}\right)$ assuming independence of I_{TR} and I_{RT} or SIR_{TR} and SIR_{RT} instead of

$$P_{\text{suc}} = P\left(\mathsf{I}_{TR} > B_{TR},\, \mathsf{I}_{RT} > B_{RT}\right)$$

as in (4.5), using Thereom 2.4.1, we will get exactly the same bounds in Theorem 4.2.3 with only a difference in the constants. This is indeed surprising and tells us that there is only a constant performance difference assuming independent SIRs in two directions instead of the actual correlated SIRs with two-way communication in wireless networks. The technique used in deriving Theorem 4.2.3 is fairly novel in its use for wireless networks and surprisingly results in tight bounds on the transmission capacity since typically, Cauchy–Schwarz inequality provides a loose upper bound. The FKG inequality allows us to derive a matching lower bound (upto a constant). This technique is also expected to be useful for cases when exact expressions for transmission capacity cannot be obtained.

Requiring that transmissions be successful in both directions, there is a loss in the two-way transmission capacity compared to the one-way transmission capacity. In Fig. 4.3, we quantify this loss for $B_{TR} = 1$ Mbits, $B_{RT} = 0.03$ Mbits, $F = 1.1$ MHz, $F_{TR} = 1$ MHz, and $F - F_{TR} = 0.01$ MHz by plotting the two-way transmission capacity with respect to the outage probability constraint ϵ. For plotting one-way transmission capacity, we use rate $B = B_{TR} + B_{RT} = 1.03$ Mbits with total bandwidth $F = 1.01$ MHz and $d = 5\,\text{m}$ and $\alpha = 4$. Fig. 4.3 shows that the two-way transmission

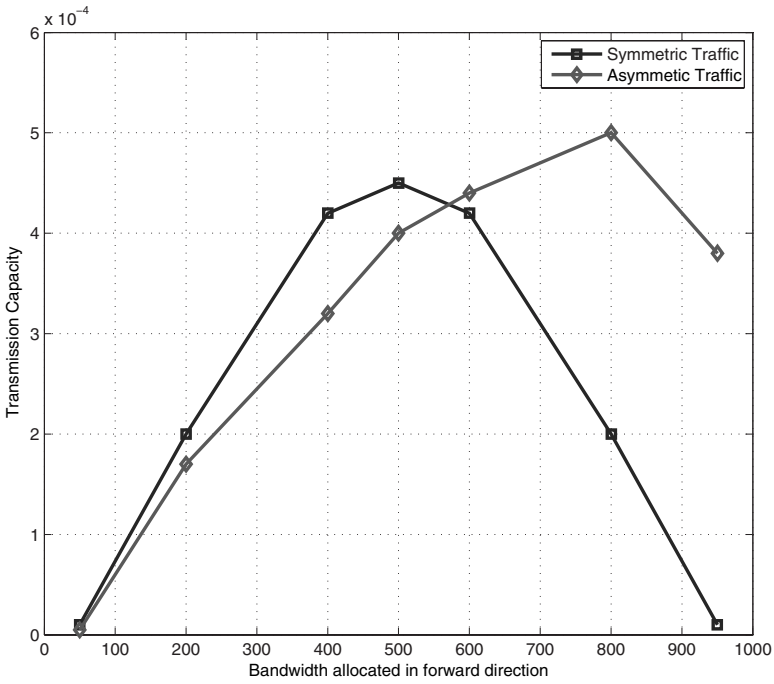

Figure 4.4: Two-way transmission capacity as a function of bandwidth allocation.

capacity is approximately half the one-way transmission capacity, thus accounting for the two-way outage constraint severely impacts the transmission capacity and the network design.

The two-way transmission capacity found in Theorem 4.2.3 has been derived for a fixed bandwidth allocation in the two directions. Given different rate requirement B_{TR}, B_{RT}, in the two directions, it is important to find the optimal bandwidth allocation that maximizes the two-way transmission capacity. We accomplish this in the next section, where given the rate requirements B_{TR} and B_{RT}, success probability $1-\epsilon$, and total bandwidth F_{total}, we find the optimal bandwidth allocation.

4.2.4 Two-Way Bandwidth Allocation

In Section 4.2.1, we derived the two-way transmission capacity of ad hoc networks within a constant as a function of bandwidth F_{TR} and F_{RT} allocated to the $T_0 \to R_0$ and $R_0 \to T_0$ communication, respectively. Since the total bandwidth F_{total} is finite, an important question to answer is what is the optimal bandwidth allocation between the two directions that maximizes the transmission capacity? For the special case of equal rate requirement in both directions, that is, $B_{TR} = B_{RT}$, equal bandwidth allocation is optimal, that is, $F_{TR} = F_{RT} = F_{\text{total}}/2$. For the non-symmetric case, however, the answer is not obvious and is derived in the following theorem.

Theorem 4.2.4 *The optimum bidirectional bandwidth allocation that maximizes the transmission capacity with two-way communication is $F^{\star}_{TR} = x^{\star}$ and $F^{\star}_{RT} = F_{\text{total}} - x^{\star}$, where $x^{\star}$ is the unique*

positive solution to the following equation:

$$\frac{1}{B_{TR}}\mathrm{g}\left(\frac{B_{TR}}{x}\right)-\frac{1}{B_{RT}}\mathrm{g}\left(\frac{B_{RT}}{F_{\text{total}}-x}\right)=0, \tag{4.12}$$

where $\mathrm{g}(t)=t^2 2^t(2^t-1)^{(\delta-1)}$ *for* $0<t<F_{\text{total}}$, *and* $\delta=\frac{2}{\alpha}$.

Proof: Neglecting the constants c_1 and c_2, from Theorem 4.2.3, the two-way transmission capacity is

$$\begin{aligned} C_{tw} &= (1-\epsilon)\frac{-\ln(1-\epsilon)\left(\frac{B_{TR}+B_{RT}}{F_{\text{total}}}\right)}{d^2\left(\left(2^{\frac{B_{TR}}{F_{TR}}}-1\right)^{\frac{2}{\alpha}}+\left(2^{\frac{B_{RT}}{F_{RT}}}-1\right)^{\frac{2}{\alpha}}\right)}, \\ &= (1-\epsilon)\frac{-\ln(1-\epsilon)\left(\frac{B_{TR}+B_{RT}}{F_{\text{total}}}\right)}{d^2\left(\left(2^{\frac{B_{TR}}{F_{TR}}}-1\right)^{\frac{2}{\alpha}}+\left(2^{\frac{B_{RT}}{F_{\text{total}}-F_{TR}}}-1\right)^{\frac{2}{\alpha}}\right)}, \end{aligned} \tag{4.13}$$

where in last equality we have substituted $F_{RT}=F_{\text{total}}-F_{TR}$. To derive the optimal bandwidth partitioning, that is, the optimal F_{TR} that maximizes C_{tw}, we need to minimize the denominator in (4.13), $\left(\left(2^{\frac{B_{TR}}{F_{TR}}}-1\right)^{\frac{2}{\alpha}}+\left(2^{\frac{B_{RT}}{F_{\text{total}}-F_{TR}}}-1\right)^{\frac{2}{\alpha}}\right)$, with respect to F_{TR}. Let $\delta=\frac{2}{\alpha}$ and

$$f(x)=\left(\left(2^{\frac{B_{TR}}{x}}-1\right)^{\delta}+\left(2^{\frac{B_{RT}}{F_{\text{total}}-x}}-1\right)^{\delta}\right).$$

Thus, the problem we need to solve is $\min_{x\in(0,F_{\text{total}})} f(x)$.

The first derivative of $f(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x)=\delta\log_e 2\left(-\frac{1}{B_{TR}}\mathrm{g}\left(\frac{B_{TR}}{x}\right)+\frac{1}{B_{RT}}\mathrm{g}\left(\frac{B_{RT}}{F_{\text{total}}-x}\right)\right),$$

where $\mathrm{g}(t)=t^2 2^t(2^t-1)^{(\delta-1)}$ for $t\geq 0$, while the second derivative of $f(x)$ is

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x)=\delta\log_e 2\left(\frac{1}{x^2}\mathrm{g}\left(\frac{B_{TR}}{x}\right)+\frac{1}{(F_{\text{total}}-x)^2}\mathrm{g}\left(\frac{B_{RT}}{F_{\text{total}}-x}\right)\right).$$

Since $\mathrm{g}(t)$ is monotonically increasing in t for $t\geq 0$, we have $\mathrm{g}(t)>\mathrm{g}(0)=0$ for all $t>0$. Therefore, $\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x)>0$ for all $x\in(0,F_{\text{total}})$. Thus $f(x)$ is a convex function over $(0,F_{\text{total}})$. Differentiating $f(x)$ and equating it to zero, results in the optimal solution $x^\star$ that satisfies

$$\frac{1}{B_{TR}}\mathrm{g}\left(\frac{B_{TR}}{x}\right)-\frac{1}{B_{RT}}\mathrm{g}\left(\frac{B_{RT}}{F_{\text{total}}-x}\right)=0.$$

□

For asymmetric traffic requirement in two directions, that is, $B_{TR}\neq B_{RT}$, allocating bandwidths proportional to the desired rate in each direction $F_{TR}=\frac{B_{TR}}{B_{TR}+B_{RT}}F_{\text{total}}$ does not

satisfy (4.12). Thus an important conclusion we derive from Theorem 4.2.4 is that the proportional bandwidth allocation policy is not optimal for asymmetric traffic for maximizing the transmission capacity and (4.12) must be satisfied to find the optimal policy. If the traffic is symmetric, that is, $B_{TR} = B_{RT}$, the intuitive strategy of allocating equal bandwidths for two directions with $F_{TR} = F_{RT} = F_{\text{total}}/2$ is optimal.

In Fig. 4.4, we plot the transmission capacity as a function of the bandwidth allocated in the forward channel from $T_n \to R_n$, F_{TR}, for both the symmetric traffic with $B_{TR} = B_{RT} = 1$ Mbits, and asymmetric traffic with $B_{TR} = 1$ Mbits and $B_{RT} = 0.05$ Mbits, with total bandwidth $F_{\text{total}} = 1$ MHz, $d = 5$ m and $\alpha = 4$, and outage probability constraint of $\epsilon = .1$. For both the cases, we see from Fig. 4.4 that the transmission capacity is maximized at the optimal bandwidth allocation derived in Theorem 4.2.4.

Next, we present one application of two-way transmission capacity formulation presented in this section to quantify the impact of limited feedback on the transmission capacity with multiple antennas and CSIT. Typically, a genie-aided error-free feedback is assumed to show remarkable capacity gains using multiple antennas (Section 3.4). These results are misleading, since they do not account for the resources used for feedback, for example, the number of bits of feedback that determine the quality of feedback and the time taken in the reverse link for feedbacking the required bits. Using the results derived in this section, we analyze the performance of multiple antennas with CSIT under practical feedback in terms of the transmission capacity and show that practical feedback severely limits the promised gains (see below).

4.3 Effect of Limited Feedback on Two-Way Transmission Capacity with Beamforming

In this section, we quantify the effect of practical feedback on the multiple antenna transmission capacity with CSIT. We consider the case where each transmitter T_n is equipped with N antennas, while each receiver R_n has a single antenna. All other system parameters and assumptions remain the same as defined in Section 3.4.

For simplicity, we assume that each transmitter sends only one data stream and uses beamforming from all its N antennas for its transmission, where the signal is transmitted along the strongest eigenmode of the channel between the transmitter and the receiver. The received signal model is given by (3.24) with $k = 1$. For this case, from Theorem 3.4.2 we know that with the genie-aided feedback, the transmission capacity scales as $N^{\frac{2}{\alpha}}$. Theorem 3.4.2 though does not account for the resources used for feeding back the channel coefficients that are required for beamforming, and therefore overestimates the transmission capacity.

Limited feedback techniques [3] are commonly used in practical systems to feedback the channel coefficients, where a codeword from the codebook (pre-decided by the transmitter and the receiver) is fed back by the receiver to the transmitter that is closest to the actual channel coefficients under some chosen metric. Clearly, larger the codebook size, better is the quality of the feedback, and consequently better is the data rate from the transmitter to the receiver with beamforming. With a codebook size of 2^F codewords, each codeword requires F bits of feedback. We assume an FDMA system, where the transmitter and receiver use different portions of the bandwidth for transmitting data to the receiver and feedbacking the channel coefficient information, respectively. Use of a large codebook at the receiver increases the precision of the channel coefficient information but at the same time requires larger bandwidth for the feedback channel, thereby restricting the bandwidth allocated for data transmission from the transmitter to

the receiver. Thus, there is a three-fold tradeoff between the bandwidth allocation for the two directions, the size of the codebook, and the transmission capacity.

In this section, we want to quantify this tradeoff and evaluate the effect of practical feedback mechanism on the transmission capacity with beamforming. Towards that end, we will use the two-way communication model of Section 4.2. In particular, we let F_{total} be the total bandwidth that is divided between the $T_0 \to R_0$ (F_{TR}) and $R_0 \to T_0$ ($F_{\text{total}} - F_{TR}$) communication, similar to Section 4.2. As before, we assume that each transmitter–receiver pair is at a distance d from each other.

The received signal at typical receiver R_0 over bandwidth F_{TR} is

$$y_0 = d^{-\alpha/2}\mathbf{h}_{00}^{\dagger}\mathbf{b}_0 x_0 + \sum_{n:T_n \in \Phi \setminus \{T_0\}} d_{T_n}^{-\alpha/2}\mathbf{h}_{0n}^{\dagger}\mathbf{b}_0 x_n, \tag{4.14}$$

where $\mathbf{b}_n$ is the beamforming vector used by transmitter T_n, $\mathbf{h}_{0n}^{\dagger} \in \mathbb{C}^{1\times N}$ is the channel between T_n and R_0, d_{T_n} is the distance between T_n and R_0, x_0 and x_n are the data symbols transmitted from T_0 and T_n, respectively. We assume a unit power transmission from all transmitters and ignore the additive white Gaussian noise.

The received signal at a typical transmitter T_0 with N antennas, corresponding to the feedback by a typical receiver R_0, over bandwidth $F_{\text{total}} - F_{TR}$ is

$$\mathbf{z}_0 = d^{-\alpha/2}\mathbf{g}_{00}u_0 + \sum_{n:R_n \in \Phi \setminus \{R_0\}} d_{R_n}^{-\alpha/2}\mathbf{g}_{0n}u_n, \tag{4.15}$$

where $\mathbf{g}_{0n} \in \mathbb{C}^{N\times 1}$ is the channel vector between R_n and T_0, d_{R_n} is the distance between R_n and T_0, and u_0 and u_n are the feedback signals transmitted by R_0 and R_n, respectively. To decode u_0, receiver R_0 employs maximum ratio combining and multiplies $\frac{\mathbf{g}_{00}^{\dagger}}{|\mathbf{g}_{00}|}$ to the received signal $\mathbf{z}_0$, and we get

$$\frac{\mathbf{g}_{00}^{\dagger}}{|\mathbf{g}_{00}|}\mathbf{z}_0 = d^{-\alpha/2}|\mathbf{g}_{00}|u_0 + \sum_{n:R_n \in \Phi \setminus \{R_0\}} d_{R_n}^{-\alpha/2}\mathbf{g}_{0n}\frac{\mathbf{g}_{00}^{\dagger}}{|\mathbf{g}_{00}|}u_n. \tag{4.16}$$

If genie-aided feeback (channel coefficients are exactly known at the transmitter) is available, the optimal beamformer $\mathbf{b}_n$ is known to be $\mathbf{b}_n = \mathbf{h}_0^*$. In practice, however, only a finite number of bits are available for feedback. With F bits of feedback, that is, codebook size of 2^F, using a Grassmannian codebook, the SNR degradation compared to the ideal feedback with N transmit antennas is [4], $|\mathbf{h}_{00}|^2\left(1 - \left(\frac{1}{F}\right)^{\frac{1}{N-1}}\right)$ for some constant c. Thus, from (4.14), the SIR for T_0 to R_0 communication with F bits of feedback is

$$\mathsf{SIR}_{TR} = \frac{d^{-\alpha}|\mathbf{h}_{00}|^2\left(1 - c\left(\frac{1}{F}\right)^{\frac{1}{N-1}}\right)}{\sum_{T_n \in \Phi\setminus\{T_0\}} d_{T_n}^{-\alpha}|\mathbf{h}_{0n}^{\dagger}\mathbf{b}_0|^2}, \tag{4.17}$$

where $|\mathbf{h}_{00}|^2$ is the signal power if $\mathbf{b}_n = \mathbf{h}_0^*$. The corresponding mutual information from T_0 to R_0 using bandwidth F_{TR} is

$$\mathsf{I}_{TR} = F_{TR}\log\left(1 + \mathsf{SIR}_{TR}\right).$$

Similarly, from (4.15), the SIR for the feedback link from R_0 to T_0 is

$$\mathsf{SIR}_{RT} = \frac{d^{-\alpha}|\mathbf{g}_{00}(1)|^2}{\sum_{T_n \in \Phi\setminus\{T_0\}} d_{R_n}^{-\alpha}\left|\mathbf{g}_{0n}\frac{\mathbf{g}_{00}^{\dagger}}{|\mathbf{g}_{00}|}\right|^2},$$

and thus with bandwidth $F_{\text{total}} - F_{TR}$, the mutual information of the feedback link is

$$\mathsf{I}_{RT} = (F_{\text{total}} - F_{TR}) \log (1 + \mathsf{SIR}_{RT}).$$

Let the target rate for transmission from T_0 to R_0 be B_{TR} bits/sec/Hz, and R_0 to T_0 be F bits/sec/Hz because of F bits/sec/Hz of feedback. Similar to Section 4.2, we define the success probability as the probability that communication in both directions is successful simultaneously, that is,

$$P_{\text{suc}} = \mathbb{P}\left(\mathsf{I}_{TR} \geq B_{TR}, \mathsf{I}_{RT} \geq F\right).$$

Consequently, the two-way transmission capacity with feedback is defined as

$$C_F = \frac{\lambda_F (1-\epsilon) B_{TR}}{F_{\text{total}}} \quad \text{bits/sec/Hz/m}^2,$$

where $\lambda_F = \sup\{\lambda : P_{\text{suc}} \geq 1-\epsilon)\}$ is the largest density for which the outage probability constraint of ϵ is met.

As shown by Fig. 4.2, in a two-way communication model, where the transmitter locations are modeled as a PPP, the interference received in both directions is correlated. Therefore, deriving the success probability in closed form is a hard problem. Thus, similar to Section 4.2 to derive a meaningful insight into the dependence of bandwidth allocation, and feedback bits on two-way transmission capacity, we derive a lower bound on the success probability using the FKG inequality as follows.

From Definition 2.8.7, it is easy to see that the success events in two directions $\{\mathsf{I}_{TR} \geq B_{TR}\}$, and $\{\mathsf{I}_{RT} \geq F\}$, respectively, are decreasing events. Thus, invoking Lemma 2.8.8,

$$P_{\text{suc}} \geq \mathbb{P}\left(\mathsf{I}_{TR} \geq B_{TR}\right) \mathbb{P}\left(\mathsf{I}_{RT} \geq F\right). \tag{4.18}$$

By definition,

$$\begin{aligned}
\mathbb{P}\left(\mathsf{I}_{TR} \geq B_{TR}\right) &= \mathbb{P}\left(F_{TR} \log\left(1 + \mathsf{SIR}_{TR}\right) \geq B_{TR}\right), \\
&\overset{(a)}{=} \mathbb{P}\left(\mathsf{SIR}_{TR} \geq \beta_1\right), \\
&\overset{(b)}{=} \mathbb{P}\left(\frac{d^{-\alpha} |\mathbf{h}_{00}|^2 \left(1 - c\left(\frac{1}{B}\right)^{\frac{1}{N-1}}\right)}{\sum_{T_n \in \Phi \setminus \{T_0\}} d_{T_n}^{-\alpha} |\mathbf{h}_{0n}^{\dagger} \mathbf{b}_0|^2} \geq \beta_1\right), \\
&\overset{(c)}{\geq} \mathbb{P}\left(\frac{d^{-\alpha} |\mathbf{h}_{00}|^2}{\sum_{T_n \in \Phi \setminus \{T_0\}} d_{T_n}^{-\alpha} |\mathbf{h}_{0n}^{\dagger} \mathbf{b}_0|^2} \geq \beta_1 / \kappa\right),
\end{aligned}$$

where (a) follows by defining $\beta_1 = \left(2^{\frac{B_{TR}}{F_{TR}}} - 1\right)$, (b) follows by substituting for SIR_{TR} (4.17), and (c) follows by defining $\kappa = \left(1 - c\left(\frac{1}{B}\right)^{\frac{1}{N-1}}\right)$.

Note that $|\mathbf{h}_{00}|^2 \sim \chi^2(2N)$, since it is the norm of an N length independent Gaussian vector. Moreover, since $\mathbf{h}_{0n}^{\dagger}$ and $\mathbf{b}_0$ are independent, similar to Lemma 3.3.1, $|\mathbf{h}_{0n}^{\dagger} \mathbf{b}_0|^2 \sim \chi^2(2)$. Thus, we can now make use of Theorem 3.3.10, where we obtained lower bound on the success probability

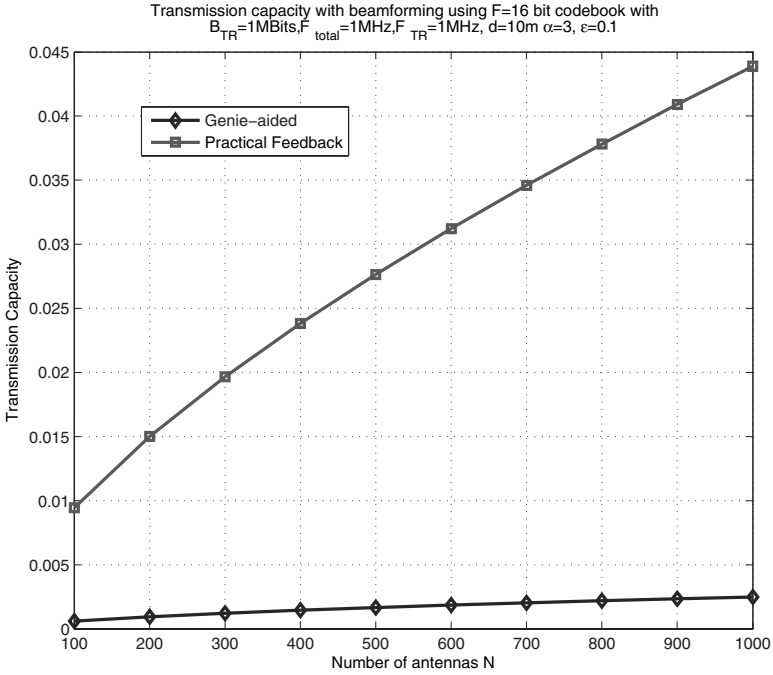

Figure 4.5: Comparison of transmission capacity performance of beamforming with genie-aided and practical feedback as a function of number of transmit antennas N.

with multiple antennas, by specializing for single data stream $k = 1$, and signal power $s = |\mathbf{h}_{00}|^2 \sim \chi^2(2N)$ and interference power $\mathsf{pow}_n = |\mathbf{h}_{0n}^\dagger \mathbf{b}_0|^2 \sim \chi^2(2)$ to get

$$\mathbb{P}\left(\mathsf{I}_{TR} \geq B_{TR}\right) \quad \geq \quad 1 - c\frac{(\beta_1/\kappa)\lambda^{\frac{\alpha}{2}}}{d^{-\alpha}N}, \tag{4.19}$$

where c is a constant.

Similarly,

$$\mathbb{P}\left(\mathsf{I}_{RT} \geq F\right) \quad = \quad \mathbb{P}(\mathsf{SINR}_{RT} \geq \beta_2),$$

where $\beta_2 = \left(2^{\frac{F}{F_{\text{total}} - F_{TR}}}\right)$.

For the $R_0 \rightarrow T_0$ transmission, from (4.18), the signal power $s = |\mathbf{g}_{00}|^2 \sim \chi^2(2N)$, while the interference power is $\mathsf{pow}_n = \left|\mathbf{g}_{0n}\frac{\mathbf{g}_{00}^\dagger}{|\mathbf{g}_{00}|}\right|^2 \sim \chi^2(2)$ from Lemma 3.3.2, since $\mathbf{g}_{0n}$ and $\mathbf{g}_{00}$ are independent. Thus, from Theorem 3.3.10, specializing for single data stream transmission $k = 1$, and no interference cancelation $m = 0$, we get

$$\mathbb{P}\left(\mathsf{I}_{RT} \geq F\right) = 1 - c\frac{\beta_2\lambda^{\frac{\alpha}{2}}}{d^{-\alpha}N}. \tag{4.20}$$

Thus, from (4.19) and (4.20), using (4.18),

$$P_{\text{suc}} \geq 1 - c\lambda^{\frac{\alpha}{2}} \left(\frac{(\beta_1)}{\kappa N} + \frac{\beta_2}{N} \right) + \mathcal{O}(\lambda^{\frac{\alpha^2}{4}}).$$

Let $P_{\text{suc}} = 1 - \epsilon$, then for small λ, the transmission capacity accounting for feedback is

$$C_F \geq \frac{(1-\epsilon)\epsilon N^{\frac{2}{\alpha}}}{cd^2[(\beta_1/\kappa) + \beta_2]^{\frac{1}{\alpha}}} \frac{B_{TR}}{F_{\text{total}}} \text{ bits/sec/Hz/m}^2. \tag{4.21}$$

Thus, as a function of F (the precision of channel coefficient feedback), the transmission capacity (4.21) increases as $F^{\frac{1}{(N-1)\alpha}}$ (using definition of κ) because of the improvement in signal strength, however, decreases as $2^{\frac{-F}{\alpha}}$ (using definition of β_3), because of the stringent requirement of supporting F bits on the feedback link that corresponds to having SIR more than β_2. This result quantifies the degradation in transmission capacity with beamforming due to practical limited feedback, compared to assuming a genie-aided feedback in Section 3.4. The practical feedback requirement not only decreases the available bandwidth for transmitter to receiver communication but also degrades the overall performance due to the successful reception requirement of the feedback bits. In Fig. 4.5, we compare the transmission capacity performance of beamforming with genie-aided feedback and practical finite rate feedback, for total bandwidth $F_{\text{total}} = 1\,\text{MHz}$, out of which $F_{TR} = 1\,\text{MHz}$ is used for data transfer at rate $B_{TR} = 1\,\text{MBits}$, and the rest $F_{\text{total}} - F_{TR} = .1\,\text{MHz}$ is used for feedbacking a $F = 16$ bit codebook. We can see there is massive loss in transmission capacity with the practical feedback strategy.

Similar to Section 4.2.4, for a fixed value of F and B_{TR}, the optimal bandwidth allocation F_{TR} that maximizes the transmission capacity lower bound with beamforming can be computed using Theorem 4.2.4, since here again, the optimization problem is convex. For a fixed value of F_{TR} and B_{TR}, finding the optimal F is slightly complicated since the lower bound is not a convex function of F, however, the problem is a single variable problem and can be solved easily by using techniques like bisection.

4.4 Reference Notes

This chapter is based on results from [2].

Bibliography

[1] T. Cover and J. Thomas. 2004. *Elements of Information Theory*. John Wiley and Sons.

[2] R. Vaze, K. Truong, S. Weber, and R. Heath. 2011. "Two-way transmission capacity of wireless ad-hoc networks." *IEEE Trans. Wireless Commun.* 10 (6): 1966 –75.

[3] D. Love, J. Heath, R.W., and T. Strohmer. 2003. "Grassmannian beamforming for multiple-input multiple-output wireless systems." *IEEE Trans. Inf. Theory* 49 (10): 2735–47.

[4] B. Mondal and R. Heath. 2006. "Performance analysis of quantized beamforming MIMO systems." *IEEE Trans. Signal Process.* 54 (12): 4753–66.

Chapter 5

Performance Analysis of Cellular Networks

5.1 Introduction

In this chapter, we consider cellular wireless networks and apply the tools from stochastic geometry to get some critical insights and almost closed-form results for several important performance measures, such as connection probability, mean rate of communication, call-drop probability, that for long have evaded analytical tractability. Traditionally, these parameters are either computed for very simple and unrealistic models or using large scale Monte–Carlo simulations that are environment-specific.

The breakthrough is made possible because in the modern paradigm, many different types of basestations are overlaid on top of each other, and the overall basestation deployment closely resembles a uniformly random basestation deployment. In the new paradigm, three different types of basestations, the macro basestation, the femto basestation, and the pico basestation, all operate at the same frequency. The location of macro basestations is controlled by cell operators, while the femto and pico basestations are deployed by users in an arbitrary manner with no centralized control over their locations. Thus, the overall basestation locations resembles a uniform deployment model, where the basestations are deployed uniformly at random locations within the area of interest. Modeling cellular network with random basestation locations enables us to explicitly find, for example, the connection probability of any user, the average transmission rate or the call drop probability using stochastic geometry results.

In this chapter, we also address another limitation made for cellular network analysis of modeling the shadowing loss because of blockages (trees or buildings) as a single loss-parameter at the receiver, which is independent of the length of the transmitter–receiver link. Although this assumption helps in the analysis and simulations, it is grossly inaccurate, since shadowing loss is distance-dependent. Larger the distance between the basestation and the mobile user, larger is the number of buildings and trees obstructing the communication. Moreover, the loss is also different for signals from different basestations.

To overcome this limitation, we consider a propagation model that assumes that the blockages are located uniformly randomly in the field of interest. Under this propagation model, together with random basestation locations, we find the distribution of the number of obstructions a typical user

experiences from any basestation. This enables us to find the distribution of the signal-to-interference-plus-noise ratio (SINR) and allows us to derive the connection probability, average transmission rate and so on.

We show that the distance-dependent shadowing model is fundamentally different than the single shadowing parameter model, and many of the conclusions drawn for the latter model do not hold in general. We also show a surprising result that if a user connects to its nearest basestation, then the connection probability of the user with the distance-dependent shadowing model is larger than with the single shadowing parameter model, thus leading to a better transmission rate.

5.2 Random Cellular Network

Traditional deployment of basestations consists of macro basestations that are placed on a regular hexagonal grid by cell operators for maximizing coverage. In addition to macro basestations, in a modern paradigm, wireless networks also include femto and pico basestations, which are low-powered and are deployed by users to improve their coverage and throughput. Since user locations these femto or pico basestations are located arbitrarily in the area of interest. Considered together, the location of these different layers of overlaid basestations resemble more to a *random* deployment of basestations rather than a fixed regular deployment. For example, see Fig. 5.1 (a) and (b), where in Fig. 5.1 (a), we illustrate the locations of a practical basestation deployment that consists of overlaid macro (dots), femto (circles), and pico basestations (squares), while in Fig. 5.1 (b), we show a random basestation deployment with the number of basestations equal to the sum of the three types of basestations in Fig. 5.1 (a). We can see that Fig. 5.1 (a) and (b) resemble each other if we do not

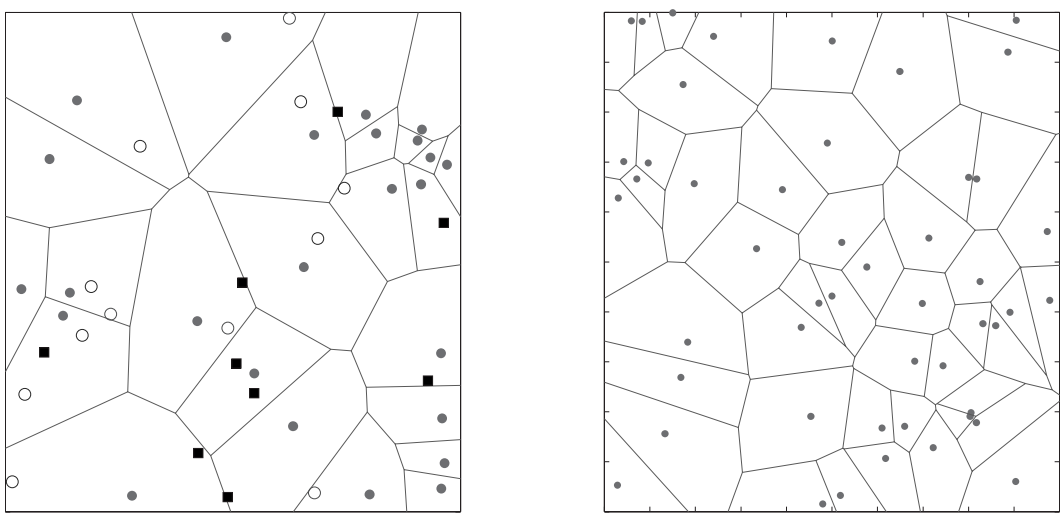

Figure 5.1: (a) Multi-tier wireless network with macro (dots), femto (circles), and pico basestations (squares) versus (b) the random cellular network deployment with identical density. Fig. (a) shows the Voronoi regions of macro basestations, while (b) shows the the Voronoi regions of all basestations combined.

consider the basestation type. Thus, it is reasonable to model these overlaid basestation locations as uniformly randomly distributed in the area of interest.

In this section, we consider a *random* cellular network model that consists of basestations $\{T_n\}$ whose locations are distributed according to a homogeneous Poisson point process (PPP) $\Phi = \{T_n\}$ with density λ, identical to the model considered in Section 2.2. This is reasonable since with a PPP, given the number of nodes lying in the given area, the node locations are uniformly distributed in the given area. In contrast to Section 2.2, however, here the mobile users are located according to an independent PPP with density λ_R, and multiple mobiles can be associated to a single basestation. There are many possible choices for basestation association for each user, such as connecting to the nearest basestation or the basestation with the largest signal power or SINR. We will restrict ourselves to analyzing the most widely used case, where each user associates/connects to its nearest basestation in Φ. Therefore, by the definition of Voronoi regions with respect to basestation locations, all users in a Voronoi region connect to its representative basestation.

As we will see, assuming randomly distributed basestation locations simplifies the analysis, however, it potentially allows the case where two basestations are very close or very far from each. Analysis with some "realistic" correlations among the basesation locations, such as no two are closer than a fixed distance and so on, is also possible using stochastic geometry tools.

Under this random cellular network model, in downlink, the received signal at a typical user u_o located at the origin (without loss of generality) that is connected to its nearest basestation T_0 is given by

$$y = \sqrt{P} d_0^{-\alpha/2} h_0 s_0 + \sum_{m:T_m \in \Phi \setminus \{T_0\}} \sqrt{P} d_m^{-\alpha/2} h_m s_m + w, \tag{5.1}$$

where s_n is signal transmitted from T_n with power P, d_m and h_m are the distance and the channel coefficient between T_m and user u_o, respectively, $\alpha > 2$ is the path-loss exponent, and w is the AWGN with unit variance. As before, we assume that the fading gains h_m are Rayleigh distributed, that is, $|h_m|^2 \sim \exp(1)$.

Let $\mathcal{R}_{\text{near}}$ be the distance of the nearest basestation T_0 from u_o that it connects to. Then conditioning on $\mathcal{R}_{\text{near}} = r$, from (5.1), the SINR at user u_o is

$$\mathsf{SINR} = \frac{P r^{-\alpha} |h_0|^2}{\sum_{m:T_m \in \Phi \setminus \mathbf{B}(u_o,r)} P d_m^{-\alpha} |h_m|^2 + 1}, \tag{5.2}$$

since T_0 is the nearest basestation to u_o and the interference seen at u_o only comes from basestations located outside of disc $\mathbf{B}(u_o, r)$.

The connection probability at user u_o is defined to be the event that the SINR is above a threshold β,

$$P_c(\beta) = \mathbb{P}(\mathsf{SINR} > \beta). \tag{5.3}$$

In the next subsection, we present the first main result of this chapter on computing the connection probability for user u_o. Note that u_0 is a randomly chosen user, so the performance obtained by u_o is identical to any other user.

5.2.1 Connection Probability

To analyze the connection probability (5.3), we need the distribution of the distance $\mathcal{R}_{\text{near}}$ of the nearest basestation T_0 from u_o.

Proposition 5.2.1

$$\mathbb{P}(\mathcal{R}_{\text{near}} > r) = \exp\left(-\lambda\pi r^2\right), \text{ and } f_{\mathcal{R}_{\text{near}}}(r) = 2\pi\lambda r \exp\left(-\lambda\pi r^2\right).$$

Proof: Event $\mathcal{R}_{\text{near}} > r$ is equivalent to having no basestation in the disc $\mathbf{B}(u_o, r)$. Thus, $\mathbb{P}(\mathcal{R}_{\text{near}} > r)$ is equal to the void probability of the basestation location process Φ over the disc $\mathbf{B}(u_o, r)$, which is equal to $\exp\left(-\lambda\pi r^2\right)$, since Φ is a PPP with density λ. □

Next, we derive the connection probability of user u_o using Proposition 5.2.1.

Theorem 5.2.2 *The connection probability of user u_o with PPP-distributed basestation locations of density λ is*

$$P_c(\beta) = \int_0^\infty \pi\lambda \exp\left(-\frac{x^{\alpha/2}\beta}{P}\right) \exp\left(-\pi\lambda x(1 + \mathsf{c}(\beta, \alpha))\,\mathrm{d}x, \right. \tag{5.4}$$

where $\mathsf{c}(\beta, \alpha) = \beta^{2/\alpha} \int_{\beta^{-2/\alpha}}^\infty \frac{1}{1+v^{\alpha/2}}\mathrm{d}v$.

Proof: From the SINR and connection probability definitions (5.2, 5.3),

$$\begin{aligned}
P_c(\beta) &= \mathbb{E}_{\mathcal{R}_{\text{near}}}\left\{\mathbb{P}\left(\mathsf{SINR} > \beta | \mathcal{R}_{\text{near}} = r\right)\right\}, \\
&= \int_r \mathbb{P}\left(\frac{Pr^{-\alpha}|h_0|^2}{\sum_{m:T_m\in\Phi\setminus\mathbf{B}(u_o,r)} Pd_m^{-\alpha}|h_m|^2 + 1} > \beta\right) f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(a)}{=} \int_r \mathbb{E}_{\Phi,h_m}\left\{\exp\left(-\frac{r^\alpha\beta}{P}\left(\sum_{m:\ T_m\in\Phi\setminus\mathbf{B}(u_o,r)} Pd_m^{-\alpha}|h_m|^2 + 1\right)\right)\right\} \\
&\quad f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&= \int_r \exp\left(-\frac{r^\alpha\beta}{P}\right) \\
&\quad \mathbb{E}_{\Phi,h_m}\left\{\exp\left(-\frac{r^\alpha\beta\sum_{m:\ T_m\in\Phi\setminus\mathbf{B}(u_o,r)} Pd_m^{-\alpha}|h_m|^2}{P}\right)\right\} f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(b)}{=} \int_r \exp\left(-\frac{r^\alpha\beta}{P}\right) \mathbb{E}_\Phi\left\{\prod_{x\in\Phi\setminus\mathbf{B}(u_o,r)}\left(\frac{1}{1+r^\alpha\beta x^{-\alpha}}\right)\right\} f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(c)}{=} \int_r \exp\left(-\frac{r^\alpha\beta}{P}\right) \exp\left(-2\pi\lambda\int_{x>r}\left(1 - \frac{1}{1+r^\alpha\beta x^{-\alpha}}\right)x\mathrm{d}x\right) \\
&\quad f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(d)}{=} \int_r \exp\left(-\frac{r^\alpha\beta}{P}\right) \exp\left(-2\pi\lambda\int_{x>r}\left(\frac{\beta}{\beta + \left(\frac{x}{r}\right)^\alpha}\right)x\mathrm{d}x\right) \\
&\quad 2\pi\lambda r\exp\left(-\lambda\pi r^2\right)\mathrm{d}r, \\
&\overset{(e)}{=} \int_r \exp\left(-\frac{r^\alpha\beta}{P}\right) \exp\left(-\pi r^2\lambda\mathsf{c}(\beta,\alpha)\right) 2\pi\lambda r\exp\left(-\lambda\pi r^2\right)\mathrm{d}r, \\
&\overset{(f)}{=} \int_0^\infty \pi\lambda\exp\left(-\frac{x^{\alpha/2}\beta}{P}\right)\exp\left(-\pi\lambda x(1+\mathsf{c}(\beta,\alpha))\,\mathrm{d}x,\right.
\end{aligned}$$

where (a) follows by taking the expectation with respect to $|h_0|^2 \sim \exp(1)$, (b) follows by taking the expectation with respect to $|h_m|^2 \sim \exp(1)$ that are independent $\forall\ m$, (c) follows by using the probability generating function of the PPP (Theorem 2.3.6), in (d) we substitute the PDF of the nearest basestation from Prop. 5.2.1, (e) is obtained by defining $v = \left(\frac{x}{r\beta^{\frac{1}{\alpha}}}\right)^2$ and $\mathsf{c}(\beta,\alpha) = \beta^{2/\alpha}\int_{\beta^{-2/\alpha}}^{\infty} \frac{1}{1+v^{\alpha/2}}\mathrm{d}v$. Finally in (f), we replace $r^2 = x$.

□

Even though the connection probability expression derived in Theorem 5.2.2 is not in absolute closed-form, it can be computed efficiently using numerical integration methods. This is in contrast to the usual practice of finding the connection probability via extensive simulations that are computationally intensive.

For the special case of an interference-limited network, where we can ignore the first exponent term $\exp\left(-\frac{r^{\alpha/2}\beta}{P}\right)$ in (5.4) that corresponds to the AWGN contribution, evaluating the integral in (5.4) we get that

$$P_c(\beta) = \frac{1}{1+\mathsf{c}(\beta,\alpha)}. \tag{5.5}$$

The most important implication of (5.5) is that it shows that the connection probability is independent of the density of basestations λ. This is surprising, since this means that no matter how large the density of basestations is, as long as each mobile user connects to its nearest basestation, it has the same connection probability. This result essentially means that the increase in the interference because of increasing density of basestations is completely nullified by the reduced distance between any user and its nearest basestation. This phenomenon has been known in literature using simulations, but it was mathematically proven only in [1]. Fig. 5.2 depicts this graphically, where we plot the connection probability $P_c(\beta)$ as a function of the basestation density λ.

Note that this independence of connection probability with density λ is in contrast to what we observed in Section 2.4.1, where the outage probability (1-connection probability) was an increasing function of the density λ, since we assumed that each transmitter–receiver pair is at a fixed distance d independent of the density λ. However, when we allow each user to connect to its nearest basestation, essentially in the model of Section 2.4.1, it corresponds to scaling d with λ as $\Theta(\lambda^{-1/2})$, which compensates for the increase in the interference.

In Fig. 5.3, we plot the connection probability $P_c(\beta)$ as a function of the SINR threshold β for $\lambda = 1$ and path-loss exponent $\alpha = 4$, with and without noise, to show that neglecting the effect of noise has very little impact. Finally, in Fig. 5.4, we compare the connection probability for the random wireless network with PPP-distributed basestations and a square grid network, where basestations are located on a square grid with density $\lambda = 1$. We see that the random wireless network lower bounds the connection probability with the grid model, since for a typical receiver, the distances to significant interfering basestations are closer in the random basestation location model than in a square grid model. Decreasing the path-loss parameter α, the gap between the random wireless network and the grid model decreases, since for low values of α, even far off interferers matter and the interference profile with the PPP and square grid is more or less similar.

With the expression for connection probability in hand, we next characterize the mean data rate obtained by any user in the network. Finding mean data rate is important for both the network

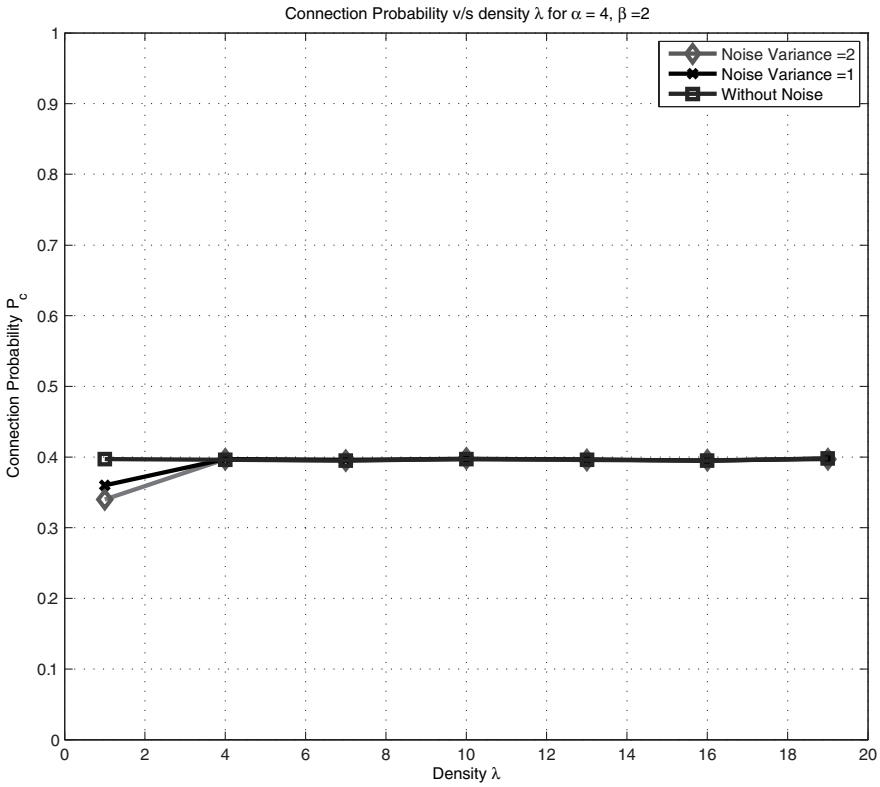

Figure 5.2: Connection probability $P_c(\beta)$ as a function of the basestation density λ. For the no noise case, $P_c(\beta)$ is invariant to λ. With additive noise, $P_c(\beta)$ does depend on λ, but the dependence is very minimal, and no noise assumption is fairly accurate.

operator and the mobile user, since it directly determines the operator's revenue and pricing models and determining utility-payoff tradeoff for each user.

5.2.2 Average Rate per User

Using Shannon's capacity formula, we define the average rate to be

$$\mathbb{E}\{B\} = \mathbb{E}\left\{\log(1+\mathsf{SINR})\right\}, \tag{5.6}$$

and derive an expression for $\mathbb{E}\{B\}$ explicitly in Theorem 5.2.4.

Remark 5.2.3 *If we insist on the user being connected, that is, with* $\mathsf{SINR} > \beta$*, then the average rate it achieves when it connects to its nearest basestation is defined as*

$$\mathbb{E}\{B_c\} = P_c(\beta)\,\mathbb{E}\left\{\log(1+\mathsf{SINR})|\mathsf{SINR}>\beta\right\}.$$

We can find $\mathbb{E}\{B_c\}$ *similar to Theorem 5.2.4.*

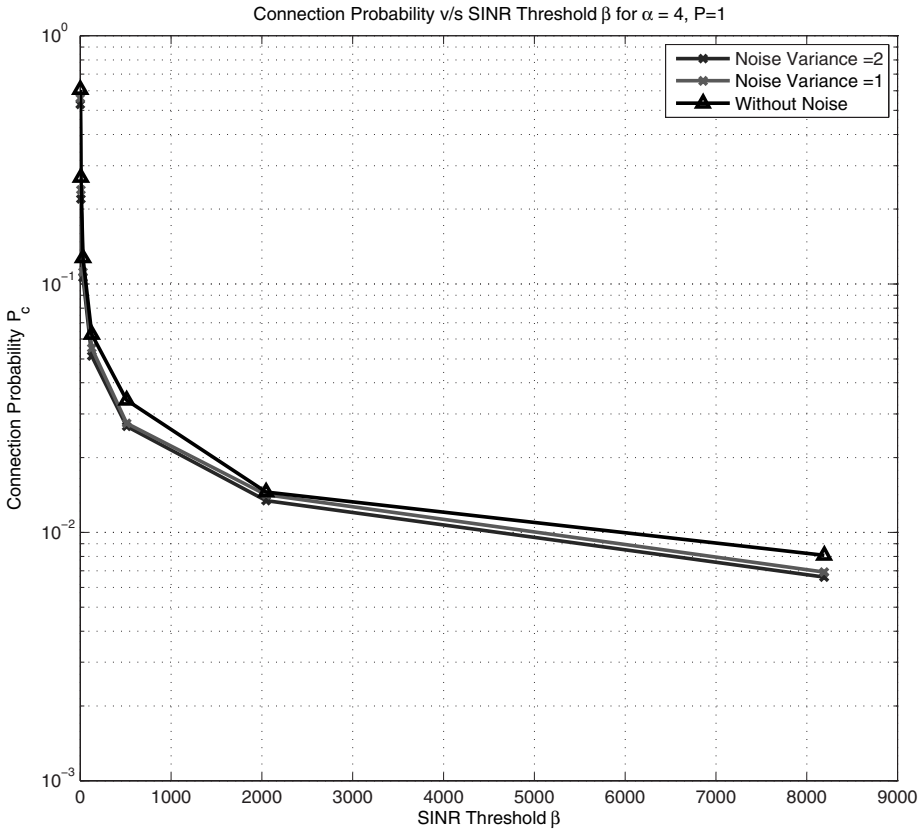

Figure 5.3: Connection probability $P_c(\beta)$ as a function of SINR threshold β for $\lambda = 1$ and path-loss exponent $\alpha = 4$.

Theorem 5.2.4 *The average rate each user gets by connecting to its nearest basestation is*

$$\begin{aligned}\mathbb{E}\{B\} &= \int_r \int_t \exp\left(-\pi\lambda r^2(2^t-1)^{2/\alpha}\int_{(2^t-1)^{-2/\alpha}}^{\infty}\frac{1}{1+x^{\alpha/2}}\mathrm{d}x\right)\\ &\quad \exp\left(-r^{\alpha}(1/P)(2^t-1)\right)\mathrm{d}t f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r. \end{aligned} \tag{5.7}$$

Proof: From the average rate definition (5.6) and SINR definition (5.2),

$$\begin{aligned}\mathbb{E}\{B\} &= \mathbb{E}\left\{\log_2(1+\mathsf{SINR})\right\},\\ &= \int_r \mathbb{E}\left\{\log_2\left(1+\frac{Pr^{-\alpha}|h_0|^2}{\sum_{m:T_m\in\Phi\setminus\mathbf{B}(u_o,r)}Pd_m^{-\alpha}|h_m|^2+1}\right)\right\} f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r,\\ &\overset{(a)}{=} \int_r\int_t \mathbb{P}\left(\log\left(1+\frac{Pr^{-\alpha}|h_0|^2}{\sum_{m:T_m\in\Phi\setminus\mathbf{B}(u_o,r)}Pd_m^{-\alpha}|h_m|^2+1}\right)>t\right)\mathrm{d}t\\ &\quad f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r,\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=} \int_r \int_t \mathbb{P}\left(|h_0|^2 > r^\alpha \left(I_r + \frac{1}{P}\right)(2^t-1)\right) dt f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(c)}{=} \int_r \int_t \mathbb{E}\left\{\exp\left(-r^\alpha \left(I_r + \frac{1}{P}\right)(2^t-1)\right)\right\} \mathrm{d}t f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&= \int_r \int_t \mathbb{E}\left\{\exp\left(-r^\alpha (2^t-1) I_r\right)\right\} \exp\left(-\frac{r^\alpha (2^t-1)}{P}\right) \mathrm{d}t f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r, \\
&\overset{(d)}{=} \int_r \int_t \exp\left(-\pi\lambda r^2 (2^t-1)^{2/\alpha} \int_{(2^t-1)^{-2/\alpha}}^{\infty} \frac{1}{1+x^{\alpha/2}}\mathrm{d}x\right) \\
&\quad \exp\left(-\frac{r^\alpha (2^t-1)}{P}\right) \mathrm{d}t f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r,
\end{aligned}
$$

where (a) follows by the equivalent definition of expectation $\mathbb{E}\{X\} = \int \mathbb{P}(X > t)dt$, (b) follows by defining $I_r = \sum_{m:T_m \in \Phi \setminus \mathbf{B}(u_o,r)} d_m^{-\alpha} |h_m|^2$, (c) follows by taking the expectation with respect to $|h_0|^2 \sim \exp(1)$, and finally (d) follows similarly to finding $\mathbb{E}\left\{\exp\left(-r^\alpha (2^t-1) I_r\right)\right\}$ in the proof of Theorem 5.2.2. □

Again considering the special case of interference-limited network, where we can ignore the last exponent term $\exp\left(-\frac{r^\alpha (2^t-1)}{P}\right)$ in (5.7) that corresponds to the AWGN contribution, computing the integrals, the average rate

$$
\mathbb{E}\{B\} = \int_t \frac{1}{1 + (\exp(t)-1)^{2/\alpha} \int_{(\exp(t)-1)^{-2/\alpha}}^{\infty} \frac{1}{1+x^{\alpha/2}}\mathrm{d}x} \mathrm{d}t,
$$

which most importantly is independent of the density of basestations λ. Hence, the per-user rate is not affected by the density of the basestations. Thus, if there are a large number of users in the system, each basestation has at least one user in its Voronoi cell, and the overall rate or the network capacity scales linearly with λ, since the per-user rate is constant with respect to λ.

The average rate expression derived in Theorem 5.2.4 involves computation of three integrations but is still a significant reduction in complexity compared to large-scale simulations. Also, it gives the true average and not the empirical average (simulation output). Thus, Theorem 5.2.4 is a great asset in the hands of network designers as it can be used for efficient deployment.

One limitation of the analysis presented in this section is that we have ignored the distance-dependent shadowing effects over the wireless link. One expects that larger the length between the basestation and the user, weaker is the received signal because of larger number of blockages (trees/buildings) encountered. Typically, to incorporate shadowing losses, a single fading coefficient is multiplied to the sum of the received signals from different basestations, which is assumed to be log-normal distributed instead of Rayleigh fading, without any dependence on the distance between the basestation and the mobile user. Even though we have considered a Rayleigh fading channel, the results of this section remain more or less the same even with a single log-normal distributed fading model. The real challenge is to incorporate distance-independent shadowing losses. In the next section, we extend the network model to allow for distance-dependent shadowing losses and show that there is a fundamental difference in terms of connection probability and average rate for each user in a random wireless network with and without the distance-dependent shadowing model.

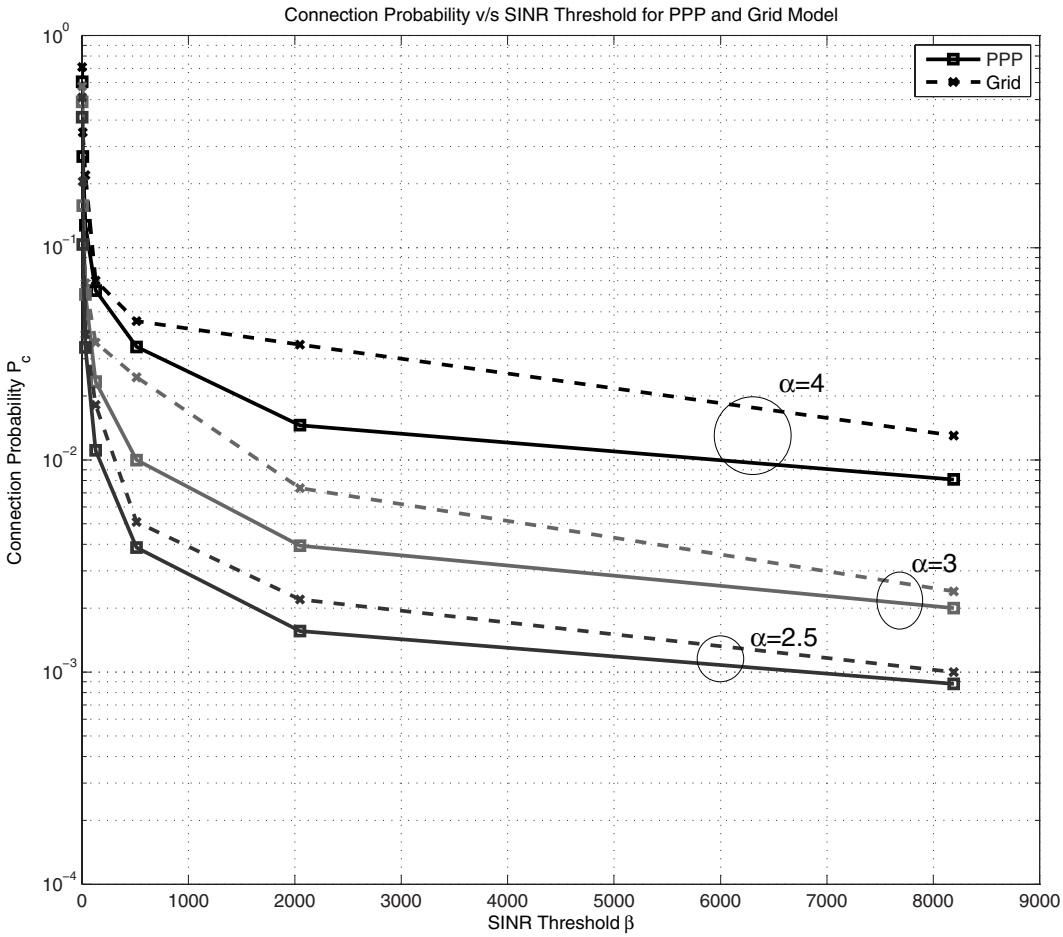

Figure 5.4: Comparing the connection probability $P_c(\beta)$ as a function of SINR threshold β for $\lambda = 1$ and path-loss exponent $\alpha = 4$ for the random wireless network and a square grid network.

5.3 Distance-Dependent Shadowing Model

Similar to Section 5.2, in this section, we assume a random cellular network, where the basestation locations are modeled by a PPP with density λ. In addition, to model the distance-dependent shadowing losses, we let the locations of blockages (buildings/trees) to be also distributed as a PPP with density μ, where each blockage is assumed to have i.i.d. length, breadth, and height. Assuming PPP distribution on the blockage locations corresponds to assuming that blockages are located uniformly randomly in a given area. Even though blockages are not always uniformly located in an urban environment, however, assuming uniformly random locations is a useful abstraction to begin with. Under this model, we will present an exact analysis of the connection probability and the average rate for a typical user located at the origin.

In this section, buildings/blockages in an urban area are modeled as a process of cuboids with random length, width, height, and orientations, whose centers form a PPP. Alternative approaches

to model the effect of blockages on wireless propagation include using ray tracing to perform site-specific simulations [2, 3] or using a grid model for blockages, where each grid point is occupied by a blockage with a certain probability [4–6]. The first approach is cumbersome, while the second one is more suited for Manhattan-type geographies and not for general urban wireless propagation scenario.

5.3.1 Cellular Network Model with Randomly Located Blockages

We consider a random cellular network, defined in Section 5.2, where basestation locations $\{T_n\}$ are distributed according to a PPP with density λ. We model blockages using a Boolean scheme of random cuboids (to model buildings), which is defined as follows.

Definition 5.3.1 *Let $\Phi_B \in \mathbb{R}^2$ be a PPP with density μ. Let $\mathcal{S}$ be a collection of cuboids with i.i.d. lengths L_k, widths W_k, and heights H_k, with PDF $f_L(x)$, $f_W(x)$, and $f_H(x)$, respectively. A Boolean scheme is defined by independently sampling cuboids from $\mathcal{S}$ and placing the centers $\mathbf{C}_k$[1] of these cuboids in $\mathbb{R}^2$ at points generated by the point process Φ_B, where the orientation O_k of each cuboid is distributed over $[0, 2\pi)$ with PDF $f_{\mathsf{O}}(x)$. Thus, a Boolean scheme is represented by $\{\mathbf{C}_k, L_k, W_k, H_k, \mathsf{O}_k\}$.*

We let the blockage process to be a cuboid Boolean scheme

$$\mathbf{\Sigma} = \{\mathbf{C}_k, L_k, W_k, H_k, \mathsf{O}_k\}\,.$$

We consider a typical mobile user u_o, located at the origin and denote the link between basestation T_n to the typical user u_o as $\mathcal{L}_n$, and the distance between them to be $|\mathcal{L}_n| = d_n$.

For ease of exposition, we only consider a rectangle Boolean scheme

$$\mathbf{\Sigma}_{\text{Rect}} = \{\mathbf{C}_k, L_k, W_k, \mathsf{O}_k\},$$

where height $H_k = 0$ for all blockages. Heights can also be incorporated in the model for which case the results are only scaled by a parameter that depends only on the distribution of heights [7]. In Fig. 5.5, we illustrate a typical network realization with randomly located rectangular blockages, where we can see that the nearest basestation of any receiver need not not have the largest signal power.

Definition 5.3.2 *A location $\mathbf{Y} \in \mathbb{R}^2$ is defined to be indoor or contained by a blockage if there exists a blockage $D \in \{\mathbf{C}_k, L_k, W_k, \mathsf{O}_k\}$, such that $\mathbf{Y} \in D$.*

Assumption 5.3.3 *We model the penetration power loss caused by any blockage b as $\mathsf{p} \in [0, 1]$, irrespective of its shape, size, or any other specific property. Thus, the signal power is $\mathsf{p}P$ after crossing one blockage if transmit power was P. Results could be easily generalized by assuming a distribution for p, and taking the expectation of the results derived with fixed p.*

Definition 5.3.4 *The special case of $\mathsf{p} = 0$, that is, no signal can pass through any blockage is defined as the impenetrable case.*

[1] The center does not necessarily have to be the geographic center of the object: any well defined point will suffice.

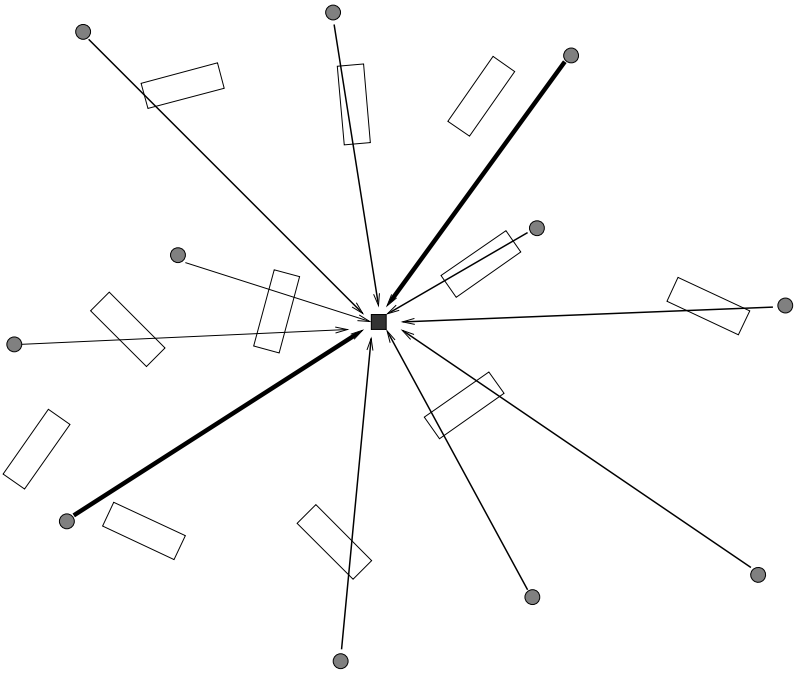

Figure 5.5: Circle nodes are basestations and the square node is the receiver. The blockage process is described by randomly oriented rectangles, and the thickness of the line between basestations and receiver indicates the relative signal strength at the receiver that is inversely proportional to the number of blockages crossing the link.

The impenetrable case is specially useful for obtaining closed-form results that allow us to draw critical insights.

Let b_n be the number of blockages on link $\mathcal{L}_n$ between basestation T_n and user u_o. Then, in downlink, the received signal at a typical user u_o is given by

$$y = d_n^{-\alpha/2}\sqrt{P}\sqrt{\mathsf{p}}^{\mathsf{b}_n}h_n s_n + \sum_{m:T_m\in\Phi\setminus\{T_n\}} \sqrt{P}\sqrt{\mathsf{p}}^{\mathsf{b}_m} d_m^{-\alpha/2}h_m s_m, \tag{5.8}$$

where s_n is signal transmitted from T_n, d_m and h_m are the distance and the channel coefficient between T_m and user u_o, respectively. For ease of exposition, we have assumed an interference-limited network and ignored the AWGN contribution. This is also reasonable since noise has a very small impact on the performance as shown in Fig. 5.3. We will once again assume that each fading gain h_m is i.i.d. and Rayleigh distributed.

Thus, the SIR at user u_o, when it is connected to basestation T_n is

$$\mathsf{SIR}_n \;=\; \frac{|h_n|^2\mathsf{p}^{\mathsf{b}_n}d_n^{-\alpha}}{\sum_{m:T_m\in\Phi,m\neq n}|h_\ell|^2\mathsf{p}^{\mathsf{b}_m}d_m^{-\alpha}}. \tag{5.9}$$

Similar to Section 5.2, we are interested in finding the connection probability $\mathbb{P}(\mathsf{SIR}>\beta)$ and mean achievable rate with this distance-dependent blockage model. Toward that end, we first derive the distribution of the number of blockages on link $\mathcal{L}_n$, b_n, as follows.

5.3.2 Distribution of the Number of Blockages on link $\mathcal{L}_n$

In this section, we show that the number of blockages b_n on link $\mathcal{L}_n$ with length d is a Poisson random variable for any n with mean proportional to d. For simplicity, we drop the index n from b_n and other variables, since we are working with only one link $\mathcal{L}_n$.

Define a sub-collection of blockages

$$\Sigma(\mathbf{C}_k, \ell, w, \theta) \subset \mathbf{\Sigma}_{\text{Rect}},$$

that consists of all blockages with lengths $(\ell, \ell+\mathrm{d}\ell)$, widths $(w, w+\mathrm{d}w)$, and orientations $(\theta, \theta+\mathrm{d}\theta)$. $\Sigma(\mathbf{C}_k, \ell, w, \theta)$ can be equivalently defined as a new blockage process

$$\{\mathbf{C}_k, L_k, W_k, \mathsf{O}_k\},$$

where centers $\mathbf{C}_k \in \Phi_B(\ell, w, \theta)$, and $\Phi_B(\ell, w, \theta) \subset \Phi_B$ is a thinned version of the PPP Φ_B, where each center/point of PPP Φ_B is retained if its blockage has length $(\ell, \ell + \mathrm{d}\ell)$, width $(w, w + \mathrm{d}w)$, and orientation $(\theta, \theta + \mathrm{d}\theta)$ and dropped otherwise. Note that $\Phi_B(\ell, w, \theta)$ is also a PPP since the thinning is random as $\{L_k\}$, $\{W_k\}$, and $\{\mathsf{O}_k\}$ are i.i.d. random variables. Moreover, if $(\ell_1, w_1, \theta_1) \neq (\ell_2, w_2, \theta_2)$, then $\Phi_B(\ell_1, w_1, \theta_1)$ and $\Phi_B(\ell_2, w_2, \theta_2)$ are independent processes.

We summarize this in the next Lemma.

Lemma 5.3.5 *$\Phi_B(\ell, w, \theta)$ is a PPP with density*

$$\mu_{\ell,w,\theta} = \mu f_L(\ell)\mathrm{d}\ell f_W(w)\mathrm{d}w f_{\mathsf{O}}(\theta)\mathrm{d}\theta,$$

and, if $(\ell_1, w_1, \theta_1) \neq (\ell_2, w_2, \theta_2)$, then $\Phi(\ell_1, w_1, \theta_1)$ and $\Phi(\ell_2, w_2, \theta_2)$ are independent PPPs.

Definition 5.3.6 *Let $J_d(\ell, w, \theta)$ be the number of blockages which belong to the subset $\Sigma(\mathbf{C}_k, \ell, w, \theta)$ and cross the link $\mathcal{L}$ of length d.*

Lemma 5.3.7 *$J_d(\ell, w, \theta)$ is a Poisson random variable with mean*

$$\mathbb{E}[J_d(\ell, w, \theta)] = \mu_{\ell,w,\theta}(d(\ell|\sin\theta| + w|\cos\theta|) + \ell w),$$

where d is the length of the link $\mathcal{L}$.

Proof: We use the alternate characterization of sub-collection of blockages $\Sigma(\mathbf{C}_k, \ell, w, \theta)$ as a blockage process

$$\{\mathbf{C}_k, L_k, W_k, \mathsf{O}_k\},$$

where centers $\mathbf{C}_k \in \Phi_B(\ell, w, \theta)$. As shown in Fig. 5.6, a rectangle belonging to $\{\mathbf{C}_k, L_k, W_k, \mathsf{O}_k\}$ intersects the link $\mathcal{L}$ if and only if its center belonging to $\Phi_B(\ell, w, \theta)$ falls in the region defined by vertices ABCDEF. Note that each vertex is the center of the rectangles with width w, length ℓ, and orientation θ. Hence, $J_d(\ell, w, \theta)$ equals the number of points of $\Phi_B(\ell, w, \theta)$ falling in the region ABCDEF. Let the area of region ABCDEF be $\nu(\ell, w, \theta)$, then

$$\begin{aligned}
\nu(\ell, w, \theta) &= d\,|\sin(\phi + \theta)|\sqrt{w^2 + \ell^2} + w\ell, \\
&= d\left(|\sin(\theta)|\frac{\ell}{\sqrt{w^2 + \ell^2}} + |\cos(\theta)|\frac{w}{\sqrt{w^2 + \ell^2}}\right)\sqrt{w^2 + \ell^2} + w\ell,
\end{aligned}$$

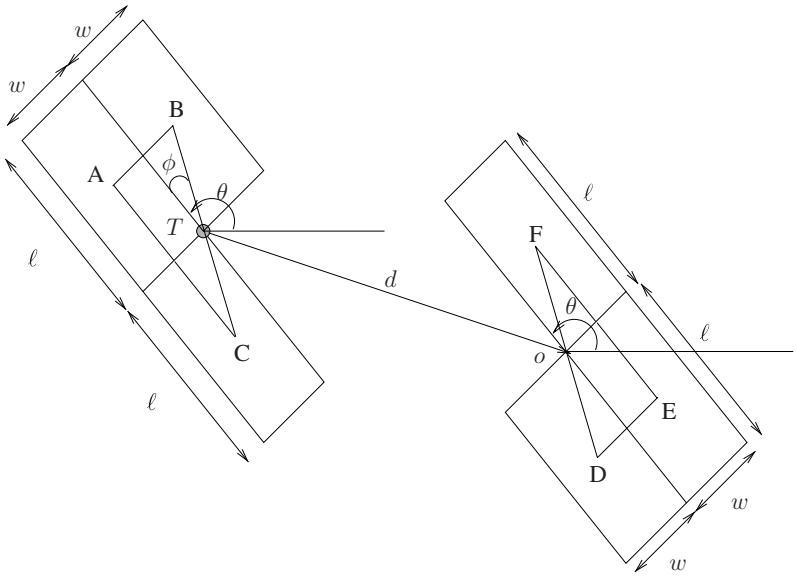

Figure 5.6: Link $\mathcal{L}$ of length d between basestation T and mobile user at o. Any rectangle of $\Sigma(\mathbf{C}_k, \ell, w, \theta)$ intersects $\mathcal{L}_n$ only if its center lies in the region defined by vertices $ABCDEF$, where each vertex is the center of the six rectangles of length ℓ and width w.

$$= \quad d\,(\ell|\sin\theta| + w|\cos\theta|) + \ell w,$$

where θ and ϕ are the angles depicted in Fig. 5.6. By Lemma 5.3.5, $\Phi_B(\ell, w, \theta)$ is a PPP of density $\mu_{\ell,w,\theta}$, thus, the number of points of $\Phi_B(\ell, w, \theta)$ falling in the region ABCDEF is a Poisson random variable with mean $\mu_{\ell,w,\theta}\ \nu(\ell, w, \theta)$. Consequently, $J_d(\ell, w, \theta)$ is a Poisson variable with mean

$$\begin{aligned}\mathbb{E}[J_d(\ell, w, \theta)] &= \mu_{\ell,w,\theta}\ \nu(\ell, w, \theta) \\ &= \mu_{\ell,w,\theta} \times (d(\ell|\sin\theta| + w|\cos\theta|) + \ell w).\end{aligned}$$ □

Recall that b, the total number of blockages crossing the link $\mathcal{L}$, is $\mathsf{b} = \int_{\ell,w,\theta} J_d(\ell, w, \theta)$. Using Lemma 5.3.7, we calculate the distribution of b in the following theorem.

Theorem 5.3.8 *The number of blockages* b *on a link* $\mathcal{L}$ *of length* d *is a Poisson random variable with the mean* $\mathsf{L}_b d + \mu\nu_b$*, where* $\mathsf{L}_b = \frac{2\mu(\mathbb{E}[W]+\mathbb{E}[L])}{\pi}$*, and* $\nu_b = \mathbb{E}[L]\mathbb{E}[W]$ *is the expected area of any blockage.*

Proof: By Lemma 5.3.5 and Lemma 5.3.7, $J_d(\ell, w, \theta)$ are independent Poisson random variables for different values of the tuple (ℓ, w, θ). By definition, the total number of blockages on any link of length d is

$$\mathsf{b} = \int_{\ell,w,\theta} J_d(\ell, w, \theta).$$

Since the superposition of independent Poisson random variables is also Poisson distributed, b is a Poisson distributed random variable, since each $J_d(\ell, w, \theta)$ is independent and Poisson distributed.

The expectation of b can be computed as follows.

$$\begin{aligned}\mathbb{E}\{\mathsf{b}\} &= \mathbb{E}\left\{\int_{\ell,w,\theta} J_d(\ell,w,\theta)\right\},\\ &= \int_L\int_W\int_{\mathsf{O}} \mu\left(d(\ell|\sin\theta|+w|\cos\theta|)+\ell w\right)\, f_L(\ell)\mathrm{d}\ell\, f_W(w)\mathrm{d}w\, \frac{1}{2\pi}\mathrm{d}\theta,\\ &= \frac{2\mu(\mathbb{E}[L]+\mathbb{E}[W])}{\pi}d+\mu\mathbb{E}[L]\mathbb{E}[W],\\ &= \mathsf{L}_b d+\mu\nu_b,\end{aligned}$$

where $\mathsf{L}_b = \frac{2\mu(\mathbb{E}[L]+\mathbb{E}[W])}{\pi}$ and $\nu_b = \mathbb{E}[L]\mathbb{E}[W]$. □

Thus, we get the intuitive result that the average number of blockages on a link is proportional to the length of the link d, since longer the link, more blockages are likely to appear on that link. The two parameters L_b and ν_b correspond to the expected perimeter of any blockage times the blockage density divided by π, and the expected area occupied by any blockage, respectively.

Next, we present some simple but useful corollaries of Theorem 5.3.8 as follows.

Corollary 5.3.9 *The probability that a link of length d admits line-of-sight propagation, that is, no blockages cross the link, is $\mathbb{P}(\mathsf{b}=0) = \exp\left(-(\mathsf{L}_b d+\mu\nu_b)\right)$.*

We can also evaluate the probability that a user is located inside a blockage/building in the following lemma.

Corollary 5.3.10 *The probability that a location in $\mathbb{R}^2$ is contained by a blockage is $1-\exp\left(-\mu\nu_b\right) \approx \mu\nu_b$, for small $\mu\nu_b$.*

Proof: A location is contained in a blockage if there is at least one blockage on link $\mathcal{L}$ no matter how small the length d of link $\mathcal{L}$. For a link with length d, $\mathbb{P}(\mathsf{b}>0) = 1-\exp\left(-(\mathsf{L}_b d+\mu\nu_b)\right)$ from Corollary 5.3.9. Thus, taking $d\to 0$, we get that the probability that a location in $\mathbb{R}^2$ is contained by a blockage is $1-\exp\left(-\mu\nu_b\right) \approx \mu\nu_b$, for small $\mu\nu_b$. □

After deriving the distribution of the number of blockages on any link, in the next subsection, we derive the connection probability of a random cellular network under the realistic distance-dependent blockage/shadowing model.

5.3.3 Connection Probability

In this section, we analyze the connection probability with the distance-dependent blockage model. Recall that the connection probability is defined as

$$P_c(\beta) = \mathbb{P}(\mathsf{SIR}>\beta). \tag{5.10}$$

We next consider the *impenetrable blockage* case, where $\mathsf{p}=0$, that is, no signal penetrates any blockage. The results presented next can easily be extended for $\mathsf{p}\in(0,1)$, but this specific case helps in getting closed-form expression and design insights.

Definition 5.3.11 *We define a basestation T_n to be visible ($V_n=1$) to the typical user u_o, if the number of blockages $\mathsf{b}_n=0$ on direct link $\mathcal{L}_n$ between T_n and u_o, and $V_n=0$ otherwise.*

As in Section 5.2, each mobile follows the nearest basestation association rule, where it connects to its nearest *visible* basestation, if there is any. If the distance of the nearest visible basestation T_n is $\mathcal{R}_{\text{near}} = r$ from u_o, then from (5.11), the SIR at u_o is

$$\mathsf{SIR} \quad = \quad \frac{|h_n|^2 r^{-\alpha}}{\sum_{m:T_m \in \Phi, m \neq n, d_m > r} |h_\ell|^2 d_m^{-\alpha} V_m}, \tag{5.11}$$

where V_m is a Bernoulli random variable, which takes value 1 if basestation T_m is visible to u_o, and 0 otherwise.

Assumption 5.3.12 *We make the assumption that the number of blockages on links $\mathcal{L}_n$ between u_o and basestations T_n are independent. Thus, Bernoulli random variables V_m are independent. Clearly, this is not true in general, since nearby basestations are likely to have similar number of blockages, however, this assumption serves as a good abstraction and helps in understanding the fundamental performance measures of this general blockage model. Simulation results also suggest that this independence assumption is not really limiting.*

To analyze the connection probability $P_c(\beta) = \mathbb{P}(\mathsf{SIR} > \beta)$, we next find the distribution of the distance of the nearest visible basestation T_n from u_o. We already know the distribution of V_m from Corollary 5.3.9.

Distribution of the distance of the nearest visible basestation T_n from u_o

Theorem 5.3.13 *Assuming that the numbers of blockages b_n on different links $\mathcal{L}_n$ from basestation T_n to u_o are independent, the distribution of the distance $\mathcal{R}_{near}$ to the nearest visible basestation T_0 is*

$$\mathbb{P}(\mathcal{R}_{near} > r) = \exp(-2\pi\lambda U(r)), \tag{5.12}$$

where $U(r) = \frac{\exp(-\mu\nu_b)}{\mathsf{L}_b^2}\left[1 - (\mathsf{L}_b r + 1)\exp\left(-\mathsf{L}_b r\right)\right]$, and L_b and ν_b are defined in Theorem 5.3.8.

Proof: The distance $\mathcal{R}_{\text{near}}$ from u_o to the nearest visible basestation is larger than r if and only if all the basestations located within the ball $\mathbf{B}(u_o, r)$ are not visible to u_o. Since the basestations form a PPP of density λ, it follows that

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_{\text{near}} > r) =& \mathbb{P}\left(\text{all basestations in } \mathbf{B}(u_o, r) \text{ are not visible}\right), \\
=& \mathbb{E}_k\left\{\mathbb{P}\left(\text{all basestations in } \mathbf{B}(u_o, r) \text{ are not visible}\right.\right. \\
& \left.\left. \left| \# \text{ basestations in } \mathbf{B}(u_o, r) = k\right)\right\}, \right. \\
\overset{(a)}{=}& \sum_{k=0}^{\infty} \left(\mathbb{P}\left(\text{any one basestation in } \mathbf{B}(u_o, r) \text{ is not visible}\right)\right)^k \\
& \frac{\exp\left(-\lambda\pi r^2\right)\left(\lambda\pi r^2\right)^k}{k!},
\end{aligned} \tag{5.13}$$

where (a) follows from the fact that the number of basestations in $\mathbf{B}(u_o, r)$ are Poisson distributed with mean $\lambda\pi r^2$, and the assumption that V_m are independent.

Next, we compute the $\mathbb{P}\left(\text{any one basestations in } \mathbf{B}(u_o, r) \text{ is not visible}\right)$ as follows. Let basestation $T_n \in \mathbf{B}(u_o, r)$ be at a distance of $d_n \leq r$ from the origin. Since basestation process Φ is a PPP, the location of basestation T_n is uniformly distributed in $\mathbf{B}(u_o, r)$. Thus, d_n^2 is uniformly distributed between $[0, r^2]$, and PDF of d_n is $f_{d_n}(t) = \frac{2t}{r^2}$. With basestation T_n at distance of $d_n = t$ from the origin, from Corollary 5.3.9, we know that

$$\mathbb{P}\left(\text{basestation } T_n \text{ at distance } d_n = t \text{ in } \mathbf{B}(u_o, r) \text{ is not visible}\right) = \left(1 - \exp\left(-(\mathsf{L}_b t + \mu\nu_b)\right)\right).$$

Taking the expectation with respect to d_n, we get

$$\mathbb{P}\left(\text{basestation } T_n \text{ at distance } d_n = t \text{ in } \mathbf{B}(u_o, r) \text{ is not visible}\right) = \int_0^r \left(1 - \exp\left(-(\mathsf{L}_b t + \mu\nu_b)\right)\right) \frac{2t}{r^2} \mathrm{d}t. \tag{5.14}$$

Thus, using (5.14), from (5.13),

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_{\text{near}} > r) &= \sum_{k=0}^{\infty} \left(\int_0^r \left(1 - \exp\left(-(\mathsf{L}_b t + \mu\nu_b)\right)\right) \frac{2t}{r^2} \mathrm{d}t \right)^k \frac{\exp\left(-\lambda\pi r^2\right) (\lambda\pi r^2)^k}{k!}, \\
&\overset{(b)}{=} \sum_{k=0}^{\infty} \left(1 - \frac{2U(r)}{r^2}\right)^k \frac{\exp\left(-\lambda\pi r^2\right) (\mu\pi r^2)^k}{k!}, \\
&= \exp\left(-\lambda\pi r^2 \left(1 - \left(1 - \frac{2U\left(r\right)}{r^2}\right)\right)\right), \\
&= \exp\left(-2\lambda\pi U(r)\right),
\end{aligned}$$

where (b) follows from the definition of $U(r)$. □

The following Corollary is immediate.

Corollary 5.3.14 *The PDF of the distance to the nearest visible basestation from u_o, $\mathcal{R}_{near}$, is*

$$f_{\mathcal{R}_{near}}(r) = 2\pi\lambda r \exp\left(-(\mathsf{L}_b r + \mu\nu_b + 2\pi\lambda U(r))\right).$$

We are now ready to derive the connection probability $P_c(\beta)$ expression (5.10), using (5.11).

Theorem 5.3.15 *In the impenetrable blockage case $\mathsf{p} = 0$, when each user connects to its nearest visible basestation, the connection probability for any user is*

$$P_c(\beta) = \int_0^{\infty} \exp\left(-2\pi\lambda \int_r^{\infty} \left[\frac{\beta r^{\alpha} \exp\left(-(\mathsf{L}_b t + \mu\nu_b)\right)}{t^{\alpha} + \beta r^{\alpha}}\right] t \mathrm{d}t\right) f_{\mathcal{R}_{near}}(r) \mathrm{d}r, \tag{5.15}$$

where $f_{\mathcal{R}_{near}}(r)$ is the PDF of the distance to the nearest visible basestation derived in Corollary 5.3.14, and L_b and ν_b are defined in Theorem 5.3.8.

Proof: First we compute the expression for the connection probability conditioning on the distance to the nearest visible basestation $\mathcal{R}_{\text{near}}$. Given that $\mathcal{R}_{\text{near}} = r$, the expression of SIR from (5.11) is

$$\mathsf{SIR} = \frac{r^{-\alpha}|h_n|^2}{\sum_{m:T_m \in \Phi, d_m > r, m \neq n} d_m^{-\alpha} |h_m|^2 V_m},$$

where $|h_m|^2$ are Rayleigh distributed, V_m (in the impenetrable case) are independent Bernoulli random variables with parameter $\exp\left(-(\beta d_i + \mu\nu_b)\right)$ (Corollary 5.3.9), modeling whether the basestation m is visible or not. Conditioned on the nearest visible basestation to be at distance $\mathcal{R}_{\text{near}} = r$, the connection probability can be computed as follows $\mathbb{P}\left(\mathsf{SIR} > \beta | \mathcal{R}_{\text{near}} = r\right)$

$$= \mathbb{P}\left(|h_n|^2 > \beta r^{\alpha} \left(\sum_{m:T_m \in \Phi, d_m > r} d_m^{-\alpha} |h_m|^2 V_m \right) \right),$$

$$\stackrel{(a)}{=} \mathbb{E}\left\{\exp\left(-\beta r^{\alpha}\sum_{m:T_m\in\Phi,d_m>r} d_m^{-\alpha}|h_m|^2 V_m\right)\right\},$$

$$= \mathbb{E}\left\{\prod_{m:T_m\in\Phi,d_m>r} \mathbb{E}_{V_m,|h_m|^2}\left\{\exp\left(-\beta r^{\alpha} d_m^{-\alpha}|h_m|^2 V_m\right)\right\}\right\},$$

$$\stackrel{(b)}{=} \mathbb{E}\left\{\prod_{m:T_m\in\Phi,d_m>r} \left(\mathbb{E}_{|h_m|^2}\left\{\exp(-\beta r^{\alpha}|h_m|^2 d_m^{-\alpha})\right.\right.\right.$$
$$\left.\left.\left.\exp\left(-(\beta d_m+\mu\nu_b)\right)\right\}+1-\exp\left(-(\beta d_m+\mu\nu_b)\right)\right)\right\},$$

where (a) follows by taking the expectation with respect to $|h_n|^2 \sim \exp(1)$, (b) follows by taking expectation with respect to Bernoulli random variable V_m (Corollary 5.3.9) and noting that V_m are independent by assumption. Taking the expectation with respect to $|h_m|^2 \sim \exp(1)$ that are independent $\forall\, m$, $\mathbb{P}\left(\mathsf{SIR} > \beta | \mathcal{R}_{\text{near}} = r\right)$

$$\stackrel{(c)}{=} \mathbb{E}\left\{\prod_{m:T_m\in\Phi,d_m>r}\left(\frac{\exp\left(-(\beta d_m+p)\right)}{1+\beta r^{\alpha} d_m^{-\alpha}}+1-\exp\left(-(\beta d_i+\mu\nu_b)\right)\right)\right\},$$

$$= \mathbb{E}\left\{\prod_{m:T_m\in\Phi,d_m>r}\left(1-\frac{\beta r^{\alpha}\exp\left(-(\beta d_m+\mu\nu_b)\right)}{d_m^{\alpha}+\beta r^{\alpha}}\right)\right\},$$

$$\stackrel{(d)}{=} \exp\left(-2\pi\lambda\int_r^{\infty}\left[\frac{\beta r^{\alpha}\exp\left(-(\beta t+\mu\nu_b)\right)}{t^{\alpha}+\beta r^{\alpha}}\right]t\mathrm{d}t\right),$$

and finally (d) follows directly from the probability generating functional of the PPP formed by the basestations $\{T_n\}$ Theorem 2.3.6. Hence, the unconditional connection probability is given by

$$P_c(\beta) = \int_r \mathbb{P}(\mathsf{SIR} > \beta | \mathcal{R}_{\text{near}} = r) f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r,$$
$$= \int_0^{\infty} \exp\left(-2\pi\lambda\int_r^{\infty}\left[\frac{\beta r^{\alpha}\exp\left(-(\mathsf{L}_b t+\mu\nu_b)\right)}{t^{\alpha}+\beta r^{\alpha}}\right]t\mathrm{d}t\right) f_{\mathcal{R}_{\text{near}}}(r)\mathrm{d}r. \qquad \square$$

In Theorem 5.3.15, we have derived the connection probability for random cellular networks with a distance-dependent blockage model, generalizing Theorem 5.2.2. The expression derived in Theorem 5.3.15 is not in exact closed form, but it lends itself to fast numerical integration techniques.

Recall that we showed in Theorem 5.2.2 that the connection probability without considering blockages is invariant to the basestation density λ. Thus, the basestation density can be increased without affecting the per-user performance to get larger and larger total network throughput. In contrast to Theorem 5.2.2, Theorem 5.3.15 shows that accounting for more realistic distance-dependent blockages, the connection probability depends on the basestation density. Since the user connects to its nearest basestation, without blockages, the strengths of the desired signal and interference seen at any user scale by a common factor with changing basestation density. In a network with blockages, however, if the user connects to its nearest visible basestation, the scale factors for signal and interference powers are different with changing basestation density, since the

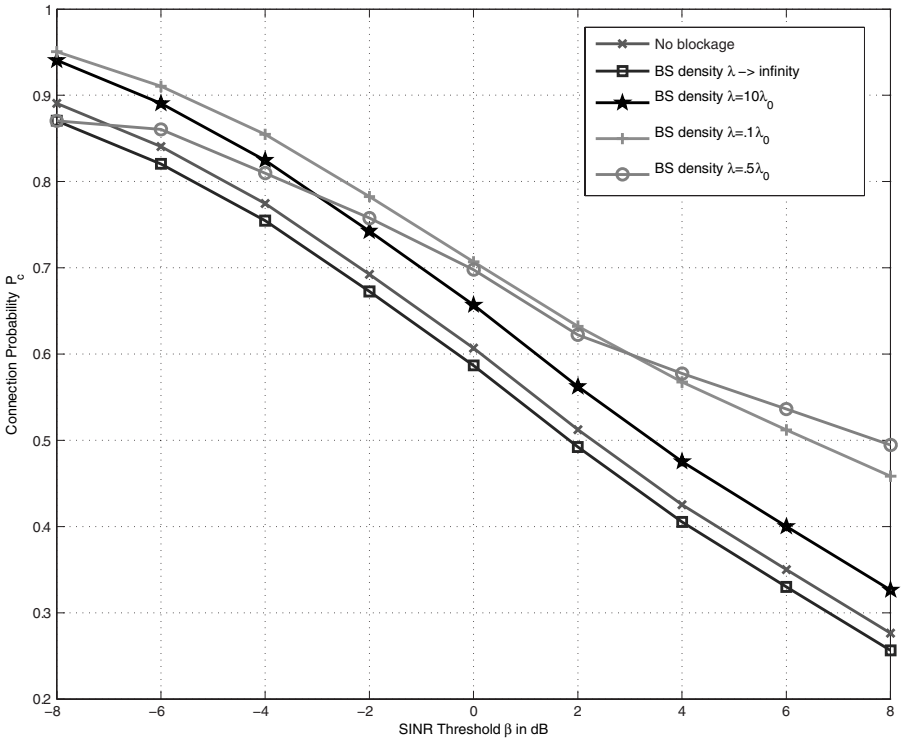

Figure 5.7: Connection probability as a function of the basestation density for $\mathbb{E}[L] = \mathbb{E}[W] = 15\,\text{m}$, $\mu_0 = 4.5 \times 10^{-4}/\text{m}^2$, $\lambda_0 = 3.5 \times 10^{-5}/\text{m}^2$, $\alpha = 4$, and $P = 1$.

interference links (that are at larger distances than the basestation to which the user connects) are likely to experience more blockages on their links on average than the link of the connected basestation. Thus, considering distance-dependent blockages is fundamentally different than the scenario that neglects the presence of blockages.

We show this behavior in Fig. 5.7, where we compare the performance of cellular networks with and without considering blockages. As expected, the connection probability is invariant to the basestation density without considering blockages. While considering blockages, Fig. 5.7 shows an interesting result that including blockages in the system model often helps improve the coverage by increasing the SINR and consequently the connection probability for reasons discussed at the end of previous section.

Another object of interest is the behavior of connection probability in the blockage case as a function of different basestation density λ. We see that the connection probability is not monotone in the basestation density, and one has to find optimum basestation density for a given blockage density either analytically or through simulations. A final point to note from Fig. 5.7 is that when the basestation density goes to infinity, the connection probability converges to the case with no blockages. This is indeed expected, since with increasing basestation density, the effect of blockages in reducing interference decreases and in the limit, the interference profile is similar to that of the no blockage case.

We next quantify the average rate each user gets when it connects to its nearest visible basestation in the distance-dependent blockage model.

5.3.4 Average Rate per User

As in Section 5.2.2, we define the average rate to be $\mathbb{E}\{B\} = \mathbb{E}\left\{\log(1+\mathsf{SINR})\right\}$ which is equal to

$$\mathbb{E}\{B\} \quad = \quad \int_t P_c(2^t-1)\mathrm{d}t, \tag{5.16}$$

by using the fact that $\mathbb{E}\{X\} = \int \mathbb{P}(X > t)dt$.

In Table 5.1, we provide simulation results for average rate, where we notice that considering blockage effects, the average rate is no longer invariant to the basestation density, and the modeling realistic blockages could help increase the achievable rate.

Table 5.1: Average Rate Comparison

Blockage density	None	Low	Intermediate	High
μ_0/λ_0	0	$0.1\mu_0/\lambda_0$	μ_0/λ_0	$10\mu_0/\lambda_0$
Average rate (bits/sec/Hz)	2.15	2.42	4.99	3.14

Note: μ_0/λ_0 is the ratio of the blockage density to the basestation density. We assume $\mathbb{E}[L] = \mathbb{E}[W] = 15\,\mathrm{m}$, $\mu_0 = 4.5 \times 10^{-4}/\mathrm{m}^2$, and $\lambda_0 = 3.5 \times 10^{-5}/\mathrm{m}^2$.

In summary, in this chapter, we have shown that the connection probability and the average rate seen by any user with and without accounting for the distance-dependent blockages are fundamentally different. We also note that the SIR with the distance-based blockage model is typically higher than without considering it, which could appear counter-intuitive at first. This result is well explained by noting that with the distance dependent-blockage model, if each user is connecting to its nearest visible (unblocked) basestation or basestation with least number of blockages, then automatically either the other interfering basestations are invisible or have much larger number of blockages on their paths, thus reducing the interference power. Numerical results also confirm this assertion by showing better connection probabilities for the distance-dependent blockage model. Thus, not considering distance-dependent blockages might actually underestimate the performance of the real-world wireless networks.

5.4 Reference Notes

Traditionally, for performance evaluation in wireless networks, either a simple one-dimensional Wyner model [8, 9], where interference is received from only two nearest basestations, or a hexagonal or square grid model is assumed for basestation locations [4–6]. Clearly, the efficacy of the Wyner model is limited in a realistic two-dimensional network, and the analysis with the grid model is limited to special cases of mobile locations such as those close to cell boundaries, which could underestimate the actual system performance. The idea of using randomly located basestations for analyzing the performance of cellular networks was first proposed in [1], where both the connection probability and the average rate expressions were derived. Extensions to modeling multi-tier wireless cellular networks can be found in [10]. The random cellular network is easily extendable to incorporate more flexibility such as in [11], where a hybrid network model is considered where only the interferers are modeled as a PPP outside a fixed-size cell to characterize the site-specific performance of cells with different sizes, rather than the aggregate performance

metrics of the entire system. The random network model also lends itself for analysis of complicated network performance indicators such as connectivity between all pairs of nodes [12]. The random cellular network model can be rigorously shown to lower bound a hexagonal grid network in terms of certain performance metrics [13].

Modeling blockages as a point process with independent and identically distributed length, breadth, and height were proposed in [7] and extensively studied in [14], where the distribution of number of blockages seen by any link, the SINR distribution, and expressions for connection probability and average rate were derived. More detailed results with this model can be found in [14]. Earlier a grid model [4–6] or ray tracing model [3, 15–17] was proposed to tackle the problem of modeling distance-dependent signal losses in wireless networks.

Bibliography

[1] J. G. Andrews, F. Baccelli, and R. K. Ganti. 2011. "A tractable approach to coverage and rate in cellular networks." *IEEE Trans. Commun.* 59 (11): 3122–34.

[2] K. Rizk, J. Wagen, and F. Gardiol. 1997. "Two-dimensional ray-tracing modeling for propagation prediction in microcellular environments." *IEEE Trans. Veh. Technol.* 46 (2): 508–18.

[3] K. R. Schaubach, N. J. Davis IV, and T. S. Rappaport. 1992. "A ray tracing method for predicting path loss and delay spread in microcellular environments." In *IEEE Vehicular Technology Conference, 1992*. IEEE. 932–35.

[4] M. Franceschetti, J. Bruck, and L. J. Schulman. 2004. "A random walk model of wave propagation." *IEEE Trans. Antennas Propag.* 52 (5): 1304–17.

[5] S. Marano, F. Palmieri, and G. Franceschetti. 1999. "Statistical characterization of ray propagation in a random lattice." *JOSA A*. 16 (10): 2459–64.

[6] S. Marano and M. Franceschetti. 2005. "Ray propagation in a random lattice: A maximum entropy, anomalous diffusion process." *IEEE Trans. Antennas Propag.* 53 (6): 1888–96.

[7] T. Bai, R. Vaze, and R. Heath. 2012. "Using random shape theory to model blockage in random cellular networks." In *International Conference on Signal Processing and Communications (SPCOM), 2012*. IEEE. 1–5.

[8] A. D. Wyner. 1994. "Shannon-theoretic approach to a Gaussian cellular multiple-access channel." *IEEE Trans. Inf. Theory* 40 (6): 1713–27.

[9] S. Shamai and A. D. Wyner. 1997. "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels i & ii." *IEEE Trans. Inf. Theory* (6): 1895–11.

[10] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo *et al.*. 2012. "Heterogeneous cellular networks: From theory to practice." *IEEE Commun. Mag.* 50 (6): 54–64.

[11] R. Heath, M. Kountouris, and T. Bai. 2013. "Modeling heterogeneous network interference using Poisson point processes." *IEEE Trans. Signal Process* 61 (16): 4114–26.

[12] M. Haenggi. 2009. "Outage, local throughput, and capacity of random wireless networks." *IEEE Trans. Wireless Commun.* 8 (8): 4350–59.

[13] T. X. Brown. 2000. “Cellular performance bounds via shotgun cellular systems.” *IEEE J. Sel. Areas Commun.* 18 (11): 2443–55.

[14] T. Bai, R. Vaze, and R. Heath. 2014. “Analysis of blockage effects on urban cellular networks.” *IEEE Trans. Wireless Commun.* [Online]. Available: http://arxiv.org/abs/1309.4141

[15] C. R. Anderson and T. S. Rappaport. 2011. “In-building wideband partition loss measurements at 2.5 and 60 ghz.” *IEEE Trans. Commun.* 3 (3): 922–28.

[16] Z. Pi and F. Khan. 2011. “An introduction to millimeter-wave mobile broadband systems.” *IEEE Commun. Mag.* 49 (6): 101–07.

[17] D. B. Taylor, H. S. Dhillon, T. D. Novlan, and J. G. Andrews. 2012. “Pairwise interaction processes for modeling cellular network topology.” In *IEEE Global Communications Conference (GLOBECOM), 2012*. IEEE, 2012. 4524–29.

Chapter 6

Delay Normalized Transmission Capacity

6.1 Introduction

In the first part of this book, we considered the single-hop model of a wireless network, where each source–destination pair is at a fixed distance from each other and performs direct communication between themselves. For the single-hop model, we introduced the concept of transmission capacity to characterize the spatial capacity and derived it for many different strategies such as scheduling, multiple antenna transmission, and two-way communication.

Even though the single-hop model is elegant, it is not realistic in a large wireless network, where the primary mode of operation is hop-by-hop communication from each source to its destination using multiple relay nodes. Three important distinctions between single-hop and multi-hop communications are i) routing protocol, finding optimal routes for each source and destination, ii) retransmissions, any relay can forward packet only after it receives it successfully, which potentially requires retransmissions, and iii) shared relay nodes that are used by many source–destination pairs.

In this second part of the book, we exclusively consider the multi-hop communication model for wireless networks. In this chapter, we begin by extending the transmission capacity framework to allow retransmissions using ARQ over multiple hops, and define a new notion called the delay normalized transmission capacity that normalizes the end-to-end throughput with the expected number of retransmissions required for successful delivery of any packet at the destination. We analyze the delay normalized transmission capacity for a simple multi-hop model, where each source–destination pair is at a fixed distance from each other and there are dedicated relays between each source and its destination. Even though this model is simple, it still allows us to capture the interplay of three important quantities: throughput, delay, and reliability, in a multi-hop wireless network.

Even with this simple model, the analysis of the delay normalized transmission capacity is quite involved, and the exact derived expressions are not in the simplest closed form. To get more insights, we also derive useful bounds on the delay normalized transmission capacity that allow us to answer some key questions such as the optimal per-hop retransmission allocation under an end-to-end retransmissions constraint and optimal number of hops to use between each source and

destination. Many shorter hops increase the per-hop success probability but entail larger end-to-end delay, and hence its important to find the optimal number of hops for effective operation of a multi-hop network. Using delay normalized transmission capacity as the metric, we show that for small densities of wireless network, it is advantageous to communicate over small number of hops.

Finally, we consider a more general multi-hop wireless network, where all source–destination pairs are at a random distance from each with no dedicated relays in between them. We show a surprising negative result that the expected number of retransmissions required for a packet to be successfully received at any node is infinite, when node locations are distributed as a PPP. Thus, the rate of transmission between any source–destination pair tends to be zero. This is a serious limitation and we show that it exists even if there is no interference, that is, there are no other active transmitters in the network. Thus, it is the AWGN and the PPP-distributed node locations that limit the ability of any transmitter to send its packet successfully to any other node in the network in finite expected time.

To make the model non-degenerate, either one has to either neglect the AWGN contribution (or assume that the AWGN variance is zero) or assume that nodes are located in a bounded area. Neglecting the AWGN contribution, we define spatial progress capacity with PPP-distributed node locations that has identical scaling with respect to the density of nodes as the throughput capacity defined in Chapter 9, where the nodes are distributed uniformly at random in a bounded area.

6.2 Delay Normalized Transmission Capacity

With multi-hop communication, each packet is transmitted hop by hop toward its destination, where in any hop similar to the single-hop case, we define that any packet is successfully decoded/received at any relay node at time t if the signal-to-interference-plus-noise ratio (SINR) at time t is larger than the threshold β. We assume that each relay employs a decode and forward protocol, where a packet is forwarded onto the next hop once it is successfully decoded. With PPP-distributed nodes, the SINR seen at any relay node at any time slot is a random variable, and hence potentially, multiple transmission attempts are required for a packet to be successfully decoded at any relay node.

To model retransmissions, we consider an automatic-repeat-request (ARQ)-based transmission strategy, where a packet is repeatedly transmitted by each relay node until it is successfully received at the next hop relay or if the maximum retransmissions constraint is exhausted. Thus, the transmission delay for any packet is the number of *retransmissions* required to transmit the packet successfully.

To capture a notion of capacity in multi-hop networks with ARQ protocol, the *delay normalized* transmission capacity is defined as the number of successfully delivered packets in the network per second per Hertz multiplied by the spectral efficiency, that is,

$$C_d = \lim_{t\to\infty} \frac{\text{number of successful packets at any destination node until time } t}{t} B\lambda$$

bits/sec/Hz/m^2, where B bits/sec/Hz is per-hop rate of transmission corresponding to SINR threshold β, and λ is the density of transmitters in the network.

Let $N_{\rm h}$ be the number of hops between any source and its destination, and let D_n be the maximum number of retransmissions used on hop n. Then assuming an end-to-end retransmissions

constraint of D, $\sum_n D_n \leq D$, with the ARQ protocol between any source and its destination, from the renewal reward theorem [1], the delay normalized transmission capacity $C_{\rm d}$ is also equal to

$$C_d = \frac{P_s \lambda B}{\mathbb{E}\{\sum_{n=1}^{N_h} M_n\}}, \text{ bits/sec/Hz/m}^2 \tag{6.1}$$

where $P_{\rm s}$ is the probability that a packet is successfully decoded by any destination within D end-to-end retransmissions, and M_n is the random variable denoting the number of retransmissions used at hop n, $M_n \leq D_n+1$, and $\mathbb{E}\{\sum_{n=1}^{N_h} M_n\}$ is the total expected end-to-end number of retransmissions or end-to-end transmission delay.

The delay normalized transmission capacity C_d quantifies the end-to-end rate that can be supported by λ simultaneous transmissions/unit area, with success probability P_s, and maximum delay $D + N_h$ (where one original transmission is counted for each hop other than the D end-to-end retransmissions). The delay normalized transmission capacity not only gives us a capacity metric but also captures the throughput-delay-reliability (T-D-R) trade-off of wireless networks, where throughput $= C_d$, delay $= \mathbb{E}\{\sum_{n=1}^{N_h} M_n\}$, and reliability $= P_s$.

Remark 6.2.1 *Compared to the single-hop definition of transmission capacity (Definition 2.2.4) without ARQ protocol, here we have normalized the transmission capacity with expected delay to account for the loss in capacity due to retransmissions.*

We first analyze the delay normalized transmission capacity for a single-hop network in Section 6.2.1, and then build upon it to extend the analysis to the multi-hop network case in Section 6.3.1. Compared with the transmission capacity framework for single-hop network studied in Section 2.2, which took a single time slot view of transmission, with the ARQ protocol, we have to consider time correlations of SINRs to characterize the delay normalized transmission capacity of a single-hop network, which is done as follows.

6.2.1 Single Hop Transmission with ARQ Protocol

In this section, we consider a single-hop wireless network, where source locations are distributed as a PPP Φ with density λ and each source–destination pair $\mathcal{S} - \mathcal{D}$ is at a fixed distance d. We consider a typical source–destination pair $\mathcal{S}_0 - \mathcal{D}_0$, where the received signal at the typical destination $\mathcal{D}_0$ at time t is given by

$$y(t) = d^{-\alpha/2} h_{00}(t) x_0(t) + \sum_{s:T_s(t)\in\Phi\setminus\{\mathcal{S}_t\}} \mathbf{1}_{T_s(t)} d_s^{-\alpha/2} h_{0s}(t) x_s(t), \tag{6.2}$$

where $\mathbf{1}_{T_s(t)}$ is the indicator variable denoting whether transmitter T_s is active at time t, $h_{0s}^n(t) \in \mathbb{C}$ is the channel coefficient between $T_s(t)$ and $\mathcal{D}_0$, $x_s(t) \sim \mathcal{CN}(0,1)$ is the signal transmitted from source T_s in time slot t, and for ease of exposition we have neglected the additive noise. As before, we assume that each $h_{0s}^n(t)$ is i.i.d. with Rayleigh distribution.

Any packet from the typical source $\mathcal{S}_0$ is defined to be successfully received at the typical destination $\mathcal{D}_0$ in time slot t if

$$\mathsf{SIR}(t) = \frac{d_n^{-\alpha} |h_{00}(t)|^2}{\sum_{s:T_s(t)\in\Phi\setminus\{\mathcal{S}_0\}} \mathbf{1}_{T_s(t)} d_s^{-\alpha} |h_{0s}(t)|^2} > \beta. \tag{6.3}$$

Each packet is retransmitted (at most D times) until it is successfully received at its corresponding destination, or dropped otherwise. We assume that each source uses an ALOHA protocol with

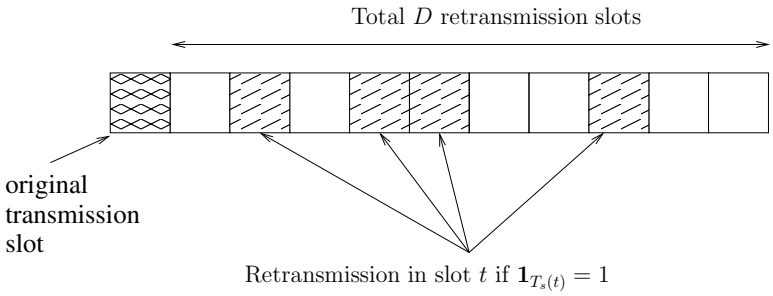

Figure 6.1: Retransmission strategy where in any slot, retransmission (shaded square) is made if $\mathbf{1}_{T_s(t)} = 1$ and no attempt is made (empty square) otherwise.

parameter p for transmitting/retransmitting any packet at each time t irrespective of the number of previously failed retransmission attempts for that packet.

Recall that the success probability P_s is the probability that the packet is correctly received by the typical destination within D retransmissions. Because of stationarity of the PPP, we will restrict our attention to the typical source–destination pair for deriving P_s. For any packet, we count slot 0 as the slot in which the source makes the first transmission attempt. Hence, starting from slot 0, there are a total of $D + 1$ potential transmission slots for each packet, where in each slot, retransmission happens if the ALOHA protocol parameter $\mathbf{1}_{T_s(t)} = 1$. In Fig. 6.1, we describe the retransmission strategy.

We first derive an expression for the success probability P_s and use that to find the delay normalized transmission capacity C_d (6.1). Let E be the event that a packet is successfully received within D retransmissions, that is, $P_s = \mathbb{P}(E)$, and

$$E = \cup_{j=0}^{D} E_j,$$

where event

$$E_j = \{\text{a packet is received correctly for the first time in the } j\text{th retransmission slot}\}. \tag{6.4}$$

Event E_j means that the packet is successfully received in the jth retransmission slot and in all the previous j slots (including the first original slot), whenever a transmission attempt was made by the ALOHA protocol, the packet reception was not successful.

Let $E_{kj}, k = 0, 1, 2, \ldots, j$, be the event denoting k unsuccessful attempts with the ALOHA protocol before the successful reception at the jth retransmission slot, that is,

$$E_{kj} = \{\text{failures in } k \text{ previous attempts and success in the } j\text{th retransmission slot}\},$$

for $k = 0, 1, 2, \ldots, j$. Since $\mathsf{SIR}(t)$ (6.3) is identically distributed $\forall\ t$, event E_{kj} only depends on how many failures have happened before retransmission slot j, and not where those failures happened. For example,

$$\mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \mathsf{SIR}(t) \leq \beta\right) = \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \mathsf{SIR}(n) \leq \beta\right),\ t \neq n,$$

since the channel coefficients are independent across time slots, and in any time slot each source transmits with probability p independent of others.

Thus,

$$P_s^{k,j} = \mathbb{P}(E_{kj}) = \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta,\ \ldots,\ \mathsf{SIR}(k) \leq \beta,\ \mathsf{SIR}(j) > \beta\right). \tag{6.5}$$

Clearly, $E_j = \cup_{k=0,1,2,\ldots,j} E_{kj}$. Then, the probability of event E_j, defined as P_s^j is equal to

$$P_s^j = \mathbb{P}(\cup_{k=0,1,2,\ldots,j} E_{kj}).$$

The events

$$\cup_{k=0,1,2,\ldots,j} E_{kj}$$

are mutually exclusive for any j, $j = 0, \ldots, D$, hence, the success probability $P_s = \mathbb{P}(E)$ is

$$P_s \quad = \quad \sum_{j=0}^{D} P_s^j. \tag{6.6}$$

With ALOHA protocol, k retransmission attempts are made before the jth retransmission slot with probability $\binom{j}{k} p^k (1-p)^{j-1-k}$, hence

$$P_s^j \quad = \quad \sum_{k=0}^{j} \binom{j}{k} p^k (1-p)^{j-k} p P_s^{k,j}, \tag{6.7}$$

by accounting for $k = 0$ or 1, or , $\ldots, j$ failures before success at the the jth retransmission slot, where extra p is because of the original transmission in slot 0 that happens with probability p. Computing the joint probability $P_s^{k,j}$ in (6.7), assuming that the typical destination $\mathcal{D}_0$ is located at the origin, the expression for P_s is given by the following proposition.

Proposition 6.2.2 *The success probability P_s is given by*

$$\begin{aligned} P_s \quad = \quad & \sum_{j=0}^{D} \sum_{k=0}^{j} \binom{j}{k} p^k (1-p)^{j-k} p \sum_{\ell=0}^{k} (-1)^\ell \binom{k}{\ell} \times \\ & \exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \left(\frac{p}{1 + d^\alpha \beta x^{-\alpha}} + 1 - p \right)^{\ell+1} \mathrm{d}x \right). \end{aligned}$$

Proof: Defining $I_\Phi(\ell) = \sum_{s:T_s \in \Phi \setminus \{\mathcal{S}_0\}} \mathbf{1}_{T_s(\ell)} d_s^{-\alpha} |h_{0s}(\ell)|^2$ as the interference received at $\mathcal{D}_0$ in time slot ℓ. Then from (6.5),

$$\begin{aligned} P_s^{k,j} \quad &= \quad \mathbb{P}\left(\mathsf{SIR}(\ell) \leq \beta_{,\ell=1,2,\ldots,k,},\ \mathsf{SIR}(j) > \beta\right), \\ &= \quad \mathbb{P}\left(\frac{d^{-\alpha}|h_{00}(1)|^2}{I_{\Phi(1)}} \leq \beta, \ldots, \frac{d^{-\alpha}|h_{00}(k)|^2}{I_{\Phi(k)}} \leq \beta,\ \frac{d^{-\alpha}|h_{00}(j)|^2}{I_{\Phi(j)}} > \beta \right), \\ &\overset{(a)}{=} \quad \mathbb{E}\left\{ \prod_{\ell=1}^{k} \left(1 - \exp\left(-\frac{\beta I_\Phi(\ell)}{d^{-\alpha}} \right) \right) \exp\left(-\frac{\beta I_\Phi(j)}{d^{-\alpha}} \right) \right\}, \\ &\overset{(b)}{=} \quad \mathbb{E}\left\{ \prod_{\ell=1}^{k} \left(1 - \prod_{s:T_s \in \Phi \setminus \{\mathcal{S}_0\}} \left(\frac{1}{1 + d^\alpha \beta \mathbf{1}_{T_s(\ell)} d_s^{-\alpha}} \right) \right) \right. \end{aligned}$$

$$\prod_{s:T_s\in\Phi\setminus\{\mathcal{S}_0\}}\left(\frac{1}{1+d^{\alpha}\beta\mathbf{1}_{T_s(j)}d_s^{-\alpha}}\right)\Bigg\}, \tag{6.8}$$

where (a) follows by taking the expectation with respect to $h_{00}(\ell),\ \ell=1,\ldots,k+1,\ j$, since $|h_{00}(\ell)|^2$ are i.i.d. $\sim \exp(1)$, (b) follows by using the definition of $I_{\Phi}(\ell)$ and taking the expectation with respect to $h_{0s}(\ell)$, where $|h_{0s}(\ell)|^2\sim\exp(1)$ and i.i.d. for each ℓ, and d_s is the distance of node T_s from $\mathcal{D}_0$ located at the origin.

Now we take the expectation with respect to the ALOHA protocol indicator function $\mathbf{1}_{T_s(\ell)}$ which is 1 with probability p and 0 otherwise. Note that

$$\frac{1}{1+d^{\alpha}\beta\mathbf{1}_{T_s(\ell)}d_s^{-\alpha}}=\frac{\mathbf{1}_{T_s(\ell)}}{1+d^{\alpha}\beta d_s^{-\alpha}}+1-\mathbf{1}_{T_s(\ell)}.$$

Thus,

$$\mathbb{E}\left\{\frac{1}{1+d^{\alpha}\beta\mathbf{1}_{T_s(\ell)}d_s^{-\alpha}}\right\}=\frac{p}{1+d^{\alpha}\beta d_s^{-\alpha}}+1-p. \tag{6.9}$$

Substituting (6.9) into (6.8), and replacing $d_s=x$, we get

$$\begin{aligned}
P_s^{k,j} &= \mathbb{E}_{\Phi}\left\{\prod_{\ell=1}^{k}\left(1-\prod_{x\in\Phi\setminus\{\mathcal{S}_0\}}\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)\right)\right.\\
&\qquad\left.\prod_{x\in\Phi\setminus\{T_0\}}\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)\right\},\\
&\overset{(c)}{=} \mathbb{E}\left\{X_{\Phi}(1-X_{\Phi})^k\right\},\\
&\overset{(d)}{=} \sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\mathbb{E}\left\{X_{\Phi}^{\ell+1}\right\},
\end{aligned}$$

where (c) follows by defining

$$X_{\Phi}=\prod_{x\in\Phi\setminus\{\mathcal{S}_0\}}\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right),$$

(d) follows from linearity of expectation and using the Binomial expansion, and finally the result follows since

$$\mathbb{E}_{\Phi}\left\{X_{\Phi}^{\ell+1}\right\}=\exp\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)^{\ell+1}\mathrm{d}x\right)$$

using the probability-generating function of PPP (Theorem 2.3.6). □

Now, using the derived expression for the success probability P_s, we find the expected number of retransmissions required. Recall that M is the random variable denoting the total number of

transmissions (one plus the number of retransmissions) required for each packet that takes values in $[0 : D+1]$. The event $\{M = j\}$ is the same as event E_j (6.4) for $j \leq D-1$, hence

$$\mathbb{P}(M = j) = P_s^j,\ j = 0, 1, 2, \ldots, D-1. \tag{6.10}$$

For the special case of having $M = D+1$ transmissions or D retransmissions, one has to also take into account the delay incurred by packets that are not decoded correctly even after D retransmissions, and hence

$$\mathbb{P}(M = D) = P_s^D + \sum_{j=0}^{D-1} P_s^j = P_s^D + (1 - P_s), \tag{6.11}$$

P_s^{D+1} is the probability of successfully receiving the packet by the $D+1\text{st}$ transmission or $D\text{th}$ retransmission. Using Proposition 6.2.2, from (6.10) and (6.11), the expected number of retransmissions $\mathbb{E}\{M\} = \sum_{j=0}^{D}(j+1)\mathbb{P}(M = j)$ is derived as follows.

Proposition 6.2.3 *The expected delay $\mathbb{E}\{M\}$ in a single-hop wireless network with at most D retransmissions is*

$$\begin{aligned}\mathbb{E}\{M\} \quad = \quad & \sum_{j=0}^{D}\sum_{k=0}^{j}\sum_{\ell=0}^{k} j\binom{j}{k}p^{k+1}(1-p)^{j-1-k}(-1)^{\ell}\binom{k}{\ell} \times \\ & \exp\left(-\lambda\int_{\mathbb{R}^2} 1 - \left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}} + 1 - p\right)^{\ell+1} \mathrm{d}x\right) + (D+1)(1-P_s).\end{aligned}$$

Propositions 6.2.2 and 6.2.3 allow us to derive an explicit expression for the delay normalized transmission capacity as follows, whose proof directly follows from Definition 6.1.

Theorem 6.2.4 *The delay normalized transmission capacity of a single-hop wireless network with at most D retransmissions is*

$$C_d = \lambda B \frac{P_s}{\mathbb{E}\{M\}},$$

where P_s is given by Proposition 6.2.2, and $\mathbb{E}\{M\}$ is given by Proposition 6.2.3.

Proposition 6.2.2 and Theorem 6.2.4 give exact expressions for the success probability and the delay normalized transmission capacity, respectively, of a wireless network with single-hop transmission, with a retransmissions constraint of D. Because of correlation among $\mathsf{SIR}s$ across different time slots with PPP-distributed source locations, the derived expressions are complicated and do not allow for a simple closed-form expression for P_s and C_d, as a function of D. Thus, precluding a derivation of a simple expression for the T-D-R trade-off of a wireless network. To get more insights on the dependence of P_s and C_d on D (to obtain simple T-D-R trade-off), we next derive tight lower and upper bounds on P_s^j, and consequently on P_s, and the delay normalized transmission capacity C_d.

Bounds on the Delay Normalized Transmission Capacity

For deriving the bounds, we will make use of the FKG inequality (Lemma 2.8.8).

Upper bound on the success probability P_s.

Proposition 6.2.5 *The success probability P_s with single-hop transmission in a wireless network with at most D retransmissions is upper bounded by*

$$P_s \leq 1 - (pq + 1 - p)^{D+1},$$

where $q = \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta\right) = 1 - \exp\left(-\frac{\lambda 2\pi^2 d^2 \beta^{\frac{2}{\alpha}} csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right)$.

Proof: Recall that each source with an ALOHA protocol retransmits with probability p in each time slot. Let source $\mathcal{S}_0$ make total k attempts with ALOHA protocol to transmit a packet to the destination $\mathcal{D}_0$ within the total $D+1$ slots, $k = 1, 2, \ldots, D+1$. Then the event $\{$success in at most D retransmissions$\}$ is also equal to the complement of the event $\{$failures in all k attempts$\}$ for $k = 1, 2, \ldots, D+1$. Then counting for all possible choices of k, we have

$$P_s \quad = \quad 1 - \sum_{k=1}^{D+1} \binom{D+1}{k} p^k (1-p)^{D+1-k} P_f^k, \tag{6.12}$$

where

$$P_f^k = \mathbb{P}(\text{failure in } k \text{ attempts}) = \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right),$$

since each $\mathsf{SIR}(t)$ is identically distributed, it does not matter where those k failures happen. From Definition 2.8.7, it follows that $\{\mathsf{SIR}(1) \leq \beta\}$ is an increasing event. Thus, using the FKG inequality (Lemma 2.8.8),

$$P_f^k \quad \geq \quad \prod_{j=1}^{k} \mathbb{P}\left(\mathsf{SIR}(j) \leq \beta\right), \tag{6.13}$$

where $\mathbb{P}\left(\mathsf{SIR}(t) \leq \beta\right) = \mathbb{P}\left(\mathsf{SIR}(m) \leq \beta\right)$, $t \neq m$. Let $q = \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta\right)$. From Theorem 2.4.1,

$$q = 1 - \exp\left(-\frac{p\lambda 2\pi^2 d^2 \beta^{\frac{2}{\alpha}} csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right). \tag{6.14}$$

Substituting (6.13) into (6.12),

$$P_s \leq 1 - \sum_{k=1}^{D+1} \binom{D+1}{k} p^k (1-p)^{D+1-k} q^k = 1 - (pq + 1 - p)^{D+1}.$$

□

Lower bound on the success probability P_s.

Proposition 6.2.6 *The success probability P_s with single-hop transmission in a wireless network with at most D retransmissions is lower bounded by*

$$P_s \quad \geq \quad P_s^{D,D+1} \left(\frac{1 - (pq + 1 - p)^{D+1}}{1 - q}\right), \text{ where} \tag{6.15}$$

$P_s^{D,D+1} = \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta | \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right)$.

Proof: From (6.7), the success probability

$$P_s = \sum_{j=0}^{D} \sum_{k=0}^{j} \binom{j}{k} p^k (1-p)^{j-k} p \mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta, \ \mathsf{SIR}(j) > \beta\right). \tag{6.16}$$

Note that $\mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta, \ \mathsf{SIR}(j) > \beta\right)$

$$\begin{aligned} &= P_f^k \mathbb{P}\left(\mathsf{SIR}(j) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right), \\ &\overset{(a)}{\geq} q^k \mathbb{P}\left(\mathsf{SIR}(j) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right), \end{aligned} \tag{6.17}$$

where (a) follows from (6.13), and $q = \mathbb{P}(\mathsf{SIR}(j) \leq \beta)$ for any $j = 1, \ldots, k$.

Because of positive temporal correlation of SIRs (Example 2.7.3), the probability of success given j failures is more than the probability of success given $j+1$ failures, hence

$$\mathbb{P}\left(\mathsf{SIR}(j) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right) \geq \mathbb{P}\left(\mathsf{SIR}(j) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k+1) \leq \beta\right).$$

Therefore, since $\mathsf{SIR}(j)$ are identically distributed for all j, for $D \geq k+1$,

$$\mathbb{P}\left(\mathsf{SIR}(j) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right) = \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right), \tag{6.18}$$

and

$$\mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(k) \leq \beta\right) \geq \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right). \tag{6.19}$$

Thus, from (6.16), (6.17), (6.19), we get the following lower bound on the success probability P_s

$$\begin{aligned} P_s &\geq \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right) \\ &\quad \sum_{j=0}^{D} \sum_{k=0}^{j} \binom{j}{k} p^k (1-p)^{j-k} p q^k, \\ &= \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right) \\ &\quad \frac{1 - (pq + 1 - p)^{D+1}}{1 - q}. \end{aligned} \tag{6.20}$$

Thus, finishing the proof. □

For small values of D, we can analytically show (Example 6.2.7) that

$$P_S^{D,D+1} = \mathbb{P}\left(\mathsf{SIR}(D+1) > \beta \mid \mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right) \approx c(1-q), \tag{6.21}$$

where c is a constant.

Example 6.2.7 *By definition,* $\mathbb{P}\left(\mathsf{SIR}(D+1) \geq \beta \mid \mathsf{SIR}(1) \leq \beta \ \ldots, \ \mathsf{SIR}(D) \leq \beta\right)$

$$\begin{aligned} &= \frac{\mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta, \ \mathsf{SIR}(D+1) > \beta\right)}{\mathbb{P}\left(\mathsf{SIR}(1) \leq \beta, \ \ldots, \ \mathsf{SIR}(D) \leq \beta, \right)}, \\ &\overset{(a)}{=} \frac{\sum_{\ell=0}^{D}(-1)^{\ell}\binom{D}{\ell}\exp\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)^{\ell+1}\mathrm{d}x\right)}{\sum_{\ell=0}^{D}(-1)^{\ell}\binom{D}{\ell}\exp\left(-\lambda\int_{\mathbb{R}^2}1-\left(\frac{p}{1+d^{\alpha}\beta x^{-\alpha}}+1-p\right)^{\ell}\mathrm{d}x\right)}, \end{aligned}$$

where (a) *follows from Proposition 6.2.2. Hence, for* $D = 2$*, by computing the integral for* $\ell = 1, 2$*, with* $c_2 = p\lambda 2\pi d^2\beta^{\frac{2}{\alpha}}$*,* $\mathbb{P}\left(\mathsf{SIR}(3) \geq \beta \mid \mathsf{SIR}(1) \leq \beta, \ \mathsf{SIR}(2) \leq \beta\right)$

$$\begin{aligned} &= \frac{\exp\left(-\frac{\pi c_2 csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right) - 2\exp\left(-c_2\left(2-p\pi\frac{\alpha-2}{\alpha^2}\pi csc\left(\frac{2\pi}{\alpha}\right)\right)\right)}{1-2\exp\left(-\frac{c_2\pi csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right)+\exp\left(-c_2\left(2-p\pi\frac{\alpha-2}{\alpha^2}\pi csc\left(\frac{2\pi}{\alpha}\right)\right)\right)} \\ &\quad + \frac{\exp\left(-\frac{c_2\left(-3\alpha(-2+p)p+2p^2+\alpha^2(3+\mathbb{P}(p-3))\right)csc\left(\frac{2\pi}{\alpha}\right)}{\alpha^3}\right)}{1-2\exp\left(-\frac{c_2\pi csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right)+\exp\left(-c_2\left(2-p\pi\frac{\alpha-2}{\alpha^2}\pi csc\left(\frac{2\pi}{\alpha}\right)\right)\right)}, \\ &\approx \exp\left(-\lambda\beta^{\frac{2}{\alpha}}cd^2\right) = 1-q. \end{aligned}$$

Similarly, for $D = 3$, it can be shown that

$$\mathbb{P}\left(\mathsf{SIR}(4) \geq \beta \mid \mathsf{SIR}(1) \leq \beta, \ \mathsf{SIR}(2) \leq \beta, \ \mathsf{SIR}(3) \leq \beta\right) \approx 1-q.$$

Thus, from Proposition 6.2.6,

$$P_s \geq c\left(1-(pq+1-p)^{D+1}\right), \tag{6.22}$$

for small values of D, where c is a constant. Comparing the lower bound on P_s (6.22) with the upper bound on P_s from Proposition 6.2.5, we get that our derived lower and upper bounds on P_s are tight upto a constant for small values of D. For higher values of D also, the bounds can be shown to be tight using simulations in the sparse network regime, that is, small λ or $\lambda \to 0$. To get a better sense of our derived upper and lower bounds, in Fig. 6.2, we plot the success probability P_s as a function of maximum retransmissions constraint D together with the upper and lower bound for density $\lambda = 0.1$, SIR threshold $\beta = 3$, and path-loss exponent $\alpha = 3$. Clearly, the simulated success probability closely follows the derived upper and lower bound.

Using approximation (6.22), in light of tight bounds from Propositions 6.2.5 and 6.2.6, we have that the success probability of any packet within D retransmissions in a single-hop wireless network is

$$P_s \approx c\left(1-(pq+1-p)^{D+1}\right), \tag{6.23}$$

where $c \leq 1$ is a constant. The approximate success probability expression reveals that even though the success/failure of packet decoding is correlated across time slots, for small D or in a sparse network, the success probability P_s is equal to cP_s^{indep}, where $c \leq 1$ is a constant, and P_s^{indep} is the success probability in at most D retransmissions if the success/failure of packet decoding

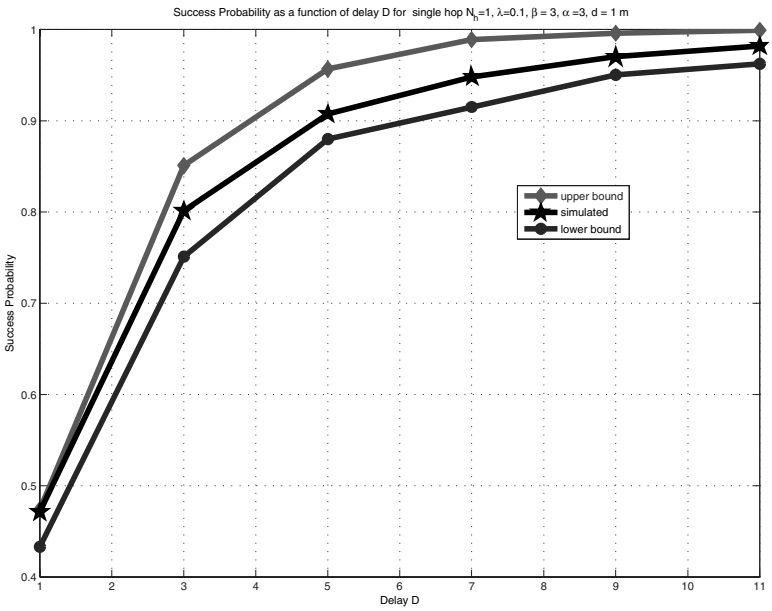

Figure 6.2: Success probability as a function of D for $N_h = 1$.

is independent across time slots. Thus, an important conclusion we draw is that the independence assumption on SIRs across time slots, if made for simpler analysis, is not too limiting. Using (6.23), we can get an approximate delay normalized transmission capacity expression for the single-hop wireless network as follows.

Approximate Delay Normalized Transmission Capacity: Recall that the expected number of retransmissions $\mathbb{E}\{M\} = \sum_{j=0}^{D}(j+1)\mathbb{P}(M=j)$, where $\mathbb{P}(M=j) = P_s^j,\ j = 0, 1, 2, \ldots, D$, and $\mathbb{P}(M=D) = P_s^D + (1-P_s)$ and P_s^j, the probability of successful reception in the jth retransmission slot for the first time, is given by (6.7). Now we use the expansion of P_s^j from (6.5) in terms of P_s^{kj}, the probability of k unsuccessful attempts before the successful reception in the jth slot for the first time. Hence, from (6.5) and (6.7), $\mathbb{E}\{M\}$

$$
\begin{aligned}
&= \sum_{j=0}^{D}(j+1)\mathbb{P}(M=j), \\
&= \sum_{j=0}^{D}(j+1)\sum_{k=0}^{j}\binom{j}{k}p^k(1-p)^{j-1-k}p\mathbb{P}\left(\mathsf{SIR}(1)\leq\beta,\ \ldots,\ \mathsf{SIR}(k)\leq\beta,\ \mathsf{SIR}(j)>\beta\right).
\end{aligned}
$$

Following approximation (6.21) used for deriving (6.23), using simple algebra we get that

$$\mathbb{E}\{M\} = c\left[\frac{1-(pq+1-p)^{D+1}}{(1-q)}\right] + (D+1)(1-c). \tag{6.24}$$

Therefore, from (6.22) and (6.24), the approximate delay normalized transmission capacity (6.1) is given by

$$C_d \approx \frac{c(1-(pq+1-p)^{D+1})\lambda B}{c\left[\frac{1-(pq+1-p)^{D+1}}{(1-q)}\right]+(D+1)(1-c)}, \tag{6.25}$$

for some constant $c \leq 1$.

The approximate expression (6.25) for the delay normalized transmission capacity in a single-hop network allows us to get useful insights on the interplay between the three inter-related quantities, throughput, delay, and reliability, and tells us how they can be traded off against each other. We get the intuitive result that as we allow more retransmissions, the success probability increases but at the cost of expected delay. The throughput expression (6.25), which is the ratio of reliability and expected delay, captures the trade-off completely and tells us how the *real* quantity of interest, the rate of transmission, increases with increasing number of allowed retransmissions.

After discussing at length, the techniques used to derive the single-hop delay normalized capacity presented in this section; we next build upon it to find the delay normalized transmission capacity of a multi-hop wireless network.

6.3 Fixed Distance Dedicated Relays Multi-Hop Model with ARQ

In this section, we consider a simple fixed distance multi-hop wireless network with dedicated relays for each source–destination pair and derive its delay normalized transmission capacity. We assume that each source–destination pair is at a fixed distance d from each other with multiple dedicated relays helping their communication. A more realistic model of random source–destination distance with non-dedicated relays is discussed in Section 6.4.

Compared to other parts of this book, in this section, we make a distinction between transmitters and sources, since there are many nodes that will only relay other sources' information. The location of source nodes $\mathcal{S}_m,\ m \in \mathbb{N}$ is assumed to be distributed as a homogenous PPP Φ on a two-dimensional plane with density λ_0. The destination $\mathcal{D}_m$ corresponding to source $\mathcal{S}_m$ is at a distance of d from it in a random direction. We assume that there are N_h-1 relays $\mathcal{R}_{n,m}, n = 1, 2, \ldots, N_h-1$ (N_h hops) in between each source $\mathcal{S}_m$ and its intended receiver $\mathcal{D}_m\ \forall\ m$, not necessarily on a straight line between $\mathcal{S}_m$ and $\mathcal{D}_m$, with inter-hop distance d_n. For simplicity, we assume that all the $N_h - 1$ relays on the $\mathcal{S}_m$-$\mathcal{D}_m$ link are dedicated and cannot be used by any other source–destination pair. Thus, link m is described by the set of nodes $\{\mathcal{S}_m, \mathcal{R}_{1,m}, \ldots, \mathcal{R}_{N_h-1,m}, \mathcal{D}_m\}$. A schematic of the considered model is shown in Fig. 6.3.

Transmission Strategy: We assume all nodes in the network to a have a single antenna each. The transmission happens hop by hop using an ARQ protocol under a maximum end-to-end retransmissions constraint of D, for each packet transmission between $\mathcal{S}_m$ and $\mathcal{D}_m,\ \forall\ m$. For simplicity, the same packet is assumed to be retransmitted (at most D times) with every NACK on any hop, without any incremental redundancy or rate adaptation.

Assumption 6.3.1 *We assume a simple transmission protocol that allows only one active packet on each link, that is, the source waits to transmit the next packet until the previous packet has been received by the destination, or the delay constraint has been violated. This assumption is critical for the ease of analysis. In a practical system, however, pipelining can significantly increase the*

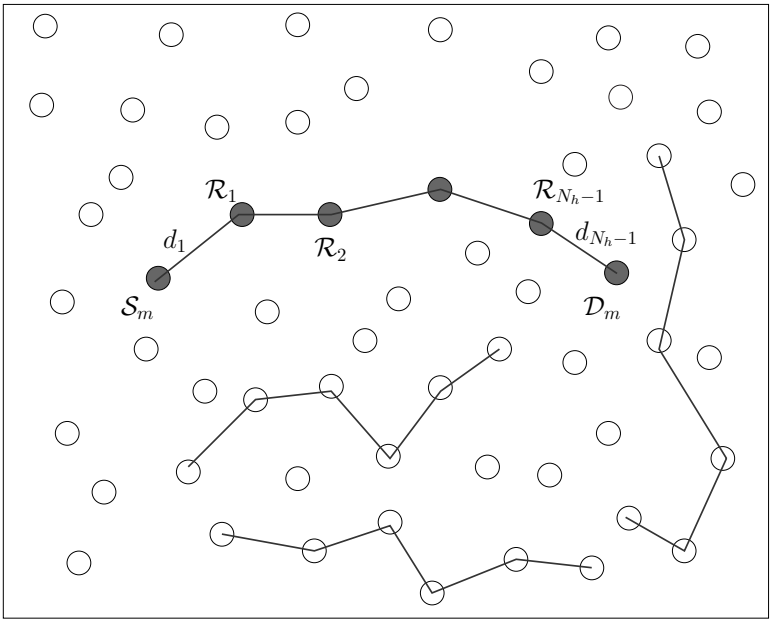

Figure 6.3: Schematic of the system model where connected lines depict a path between a source and its destination.

throughput, and analysis with more than one active packet is an important topic, however, that remains unsolved.

Let the transmitter and the receiver on link m between source $\mathcal{S}_m$ and its destination $\mathcal{D}_m$ in time slot t be $T_m(t)$ and $R_m(t)$, respectively, where

$$T_m(t) \in \{\mathcal{S}_m, \mathcal{R}_{1,m}, \ldots, \mathcal{R}_{N_h-1,m}\},$$

and

$$R_m(t) \in \{\mathcal{R}_{1,m}, \ldots, \mathcal{R}_{N_h-1,m}, \mathcal{D}_m\} \backslash \{T_m(t)\}.$$

The set of interfering nodes for receiver $R_m(t)$ at time slot t is $\Phi(t) = \{\Phi(t) \backslash T_m(t)\}$, where $\Phi(t) = \{T_k(t)\}$.

Remark 6.3.2 *When the active transmitter process is a PPP, there is spatial correlation, and the success/failure of packet decoding at different receivers is correlated (as discussed in Section 2.7). Therefore, retransmission of packets depending on the NACK in a multi-hop network introduces correlation among the active transmitter process $\Phi(t)$, and strictly speaking, $\Phi(t)$ is no longer a random thinning of PPP, and consequently not a PPP.*

Assumption 6.3.3 *To facilitate analysis, we assume that the success/failure of packet decoding on different links with ARQ is independent, and therefore the active transmitter process $\Phi(t)$ at any time t is a PPP. There are two justifications for this assumption, i) the decoupling argument [2], which states that in a large network, the success probability over different links is independent using the mean-field type argument, and ii) the fact that spatial correlation coefficient in a PPP*

network is zero with the $x^{-\alpha}$ path-loss model (Example 2.7.3). We do, however, take into account the spatial/temporal correlations over different hops on the same link.

Similar to the single-hop model, we consider a slotted ALOHA protocol, where each transmitter (source or any relay) attempts to transmit its packet (either new or retransmission) with an access probability p, independently of all other transmitters. Consequently, the active transmitter process $\Phi(t)$ is a homogenous PPP on a two-dimensional plane with density $\lambda = p\lambda_0$.

We consider a typical link $\{\mathcal{S}_0, \mathcal{R}_{1,0}, \ldots, \mathcal{R}_{N_h-1,0}, \mathcal{D}_0\}$ and refer to it as $\{\mathcal{S}_0, \mathcal{D}_0\}$ link. Recall that at time slot t only one of the nodes of link $\{\mathcal{S}_0, \mathcal{D}_0\}$ is transmitting, let that be the nth relay $\mathcal{R}_{n,0}$ ($n = 0$ corresponds to the source $\mathcal{S}_0$), that is, $T_0(t) = \mathcal{R}_{n,0}$. Then the received signal over the nth hop at the $n + 1$th relay (defined $R_0(t)$) of link $\{\mathcal{S}_0, \mathcal{D}_0\}$ at time slot t is

$$y_{0n}(t) = d_n^{-\alpha/2} h_{00}^n(t) x_{0n}(t) + \sum_{s:T_s(t)\in\Phi(t)\setminus\{T_0(t)\}} \mathbf{1}_{T_s(t)} d_s^{-\alpha/2} h_{0s}^n(t) x_{sn}(t), \tag{6.26}$$

where $h_{0s}^n(t) \in \mathbb{C}$ is the channel coefficient between the active transmitter $T_s(t)$ of the $\{\mathcal{S}_s, \mathcal{D}_s\}$ link and the active receiver $R_0(t)$ of the $\{\mathcal{S}_0, \mathcal{D}_0\}$ link, d_n is the nth hop distance between $T_0(t)$ and $R_0(t)$ and d_s is the distance from $T_s(t) \in \Phi(t)\setminus\{T_0(t)\}$ to $R_0(t)$, $x_{sn}(t) \sim \mathcal{CN}(0, 1)$ is the signal transmitted from transmitter $T_s(t)$ in time slot t, and $\mathbf{1}_{T_s(t)} = 1$ if transmitter $T_s(t)$ is active at time t which happens with probability p because of the ALOHA protocol. As before, we assume that each $h_{0s}^n(t)$ is i.i.d. with Rayleigh distribution and ignore the AWGN contribution.

Let $\mathsf{SIR}_n(t)$ denote the SIR between the active transmitter and receiver of link $\{\mathcal{S}_0, \mathcal{D}_0\}$ at time slot t on hop n. With the received signal model (6.26),

$$\mathsf{SIR}_n(t) = \frac{d_n^{-\alpha}|h_{00}^n(t)|^2}{\sum_{s:T_s(t)\in\Phi(t)\setminus\{T_0(t)\}} \mathbf{1}_{T_s(t)} d_s^{-\alpha}|h_{0s}^n(t)|^2}. \tag{6.27}$$

We assume that the rate of transmission for each hop is $B = \log(1 + \beta)$ bits/sec/Hz. Therefore, a packet transmitted by transmitter $T_0(t)$ can be successfully decoded at receiver $R_0(t)$ in time slot t if $\mathsf{SIR}_0(t) > \beta$. Then the *delay normalized* transmission capacity C_d of a wireless network with multi-hop transmission as defined in (6.1) is the number of successfully delivered packets in the network per second per Hertz multiplied by the spectral efficiency.

Remark 6.3.4 *The definition of delay normalized transmission capacity C_d takes a holistic view of the network by coupling the interdependent metrics, namely throughput, delay, and success probability. Some earlier work [3–8] focussed on finding routing protocols for minimizing delay or minimizing transmit energy or maximizing reliability.*

6.3.1 Deriving Delay Normalized Transmission Capacity

In this subsection, we derive the delay normalized transmission capacity of the fixed distance dedicated relays multi-hop wireless network model. Let D_n be the maximum number of retransmissions used on hop n. Since the total end-to-end retransmissions are constrained by D, we have $\sum_{n=1}^{N_h} D_n \leq D$. The success probability, as before, is defined as the probability that the packet is successfully received at the destination within D end-to-end retransmissions. Let $E_n(j_n)$

denote the event that a tagged packet sent from the typical source $\mathcal{S}_0$ is received correctly for the first time in the j_nth retransmission slot over hop n of link $\{\mathcal{S}_0, \mathcal{D}_0\}$. Thus, defining

$$P_s^{j_1 j_2 \cdots j_{N_h}} = \mathbb{P}(\cap_n E_n(j_n)),$$

the success probability is given by

$$P_s = \sum_{j_1=0}^{D_1} \sum_{j_2=0}^{D_2} \cdots \sum_{j_{N_h}=0}^{D_{N_h}} P_s^{j_1 j_2 \cdots j_{N_h}}.$$

Note that $\mathsf{SIR}_n(t)$ (6.27) is identically distributed $\forall\ t,\ n$, thus, $P_s^{j_1 \cdots j_{N_h}}$ only depends on how many failures have happened before the success at the j_nth retransmission slot on hop n, and not where those failures happened. Let ℓ_n be the number of failed attempts on hop n before success at the j_nth slot. Then we can write,

$$\begin{aligned} P_s^{j_1 j_2 \cdots j_{N_h}} \quad = \quad & \sum_{\ell_1=0}^{j_1} \cdots \sum_{\ell_N=0}^{j_{N_h}} \prod_{n=1}^{N_h} \binom{j_n}{\ell_n} p^{\ell_n} (1-p)^{j_n - \ell_n} \\ & \mathbb{P}\left(\mathsf{SIR}_1(k_1) \leq \beta, k_1 = 1, 2, \ldots, \ell_1,\ \mathsf{SIR}_1(j_1) > \beta, \ldots,\right. \\ & \left.\mathsf{SIR}_n(k_n) \leq \beta, k_n = 1, 2, \ldots, \ell_n,\ \mathsf{SIR}_1(j_n) > \beta\right), \end{aligned}$$

since at each slot any relay makes a transmission attempt with probability p independently. Note that the joint probability in the RHS is similar to (6.5), and thus using the derivation of the proof of Proposition 6.2.2 where we derive an expression for (6.5), we get the exact expression for the success probability as follows.

Proposition 6.3.5

$$\begin{aligned} P_s \quad = \quad & \sum_{j_1=0}^{D_1} \cdots \sum_{j_{N_h}=0}^{D_{N_h}} \sum_{\ell_1=0}^{j_1} \cdots \sum_{\ell_N=0}^{j_{N_h}} \prod_{n=1}^{N_h} \binom{j_n}{\ell_n} p^{\ell_n} (1-p)^{j_n - 1 - \ell_n} \times \\ & \sum_{r_1=0}^{\ell_1} \cdots \sum_{r_N=0}^{\ell_N} \prod_{n=1}^{N_h} (-1)^{r_n} \binom{\ell_n}{r_n} \\ & \exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \prod_{n=1}^{N_h} \left(\frac{p}{1 + d_n^{\alpha} \beta x^{-\alpha}} + 1 - p\right)^{r_n + 1} \mathrm{d}x\right). \end{aligned}$$

The expected end-to-end delay $\mathbb{E}\{M\}$ can be computed easily by using the linearity of expectation, $\mathbb{E}\{M\} = \sum_{n=1}^{N_h} \mathbb{E}\{M_n\}$, where $\mathbb{E}\{M_n\}$ is given by Proposition 6.2.3 by replacing D_n with D. Thus, we get the main result of this section on the delay normalized transmission capacity of a multi-hop network.

Theorem 6.3.6 *The delay normalized transmission capacity of a wireless network with N_h-hop transmission with dedicated relays and end-to-end retransmission constraint of D is*

$$C_d = \frac{\lambda B P_s}{\sum_{n=1}^{N_h} \mathbb{E}\{M_n\}},$$

where P_s is given by Proposition 6.3.5, and $\mathbb{E}\{M_n\}$ is given by Proposition 6.2.3.

Similar to the single-hop case (Section 6.2.1), because of correlated SIRs across time slots and different hops, we again see that finding a closed-form expression for the success probability P_s and the delay normalized transmission capacity in terms of the end-to-end retransmissions constraint of D is not possible. The complicated derived expression for the delay-normalized transmission capacity of Theorem 6.3.6 limits our ability to find the optimal allocation of number of per-hop retransmissions D_n's, under the end-to-end retransmissions constraint of D, that is, $\sum_{n=1}^{N_h} D_n \leq D$, and the optimal number of hops N_h to use for transmission between any source and its destination. Both these questions are rather important for effective operation of the wireless network.

To gain more insight into the delay normalized transmission capacity, and facilitate finding the optimal D_n's and N_h, we derive a lower bound on P_s, and the delay normalized transmission capacity as follows.

6.3.2 Lower Bound on the Delay Normalized Transmission Capacity

By definition, the probability of successful reception of packet in at most D end-to-end retransmissions $P_s = P(\cap_{n=1,\ldots,N_h} S_{D_n})$, where

$$S_{D_n} = \{\text{success in} \leq D_n \text{ retransmissions on the } n\text{th hop}\}. \tag{6.28}$$

Event S_{D_n} is a decreasing event (Definition 2.8.7), since increasing the number of interferers decreases the probability of event S_{D_n}. Therefore, from the FKG inequality (Lemma 2.8.8), we get the following lower bound.

Lemma 6.3.7 *The success probability in a wireless network with N_h hop communication is lower bounded as $P_s \geq c^n \prod_{n=1}^{N_h} \left(1-(pq_{d_n}+1-p)^{D_n+1}\right)$, where $q_{d_n} = \mathbb{P}(\mathsf{SIR}_n \leq \beta)$ is the probability of failure in any one attempt on the nth hop, and $c \leq 1$ is a constant.*

Proof: $P_s = \mathbb{P}\left(\cap_{n=1,\ldots,N_h} S_{D_n}\right)$. Since S_{D_n} (6.28) is a decreasing event for each $n = 1, \ldots, N_h$,

$$P_s \geq \prod_{n=1}^{N_h} \mathbb{P}\left(S_{D_n}\right)$$

from the FKG inequality (Lemma 2.8.8). Result follows by substituting for the lower bound on $\mathbb{P}\left(S_{D_n}\right)$ from Proposition 6.2.6. □

To get some numerical insight, in Fig. 6.4, we plot the success probability P_s and the derived lower bound as a function of end-to-end retransmission constraints of D for a two-hop $N_h = 2$ network, for density $\lambda = 0.1$, SIR threshold $\beta = 3$, with $d_1 = d_2 = 1m$ and equally dividing the retransmissions constraint over two hops, $D_1 = D_2 = D/2$ and path-loss exponent $\alpha = 3$. Clearly, the lower bound appears to be tight.

Remark 6.3.8 *The lower bound on the success probability (Lemma 6.3.7) corresponds to the case when the success events on each hop are independent. Since the spatial correlation coefficient of interference in a PPP is zero with path-loss model of $x^{-\alpha}$ (Example 2.7.3), the derived lower bound is expected to be tight (can be verified using simulations).*

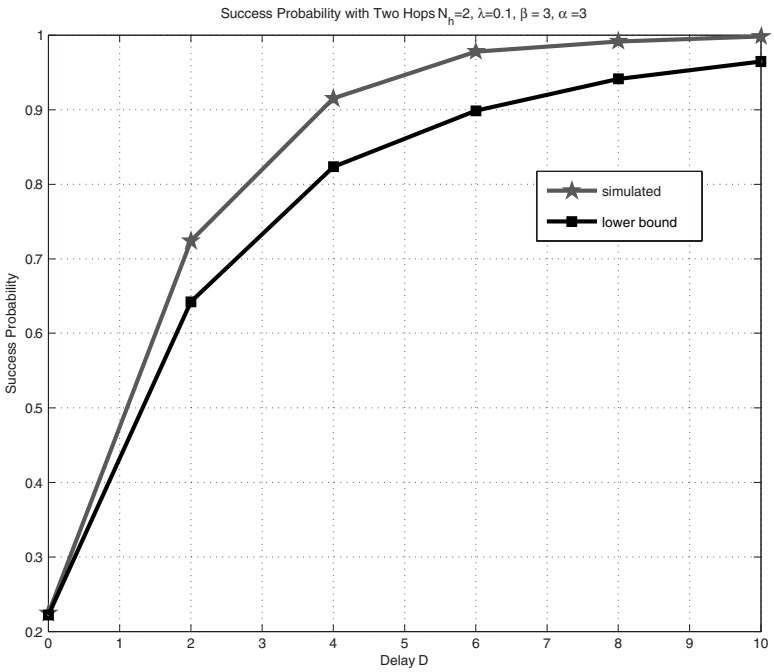

Figure 6.4: Success probability as a function of number of retransmissions D for a two-hop network $N_h = 2$ with $d_1 = d_2 = 1m$ and equally dividing the retransmissions constraint over two hops, $D_1 = D_2 = D/2$.

Now we move on to upper bound the expected end-to-end delay. Recall that the end-to-end retransmissions (delay) is $M = \sum_{n=1}^{N_h} M_n$, and by linearity of expectation $\mathbb{E}\{M\} = \sum_{n=1}^{N_h} \mathbb{E}\{M_n\}$, where $\mathbb{E}\{M_n\} = \sum_{j=0}^{D_n}(j+1)\mathbb{P}(M_n = j)$. Using (6.24), we get that

$$\mathbb{E}\{M_n\} \leq \left[\frac{c\left(1-(pq_{d_n}+1-p)^{D_n+1}\right)}{(1-q_{d_n})} + (D_n+1)(1-c) \right]. \tag{6.29}$$

Using Lemma (6.3.7) and (6.29), we obtain the following theorem.

Theorem 6.3.9 *The delay normalized transmission capacity of a wireless network with multi-hop transmission, under an end-to-end retransmissions constraint of D is lower bounded as*

$$C_d \geq \frac{\lambda B c^n \prod_{n=1}^{N_h}\left(1-(pq_{d_n}+1-p)^{D_n+1}\right)}{\sum_{n=1}^{N_h} \frac{c(1-(pq_{d_n}+1-p)^{D_n+1})}{(1-q_{d_n})} + (D_n+1)(1-c)} \text{ bits/sec/Hz/m}^2.$$

The lower bound (Theorem (6.3.9)) for the delay normalized capacity of multi-hop wireless network is far simpler to understand than the exact expression (Theorem (6.3.6)). It directly tells us the dependence of several important parameters of interest, for example, number of hops N_h, number of retransmission attempts to be made at hop n, and D_n, on the delay normalized transmission capacity.

Example 6.3.10 *For the special case of $D \to \infty$ that corresponds to no retransmissions constraint,*

$$C_d \propto \frac{1}{\sum_{n=1}^{N_h} \frac{1}{(1-q_{d_n})}}, \tag{6.30}$$

where we have dropped the additive term $(D_n+1)(1-c)$ in the denominator, since with $D=\infty$ there is no last retransmission. This is an intuitive result, since over the nth hop it takes on average $\frac{1}{1-q_{d_n}}$ attempts to receive any packet successfully with $D_n=\infty$, where $q_{d_n}=\mathbb{P}(\mathsf{SIR}_n \leq \beta)$ is the outage probability on the nth hop in a single transmission. Accounting for total N_h hops, we get the result. If all hops have the same outage probability, that is, $q_{d_n}=q,\ \forall\ n$, then from (6.30)*,*

$$C_d = \frac{q}{N_h}$$

for any n.

The delay normalized transmission capacity or the T-D-R trade-off (Theorem 6.3.6) of a multi-hop wireless network derived in this section allows us to fix any two of the QoS parameters (among throughput, delay, and reliability) and find the fundamental limit on the third parameter. The T-D-R trade-off defines a three-dimensional region and allows the network designer to choose any point on the boundary of the three-dimensional region. The derived delay normalized transmission capacity expression is easy to compute numerically, however, it is not amenable for finding direct analytical relations between the three quantities. To get more insights into the inter-dependence of throughput, delay, and reliability, leveraging FKG inequality, we first showed that the end-to-end success probability is lower bounded by the product of the success probabilities on each hop, and using that lower bound we then derived a lower bound on the delay normalized transmission capacity for simpler interpretation of the T-D-R trade-off.

Next, using the delay normalized transmission capacity framework, we answer some important questions in a multi-hop network, such as how to optimally divide the total retransmissions constraint D into per-hop retransmissions constraint D_n's, and find the optimal number of hops/relays to use between any source and destination.

Optimal Per-Hop Retransmissions

In this section, we first derive a lower bound that is simpler than Theorem 6.3.9 on the delay normalized transmission capacity, and then find the optimal per-hop maximum retransmissions D_n's under the end-to-end retransmissions constraint D that maximizes the lower bound. Note that the number of retransmissions allocated on each hop critically depends on the hop distance d_n, and consequently on the one slot failure probability on each hop $q_{d_n}=\mathbb{P}(\mathsf{SIR}_n \leq \beta)$. We have to take recourse in working with a simple lower bound on the delay normalized transmission capacity because the exact expression is a complicated function of q_{d_n}'s, and not directly amenable for analysis.

By definition, the number of retransmissions M_n used on hop n satisfy $M_n \leq D_n+1$. Thus, $\sum_{n=1}^{N_h} \mathbb{E}\{M_n\} \leq \sum_{n=1}^{N_h} D_n + 1 = D + N_h$. Then using the lower bound on the success probability P_s with end-to-end retransmission constraint of D (Lemma 6.3.7), the delay normalized transmission capacity (6.1) can be lower bounded as follows

$$C_d \geq \max_{D_n,\ \sum_{n=1}^{N_h} D_n \leq D} \frac{\lambda B c^n \prod_{n=1}^{N_h} \left(1-(pq_{d_n}+1-p)^{D_n+1}\right)}{D+N_h}. \tag{6.31}$$

Proposition 6.3.11 *The $D_n^\star$'s that maximize the lower bound (6.31) on the delay normalized transmission capacity satisfy*

$$D_n^\star + 1 = \frac{\ln\left(\frac{\zeta}{\ln(\hat{q}_{d_n})+\zeta}\right)}{\ln(\hat{q}_{d_n})},$$

where ζ is the Lagrange multiplier that satisfies $\sum_{n=1}^{N_h} D_n = D$. For equidistant hops $d_n = d/N$, $\forall\, n$, $D_n^\star = D/N$.

Proof: Let $\hat{q}_{d_n} = 1 - p + pq_{d_n}$, from (6.31), the objective function is

$$\max_{D_n, \sum_{n=1}^{N_h} D_n = D} \prod_{n=1}^{N_h} (1 - (\hat{q}_{d_n})^{D_n+1}).$$

Since $\log$ is a monotone function, an equivalent problem is

$$\max_{D_n, \sum_{n=1}^{N_h} D_n = D} \sum_{n=1}^{N_h} \ln(1 - \hat{q}_{d_n}^{D_n+1}). \tag{6.32}$$

It is easy to verify that the objective function (6.32) is a concave function of D_n's. Using Lagrange multiplier ζ, we can write the Lagrangian as

$$\mathbb{L} = \sum_{n=1}^{N_h} \ln(1 - \hat{q}_{d_n}^{D_n+1}) + \zeta(\sum_{n=1}^{N_h} D_n - D).$$

Differentiating with respect to D_n, and equating it to zero, we have

$$\frac{\mathrm{d}\mathbb{L}}{\mathrm{d}D_n} = \frac{-\ln(\hat{q}_{d_n})\hat{q}_{d_n}^{D_n+1}}{1 - \hat{q}_{d_n}^{D_n+1}} + \zeta = 0.$$

Thus, optimal $D_n^\star + 1 = \frac{\ln\left(\frac{\zeta}{\ln(\hat{q}_{d_n})+\zeta}\right)}{\ln(\hat{q}_{d_n})}$. An explicit solution for the optimal ζ cannot be found easily, hence we need to use an iterative algorithm to find optimal ζ. At each step ζ is increased if $\sum_{n=1}^{N_h} D_n \leq D$, or decreased if $\sum_{n=1}^{N_h} D_n > D$, similar to the Waterfilling solution [17]. For equidistant hops $d_n = d_{n+1},\ \forall\, n$, $\hat{q}_d = \hat{q}_{d_n}\ \forall\, n$, the optimal

$$\zeta = \frac{\ln(\hat{q}_d)\exp\left(\frac{\ln(\hat{q}_d)(D+N_h)}{N}\right)}{1 - \exp\left(\frac{\ln(\hat{q}_d)(D+N_h)}{N}\right)},$$

and $D_n^\star = \frac{D}{N_h}, \forall\, n$, if D is a multiple of N_h. □

To find the optimal number of per-hop retransmissions D_n's, we derived a simple lower bound on the delay normalized transmission capacity that is concave in D_n's, and derived the sufficient conditions for optimality using the KKT conditions. The optimal solution tries to give larger number of retransmissions on longer hops with smaller success probabilities, so as to minimize the difference in per-hop success probabilities across different hops. This is an intuitive result since if the per-hop success probabilities are very different, then the end-to-end success probability will be limited by the longest hop with the smallest success probability. For the special case of equidistant hops, $d_n = \frac{d}{N_h}$,

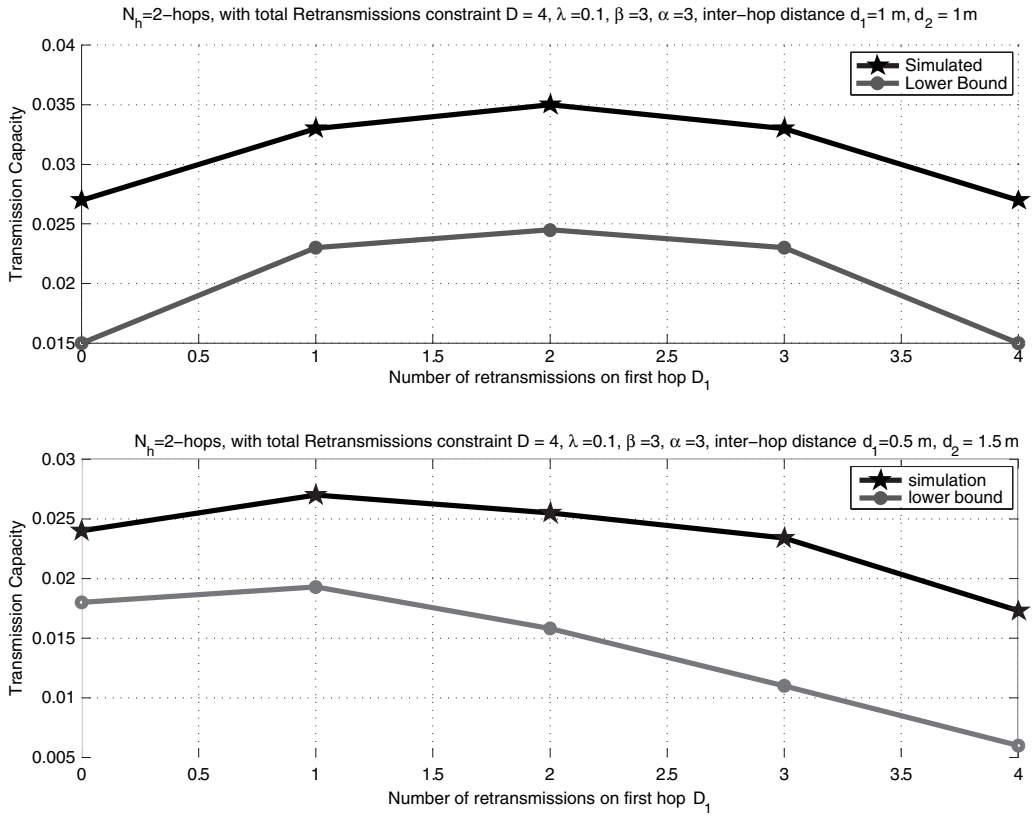

Figure 6.5: Delay normalized transmission capacity as a function of retransmissions used on first hop D_1 with total retransmissions $D = 4$, for equidistant and non-equidistant two hops $N_h = 2$.

we get the expected result that equally distributing D (the end-to-end delay constraint) among the N_h hops, maximizes the delay normalized transmission capacity.

To get numerical insights, in Fig. 6.5, we plot the exact delay normalized transmission capacity, and the derived lower bound (6.31) for two-hop communication $N_h = 2$, as a function of the number of retransmissions used on hop 1 (D_1), under an end-to-end retransmissions constraint of $D = 4$, for equidistant hops $d_1 = d_2 = 1\,\text{m}$, and non-equidistant hops $d_1 = 0.5\,\text{m},\ d_2 = 1.5\,\text{m}$, respectively, with $\lambda = 0.1$. The delay normalized transmission capacity (simulated and the lower bound) is maximized at $D_1 = D_2 = 2$ for $d_1 = d_2 = 1\,\text{m}$, and $D_1 = 1,\ D_2 = 3$ for $d_1 = 0.5\,\text{m},\ d_2 = 1.5\,\text{m}$ which is in accordance with Proposition 6.3.11.

After finding the optimal per-hop retransmissions allocation, in the next section, we answer a related important question on finding the optimal number of hops to use in a multi-hop wireless network.

Optimal Number of Hops N_h

In a multi-hop network, how many hops to use for transmitting information from the source to the destination is an important question to answer. Increasing the number of hops N_h between source and its destination decreases the distance between any two hops, and consequently increases the success probability on each hop, however, at the cost of increased end-to-end delay. Thus, we consider the delay normalized transmission capacity to find the optimal number of hops, since this metric inherently captures the incurred delay together with the success probability on each hop.

In [10], a detailed discussion and quantitative comparison of the performance of short hops versus long hops has been made for twelve different metrics. Most notable among these metrics are; overall energy efficiency, end-to-end reliability, routing overhead, and path efficiency, which is defined to be the ratio of total distance traversed by the packet and the actual distance between the source and destination. Overall, it is argued that the conventional wisdom of short hops being better than long hops is unfounded, and long hop (fewer number of hops) communication could be more efficient. We show that the conclusion of using fewer number of hops of [10] also holds in terms of maximizing the delay normalized transmission capacity, and in particular a single-hop $N_h = 1$ is optimal for small node densities λ. For analysis, we consider that each relay is located on a straight line between the source and the destination.

Proposition 6.3.12 *With equidistant hops, for a sparse network, where* $\exp(-\lambda\kappa) \approx 1 - \lambda\kappa$*, and* $\kappa = c\left(\frac{d}{N}\right)^2 \beta^{\frac{2}{\alpha}}$*, single-hop transmission (*$N_h = 1$*) maximizes the lower bound (6.31) on the transmission capacity.*

Proof: With equidistant relays located on a straight line between the source and the destination, the inter-hop distance $d_n = d/N,\ \forall\ n$, where d is the distance between the source and destination. Thus, the outage probability on each hop is equal and given by $q_{d_n} = q_{d/N}$, and the optimal $D_n = \lfloor D/N \rfloor, \forall\ n$ (Proposition 6.3.11). Hence, the lower bound on the delay normalized transmission capacity (6.31) specializes to

$$C_d \geq \frac{\lambda B \prod_{n=1}^{N_h} \left(1 - q_d^{\lfloor D/N_h \rfloor + 1}\right)}{D + N_h}. \tag{6.33}$$

Recall from (6.14) that $q_d = 1 - \exp\left(-\lambda c \left(\frac{d}{N_h}\right)^2 \beta^{\frac{2}{\alpha}}\right)$ for a constant c. Thus, to find the optimal number of hops N that maximize the lower bound (6.33), we need to solve the following optimization problem

$$\max_{N_h} \frac{\left(1 - \left(1 - \exp\left(-c\lambda \left(\frac{d}{N_h}\right)^2 \beta^{\frac{2}{\alpha}}\right)\right)^{\lfloor D/N_h \rfloor + 1}\right)^N}{D + N_h}.$$

Using the Taylor series expansion of $\exp\left(-\lambda c \left(\frac{d}{N}\right)^2 \beta^{\frac{2}{\alpha}}\right)$, and keeping only the first two terms, the optimization problem is

$$= \max_{N_h} \frac{\left(1 - \left(c\lambda \left(\frac{d}{N_h}\right)^2 \beta^{\frac{2}{\alpha}}\right)^{\lfloor D/N_h \rfloor + 1}\right)^N}{D + N_h},$$

$$= \max_{N_h} \frac{1}{D+N_h} - \frac{\left(c\lambda d^2 \beta^{\frac{2}{\alpha}}\right)^{\lfloor D/N_h \rfloor + 1} \left(\frac{1}{N_h}\right)^{\frac{D}{N_h}}}{D+N_h} + \mathcal{O}(\lambda^2). \tag{6.34}$$

Note that for small $\lambda d^2 \beta^{\frac{2}{\alpha}}$ for which the Taylor series expansion is valid, (6.34) is a decreasing function of N_h; thus, $N_h = 1$ maximizes the delay normalized transmission capacity by maximizing (6.34).

□

Remark 6.3.13 *For $\lambda \to 0$, from Proposition 6.3.5, it is easy to see that the success probability $P_s \to 1$, and on average only one transmission is required for successful transmission on each hop. Hence, the delay normalized transmission capacity $C_d \approx \frac{1}{N_h}$ for $\lambda \to 0$, and clearly, a single-hop ($N_h = 1$) maximizes the delay normalized transmission capacity. Proposition 6.3.12 shows that even for values of λ for which $\exp(-\lambda\kappa) \approx 1 - \lambda\kappa$, ($\kappa$ is a constant), $N_h = 1$ maximizes the delay normalized transmission capacity.*

Proposition 6.3.12 shows that in a sparse network regime with equidistant hops, it is optimal to transmit over a single-hop. The physical interpretation of this result is that in a sparse network with few interferers, the decrease in transmission capacity due to the end-to-end delay (linear in N_h) outweighs the increase in per-hop success probability because of the reduced per hop distance $\left(\frac{d}{N_h}\right)$. In general, however, using single-hop is not optimal, and finding optimal number of hops remains an open problem, and one has to resort to extensive simulations to answer this question.

To get a better understanding, in Fig. 6.6, we plot the delay normalized transmission capacity as the function of the number of hops N_h with end-to-end retransmissions constraint $D = 10$ for transmission density $\lambda = 0.1$ and $\lambda = 0.5$, respectively. From Fig. 6.6, we can see that for smaller densities, for example, $\lambda = 0.1$, single-hop transmission is optimal (as derived in Proposition 6.3.12), however, as we increase to $\lambda = 0.5$, that is no longer true.

In the next section, we generalize the fixed source–destination distance with dedicated relays model of wireless networks studied in this section, to allow for random source–destination distances and non-dedicated relays, which is fundamentally different from the dedicated relays model.

6.4 Shared Relays Multi-Hop Communication Model

In Section 6.3, we considered a fixed distance multi-hop communication model, where the source–destination distance is fixed, and there are N_h dedicated relays between them. In this section, we generalize that earlier restrictive model to allow for random distances between each source and its destination [11], and more importantly there are no dedicated relays for any source–destination pair. Each-source destination pair is chosen uniformly randomly from the set of nodes whose locations are distributed as a PPP, with the expected source–destination distance of W. We let each node to be a source or destination for some other node, and all nodes act as relays for communication between multiple source–destination pairs.

We consider the location of nodes $\{L_n\} = \Phi$ to follow a PPP with density λ_0. At time t, each node n is tagged with a indicator random variable $\mathbf{1}_n(t) \in \{0, 1\}$, where if $\mathbf{1}_n(t) = 1$ it acts as a transmitter, otherwise as a receiver. Note that each node at time t can be acting as a relay or a source or a destination. Each $\mathbf{1}_n(t)$ is independent for all n, and t, with $\mathbb{E}\{\mathbf{1}_n(t)\} = p$. The set of transmitting (receiving) nodes at time t is denoted as $\Phi_T(t) = \{L_n \in \Phi : \mathbf{1}_n(t) = 1\}$ ($\Phi_R(t) = \{L_n \in \Phi : \mathbf{1}_n(t) = 0\}$).

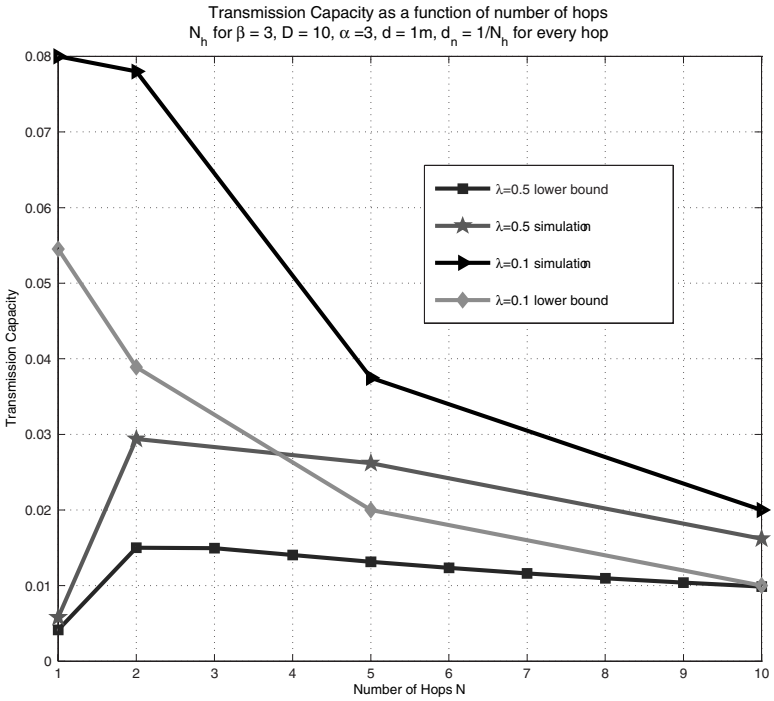

Figure 6.6: Delay normalized transmission capacity as a function of number of hops N_h for $\lambda = 0.1$ and $\lambda = 0.5$.

Under this model, the received signal at $R_j \in \Phi_R(t)$ is given by

$$y_j(t) = d_{ij}^{-\alpha/2} h_{ij}(t) x_i(t) + \sum_{m:T_m \in \{\Phi_T(t) \setminus \{T_i\}\}} d_{mj}^{-\alpha/2} h_{mj}(t) x_m(t) + w, \tag{6.35}$$

where the first term is the signal received from $T_i \in \Phi_T(t)$, $x_m(t) \in \mathcal{CN}(0,1)$ is the transmitted signal from node m at time t, d_{mj} is the distance between T_m and R_j, $h_{mj}(t)$'s are the fading coefficients between T_m and R_j that are i.i.d. Rayleigh-distributed random variable, and w is the AWGN with variance N_0. This is the only place in this book, where considering or neglecting AWGN produces dramatically different results. Thus, we consider the AWGN variance explicitly.

We assume that the transmission between transmitter $T_i \in \Phi_T(t)$ and $R_j \in \Phi_R(t)$ is successful (denoted by indicator function $e_{i,j}(t)$) at time t if the

$$\mathsf{SINR}_{ij}(t) = \frac{d_{ij}^{-\alpha} |h_{ij}(t)|^2}{\mathsf{N}_0 + \sum_{m:T_m \in \{\Phi_T(t) \setminus \{T_i\}\}} d_{mj}^{-\alpha} |h_{mj}(t)|^2}$$

between them is greater than a threshold β, where $h_{ij}(t)$ is the channel coefficient, and d_{ij} is the distance between nodes T_i and R_j at time t, respectively. Thus, $e_{i,j}(t) = 1$ if $\mathsf{SINR}_{ij}(t) > \beta$, and zero otherwise.

Since SINR_{ij}'s are random variables, multiple transmissions are required for a packet sent by T_i to be successfully received at R_j. To further the discussion on the total number of transmissions used by a packet to reach its destination successfully, we define the following two quantities.

Definition 6.4.1 *Let the minimum time taken by any packet to successfully reach R_j from T_i be*

$$M_{ij} = \min\{k > 0 : e_{i,j}(k) = 1\}.$$

Definition 6.4.2 *The exit time is defined as the minimum time taken by any packet to be successfully received at any other node of Φ from T_i, that is,*

$$M_i = \min_{s \in \Phi \backslash \{T_i\}} M_{is}.$$

The exit time M_i is the minimum time required for a packet to leave T_i and be successfully received at any other node and clearly a lower bound on the time taken by a packet to reach its destination. We next show a negative result that as long as the noise variance $N_0 > 0$, the expected exit time is infinite. Thus, in a sense, it takes infinite time for a packet to reach its destination, making the rate of transmission close to zero for any source–destination pair.

Theorem 6.4.3 *With PPP-distributed node locations, for AWGN variance $N_0 > 0$, the expected exit time $\mathbb{E}\{M_i\} = \infty$.*

Proof: To prove the result, we will neglect interference in (6.35) and replace the $\mathsf{SINR}_{ij}(t)$ with $\mathsf{SNR}_{ij}(t)$, where

$$\mathsf{SNR}_{ij}(t) = \frac{d_{ij}^{-\alpha}|h_{ij}(t)|^2}{\mathsf{N}_0}.$$

Neglecting interference, we denote all relevant quantities with a hat, that is, $\hat{e}_{i,j}(t) = 1$ if $\mathsf{SNR}_{ij}(t) > \beta$ and zero otherwise, similarly for $\hat{M}_{ij}$ and the exit delay $\hat{M}_i$ etc. The basic idea behind neglecting interference is the lower bound

$$\hat{M}_i \leq M_i$$

since $\mathsf{SINR}_{ij}(t) \leq \mathsf{SNR}_{ij}(t)$, and hence if there is no interference, it takes less time for a packet to exit T_i.

To complete the proof, we show that

$$\mathbb{E}\{\hat{M}_i\} = \infty.$$

To show that $\mathbb{E}\{\hat{M}_i\} = \infty$, it is sufficient to show that $\mathbb{P}\left(\hat{M}_i > k\right) > \frac{1}{k}$ for $k \in \mathbb{N}$ as follows.

We begin by conditioning over the PPP Φ, where the remaining randomness, induced by the channel coefficient $h_{mj}(t)$, is independent across different users and time slots. Essentially, given Φ, $\mathsf{SNR}_{ij}(t)$ are independent for $\forall\ j$ and $\forall\ t$. In any slot, the transmission from T_i to node L_j is unsuccessful if L_j is in transmitting mode $\mathbf{1}_j(t) = 1$ ($L_j \in \Phi_T$) or $\mathsf{SNR}_{ij}(t) \leq \beta$ if $\mathbf{1}_j(t) = 0$ with $L_j \in \Phi_R$.

Thus, we can write

$$\begin{aligned}
\mathbb{P}\left(\hat{M}_i > k|\Phi\right) &= \mathbb{P}\left(1 \leq t \leq k, \forall\ L_j \in \Phi\backslash\{T_i\} : \mathbf{1}_j(t) = 1 \text{ or } \mathsf{SNR}_{ij}(t) \leq \beta|\Phi\right), \\
&\overset{(a)}{=} \left(\mathbb{P}\left(\forall\ L_j \in \Phi\backslash\{T_i\}, \mathbf{1}_j(t) = 1 \text{ or } |h_{ij}(t)|^2 \leq d_{ij}^{\alpha}\mathsf{N}_0\beta\right)\right)^k, \\
&\overset{(b)}{=} \prod_{L_j \in \Phi\backslash\{T_i\}} \left(p + (1-p)\mathbb{P}\left(|h_{ij}(t)|^2 \leq d_{ij}^{\alpha}\mathsf{N}_0\beta\right)\right)^k,
\end{aligned}$$

$$\stackrel{(c)}{=} \prod_{L_j \in \Phi \setminus \{T_i\}} \left(p + (1-p)\left(1 - \exp\left(-d_{ij}^{\alpha} \mathsf{N}_0 \beta\right)\right)\right)^k,$$

$$= \exp\left(k \sum_{L_j \in \Phi \setminus \{T_i\}} \log\left(p + (1-p)\left(1 - \exp\left(-d_{ij}^{\alpha} \mathsf{N}_0 \beta\right)\right)\right)\right), \tag{6.36}$$

where (a) follows from the definition of $\mathsf{SNR}_{ij}(t)$, and since given Φ, $\mathsf{SNR}_{ij}(t)$ and $\mathbf{1}_j(t)$ are i.i.d. for t, (b) follows by taking the expectation with respect to $\mathbf{1}_j(t)$ since $\mathbb{E}\{\mathbf{1}_j(t)\} = p$, (c) follows since $|h_{ij}(t)|^2 \sim \exp(1)$. Now we make use of the probability-generating functional (Theorem 2.3.6) to uncondition (6.36) over Φ, as follows.

$$\begin{aligned}
\mathbb{P}\left(\hat{M}_i > k\right) &= \mathbb{E}_{\Phi}\left\{\mathbb{P}\left(\hat{M}_i > k | \Phi\right)\right\}, \\
&\stackrel{(d)}{=} \exp\left(-2\pi\lambda_0 \int_{v>0} \left(1 - \left(1 - (1-p)\exp\left(-v^{\alpha}\mathsf{N}_0\beta\right)\right)^k\right) v \mathrm{d}v\right), \\
&\stackrel{(e)}{=} \exp\left(-\pi\lambda_0 \int_{v>0} \left(1 - (1 - f(v))^k\right) \mathrm{d}v\right),
\end{aligned} \tag{6.37}$$

where (d) follows from (6.36) using Theorem 2.3.6 and replacing d_{ij} by v, and (e) follows by defining $f(v) = (1-p)\exp\left(-v^{\alpha}\mathsf{N}_0\beta\right)$. Let v_k be a unique solution of $f(v) = \frac{1}{k}$, that is,

$$v_k = \frac{1}{(\mathsf{N}_0\beta)^{2/\alpha}} \log(k(1-p)))^{2/\alpha}. \tag{6.38}$$

Since $f(v) \to 0$ as $v \to \infty$ and $v_k \to \infty$ as $k \to \infty$, there exists a constant $\kappa \leq \infty$ such that for all $k \geq \kappa$ and $v \geq v_q$,

$$(1 - f(v)) \geq \exp\left(-f(v)\right).$$

Hence, with some manipulations, we can get that for all $k \geq \kappa$,

$$\int_{v>0} \left(1 - (1 - f(v))^k\right) \mathrm{d}v \quad \leq \quad v_k + \int_{u=0}^{\infty} k f(u + v_k) \mathrm{d}u. \tag{6.39}$$

Using the definition of $f(v)$, we get

$$\begin{aligned}
\int_{v>0} k f(u+v_k)\mathrm{d}u &= \int_{u=0}^{\infty} k(1-p)\exp\left(-(u+v_k)^{\alpha}\mathsf{N}_0\beta\right)\mathrm{d}u, \\
&\leq \int_{u=0}^{\infty} k(1-p)\exp\left(-(u+v_k^{\alpha})\mathsf{N}_0\beta\right)\mathrm{d}u, \\
&= \frac{1}{\mathsf{N}_0\beta},
\end{aligned} \tag{6.40}$$

where the second inequality follows since for large enough k, $(u+v_k)^{\alpha} \geq u + v_k^{\alpha}$, and the last equality follows, since by definition $f(v) = (1-p)\exp\left(-v^{\alpha}\mathsf{N}_0\beta\right) = \frac{1}{k}$.

Using (6.40) in (6.39), we get for $k \geq \kappa$

$$\int_{v>0} \left(1-(1-f(v))^k\right) \mathrm{d}v \quad \leq \quad v_k + \frac{1}{\mathsf{N}_0\beta}. \tag{6.41}$$

This implies that for $k \geq \kappa$,

$$\begin{aligned} \exp\left(-\pi\lambda_0 \int_{v>0} \left(1-(1-f(v))^k\right) \mathrm{d}v\right) \quad &\geq \quad \exp\left(-\pi\lambda_0 \left(v_k + \frac{1}{\mathsf{N}_0\beta}\right) \mathrm{d}v\right), \\ &\geq \quad \frac{1}{k}, \end{aligned} \tag{6.42}$$

where the last inequality follows since for $\alpha > 2$, from (6.38)

$$v_k \leq \frac{\log k}{\pi\lambda_0} - \frac{1}{\mathsf{N}_0\beta}.$$

Substituting (6.42) into (6.37), we have proved that

$$\mathbb{P}\left(\hat{M}_i > k\right) > \frac{1}{k},$$

thus proving that the expected exit delay $\mathbb{E}\{\hat{M}_i\} = \infty$, and consequently $\mathbb{E}\{M_i\} = \infty$.

□

Theorem 6.4.3 tells us that on average, it takes an unbounded number of retransmissions for a packet to be successfully received at any other node. This result is essentially a manifestation of the fact that in PPP-distributed network, for a given realization of node locations, there can be large enough area/void around a typical node that contains no other node. Thus, even without considering interference, no matter how many retransmission attempts are made, for a typical node, the effect of additive noise cannot be overcome to support a minimum SNR at any of the other receiving nodes.

This clearly is a very disappointing result and does not allow for any meaningful capacity metric definition with PPP-distributed nodes, since the effective transmission rate is arbitrarily close to zero, as it takes infinitely many retransmissions for a packet to even leave its source, let alone reach the destination. Instead of assuming a network with PPP-distributed nodes, if nodes are assumed to lie uniformly at random in a bounded area, this limitation can be overcome, and meaningful capacity metric of transport and throughput capacity can be defined. In Chapter 9, we discuss and analyze the transport and throughput capacity in detail.

To define a useful notion of capacity in a network with PPP-distributed nodes, one alternative is to ignore the AWGN contribution. Ignoring the AWGN contribution, the first encouraging result one can obtain is that the mean exit time from a transmitter to its nearest neighbor is finite. To derive this result, conditioned on the PPP Φ and the distance to the nearest neighbor, note that the success probability P_s is independent across multiple time slots, and hence, the expected time for successful reception at the nearest neighbor is $\frac{1}{P_s}$. Unconditioning over Φ and the nearest neighbor distance, one can explicitly show the finiteness of the exit time. We leave the complete proof of this result as an exercise.

The finite nearest neighbor expected exit time is an encouraging result, since it allows for the possibility of the packet reaching its destination in finite time. To make this more concrete, we next define a notion of capacity called the *spatial progress capacity* and show that it is non-zero for a network with PPP-distributed nodes as long as AWGN variance is zero or AWGN contribution can be neglected.

6.4.1 Spatial Progress Capacity

The system model remains the same as defined in Section 6.4, except now we assume that the AWGN variance in (6.35) is $N_0 = 0$. Without loss of generality, let source $\mathcal{S}_0 = T_0$ be located at the origin and its destination lie on the x-axis at a random distance w from T_0, with mean W. For defining the spatial progress capacity, we will not directly require multiple transmission attempts by any node, and hence we suppress the dependence of time on all quantities of interest such as SINR, SIR, Φ_T, Φ_R, e_{ij}, and $\mathbf{1}_n$. Thus, $\Phi_T = \{L_n \in \Phi : \mathbf{1}_n = 1\}$ ($\Phi_R = \{L_n \in \Phi : \mathbf{1}_n = 0\}$). With $N_0 = 0$, from (6.35), for transmitter $T_i \in \Phi_T$ and receiver $R_j \in \Phi_R$, $e_{ij} = 1$ if $\mathsf{SIR}_{ij} > \beta$, where

$$\mathsf{SIR}_{ij} = \frac{d_{ij}^{-\alpha}|h_{ij}|^2}{\sum_{m:T_m \in \{\Phi_T \setminus \{T_i\}\}} d_{mj}^{-\alpha}|h_{mj}|^2}.$$

We assume that whenever a node has a packet to transmit, it forward it to its farthest neighbor in the direction toward the packet's destination. Therefore, the effective progress w_{eff} made in one hop by any packet toward its destination is given by the maximum distance of nodes $R_j \in \Phi_R$ from T_i such that $e_{i,j} = 1$. Therefore, for node T_i, the effective spatial progress of its packets toward their destination in one hop is

$$\mathsf{w}_{\text{eff}} = \max_{R_j \in \Phi_R} \left\{(e_{i,j} = 1)|R_j|(\cos(\theta(R_j)))^+\right\},$$

where $\theta(R_j)$ is the angle node R_j makes with the x-axis in the counter-clockwise direction. In Fig. 6.7, we pictorially illustrate the progress of packets towards their destination.

Let the expected spatial progress over one hop be $\mathsf{W}_{\text{eff}} = \mathbb{E}\{\mathsf{w}_{\text{eff}}\}$. Since the expected source–destination distance is W, it takes on average $\frac{\mathsf{W}}{\mathsf{W}_{\text{eff}}}$ hops for a packet to reach its destination. Since at each time slot, transmission happens with probability p from any node, on average $p\frac{\mathsf{W}_{\text{eff}}}{\mathsf{W}}$ packets are drained from source per slot. Thus, we define

$$C_s = p\frac{\mathsf{W}_{\text{eff}}}{\mathsf{W}}$$

as per-node spatial progress capacity of the network with PPP-distributed nodes. Next, we show that with $N_0 = 0$, $C_s = \Omega\left(\frac{1}{\sqrt{\lambda}}\right)$ similar to the throughput capacity derived in Chapter 9, where nodes are distributed uniformly at random in a bounded area with density λ.

To derive the spatial capacity C_s, we need to find a closed-form expression for W_{eff}, which is rather challenging. To simplify analysis, instead we consider $\tilde{\mathsf{w}}_{eff}$, that we obtain by replacing the indicator variable $e_{i,j}$ with its expectation in the definition of w_{eff}. Note that $\mathbb{E}\{e_{ij}\} = \mathbb{P}(\mathsf{SIR}_{ij} > \beta)$, for which we know the exact expression from Theorem 2.4.1. Hence, we define

$$\tilde{\mathsf{w}}_{eff} = \max_{R_j \in \Phi_R} \left\{\mathbb{P}(\mathsf{SIR}_{ij} > \beta)|R_j|(\cos(\theta(R_j)))^+\right\},$$

and $\tilde{\mathsf{W}}_{eff} = \mathbb{E}\{\tilde{\mathsf{w}}_{eff}\}$. One can show using Jensen's inequality that $\mathsf{W}_{eff} \geq \tilde{\mathsf{W}}_{eff}$, thus it is useful to work with $\tilde{\mathsf{W}}_{eff}$ to obtain a lower bound on the per-node spatial progress capacity. We next derive an expression for $\tilde{\mathsf{W}}_{eff}$ as follows

Proposition 6.4.4 $\tilde{\mathsf{W}}_{eff} = \frac{c}{\beta^{1/\alpha}\sqrt{\lambda_0 p}} H(p, \beta)$ *for constant c, where* $H(p, \beta)$ *is defined in the proof.*

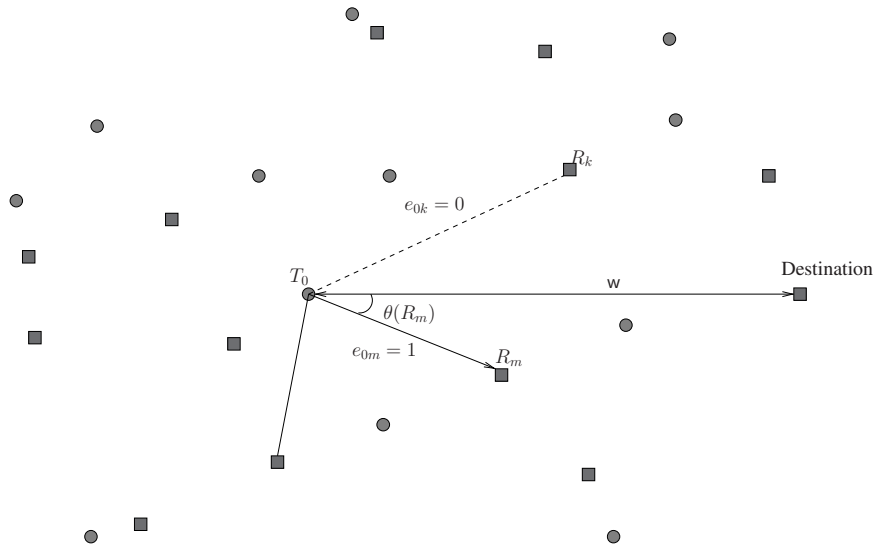

Figure 6.7: Transmission model for multi-hop communication, where dots are transmitters and squares are receivers, and the spatial progress for T_0 is the largest projection of squares on the x-axis for which $e_{0j} = 1$.

Proof: The effective per-hop spatial progress

$$\tilde{\mathsf{w}}_{eff} = \max_{R_j \in \Phi_R} \left\{ \mathbb{P}(\mathsf{SIR}_{ij} > \beta) |R_j| (\cos(\theta(R_j)))^+ \right\}$$

is of the form of an extremal shot noise process

$$\max_{x_i \in \Phi_R} g(x_i),$$

where $g(x) = p_{|x|}|x|(\cos(\theta(x)))^+$ and $p_{|x|}(.)$ is a probability measure. For the extremal shot noise process, its distribution is known via the Laplace transform [11] for a PPP Φ_R to be

$$\mathbb{P}(\max_{x_i \in \Phi_R} g(x_i) \leq z) = \mathbb{E}\left\{ \exp\left(\sum_{x_i \in \Phi_R} \ln(\mathbf{1}_{g(x_i) \leq z}) \right) \right\}. \tag{6.43}$$

Using (6.43), and the probability-generating functional (Theorem 2.3.6) for PPP Φ_R with density $(1-p)\lambda_0$, we get

$$\mathbb{P}(\tilde{\mathsf{w}}_{eff} \leq z) = \exp\left(-\lambda_0 (1-p) \int_{\mathbb{R}^2} \mathbf{1}_{g(x_i) > z} \mathrm{d}x \right).$$

After some manipulations, using the definition $\mathbb{E}\{X\} = \int_0^\infty \mathbb{P}(X > x)\mathrm{d}x$, we get

$$\tilde{\mathsf{W}}_{eff} = \mathbb{E}\{\tilde{\mathsf{w}}_{eff}\} = \frac{c}{\beta^{2/\alpha}\sqrt{\lambda_0 p}} H(p, \beta),$$

where $H(p,\beta) = \int_0^1 1 - \exp\left(\left(1-\frac{1}{p}\right)\frac{G(z)}{c_3\beta^{2/\alpha}}\right) \mathrm{d}z$ and

$$G(z) = 2\int_{t:\frac{e(t)}{\sqrt{2et}} \leq \frac{1}{z}} arc\cos\left(\frac{ze(t)}{\sqrt{2et}}\right) \mathrm{d}t.$$

□

Using the fact that $\mathsf{W}_{eff} \geq \tilde{\mathsf{W}}_{eff}$, from Proposition 6.4.4, it follows that the per-node spatial progress capacity

$$C_s = \Omega\left(\frac{1}{\sqrt{\lambda}}\right),$$

where $\lambda = p\lambda_0$.

Thus, the per-node spatial progress capacity scales as $\Omega\left(\frac{1}{\sqrt{\lambda}}\right)$, and accounting for all simultaneous transmissions across the network, the network-wide spatial progress capacity is $\Omega(\sqrt{\lambda})$. Thus, on average, $\Omega(\sqrt{\lambda})$ bits can be pushed toward their destinations in each time slot.

It is instructive to compare the per-node spatial progress capacity C_s with the per-node throughput capacity (Chapter 9). The per-node throughput capacity is defined as the per-node throughput that can be simultaneously supported by all source–destination pairs with high probability, when nodes are distributed uniformly at random in a unit area. Similar to the per-node spatial progress capacity scaling result of $\Omega\left(\frac{1}{\sqrt{\lambda}}\right)$, in Chapter 9, we will show that the per-node throughput capacity is $\Theta\left(\frac{1}{\sqrt{\lambda}}\right)$ with density λ. However, there are two main differences between the spatial progress capacity and throughput capacity framework i) spatial progress capacity is an expected rate of transmission, while the throughput capacity guarantees certain rate with high probability and, ii) to derive spatial progress capacity we needed AWGN noise variance $N_0 = 0$, while such a restriction is not required for deriving throughput capacity as long as we consider that the nodes are distributed in a bounded area.

6.5 Reference Notes

The delay normalized transmission capacity was first proposed in [12], and its detailed analysis can be found in [13]. A qualitative and quantitative performance comparison of short hops versus long hops in a wireless network can be found in [10], though it does not explicitly consider interference or SINR model. The infinite expected exit time result for shared relays model was derived in [14], while the analysis of spatial progress capacity under the no noise assumption can be found in [11]. Some related results on dynamic time-dependent connectivity properties can be found in [15].

Bibliography

[1] D. Cox. 1962. *Renewal Theory*. Methuen & Co.

[2] A. Kumar, E. Altman, D. Miorandi, and M. Goyal. 2007. "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs." 15 (3): 588–01.

[3] M. Haenggi. 2005. "On routing in random Rayleigh fading networks." *IEEE Trans. Wireless Commun.* 4 (4): 1553–62.

[4] S. Weber, N. Jindal, R. K. Ganti, and M. Haenggi. 2008. "Longest edge routing on the spatial ALOHA graph." In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*. 1–5.

[5] K. Stamatiou, F. Rossetto, M. Haenggi, T. Javidi, J. Zeidler, and M. Zorzi. 2009. "A delay-minimizing routing strategy for wireless multi-hop networks." In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on* June 2009 1–6.

[6] K. Stamatiou and M. Haenggi. 2010. "Random-access Poisson networks: Stability and delay." *IEEE Commun. Lett.* 14 (11): 1035–37.

[7] ——. 2010. "The delay-optimal number of hops in poisson multi-hop networks." In *Proc. (ISIT) Symp. IEEE Int Information Theory*. 1733–37.

[8] M. Haenggi. 2010. "Local delay in static and highly mobile Poisson networks with Aloha." In *Proc. IEEE Int Communications (ICC) Conf* 1–5.

[9] T. Cover and J. Thomas. 2004. *Elements of Information Theory*. John Wiley and Sons.

[10] M. Haenggi. 2004. "Twelve reasons not to route over many short hops." In *Proceedings of VTC2004-Fall Vehicular Technology Conference 2004 IEEE 60th*, vol. 5. 3130–34.

[11] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler. 2006. "An ALOHA protocol for multihop mobile wireless networks." *IEEE Trans. Inf. Theory* 52 (2): 421–36.

[12] J. Andrews, S. Weber, M. Kountouris, and M. Haenggi. 2010. "Random access transport capacity." *IEEE Trans. Wireless Commun.* 9 (6): 2101 –11.

[13] R. Vaze. 2011. "Throughput-delay-reliability tradeoff with ARQ in wireless ad hoc networks." *IEEE Trans. Wireless Commun.* 10 (7): 2142 –49.

[14] F. Baccelli, B. Błaszczyszyn, and M.-O. Haji-Mirsadeghi. 2011. "Optimal paths on the space-time SINR random graph." *Adv. Appl. Prob.* 43 (1): 131–50.

[15] R. Ganti and M. Haenggi. 2007. "Dynamic connectivity and packet propagation delay in ALOHA wireless networks." In *Proceedings of IEEE Asilomar*, Pacific Grove, CA.

Chapter 7

Percolation Theory

7.1 Introduction

Percolation theory studies the phenomenon of formation of unbounded connected clusters in large graphs, and percolation is defined as the event that there exists an unbounded connected component in a graph. Any wireless network can be naturally thought of as a graph, where the presence of an edge/connection between any two nodes can be defined in variety of ways. Percolation in a wireless network corresponds to having long-range connectivity, that is, nodes that are far apart in space have a connected path between them.

The objective of this chapter and the next is two-fold: study percolation properties of wireless networks and lay sufficient background required for deriving the throughput capacity of wireless networks in Chapter 9.

In this chapter, we begin by introducing basics of discrete and continuum percolation. We first consider discrete percolation over square lattice, where each edge of the square lattice is assumed to be open or closed independently of each other with probability $p_{\rm o}$ and $1-p_{\rm o}$, respectively. We show that there is a phase transition at $p_{\rm o}=p_c$, such that for $p_{\rm o}>p_c$ ($p_{\rm o}<p_c$), the probability of percolation in non-zero (zero), and find non-trivial bounds for p_c. The techniques presented to derive bounds on p_c are also helpful for obtaining percolation results for wireless networks, and finding throughput capacity of wireless networks in Chapter 9. We also state results for the discrete face percolation over a hexagonal lattice that are relevant for finding percolation regimes of wireless networks.

Next, to model a wireless communication network, we consider continuum percolation, where node locations are drawn from a spatial distribution, and connections are defined appropriately. First, we study the percolation properties of the most popular Gilbert's disc model, where two nodes are connected if they are within a fixed distance (radio range) r from each other, inspired by point-to-point wireless communication under a path-loss model. We find percolation properties of the Gilbert's disc model as a function of the density of nodes and radio range r. Then we consider the more difficult problem of ensuring connectivity in the Gilbert's disc model, where each node is required to have a connected path to every other node, that is, the graph defined by the Gilbert's disc model should be completely connected. Assuming n nodes are located in a unit disc, we find conditions on the minimum radio range $r(n)$ that ensures connectivity with high probability in the Gilbert's disc model.

We then generalize the Gilbert's disc model to better suit a heterogeneous wireless network model by allowing each node to have an i.i.d. random radio range called the Gilbert's random disc model and find its percolation properties. Finally, we consider the point-to-point wireless communication network with a path-loss plus fading model, where two nodes are connected if the signal strength between them is above a threshold. We make use of the results derived for identifying percolation properties of the Gilbert's random disc model to find the percolation properties of the wireless communication network with the path-loss plus fading model.

7.2 Discrete Percolation

In this section, we will consider percolation for graphs defined over lattices, typically referred to as discrete percolation. We begin with the simplest model of percolation over square lattice $\mathbb{Z}^2$, and follow it up with percolation over a discrete hexagonal face grid that is useful for studying percolation regimes for wireless networks.

7.2.1 Square Lattice Percolation

Consider the square lattice $\mathbb{Z}^2$ shown in Fig. 7.1, where each edge between any two adjacent lattice points is *open* (bold line) with probability p_{o} and *closed* (dotted line) with probability $1 - p_{\mathrm{o}}$ independently of all other edges. Let the connected component of any arbitrary lattice point be all the lattice points that are reachable from it using only the *open* edges. Assume that we want to ask the question : when is the probability that the size of the connected component of any arbitrary lattice point is unbounded is non-zero. This is a vital question and the main object of study in percolation theory [1, 1, 2, 4], introduced by Broadbent and Hammersley [5], where the quest is to find the smallest (critical) probability p_c, such that for $p_{\mathrm{o}} > p_c$, the probability that the connected component of any arbitrary lattice point is unbounded, is non-zero. We begin with some preliminaries that allow us to get some insights on p_c.

Definition 7.2.1 *We define that there is a path from a lattice point $x \in \mathbb{Z}^2$ to $y \in \mathbb{Z}^2$, if y can be reached from x using only the open edges. A path from x and y is represented as $x \rightarrow y$.*

Definition 7.2.2 *The connected component of any node $x \in \mathbb{Z}^2$ is defined as $\mathsf{C}_x := \{y \in \mathbb{Z}^2, x \rightarrow y\}$, with cardinality $|\mathsf{C}_x|$.*

We define $\theta_x(p_{\mathrm{o}})$, as the probability that the connected component of lattice point x is of infinite size when the edge open probability is p_{o}, that is,

$$\theta_x(p_{\mathrm{o}}) = \mathbb{P}_{p_{\mathrm{o}}}(|\mathsf{C}_x| = \infty). \tag{7.1}$$

We drop the subscript p_{o} from $\mathbb{P}_{p_{\mathrm{o}}}(.)$ whenever possible. Since the lattice $\mathbb{Z}^2$ is symmetric or translationally invariant, for the purposes of computing $\theta_x(p_{\mathrm{o}})$, there is no way of distinguishing between $\theta_x(p_{\mathrm{o}})$ and $\theta_y(p_{\mathrm{o}})$. Thus, $\theta_x(p_{\mathrm{o}})$ does not depend on the choice of the lattice point x, and we can drop the subscript x from $\theta_x(p_{\mathrm{o}})$, and just consider $\theta(p_{\mathrm{o}})$, and find conditions on edge open probability p_{o} such that $\theta(p_{\mathrm{o}}) > 0$. Similarly, we drop subscript x from the connected component, and just consider the connected component C of the origin.

Definition 7.2.3 *Event $|\mathsf{C}| = \infty$ is referred as percolation and $\theta(p_o)$ is referred to as the percolation probability.*

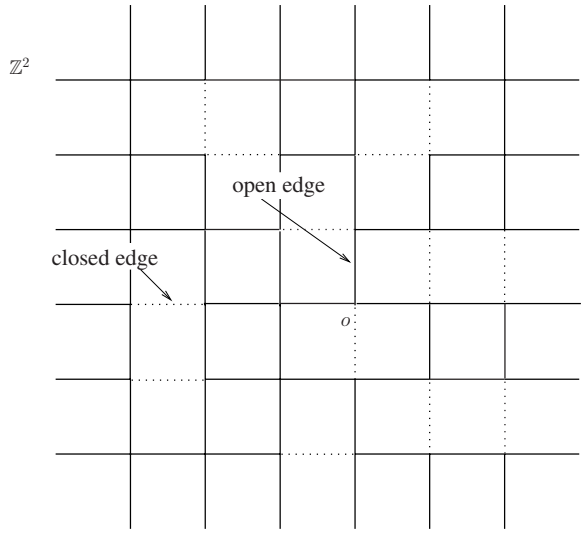

Figure 7.1: Square lattice $\mathbb{Z}^2$ with open and closed edges.

Next, we define the critical edge open probability as follows.

Definition 7.2.4 *The critical edge open probability p_c is defined as the "smallest" probability p_o for which $\theta(p_o) > 0$. Formally,*

$$p_c = \inf\{p_o : \theta(p_o) > 0\}.$$

To show that for $p_{\rm o} > p_c$, $\theta(p_{\rm o}) > 0$, while for $p_{\rm o} < p_c$, $\theta(p_{\rm o}) = 0$, we next show that $\theta(p_{\rm o})$ is non-decreasing in $p_{\rm o}$. The regime of $p_{\rm o} < p_c$ ($p_{\rm o} > p_c$) is called sub (super)-critical regime. In general, sub-critical (super-critical) regime refers to the case when the largest connected component of a graph is of bounded (unbounded) size.

Proposition 7.2.5 *The percolation probability $\theta(p_o)$ is a non-decreasing function of the open edge probability p_o.*

Proof: The proof is via a standard coupling argument. Let $0 < p_A < p_B < 1$. Then for each edge (i, j) of square lattice $\mathbb{Z}^2$, draw a random variable v_{ij} with uniform distribution in $[0, 1]$. Then consider two different subgraphs $G_A, G_B \subseteq \mathbb{Z}^2$, where in G_* only those edges of $\mathbb{Z}^2$ are present for which $v_{ij} < p_*$, $* \in \{A, B\}$. Then one can easily see that the connected component $\mathsf{C}(G_*)$ of the origin in the subgraph G_A and G_B satisfies $\mathsf{C}(G_A) \subseteq \mathsf{C}(G_B)$. Thus,

$$\mathbb{P}_{p_A}(|\mathsf{C}| = \infty) = \mathbb{P}(|\mathsf{C}(G_A)| = \infty) \leq \mathbb{P}(|\mathsf{C}(G_B)| = \infty) = \mathbb{P}_{p_B}(|\mathsf{C}| = \infty),$$

proving $\theta(p_A) \leq \theta(p_B)$ for $0 < p_A < p_B < 1$. □

Thus, using Definition 7.2.4, we know that there is a phase transition at p_c, such that for $p_{\rm o} > p_c$, the percolation probability is non-zero, while for $p_{\rm o} < p_c$, the probability of percolation is zero. Finding the exact value of p_c is a hard problem, however, has been found exactly to be equal to $\frac{1}{2}$ in [6]. Next, we present some simple techniques to find non-trivial bounds on p_c. The techniques presented here will be re-used to study percolation and connectivity regimes in wireless networks.

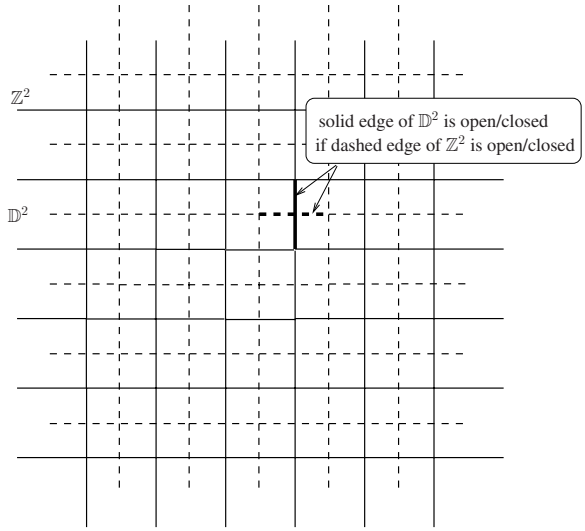

Figure 7.2: Dual lattice $\mathbb{D}^2$ of the square lattice $\mathbb{Z}^2$ is represented with dashed lines, where an edge is open/closed in $\mathbb{D}^2$ if the edge of $\mathbb{Z}^2$ intersecting it is open/closed.

Proposition 7.2.6 *For square grid percolation, $p_c \leq \frac{2}{3}$.*

Let $\mathbb{D}^2 = \mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$ be the dual lattice of $\mathbb{Z}^2$ obtained by translating each edge of $\mathbb{Z}^2$ by $(\frac{1}{2}, \frac{1}{2})$ as shown in Fig. 7.2 by dashed lines. Any edge $e' \in \mathbb{D}^2$ is defined to be *open* (*closed*) if and only if the edge $e \in \mathbb{Z}^2$ intersecting e' is open (closed).

Definition 7.2.7 *A circuit in $\mathbb{Z}^2$ or $\mathbb{D}^2$ is a self-avoiding connected path of $\mathbb{Z}^2$ or $\mathbb{D}^2$ which starts and ends at the same point. A circuit in $\mathbb{Z}^2$ or $\mathbb{D}^2$ is defined to be open/closed if all the edges on the circuit are open/closed in $\mathbb{Z}^2$ or $\mathbb{D}^2$. A pictorial representation of a closed circuit is given by Fig. 7.3.*

The following property of discrete percolation follows easily.

Lemma 7.2.8 *The connected component of the origin is finite if and only if there is a closed circuit in $\mathbb{D}^2$ surrounding the origin.*

Lemma 7.2.8 can be intuitively understood by looking at Fig. 7.3, where if there is a closed circuit of $\mathbb{D}^2$ surrounding the origin, then the origin can only be connected to lattice points of $\mathbb{Z}^2$ lying inside the closed circuit.

Thus, if we can show that the probability of having a closed circuit in $\mathbb{D}^2$ surrounding the origin is less than one, we have that the percolation probability in the square lattice $\mathbb{Z}^2$ is non-zero, $\theta(p_o) > 0$. To prove Proposition 7.2.6, we show that for $p_o > \frac{2}{3}$, the probability of having a closed circuit in $\mathbb{D}^2$ surrounding the origin is less than one, and hence the probability that the connected component of the origin is of infinite size is greater than zero.

Proof: [Proposition 7.2.6] Note that

$$\mathbb{P}(\text{having a closed circuit of length } n \text{ around the origin}) \leq$$

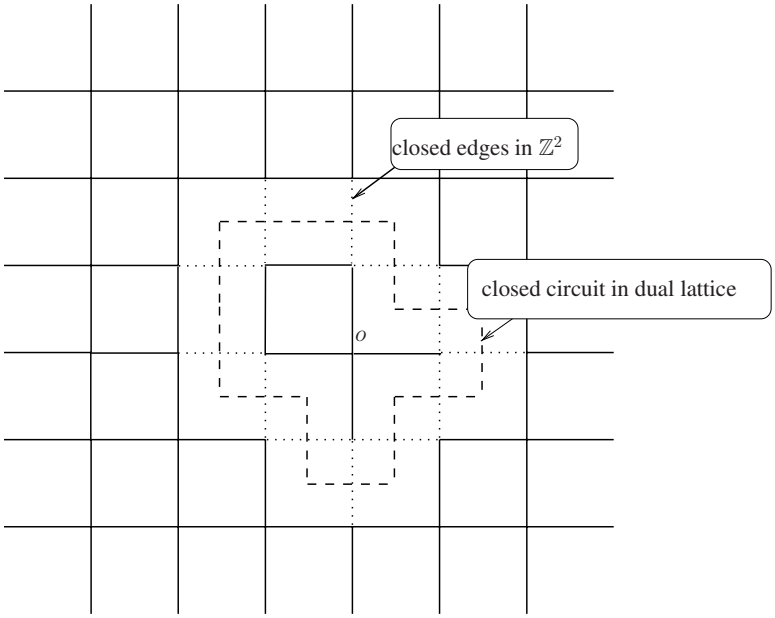

Figure 7.3: Depiction of a closed circuit of dual lattice $\mathbb{D}^2$ surrounding the origin.

$$\mathbb{E}_{p_o}\left\{|\text{ the total number of closed circuits of length } n|\right\}, \tag{7.2}$$

hence

$$\mathbb{P}(\text{having a closed circuit of around the origin}) \leq$$
$$\sum_{n=1}^{\infty} \mathbb{E}_{p_o}\left\{|\text{ the total number of closed circuits of length } n|\right\}. \tag{7.3}$$

First, we count the maximum number of circuits in $\mathbb{D}^2$ of length n surrounding the origin o. In $\mathbb{D}^2$, for a circuit of length n, there are at most n possible choices where the circuit intersects the x-axis (horizontal line containing the origin), as shown in Fig. 7.4. Moreover, for each lattice point, there are at most 3 possible choices (except for starting point when it is 4) to continue on a self-avoiding path of the circuit, as shown in Fig. 7.4. Thus, there are at most $n.4.3^{n-1}$ circuits of length n surrounding the origin o.

Since each of the edges on any circuit are closed with probability $1 - p_{\rm o}$ independently of all other edges, we have

$$\mathbb{E}_p\left\{|\text{ the total number of closed circuits of length } n|\right\} \leq 4n.3^{n-1}(1-p_{\rm o})^n. \tag{7.4}$$

Note that

$$\sum_{n=1}^{\infty} 4n.3^{n-1}(1-p_{\rm o})^n < \infty$$

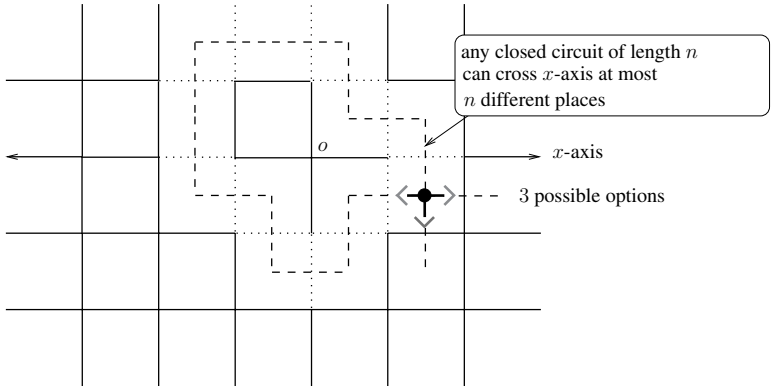

Figure 7.4: Counting the maximum number of closed circuits of length n surrounding the origin.

for $1 - p_o < \frac{1}{3}$, that is, $p_o > \frac{2}{3}$. Thus, for $p_o > \frac{2}{3}$, there exists a finite N, such that for

$$\sum_{n=N}^{\infty} 4n.3^{n-1}(1-p_o)^n \leq \frac{1}{2}$$

and

$$\sum_{n=N}^{\infty} \mathbb{E}_{p_o}\{|\text{ the total number of closed circuits of length } n|\} \leq \frac{1}{2}. \tag{7.5}$$

Therefore, similar to (7.3), we have that for $p_o > \frac{2}{3}$,

$$\mathbb{P}(\text{having a closed circuit of length } N \text{ or more around the origin}) \leq \sum_{n=N}^{\infty} \mathbb{E}_{p_o}\{|\text{ the total number of closed circuits of length } n|\} \leq \frac{1}{2}. \tag{7.6}$$

Let $B_{N/2} = [-N/2\ N/2] \times [-N/2\ N/2]$ be a square box centered at origin of $\mathbb{Z}^2$ with side length N. The probability that all edges inside box $B_{N/2}$ are open is at least $p_o^{(N+1)^2}$, since there are at most $(N+1)^2$ edges inside $B_{N/2}$.

Thus, the overall probability of not having any closed circuit in $\mathbb{D}^2$ surrounding the origin o is at least equal to the product of the probability of not having any closed circuit of finite length N or more surrounding o, and the probability that all edges inside box $B_{N/2}$ are open since that allows origin to be connected to some node outside of box $B_{N/2}$.

Since both these probabilities are non-zero, hence, using Lemma 7.2.8, we have that the probability of percolation is non-zero as long as $p_o > \frac{2}{3}$.

□

Next, we move on to show a lower bound of $\frac{1}{3}$ on the critical probability p_c.

Proposition 7.2.9 *For square grid percolation,* $p_c \geq \frac{1}{3}$.

Proof: Consider the self-avoiding connected paths of length n starting from the origin $o \in \mathbb{Z}^2$. Clearly, for a path going out from the origin, there are 4 possible choices, and thereafter at each lattice point, there are 3 choices for any self-avoiding path to proceed further. Thus,

$$\text{the total number of self-avoiding connected paths of length } n \leq 4.3^{n-1}.$$

Since each edge is open with probability $p_{\rm o}$ independently of all other edges, the expected number of self-avoiding connected paths of length n starting from the origin is

$$\mathbb{E}_{p_o}\{|\text{ the total number of self-avoiding connected paths of length } n|\} \leq 4.3^{n-1}p_{\rm o}^n. \tag{7.7}$$

Note that if there is percolation, that is, $|\mathsf{C}| = \infty$, then there has to be a self-avoiding connected path of length n starting from the origin for each n. Thus,

$$\theta(p_{\rm o}) \leq \mathbb{P}(\exists \text{ at least 1 self-avoiding connected path of length } n \text{ from } o). \tag{7.8}$$

Since trivially,

$$\mathbb{P}(\exists \text{ at least 1 self-avoiding connected path of length } n \text{ from } o) \leq$$
$$\mathbb{E}_{p_o}\{|\text{ the total number of self-avoiding connected paths of length } n|\}.$$

We have from (7.7) and (7.8), that

$$\theta(p_{\rm o}) \leq 4.3^{n-1}p_{\rm o}^n.$$

Since $4.3^{n-1}p_{\rm o}^n \to 0$ for $p_{\rm o} < \frac{1}{3}$, we have that the percolation probability $\theta(p_{\rm o})$ is zero for $p_{\rm o} < \frac{1}{3}$. □

So far we have shown that for square lattice $\mathbb{Z}^2$, the critical probability p_c satisfies $\frac{1}{3} \leq p_c \leq \frac{2}{3}$. A stronger result is known for p_c, which shows that $p_c = \frac{1}{2}$ [6]. The analysis to obtain $p_c = \frac{1}{2}$ is far more involved and requires finer analysis that is beyond the scope of this book. We state this result formally without a proof here.

Theorem 7.2.10 *For square lattice $\mathbb{Z}^2$, the critical probability for percolation is $p_c = \frac{1}{2}$.*

Remark 7.2.11 *We will not show here, but another important related result is an interesting property that whenever there is percolation in square lattice $\mathbb{Z}^2$, that is, there exists an unbounded size connected component, this component is unique. This property is useful for many of the finer analyses in percolation theory.*

We next state some important results from percolation theory on the number of connected paths of square lattice $\mathbb{Z}^2$ crossing the square box $B_{n/2}$, a square centered at origin with side n. Let $B_{n/2}(L \to R)$ (left-right crossing) denote a connected path of the square lattice $\mathbb{Z}^2$ crossing both the left and the right side of the box $B_{n/2}$. The next result states that for $p_{\rm o} > p_c$ (above criticality), there are roughly an order n (side length of box considered) edge disjoint left-right crossings of $B_{n/2}$ with high probability.

The utility of edge-disjoint left-right crossings of $B_{n/2}$ in terms of wireless networks is that they can be mapped to interference-free parallel communication paths crossing the physical area of the wireless network. Theorems 7.2.12 and 7.2.13 will be useful in analyzing the throughput capacity of wireless networks in Chapter 9, as they allow us to estimate the number of such interference-free parallel paths that exist with high probability for $p_{\rm o} > p_c$.

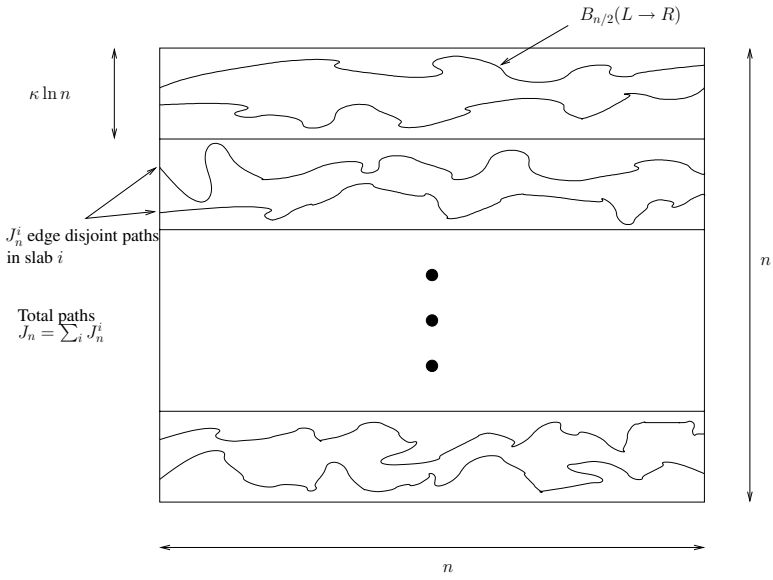

Figure 7.5: Left-right crossings of box $B_{n/2}$ by connected paths of square lattice $\mathbb{Z}^2$.

Theorem 7.2.12 *Let J_n denote the number of edge-disjoint left-right crossings of $B_{n/2}$. For $p_o > p_c$, constants δ and κ (that depend on the edge open probability p_o),*

$$\mathbb{P}(J_n \leq \delta n) \leq \exp(-\kappa n).$$

Proof: See Theorem 4.3.8. [7]. □

Actually, one can say a little more about the number of $B_{n/2}(L \to R)$ in the sense that if we divide the box $B_{n/2}$ horizontally in multiple slabs with height of each slab being $\kappa \ln n$ as shown in Fig. 7.5, there are roughly an order $\ln n$ edge-disjoint left-right crossings of $B_{n/2}$ that are completely contained inside each slab with high probability.

Formally, let the box $B_{n/2}$ be horizontally divided into slabs S_n^i of height $\kappa \ln n - \psi_n$, where $\psi_n > 0$ is chosen such that $\frac{n}{\kappa \ln n - \psi_n}$ is an integer. Let J_n^i be the number of edge-disjoint left-right crossings in slab S_n^i, and let $\mathsf{J} = \min_i J_n^i$.

Theorem 7.2.13 *For $p_o > \frac{5}{6}$, such that $2 + \kappa \ln(6(1-p_o)) < 0$, there exists a δ (depends on κ and p_o) such that*

$$\lim_{n \to \infty} \mathbb{P}(\mathsf{J} \leq \delta \ln n) = 0$$

Proof: See Theorem 4.3.9. [7]. □

After studying square lattice percolation, in the next subsection, we discuss a discrete hexagonal face percolation model that will also be useful in Section 7.3.1 for studying percolation in wireless networks.

7.2.2 Percolation on the Hexagonal Grid

Consider a hexagonal grid with side 1, as shown in Fig. 7.6. Each face of the hexagonal grid is open (closed) with probability p_{o} $(1 - p_{\mathrm{o}})$ independently of all other faces. The question once again is: for

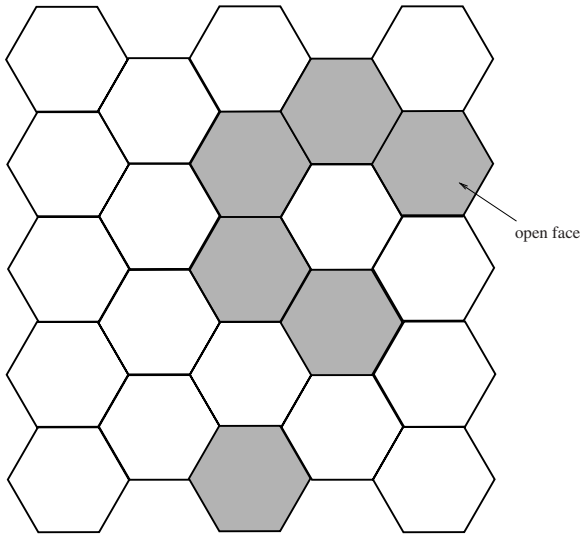

Figure 7.6: Hexagonal tilting of $\mathbb{R}^2$ with each face open (shaded)/closed independently.

what values of p_{o} does an unbounded connected component exist in the hexagonal grid. The answer is not very surprising given Theorem 7.2.10, as stated next, whose proof can be found in [1, 2].

Theorem 7.2.14 *For hexagonal grid with side* 1*, the critical probability for face percolation is* $p_c = \frac{1}{2}$.

We will make use of Theorem 7.2.14 in the next section to derive percolation regimes for graphs derived from a wireless communication model, by mapping it to an appropriate hexagonal grid.

7.3 Continuum Percolation

In this section, rather than looking at graphs generated by discrete geometry-based points as done in Section 7.2, we will let the points/nodes of the graph to lie in a region whose locations are derived from a spatial distribution. We then study percolation on a suitably defined graph over the spatially realized points, which has connections to wireless communications models. We begin with the simplest model, called the Gilbert's disc model, where nodes/point locations follow a Poisson point process (PPP), and two nodes are defined to be connected if they are within a fixed distance from each other.

7.3.1 Gilbert's Disc Model

Consider $\mathbb{R}^2$, and a spatial point process $\Phi \subset \mathbb{R}^2$, where node locations Φ follow a PPP with density λ. We define two nodes $x, y \in \Phi$ to have an edge between them if $||x - y|| \leq r$, where r is some fixed distance. This is known as the Gilbert's disc model denoted by $G_P(\lambda, r)$. This model is motivated by the path-loss model of wireless communication, where communication is deemed successful between two nodes x and y at a distance of d_{xy}, if the received signal power $d_{xy}^{-\alpha}$ is more

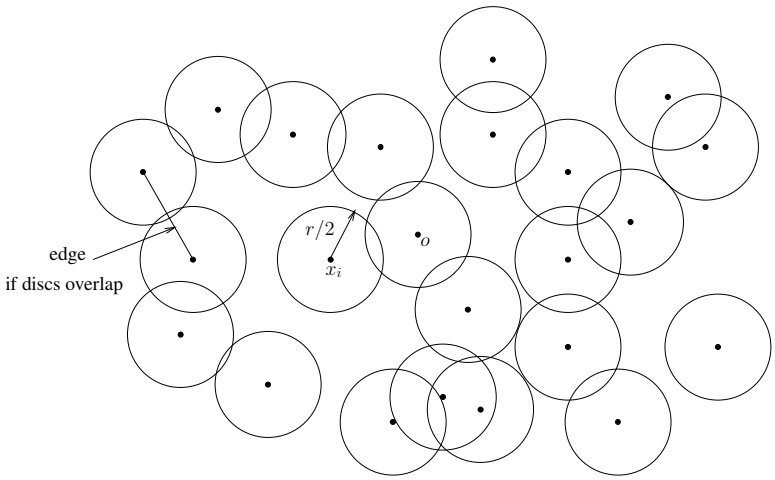

Figure 7.7: Gilbert's disc model where each node has a radio range $r/2$ and any two nodes are connected that are at a distance of less than r.

than a threshold β. Essentially, that translates to the radio range of $r = \beta^{-1/\alpha}$, such that two nodes can communicate with each other if they are inside each other's radio range.

Note that Gilbert's disc model defines an undirected graph assuming reciprocity of communication. A pictorial representation of Gilbert's disc model is as shown in Fig. 7.7, where we draw a disc of radius $r/2$ for each node, and two nodes are connected or have an edge between them if their discs overlap. We call node y to be a neighbor of node x if there is an edge between them.

We let a node of Φ to be present at the origin and consider its connected component. Recall that from Theorem 2.3.3, conditioning on a node being present at the origin, the distribution of the rest of the nodes remains unchanged. As before, percolation is defined to be the event that the connected component of the node at the origin or simply the origin is of unbounded size. Clearly, the percolation properties of Gilbert's disc model depend on both the density λ and the radius r. One can either fix λ or r, and find the conditions on the other parameter under which the graph defined by the Gilbert's disc model percolates or not. Note that by definition, percolation probability is non-decreasing in both λ and r.

Remark 7.3.1 *An important property of the Gilbert's disc model is that the expected number of neighbors given by $\lambda\pi r^2$ is an invariant for percolation, that is, for two different Gilbert disc models $G_P(\lambda_1, r_1)$ and $G_P(\lambda_2, r_2)$, their percolation properties are identical as long as $\lambda_1 r_1^2 = \lambda_2 r_2^2$. To see this equivalence, consider Fig. 7.8, where on the left side we have $G_P(\lambda, r)$, while on the right, we let radio range to be $r = 1$ and scale the distances between any two points/nodes of Φ with density λ by $\frac{1}{r}$ to get the new density to be λr^2. Thus, the figure on the right is equivalent to $G_P(\lambda r^2, 1)$, and it is easy to notice that whenever there is a connection between two nodes in $G_P(\lambda, r)$ so does in $G_P\left(\lambda r^2, 1\right)$, and vice versa.*

Next, we obtain bounds on the percolation regimes in the Gilbert's disc model as a function of its invariant λr^2 using results from the hexagonal grid percolation.

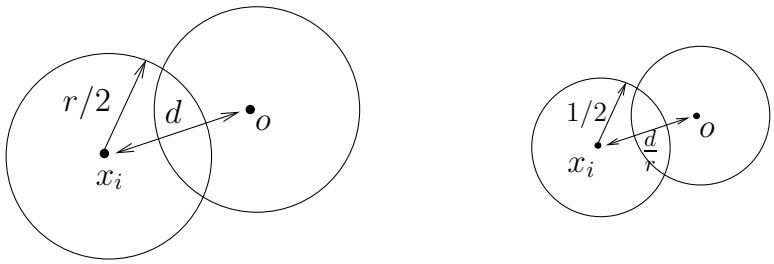

Figure 7.8: Invariance property of the Gilbert's disc model under fixed λr^2, where scaling radio range by $1/r$ and scaling distance between any two nodes also by $\frac{1}{r}$ has no effect on the connection model, where the two-sided arrow depicts an edge.

Theorem 7.3.2 *For*

$$\lambda r^2 \geq \frac{26 \log 2}{3\sqrt{3}},$$

the probability of percolation in the Gilbert's disc model is greater than 0.

Proof: Consider a hexagonal grid partition of $\mathbb{R}^2$ with side length s. Define any face to be *open* if it contains at least one node of Φ in it, and *closed* otherwise, as shown in Fig. 7.9. Clearly, since the number of nodes of Φ lying in disjoint sets are independent, faces of the hexagonal grid are open or closed independently of each other. Moreover, the probability $p_{\rm o}$ $(1 - p_{\rm o})$ of a face of the hexagonal grid being open (closed) is given by the void probability of the PPP to be $p_{\rm o} = 1 - \exp\left(-\lambda \frac{3\sqrt{3}s^2}{2}\right)$, since the area of each hexagonal face is $\frac{3\sqrt{3}s^2}{2}$.

Note that because of the hexagonal grid geometry, any two nodes of Φ in neighboring open faces in the hexagonal grid are at a distance of at most $((2\sqrt{3})^2 + 1)^{1/2} s = \sqrt{13}s$ from each other. Thus, if $\sqrt{13}s < r$ ($r/2$ is the radio range for $G_P(\lambda, r)$), the nodes of Φ in neighboring open faces are connected to each other in the Gilbert's disc model. From Theorem 7.2.14, we know that for $p_{\rm o} > 1/2$, the probability of percolation in the hexagonal grid is greater than 0. Hence, as long as the probability of any face of the hexagonal grid being open $1 - \exp\left(-\lambda \frac{3\sqrt{3}s^2}{2}\right) > 1/2$, and $\sqrt{13}s < r$, there is percolation in the Gilbert's disc model. Solving these two equations, we get that probability of percolation in the Gilbert's disc model is greater than 0 if

$$\lambda r^2 \geq \frac{26 \log 2}{3\sqrt{3}}.$$

□

We can also use this method to obtain a lower bound on the λr^2 below which the Gilbert's disc model does not percolate, as follows. From Fig. 7.9, any two nodes of Φ in non-neighboring faces of the hexagonal grid are at least s distance away. So if the hexagonal grid does not percolate (open face probability $p_{\rm o} < 1/2$), that is, the connected component of open faces is of bounded size, and $s > r$, then the connected components in the Gilbert's disc model are also of bounded size. Thus, equating the probability of open face being less than $1/2$, $1 - \exp\left(\lambda \frac{3\sqrt{3}s^2}{2}\right) < 1/2$, and $s > r$, we get that for

$$\lambda r^2 < \frac{2 \log 2}{3\sqrt{3}},$$

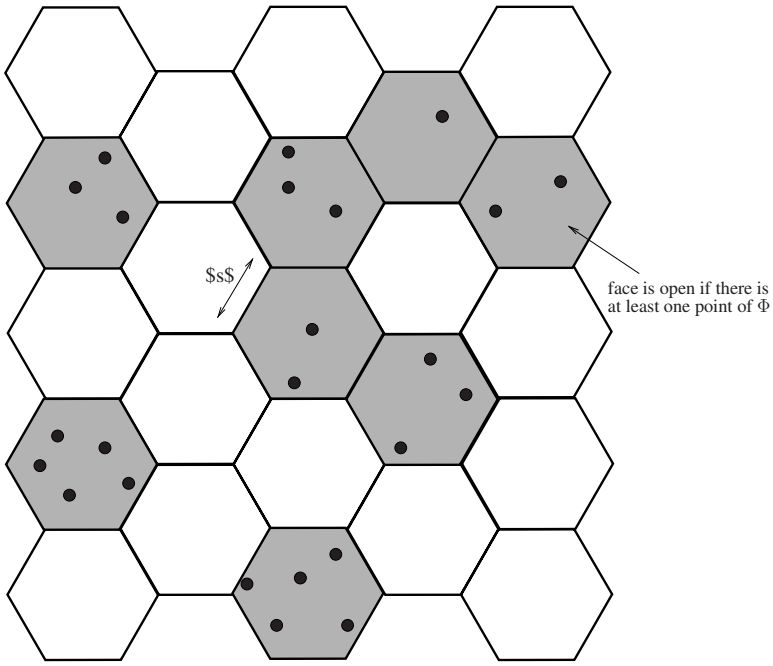

Figure 7.9: Mapping Gilbert's disc model on a hexagonal tiling of $\mathbb{R}^2$, where a face is open if it contains at least one node of Φ.

the Gilbert's disc model does not percolate.

This, however, is a disappointing result since a simple branching process argument can tell us that as long as the expected number of neighbors $\lambda\pi r^2$ of any node is less than 1, the Gilbert's disc model cannot percolate. We next present a refined version of the simple branching process argument from [8], which shows that as long as $\lambda r^2 < \frac{6}{2\pi+3\sqrt{3}}$, the Gilbert's disc model does not percolate.

Theorem 7.3.3 *If $\lambda r^2 < \frac{6}{2\pi+3\sqrt{3}}$, then the probability of percolation in the Gilbert's disc model is zero.*

Proof: Let the neighborhood set of the origin be $\mathcal{N}_0$, that is, the set of nodes lying in $\mathbf{B}(o,r)$ other than the one at origin. Let $x_1 \in \mathcal{N}_0$, and let the neighborhood set of the x_1 be $\mathcal{N}_1$. We are interested in the neighbors of x_1 in $\mathcal{N}_1 \backslash \mathcal{N}_0$, since this is really the effective new nodes that belong to the connected component of the origin contributed by x_1. $\#(\mathcal{N}_1 \backslash \mathcal{N}_0)$ corresponds to the number of nodes lying in the region of the disc $\mathbf{B}(x_1,r)\backslash\mathbf{B}(o,r)$, as represented by the shaded area in Fig. 7.10 for $x_1 \in \mathbf{B}(o,r)$. Since $x_1 \in \mathbf{B}(o,r)$, the area of the region $\mathbf{B}(x_1,r)\backslash\mathbf{B}(o,r)$ is maximized when x_1 is exactly at distance r from o, and the corresponding maximum area of shaded region in Fig. 7.10 is upper bounded by $\left(\frac{\pi}{3}+\frac{\sqrt{3}}{2}\right) r^2$. Thus,

$$\mathbb{E}\{\#(\mathcal{N}_1 \backslash \mathcal{N}_0)\} \leq \left(\frac{\pi}{3}+\frac{\sqrt{3}}{2}\right) \lambda r^2.$$

Moreover, note that $\#(\mathcal{N}_1 \backslash \mathcal{N}_0)$ is independent of $\#(\mathcal{N}_0)$, since they count the number of nodes of a PPP Φ that are defined over non-overlapping areas. The same process can be repeated for any

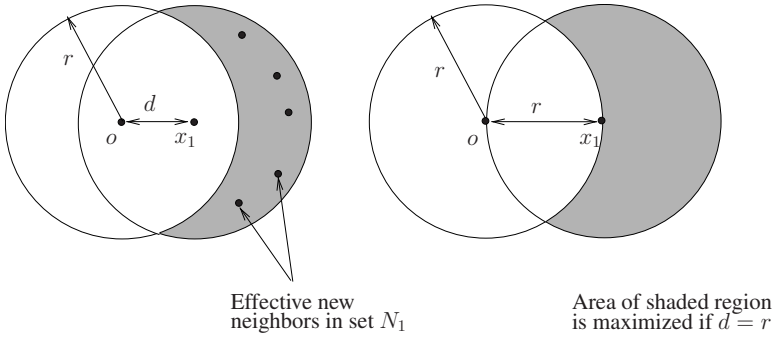

Figure 7.10: The largest region for finding new neighbors of x_1 that are not neighbors of the origin.

neighbor of x_0 or x_1 while discounting the neighbors found in $\mathcal{N}_0 \cup \mathcal{N}_1$, and for each such step, it follows that the new set of neighbors is independent of previous ones, and the expected number of new neighbors is always upper bounded by

$$\leq \left(\frac{\pi}{3} + \frac{\sqrt{3}}{2}\right) \lambda r^2,$$

since the neighbor set to be discounted is growing, for example, first it was $\mathcal{N}_0$, then it becomes $(\mathcal{N}_0 \cup \mathcal{N}_1)$ and so on.

Now we invoke the well-known branching process result that says that if the children are born independently with the expected number of children born at any step is less than 1, then the family dies out in finite time almost surely. This is true even if at any one step, the expectation is larger than 1. In our case, the number of children born at each step $\#(\mathcal{N}_0), \#(\mathcal{N}_1 \backslash \mathcal{N}_0), \ldots$ are independent, with expected value for each term $\leq \left(\frac{\pi}{3} + \frac{\sqrt{3}}{2}\right) \lambda r^2$ except for $\mathbb{E}\{\#(\mathcal{N}_0)\} = \lambda \pi r^2$. Substituting $\left(\frac{\pi}{3} + \frac{\sqrt{3}}{2}\right) \lambda r^2 < 1$, we get the result. □

Remark 7.3.4 *Similar to discrete percolation over square grid $\mathbb{Z}^2$, whenever there is percolation in the Gilbert's disc model, there is a unique unbounded size connected component.*

Percolation guarantees the existence of an unbounded connected component, however, that is sometimes not sufficient for efficient operation of a wireless network. Typically, in wireless networks, there is a strict requirement that each node should be reachable from every other node, generally known as *connectivity*. Asking for connectivity for node locations distributed as a PPP over the entire $\mathbb{R}^2$ is hopeless, and one has to restrict to a finite area of operation. In the next subsection, we discuss the connectivity properties of a Gilbert's disc model when nodes are distributed uniformly over a bounded area.

7.3.2 Connectivity in Gilbert's Disc Model

In contrast to percolation studied in Section 7.3.1, now we demand connectivity; each node should have a connected path to every other node in the network. It is easy to see that ensuring connectivity is a much more stricter condition than percolation, and finding conditions under which connectivity holds is rather challenging.

To study connectivity, we consider a unit area disc $\mathbf{B}(o,1)$, where n nodes are distributed uniformly in $\mathbf{B}(o,1)$. With reference to Section 7.3.1, this corresponds to the density $\lambda = n$. For the ease of interpretation, we use n instead of λ. Any two nodes are defined to be connected by an edge if they are within a distance of r from each other, and we refer to this graph as $G(n,r)$, where compared to Section 7.3.1, we have dropped the subscript P that represented that the node locations were derived from a PPP Φ.

It is apparent that if we keep increasing the radio range $r/2$ in $G(r,n)$, all nodes will have connected paths to each other, thereby ensuring connectivity. Since larger radio range requires larger transmission power which is at a premium for the purposes of efficiency as well as decreasing interference, the object of study is to find the minimum radio range that ensures connectivity in the wireless network. For this purpose, we make r dependent on n as $r(n)$ and represent the graph as $G(n,r(n))$.

In a seminal paper [9], Gupta and Kumar found the critical radio range required for ensuring connectivity in $G(n,r(n))$, which is stated in the next theorem.

Theorem 7.3.5 *$G(n,r(n))$ is connected for $\pi r^2(n) = \frac{\ln n + c(n)}{n}$ with probability 1 as $n \to \infty$ if and only if $c(n) \to \infty$.*

Remark 7.3.6 *We have considered what is known as dense network setting, where in a unit square, nodes are located uniformly randomly with density n. An alternate formulation is the extended network setting, where we consider a square with side $\sqrt{n}$ and unit density of nodes. An extended network is equivalent to a dense network after scaling all distances by $\sqrt{n}$. Hence, we can get the connectivity results for the extended network by scaling the critical radio range by $\sqrt{n}$.*

Theorem 7.3.5 states that as long as the radio range scales larger than $\sqrt{\frac{\ln n}{n}}$, however, small that additional scaling with n be, the whole network is connected with high probability. In converse, it says that if the radio range does not scale larger than $\sqrt{\frac{\ln n}{n}}$, the whole network cannot be connected with high probability. Thus, radio range of order $\sqrt{\frac{\ln n}{n}}$ is critical for connectivity of the Gilbert's disc model, when n nodes are distributed uniformly in a unit disc.

With $r(n) = \Theta\left(\sqrt{\frac{\ln n}{n}}\right)$, there are $\Theta(\ln n)$ neighbors for each node. Hence, one interpretation of this result is that if each node has order $\ln n$ neighbors, there is high likelihood of having connectivity in the network. This phenomenon is known to be true in many other networks, such as Erdős-Rényi graph, where the network is connected in the presence of $\ln n$ neighbors [10]. The intuitive explanation for critical $r(n) = \Theta\left(\sqrt{\frac{\ln n}{n}}\right)$ or $\ln n$ neighbors for connectivity can be obtained by considering a unit square $\mathbf{S}_1$ instead of a unit disc $\mathbf{B}(o,1)$ as follows.

Consider a unit square $\mathbf{S}_1$ and let there be n nodes uniformly distributed in $\mathbf{S}_1$. Consider a tiling of $\mathbf{S}_1$ into smaller squares s_{ij} with side $\sqrt{\frac{\kappa \log n}{n}}$, where $i,j \in \left\{1,\ldots,\left\lceil\left(\sqrt{\frac{n}{\kappa \log n}}\right)\right\rceil\right\}$. Let the number of nodes lying in s_{ij} be N_{ij}. From the Chernoff bound in Lemma 7.3.7, we have $\mathbb{P}(N_{ij} < \frac{\kappa \log n}{2}) \leq n^{-\kappa/8}$. Taking the union bound over all smaller squares s_{ij}, we get

$$\mathbb{P}\left(N_{ij} < \frac{\kappa \log n}{2} \text{ for any } i,j\right) < \kappa n^{1-\kappa/8},$$

where κ is a constant. Thus, for $\kappa > 8$, we see that each small square $s_{i,j}$ contains at least $\frac{\kappa \log n}{2}$ sensors with high probability. If the radio range is $r_s(n) = \Theta\left(\sqrt{\frac{\log n}{n}}\right)$, then this means that each node is connected to every other node in the same square, as well as nodes lying in the adjacent square. Thus, there exist paths (routed through adjacent smaller squares) between any pair of nodes lying in the unit square $\mathbf{S}_1$.

Lemma 7.3.7 *(Chernoff Bound) Let $X_1, X_2, \ldots$ be identical and independently distributed Bernoulli random variables with mean $\mathbb{E}\{X\}$, and let $X = \sum_{i=1}^{n} X_i$. Then for $0 < \delta < 1$, we have that $\mathbb{P}(X < (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mathbb{E}\{X\}}{2}\right)$.*

Conversely, if the side of small squares is $s(n) = \mathrm{o}\left(\sqrt{\frac{\kappa \log n}{n}}\right)$, then we cannot use Chernoff bound or any other tool to show that each small square contains $ns^2(n)$ nodes. Similarly, for $r(n) = \mathrm{o}\left(\sqrt{\frac{\kappa \log n}{n}}\right)$, we cannot guarantee that for all nodes there are $nr^2(n)$ neighbors with high probability.

We first show the necessary part of Theorem 7.3.5, and then follow it up with the sufficiency condition.

Proof: [Necessary part of Theorem 7.3.5] To bound the probability of disconnection (i.e., there is some node that is not reachable from some other node), we will consider the extreme case of isolated nodes, nodes that are not connected to any other node. Clearly, if the probability of having any isolated nodes does not go to zero, the probability of disconnection cannot go to zero. Even though having isolated nodes is a very small subset of all possible ways of not having connectivity, as we will see, it suffices for proving Theorem 7.3.5.

Let P_{iso} be the probability that there is at least one isolated node in $G(n, r(n))$. Then,

$$\begin{aligned}
P_{iso} &\overset{(a)}{\geq} \sum_{i=1}^{n} \mathbb{P}(\{i \text{ is the only isolated node in } G(n, r(n))\}), \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \Big(\mathbb{P}(\{i \text{ is an isolated node in } G(n, r(n))\}) \\
&\qquad - \sum_{j \neq i} \mathbb{P}(\{i \text{ and } j \text{ are isolated nodes in } G(n, r(n))\})\Big), \\
&= \sum_{i=1}^{n} \mathbb{P}(\{i \text{ is an isolated node in } G(n, r(n))\}) \\
&\qquad - \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{P}(\{i \text{ and } j \text{ are isolated nodes in } G(n, r(n))\}), \qquad (7.9)
\end{aligned}$$

where (a) follows since each of these events are exclusive, and (b) follows since

$$\begin{aligned}
\mathbb{P}(\{i \text{ is an isolated node in } G(n, r(n))\}) \leq \\
\mathbb{P}(\{i \text{ is the only isolated node in } G(n, r(n))\}) \\
+ \sum_{j=1, j \neq i}^{n} \mathbb{P}(\{i \text{ and } j \text{ are isolated nodes in } G(n, r(n))\}).
\end{aligned}$$

Throughout this proof we neglect edge effects resulting from nodes lying close to the boundary of disc $\mathbf{B}(o,1)$. For a node i located at x_i to be isolated in $G(n,r(n))$, the disc $\mathbf{B}(x_i,r(n))$ should not contain any of the other $n-1$ nodes, thus,

$$\mathbb{P}(\{i \text{ is an isolated node in } G(n,r(n))\}) = (1-\pi r^2(n))^{n-1}. \tag{7.10}$$

The event $\{i$ and j are isolated nodes in $G(n,r(n))\}$ implies that i and j are not connected to each other as well as to any other node. Thus, the distance d_{ij} between node i and j is more than $r(n)$. Considering the two exclusive events, $d_{ij} \in [r(n), 2r(n)]$ and $d_{ij} > 2r(n)$, we can write

$$\mathbb{P}(\{i \text{ and } j \text{ are isolated nodes in } G(n,r(n))\}) \approx$$
$$(4\pi r^2(n) - \pi r^2(n))(1-\frac{5}{4}\pi r^2(n))^{n-2} + (1-4\pi r^2(n))(1-2\pi r^2(n))^{n-2}, \tag{7.11}$$

where the two terms correspond to the two events, respectively.

Consider the case when $\pi r^2(n) = \frac{\ln n + c}{n}$ for a fixed c. Subsequently, we will replace c with $c(n)$. Using (7.10) and (7.11), from (7.9), we get

$$\begin{aligned}
P_{iso} \quad &\geq \quad \sum_{i=1}^{n}(1-\pi r^2(n))^{n-1} \\
&\quad -\sum_{i=1}^{n}\sum_{j\neq i}\Bigg(3\pi r^2(n)\left(1-\frac{5}{4}\pi r^2(n)\right)^{n-2} \\
&\quad +(1-4\pi r^2(n))\left(1-2\pi r^2(n)\right)^{n-2}\Bigg), \qquad (7.12)\\
&\overset{(a)}{\geq} \quad n(1-\pi r^2(n))^{n-1} - \\
&\quad n(n-1)\left(3\pi r^2(n)\left(1-\frac{5}{4}\pi r^2(n)\right)^{n-2} + \left(1-2\pi r^2(n)\right)^{n-2}\right), \\
&\overset{(b)}{\geq} \quad \theta\exp(-c) \\
&\quad -n(n-1)3\pi r^2(n)\left(\left(1-\frac{5}{4}\pi r^2(n)\right)^{n-2} + \left(1-2\pi r^2(n)\right)^{n-2}\right), \\
&\overset{(c)}{\geq} \quad \theta\exp(-c) - n(n-1)3\pi r^2(n)\left(\exp\left(-\frac{5}{4}(n-2)\pi r^2(n)\right)\right. \\
&\quad \left.+\exp\left(2(n-2)\pi r^2(n)\right)\right), \\
&\geq \quad \theta\exp(-c) - (1+\psi)\exp(-2c), \qquad (7.13)
\end{aligned}$$

where to get (a) we have dropped the term $4\pi r^2(n)(1-2\pi r^2(n))^{n-2}$, (b) follows from Lemma 7.3.8 that holds for large enough n for any fixed θ and c, for $\pi r^2(n) = \frac{\ln n + c}{n}$ (c) follows from Lemma 7.3.9 where $(1-t) \leq \exp(-t)$. Last inequality follows by evaluating the second expression of the RHS at $\pi r^2(n) = \frac{\ln n + c}{n}$ for any $\psi > 0$ and large enough n.

Lemma 7.3.8 *If $\pi r^2(n) = \frac{\ln n + c}{n}$, then for any fixed $\theta < 1$ and for large enough n*

$$n(1-\pi r^2(n))^{n-1} \geq \theta\exp(-c).$$

Proof is left as an exercise. It can found in Lemma 2.2 [9].

Lemma 7.3.9 *For any* $t \in [0,1]$, $(1-t) \leq \exp(-t)$, *and for any given* $\theta \geq 1$, *there exists a* $t_0 \in [0,1]$, *such that* $\exp(-\theta t) \leq (1-t)$ *for all* $0 \leq t \leq t_0$. *If* $\theta > 1$, *then* $t_0 > 0$.

Since having an isolated node implies disconnection, we have $P_{iso} \leq P_d$, where $P_d = 1 - P_c$, and P_c is the probability of connectivity. Thus, from (7.13), we have

$$P_d \geq \theta \exp(-c) - (1+\psi)\exp(-2c), \tag{7.14}$$

for all $n > N$ that depends on θ, ψ, and c.

The quantity c in (7.14) is actually $c(n)$ coming from $\pi r^2(n) = \frac{\ln n + c(n)}{n}$. Let $c(n)$ be bounded, and $\lim_{n\to\infty} c(n) = \bar{c}$. Then for any $\psi > 0$, $c(n) \leq \bar{c} + \psi$ for all $n > N'(\psi)$. Moreover, by definition, the probability of disconnection P_d is monotone decreasing in c, since larger c increases the radio range and decreases the chance of isolation. Hence, from (7.14),

$$P_d \geq \theta \exp(-(\bar{c}+\psi)) - (1+\psi)\exp(-2(\bar{c}+\psi)),$$

for $n > \max\{N, N'(\psi)\}$. As $n \to \infty$, we have

$$\lim_{n\to\infty} \inf P_d \geq \theta \exp(-(\bar{c}+\psi)) - (1+\psi)\exp(-2(\bar{c}+\psi)).$$

Note that this relation holds for all $\psi > 0$ and $\theta < 1$, thus, we have that the disconnection probability is lower bounded by a non-zero quantity for $n \to \infty$ as long as $c(n)$ is bounded. Thus, proving the only if condition of Theorem 7.3.5.

□

We have shown that as long as $c(n)$ is bounded, with $\pi r^2(n) = \frac{\ln n + c(n)}{n}$, the disconnection probability of finding at least one node that is not connected to some other node is non-zero. Next, we work toward showing the sufficiency conditions of Theorem 7.3.5. For this purpose, we will use the Gilbert's disc model with node locations $\Phi \subset \mathbb{R}^2$ distributed as a PPP discussed in Section 7.3.1. As before, let there be a node of Φ at the origin, and let $q_k(\lambda, r(\lambda))$ be the probability that the connected component of the origin has k nodes in it. One result that we did not discuss in Section 7.3.1, which is useful now, is as follows.

Lemma 7.3.10 $\lim_{\lambda\to\infty} \frac{1}{q_1(\lambda, r(\lambda))} \sum_{k=1}^{\infty} q_k(\lambda, r(\lambda)) = 1.$

For proof see Propositions 6.4–6.6 [1]. Lemma 7.3.10 says that as the density λ grows large, either the origin is isolated or is part of the infinite size component (that is known to be unique from Remark 7.3.4) with probability 1.

Proof: [Sufficiency part of Theorem 7.3.5] Consider the restriction of Gilbert's disc model $G_P(n, r(n))$ with node locations Φ (that is a PPP) to the unit disk $\mathbf{B}(o,1)$, with density $\lambda = n$. What we will show is that the characteristics of $G_P(n, r(n))$ and $G(n, r(n))$ (the object of interest) over $\mathbf{B}(o,1)$ are not very different for the purposes of analyzing the connectivity probability.

As a first step, we make an important observation from Lemma 7.3.10, that for large enough density of nodes n, the probability that $G_P(n, r(n))$ is disconnected is asymptotically the same as the probability of having an isolated node in $G_P(n, r(n))$. Formally, we state the result as follows.

Proposition 7.3.11 *For any $\kappa > 0$, and for sufficiently large n,*

$$P_d(G_P(n, r(n))) \leq (1+\kappa) P_{\text{iso}}(G_P(n, r(n))),$$

where $P_d(G_P(n,r(n))) = 1 - P_c(G_P(n,r(n)))$, and $P_c(G_P(n,r(n)))$ is the connectivity probability in graph $G_P(n,r(n))$.

Next, we upper bound $P_{iso}(G_P(n,r(n)))$, the probability of having at least one isolated node in $G_P(n,r(n))$.

Proposition 7.3.12 *If $\pi r^2(n) = \frac{\ln n + c(n)}{n}$, then*

$$\limsup_{n\to\infty} P_{iso}(G_P(n,r(n))) \leq \exp(-\bar{c}),$$

where $\bar{c} = \lim_{n\to\infty} c(n)$.

Proof: Since $G_P(n,r(n))$ is defined over nodes Φ distributed as a PPP with density n, the probability that the number of nodes of $G_P(n,r(n))$ lying in unit disc $\mathbf{B}(o,1)$ is equal to j is Poisson distributed with parameter n, that is,

$$\mathbb{P}(\#(G_P(n,r(n)) \in \mathbf{B}(o,1)) = j) = \exp(-n)\frac{n^j}{j!}.$$

Conditioned on $\#(G_P(n,r(n)) \in \mathbf{B}(o,1)) = j$,

$$P_{iso}(G_P(n,r(n))) = P_{iso}(G(j,r(n))),$$

since from the PPP property, conditioned on the fact that $\mathbf{B}(o,1)$ contains j nodes of $P_{iso}(G_P(n,r(n)))$, the j nodes are distributed uniformly in $\mathbf{B}(o,1)$. Hence, taking the expectation with respect to $\#(G_P(n,r(n)) \in \mathbf{B}(o,1))$,

$$P_{iso}(G_P(n,r(n))) = \sum_{j=1}^{\infty} P_{iso}(G(j,r(n)))\exp(-n)\frac{n^j}{j!}. \tag{7.15}$$

Let the expected number of isolated nodes in $G(j,r(n))$ be $\mathbb{E}_{iso}(G(j,r(n)))$. Then,

$$\begin{aligned} P_{iso}(G(j,r(n))) &\overset{(a)}{\leq} \mathbb{E}_{iso}(G(j,r(n))), \\ &\overset{(b)}{=} \mathbb{E}\left\{\sum_{i=1}^{j} \mathbf{1}_{\text{node } i \text{ is isolated in } G(j,r(n))}\right\}, \\ &\overset{(c)}{=} j\mathbb{P}(\text{node } i \text{ is isolated in } G(j,r(n))), \\ &\overset{(d)}{\sim} j(1-\pi r^2(n))^{j-1}, \end{aligned} \tag{7.16}$$

where (a) follows from the definition of the expectation, (b) follows from the linearity of expectation, (c) follows since the event that $\{\text{node } i \text{ is isolated in } G(j,r(n))\}$ is identically distributed for all nodes i, and finally (d) follows since for a node to be isolated there have to be no nodes within a disc of radius $r(n)$ around it.

Using (7.16) in (7.15), we get

$$\begin{aligned} P_{iso}(G_P(n, r(n))) &\leq \sum_{j=1}^{\infty} j(1-\pi r^2(n))^j \exp(-n)\frac{n^j}{j!}, \\ &\leq n\sum_{j=0}^{\infty}(1-\pi r^2(n))^j \exp(-n)\frac{n^j}{j!}, \\ &= n\exp(-n\pi r^2(n)), \end{aligned}$$

where the second inequality is obtained by upper bounding j by n, the total number of nodes. This completes the proof by substituting for $\pi r^2(n) = \frac{\ln n + c(n)}{n}$ and $\bar{c} = \lim_{n\to\infty} c(n)$. □

We next present a basic inequality that will be required later on.

Lemma 7.3.13 *For all $\kappa > 0$, and large enough n*

$$\sum_{j=1}^{n} \exp(-n)\frac{n^j}{j!} \geq \left(\frac{1}{2} - \kappa\right).$$

Proof left as an exercise.

Next, using Proposition 7.3.12, we upper bound the probability of disconnection in $G(n, r(n))$ as follows.

Theorem 7.3.14 *If $\pi r^2(n) = \frac{\ln n + c(n)}{n}$, then*

$$\limsup_{n\to\infty} P_d(G(n, r(n))) \leq 4\exp(-\bar{c}),$$

where $\bar{c} = \lim_{n\to\infty} c(n)$.

This result once proven shows that if $c(n) \to \infty$ as $n \to \infty$, then the disconnection probability is zero and proves the sufficiency condition of Theorem 7.3.5.

Proof: Similar to the fact noted to get (7.15) in the proof of Proposition 7.3.12, we have the following relation

$$P_d(G_P(n, r(n))) = \sum_{j=1}^{\infty} P_d(G(j, r(n)))\exp(-n)\frac{n^j}{j!}, \tag{7.17}$$

for the disconnection probability in the $G_P(n, r(n))$ and $G(j, r(n))$.

For a fixed radio range $r = r(n)$, where there are k nodes lying in $\mathbf{B}(o, 1)$, we have

$$P_d(G(k, r(n))) \leq \mathbb{P}(\{k^{th}\text{node is isolated in } G(k, r(n))\}) + P_d(G(k-1, r(n))), \tag{7.18}$$

which follows by the union bound. Recursing this relation for j steps, for $0 < j < n$, we have

$$P_d(G(n, r(n))) \leq \sum_{k=j+1}^{n} \mathbb{P}(\{k^{th}\text{node is isolated in } G(k, r)\}) + P_d(G(j, r(n))),$$

$$
\begin{aligned}
&= \sum_{k=j+1}^{n} (1-\pi r^2(n))^{k-1} + P_d(G(j,r(n))), \\
&= \frac{(1-\pi r^2(n))^j - (1-\pi r^2(n))^n}{\pi r^2(n)} + P_d(G(j,r(n))) \\
&\leq \frac{(1-\pi r^2(n))^j}{\pi r^2(n)} + P_d(G(j,r(n))), \qquad (7.19)
\end{aligned}
$$

where the second equality is obtained by computing the isolation probability of any node in $G(k, r(n))$ with total k nodes.

Substituting for $P_d(G(j, r(n)))$ in (7.17) from (7.19), we get

$$
\begin{aligned}
P_d(G_P(n,r(n))) &\geq P_d(G(n,r(n))) \sum_{j=1}^{n} \exp(-n)\frac{n^j}{j!} \\
&\quad - \sum_{j=1}^{\infty} \frac{(1-\pi r^2(n))^{j-1}}{\pi r^2(n)} \exp(-n)\frac{n^j}{j!}, \\
&\geq P_d(G(n,r(n))) \left(\frac{1}{2} - \psi\right) - \frac{\exp\left(-n\pi r^2(n)\right)}{\pi r^2(n)}, \qquad (7.20)
\end{aligned}
$$

where in the first inequality we have only kept sum until $j = n$ for the first term, and the last inequality is obtained by using Lemma 7.3.13 for all $\psi > 0$ for large enough n. □

Next, we invoke Proposition 7.3.11 that gives a relation between $P_d(G_P(n, r(n)))$, the disconnection probability and $P_{iso}(G_P(n, r(n)))$, the probability of having at least one isolated node in $G_P(n, r(n))$. Using Proposition 7.3.11 in (7.20), we get

$$
P_d(G(n,r(n))) \leq 2(1+4\psi) \left(P_{iso}(G_P(n,r(n))) + \frac{\exp\left(-n\pi r^2(n)\right)}{\pi r^2(n)} \right) \qquad (7.21)
$$

Now we use the exponential decay bound on $P_{iso}(G_P(n, r(n)))$ that we obtained in Lemma 7.3.12, the probability of having at least one isolated node in $G_P(n, r(n))$, to get

$$
P_d(G(n,r(n))) \leq 2(1+4\psi) \left(\exp(-c(n)) + \frac{\exp\left(-n\pi r^2(n)\right)}{\pi r^2(n)} \right), \qquad (7.22)
$$

for any $\psi > 0$ and large enough n. For $\pi r^2 = \frac{\ln + c(n)}{n}$, we have

$$
P_d(G(n,r(n))) \leq 2(1+4\psi) \left(\exp(-c(n)) + \frac{\exp(-c(n))}{\ln n + c(n)} \right), \qquad (7.23)
$$

Since $\psi > 0$ is arbitrary,

$$
\limsup_{n\to\infty} P_d(G(n,r(n))) \leq \exp(-\bar{c}), \qquad (7.24)
$$

where $\bar{c} = \lim_{n\to\infty} c(n)$. Since $\bar{c} \to \infty$

$$
\limsup_{n\to\infty} P_d(G(n,r(n))) = 0
$$

proving Theorem 7.3.14, and consequently the sufficiency condition in Theorem 7.3.5. □

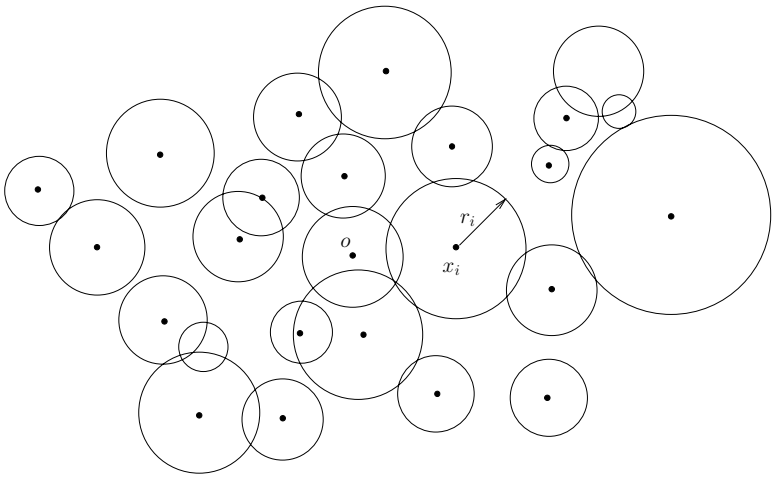

Figure 7.11: A realization of the Gilbert's random disc model, where each node x_i has radius r_i and two nodes are defined to be connected if their corresponding discs overlap.

After analyzing the percolation and connectivity properties of the Gilbert's disc model, in the next section, we study a generalized version of the Gilbert's disc model, called the Gilbert's random disc model, where each node has a disc with random radius (random radio range), and two nodes are connected if their discs overlap.

7.3.3 Gilbert's Random Disc Model

In this section, we generalize the Gilbert's disc model to allow for non-identical radio ranges. In particular, we consider the node locations to follow a PPP Φ over $\mathbb{R}^2$ with density λ. In contrast to Section 7.3.1, we let each node x to have a random radio range r_x that is i.i.d. for all x. We denote the random variable representing the radio range as ρ. This generalization essentially models non-homogeneous nodes in the network, that can potentially have different transmission powers, sophisticated signal processing algorithms, and so on. The connection model is defined by drawing discs of radius r_x around each node x, and two nodes x and y are defined to be connected if their corresponding discs overlap, as shown in Fig. 7.11. Similar to Section 7.3.1, we want to study the percolation properties of the Gilbert's random disc model as a function of λ and random radio range ρ.

We define the critical density as the minimum density for which there is percolation in the Gilbert's random disc model

$$\lambda_c(\rho) = \inf\{\lambda : |\mathsf{C}| = \infty\},$$

where C is the connected component of the origin. We first present a sub-critical regime result that shows that $\lambda_c(\rho)$ in Section 7.3.3 is at least $\frac{1}{\mathbb{E}\{\rho^2\}}$ times a constant for the Gilbert's random disc model, and then study the super-critical regime in Section 7.3.3.

Lower Bound on $\lambda_c(\rho)$

Let B_m be a square box with side $2m$ centered at origin, that is, $B_m = [-m\ m] \times [-m\ m]$. Consider a node at the origin, and let C be its connected component. Let $x_F \in \mathsf{C}$ be the farthest node from the origin in terms of the Euclidean distance.

Definition 7.3.15 *For any bounded set $\mathcal{B} \subset \mathbb{R}^2$, let $A_{\mathcal{B}}(r)$ be the event that the maximum radio range of any node x lying in $\mathcal{B}$ is less than r, that is, $A_{\mathcal{B}}(r) = \{r_x \leq r,\ \forall\, x \in \Phi \cap \mathcal{B}\}$.*

Definition 7.3.16 *The translation of set $S \subseteq \mathbb{R}^2$ by vector q is denoted by $q + S$, that is, $q + S = \{q + x : x \in S\}$.*

Definition 7.3.17 *The complement of set $S \subseteq \mathbb{R}^2$ and event E is denoted by $S^c = \mathbb{R}^2 \backslash S$ and E^c, respectively.*

Definition 7.3.18 *For a point $q \in \mathbb{R}^2$, let $E(q, r)$ denote the event that there is a path from a node $x \in \Phi \cap q + B_r$ to a node $y \in \Phi \cap q + B_{9r} \backslash q + B_{8r}$ with all the nodes on the path between x and y lying inside $B_{10r} + q$.*

Note that due to stationarity $\mathbb{P}(E(q, r)) = \mathbb{P}(E(o, r))$.

Definition 7.3.19 *Let $G(r)$ be the event that there is a node $x \in \Phi$ outside of B_{10r} such that the disc $\mathbf{B}(x, r_x)$ overlaps with B_{9r} as shown in Fig. 7.12. Formally,*

$$G(r) = \{\mathbf{B}(x, r_x) \cap B_{9r} \neq \phi, \text{ for some } x \in \Phi \backslash B_{10r}\}.$$

Let the farthest node x_F of C lie in B^c_{10r} as shown in Fig. 7.12. In addition, let event $G(r)^c$ also occur. Then there is a connected path from the origin to a node $y \in B_{9r} \backslash B_{8r}$ with all the nodes on the path lying inside B_{10r}, since there is path from the origin to x_F, and there is no node of Φ outside of B_{10r} whose disc has any overlap with B_{9r} because of $G(r)^c$. Thus, event $E(o, r)$ occurs if $x_F \in B^c_{10r}$ and $G(r)^c$ occur simultaneously, that is, $\mathbb{P}(\{x_F \in B^c_{10r}\} \cap G(r)^c) \leq \mathbb{P}(E(o, r))$. Hence, the following proposition follows.

Proposition 7.3.20

$$\mathbb{P}(x_F \in B^c_{10r}) \leq \mathbb{P}(E(o, r)) + \mathbb{P}(G(r)). \tag{7.25}$$

For percolation, the farthest node x_F of the connected component has to lie at an unbounded distance from the origin, hence

$$\mathbb{P}(|\mathsf{C}| = \infty) \leq \lim_{r \to \infty} \mathbb{P}(x_F \in B^c_{10r}),$$

since infinitely many nodes of a PPP cannot lie in a finite region. It easily follows that $\mathbb{P}(G(r)) \to 0$ as $r \to \infty$ (Proposition 7.3.28). Thus, to show that $\mathbb{P}(|\mathsf{C}| = \infty) = 0$ for some $\lambda < \lambda_c(\rho)$, from (7.25), it is sufficient to show that $\mathbb{P}(E(o, r))$ goes to zero as $r \to \infty$ for $\lambda < \lambda_c(\rho)$. The main theorem on the Gilbert's random disc model is as follows.

Theorem 7.3.21 *$\lambda_c(\rho) > \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$, that is, for $\lambda \leq \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$, where c is a constant, $\mathbb{P}(|\mathsf{C}| = \infty) = 0$.*

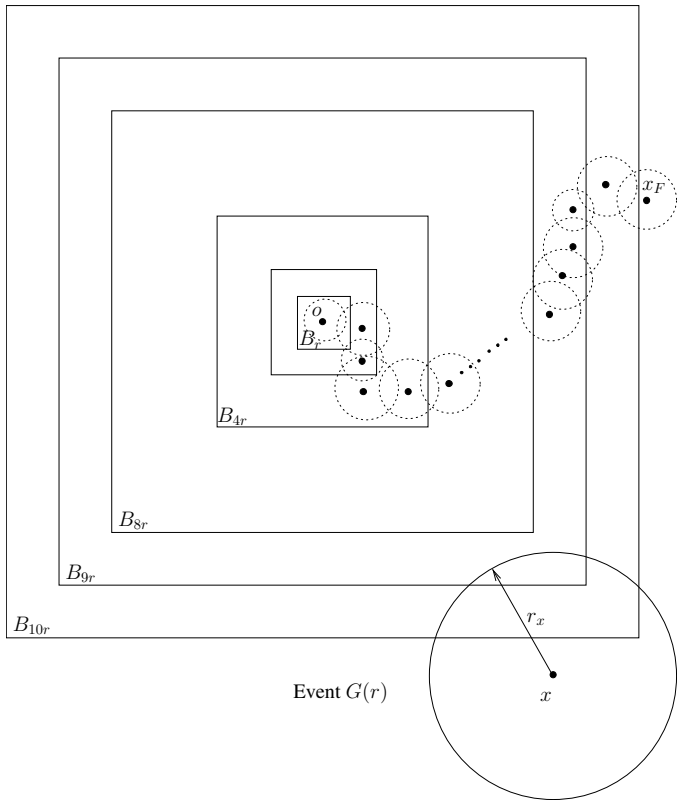

Figure 7.12: Depiction of scenario considered for obtaining Proposition 7.3.20, where the farthest node lies outside of B_{10r} and event $G(r)$.

Proof: From Proposition 7.3.20,

$$\mathbb{P}(|\mathsf{C}| = \infty) \leq \lim_{r\to\infty} \mathbb{P}(E(o,r)) + \lim_{r\to\infty} \mathbb{P}(G(r)).$$

From Proposition 7.3.28, we get $\lim_{r\to\infty} \mathbb{P}(G(r)) = 0$, and more importantly from Lemma 7.3.24, for $\lambda \leq \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$,

$$\lim_{r\to\infty} \mathbb{P}(E(o,r)) = 0.$$

Thus, proving the result. □

Theorem 7.3.21 says that if the expected number of neighbors $\pi\lambda\mathbb{E}\{\rho^2\}$ is less than a constant (exactly derivable, see proof), then the probability of percolation is zero for the Gilbert's random disc model. This is one of very few concrete bounds for the random disc models, where most often only existential results are known for percolation.

The basic idea behind proving Theorem 7.3.21 is to show that

$$\lim_{r\to\infty} \mathbb{P}(E(o,r)) = 0$$

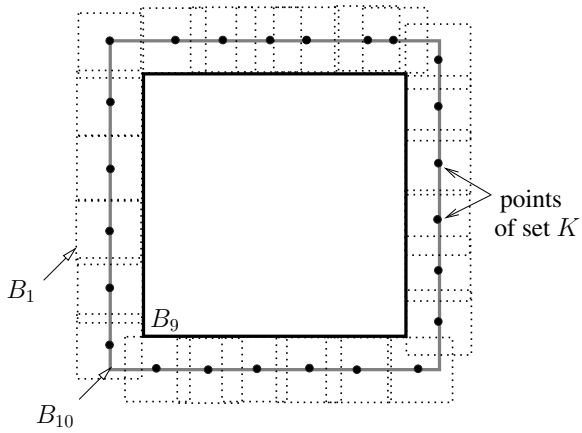

Figure 7.13: Covering of $B_{10}\backslash B_9$ by discrete points (black dots) lying on the boundary of B_{10} using boxes B_1.

for $\lambda \leq \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$. What that means is that if the expected number of neighbors $\pi\lambda\mathbb{E}\{\rho^2\}$ is less than a constant, the probability that the farthest node x_F of the connected component of the origin lies in $B_{9r}\backslash B_{8r}$ such that all nodes on the connected path lie inside B_{10r} goes to zero as r goes to infinity.

Now, we state the most important result required for proving Lemma 7.3.24 where we show that $\lim_{r\to\infty}\mathbb{P}(E(o,r)) = 0$, which in turn proves Theorem 7.3.21.

Lemma 7.3.22 $\mathbb{P}(E(o,10r)) \leq c_4\mathbb{P}(E(o,r))^2 + \mathbb{P}(A_{B_{100r}}(r)^c)$*, where* c_4 *is a constant.*

Lemma 7.3.22 is the main result required to prove Lemma 7.3.24. The idea behind proving Lemma 7.3.22 is that if events $E(o,10r)$ and $A_{B_{100r}}(r)$ occur simultaneously, that is, there is a path from a point inside of $\Phi \cap B_{10r}$ to some point in $B_{90r}\backslash B_{80r}$ where all the nodes on the connected path are inside box B_{100r}, and all nodes lying in B_{100r} have radio ranges less than r, respectively, then two independent and identically distributed disjoint events $E(kr,r)$ and $E(\ell r,r)$ occur simultaneously, where k lies on the boundary of B_{10} and ℓ lies on the boundary of B_{80}.

Proof: Let K and L be two discrete set of points that lie on the boundary of B_{10} and B_{80}, respectively, such that union of boxes B_1 centered at points of K and L contain $B_{10}\backslash B_9$ and $B_{81}\backslash B_{80}$, respectively. Formally, $B_{10}\backslash B_9 \subseteq \cup_{k\in K}(k + B_1)$ and $B_{81}\backslash B_{80} \subseteq \cup_{\ell\in L}(\ell + B_1)$. For example, see Fig. 7.13, where black dots represent the points of K on the boundary of B_{10} covering $B_{10}\backslash B_9$ using B_1. Let c_4 be the product of the cardinality of K and L.

Assume that $E(o,10r)$ and $A_{B_{100r}}(r)$ occur simultaneously. Event $E(o,10r)$ implies that there exists a node x_0 in B_{10r} which is connected to a node $x_L \in B_{90r}\backslash B_{80r}$ as shown in Fig. 7.14. Moreover, because of simultaneous occurrence of event $A_{B_{100r}}(r)$, that is, the radio range $r_x < r$ for all nodes $x \in \Phi \cap B_{100r}$, in fact, there exists a node at $\zeta \in B_{10r}\backslash B_{9r}$ which is connected to a node ζ_F in $\zeta + B_{9r}\backslash B_{8r}$ with all nodes on the connected path between ζ and ζ_F lying inside $B_{10r} + \zeta$.

Since $B_{10}\backslash B_9 \subseteq \cup_{k\in K}(k + B_1)$, we have $B_{10r}\backslash B_{9r} \subseteq \cup_{k\in K}(kr + B_1)$. Hence, $\zeta \in rk + B_r$ for some $k \in K$. See Fig. 7.14 for a pictorial description. Hence, if $E(o,10r)$ and $A_{B_{100r}}(r)$ occur, then because of the connected path between ζ and ζ_F, event $\cup_{k\in K}E(rk,r)$ happens, where $\mathbb{P}(E(rk,r)) = \mathbb{P}(E(o,r))$ for any $k \in K$.

Similarly looking at nodes around B_{80r}, it follows that if $E(o,10r)$ and $A_{B_{100r}}(r)$ occur together, then there exists a node $v \in r\ell + B_r, \ell \in L$, such that there is a connected path from v to a node in $v + B_{9r} \backslash B_{8r}$ that is completely contained in $v + B_{10r}$. Thus, event $\cup_{\ell \in L} B(r\ell, r)$ happens. Hence, if both $E(o,10r)$ and $A_{B_{100r}}(r)$ occur simultaneously, then two events $\cup_{k \in K} E(rk, r) \cap \cup_{\ell \in L} E(r\ell, r)$ happen simultaneously, where both events are translations of $E(o,r)$ and are identically distributed to $E(o,r)$, that is, $\mathbb{P}(E(r\ell, r)) = \mathbb{P}(E(o,r))$ for any $\ell \in L$.

Because of event $A_{B_{100r}}(r)$, radius of each node lying inside B_{100r} is less than r, thus the event $\cup_{k \in K} B(rk, r)$ depends only upon the nodes of Φ lying in B_{20r}, while the event $\cup_{\ell \in L} B(r\ell, r)$ depends only upon the nodes of Φ lying in B^c_{69r}. Since B_{20r} and B^c_{69r} are disjoint, the events $\cup_{\ell \in L} E(r\ell, r)$ and $\cup_{k \in K} E(rk, r)$ are also independent, and hence we get that

$$\mathbb{P}(E(o,10r) \cap A_{B_{100r}}(r)) \leq c_4 \mathbb{P}(E(o,r))^2.$$

□

Now we are ready to prove Lemma 7.3.24 that shows that $\mathbb{P}(E(0,r)) \to 0$ as $r \to \infty$ required for proving Theorem 7.3.21 though, there are few more intermediate results that are required for the proof that are stated after the proof of Lemma 7.3.24 for ease of exposition.

We first state an analytical result from [11] first.

Proposition 7.3.23 *Let f and g be two measurable, bounded and non-negative functions from $[1, +\infty]$ to $\mathbb{R}^+$. If $f(x) \leq 1/2$ for $x \in [1,10]$, and $g(x) \leq 1/4$ for $x \in [1, +\infty]$, and $f(x) \leq f(x/10)^2 + g(x)$ for $x \geq 10$, then $f(x)$ converges to 0 as $x \to \infty$ whenever $g(x)$ converges to 0 as $x \to \infty$.*

Proof: See Lemma 3.7 [11]. □

Let $M = \frac{(\mathbb{E}\{\rho^2\})^{1/2}}{10}$, $f(r) = c\mathbb{P}(E(o, Mr))$, and $g(r) = \lambda c^2 \int_{\frac{Mr}{10}}^{\infty} s^2 f_\rho(s) \mathrm{d}s$, where $f_\rho(.)$ denotes the PDF of ρ, and $c = \max\{c_1, c_2, c_3\}$, c_1, c_2, c_3 are constants defined in Propositions 7.3.27, 7.3.28, and 7.3.29, respectively.

Lemma 7.3.24 *For $\lambda \leq \frac{1}{4c^2 \mathbb{E}\{\rho^2\}}$, $f(r) \to 0$ and $\mathbb{P}(E(o,r)) \to 0$ as $r \to \infty$.*

Proof: Using definitions of $f(r)$ and $g(r)$, from Lemma 7.3.22,

$$f(r) \leq f(r/10)^2 + g(r).$$

Moreover, from Proposition, 7.3.25 and 7.3.26, we have

$$f(r) \leq \frac{1}{2}$$

for $r \in [1,10]$ and

$$g(r) \leq \frac{1}{4} \; \forall \, r.$$

Hence, using Proposition 7.3.23, it follows that $f(r) \to 0$ and consequently $\mathbb{P}(E(o,r)) \to 0$ as $r \to \infty$. □

Proposition 7.3.25 *For $\lambda \leq \frac{1}{4c^2 \mathbb{E}\{\rho^2\}}$, $f(r) \leq \frac{1}{2}$ for $r \in [1,10]$.*

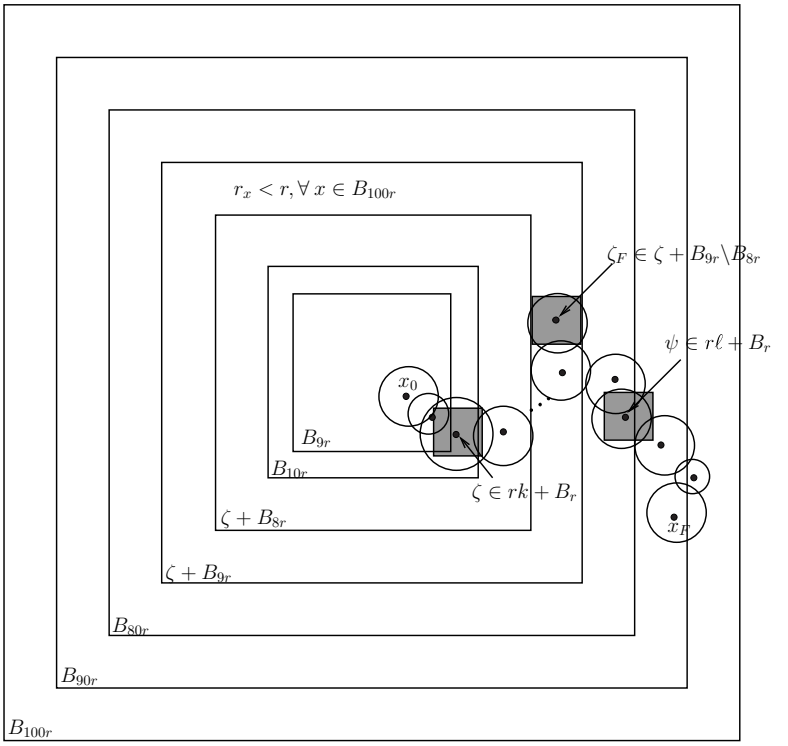

Figure 7.14: Depiction of scenario when both events $E(o, 10r)$ and $A_{B_{100r}}(r)$ occur simultaneously, giving rise to two smaller events that are i.i.d. with $E(o, r)$ around B_{10r} and B_{80r}.

Proof: From Proposition 7.3.27, $f(r) = c\mathbb{P}(E(o, Mr)) \leq \lambda c^2 M^2 r^2$. Using the definition of M, we get

$$f(r) \leq \lambda c^2 \mathbb{E}\{\rho^2\} \left(\frac{r}{10}\right)^2,$$

from which the result follows. □

Proposition 7.3.26 *For* $\lambda \leq \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$, $g(r) \leq \frac{1}{4}$, $\forall\, r \in [1, +\infty]$.

Proof: Note that $\int_{\frac{Mr}{10}}^{\infty} s^2 f_\rho(s)\mathrm{d}s \leq \mathbb{E}\{\rho^2\}$, hence

$$g(r) = \lambda c^2 \int_{\frac{Mr}{10}}^{\infty} s^2 f_\rho(s)\mathrm{d}s \leq \frac{1}{4},\ \forall\, r,$$

when $\lambda \leq \frac{1}{4c^2\mathbb{E}\{\rho^2\}}$. □

Proposition 7.3.27 $\mathbb{P}(E(o, r)) \leq \lambda c_1 r^2$, *where* c_1 *is a constant.*

Proof: Event $E(o, r)$ implies that there is at least one node of Φ in B_{10r}, thus $\#(\Phi \cap B_{10r}) > 0$. Hence, $\mathbb{P}(E(o, r)) \leq \mathbb{P}(\#(\Phi \cap B_{10r}) > 0)$. Since $\mathbb{E}\{\#(\Phi \cap B_{10r})\} = \lambda\nu(B_{10r})r^2$ is clearly greater than or equal to $\mathbb{P}(\#(\Phi \cap B_{10r}) > 0)$, the result follows. □

Proposition 7.3.28

$$\mathbb{P}(G(r)) \leq \lambda c_2 \mathbb{E}\{\rho^2 \mathbf{1}_{\{\rho > r\}}\}.$$

Proof: Following the Definition 7.3.19 of event $G(r)$, we get

$$\begin{aligned}
\mathbb{P}(G(r)) &\overset{(a)}{\leq} \mathbb{E}\left\{\#\{x \in \Phi \backslash B_{10r} : \mathbf{B}(x, r_x) \cap D_{9r} \neq \phi\}\right\}, \\
&\overset{(b)}{\leq} \lambda \int_{\mathbb{R}^2 \backslash B_{10r}} \mathbb{P}(\mathbf{B}(x, r_x) \cap D_{9r} \neq \phi) \mathrm{d}x, \\
&\overset{(c)}{\leq} \lambda \int_0^\infty \nu(B_{s+9r} \backslash B_{10r}) f_\rho(s) \mathrm{d}s, \\
&\leq \lambda \int_r^\infty \nu(B_{s+9r}) f_\rho(s) \mathrm{d}s, \\
&\leq \lambda \int_r^\infty \nu(B_{10s}) f_\rho(s) \mathrm{d}s, \\
&= \lambda 100^2 \int_r^\infty f_\rho(s) \mathrm{d}s, \\
&= \lambda 100^2 \mathbb{E}\{\rho^2 \mathbf{1}_{\{\rho > r\}}\},
\end{aligned}$$

where (a) follows since expectation is more than the probability of one of the participating event, (b) follows from the definition of PPP Φ, (c) follows since the set $G(r)$ can also be written as

$$G(r) = \{(x, r) \in (\mathbb{R}^2, (0, \infty)) : x \notin B_{10r}, x \in B_{9r+s}\},$$

and $\nu(.)$ is the Lebesgue measure. Since $\mathbb{E}\{\rho^2\}$ is finite, $\mathbb{P}(G(r)) \to 0$ as $r \to \infty$. □

Proposition 7.3.29 $\mathbb{P}(A^c_{B_{100r}}) \leq \lambda c_3 \int_r^\infty s^2 f_\rho(s) \mathrm{d}s$*, where* c_3 *is a constant, and*

$$\mathbb{P}(A_{B_{100r}}(r)^c) \to 0$$

as $r \to \infty$.

Proof: Note that $1 - \mathbb{P}(A_{B_{100r}}(r)) = \mathbb{P}(\exists$ at least one node $x \in \Phi \cap B_{100r},$ such that $r_x > r)$. Hence,

$$\begin{aligned}
1 - \mathbb{P}(A_{B_{100r}}(r)) &= \sum_{j=0}^\infty \mathbb{P}(\#(\Phi \cap B_{100r}) = j) \mathbb{P}(\{r_{x_1} > r\} \cup \ldots \cup \{r_{x_j} > r\}), \\
&\overset{(a)}{\leq} \sum_{j=0}^\infty \mathbb{P}(\#(\Phi \cap B_{100r}) = j) j \mathbb{P}(\{\rho > r\}), \\
&= \sum_{j=0}^\infty \frac{(\lambda \nu(B_{100r}))^j}{j!} e^{-\lambda \nu(B_{100r})} j \mathbb{P}(\{\rho > r\}), \\
&= \lambda \nu(B_{100}) r^2 \mathbb{P}(\{\rho > r\}), \\
&\leq \lambda \nu(B_{100}) \int_r^\infty s^2 f_\rho(s) \mathrm{d}s,
\end{aligned}$$

$$\leq \quad \lambda\nu(B_{100})\mathbb{E}\{\rho^2\mathbf{1}_{\{\rho>r\}}\},$$

where (a) is obtained by using the union bound and since $\{r_{x_j} \leq r\}$ is identically distributed $\forall\, j$. Since $\mathbb{E}\{\rho^2\}$ is finite, $\mathbb{P}(A_{B_{100r}}(r)^c) \to 0$ as $r \to \infty$.

□

Upper Bound on $\lambda_c(\rho)$

One easy way to obtain upper bounds on the critical density $\lambda_c(\rho)$ is by mapping the Gilbert's random disc model to a Gilbert's disc model as follows. Choose a radius r_0, and delete all points of the set $\{x \in \Phi : x_r < r_0\}$, that is, drop all points of the Gilbert's random disc model that have their radio ranges less than r_0. Since r_x is i.i.d., this translates to random thinning of PPP Φ and results in a new PPP Φ_{r_0} with density $\lambda\mathbb{P}(\rho \geq r_0)$. Now in addition, for points in the set $\{x \in \Phi : x_r \geq r_0\}$, we artificially make $r_x = r_0$, since reducing the radio ranges can only decrease the size of the connected component. Following this process, we get a Gilbert's disc model with fixed radio range r_0 and density $\lambda\mathbb{P}(\rho \geq r_0)$. Now, we can use the result obtained in Theorem 7.3.2 to get that as long as

$$\lambda\mathbb{P}(\rho \geq r_0)r_0^2 > \frac{26\log 2}{3\sqrt{3}},$$

the probability of percolation is non-zero. Optimizing over the choice of r_0 allows us to obtain the tightest bound with this technique.

Using the results of this section, we next derive percolation results on the path-loss plus fading model for wireless communication, where two nodes are connected if the signal power (function of both distance and fading) between them is larger than a threshold.

7.3.4 Incorporating Fading using Gilbert's Random Disc Model

In this section, we consider a more general path-loss plus fading model of wireless communication compared to the path-loss model of Section 7.3.1, where two nodes x and y can communicate with each other or have an edge between them if the received signal power $d_{xy}^{-\alpha}|h_{xy}|^2 > \beta$, where h_{xy} is the random fading coefficient, and β is the threshold. Thus, node x can connect to node y if

$$||x - y|| \leq \left(\frac{|h_{xy}|^2}{\beta}\right)^{\frac{1}{\alpha}} = r_{xy}.$$

The random connection distance (radius) $r_{xy} = \left(\frac{|h_{xy}|^2}{\beta}\right)^{\frac{1}{\alpha}}$ corresponds to the random radius r_x from the Gilbert's random disc model, however, now it depends on both nodes x and y. So even though this model seems similar to the Gilbert's random disc model, it is much harder to analyze since the edges/link connections are directional in nature, and connections are not monotonic with distance. We represent ρ_h as the random variable denoting the random radius r_{xy}.

As before, we assume that the node locations x are distributed as a PPP Φ with density λ, and assuming there is a node of Φ at the origin o, we are interested in finding the probability that the size of the connected component of the origin is unbounded. The critical density $\lambda_c(\rho_h)$ is defined correspondingly.

The technique of obtaining the lower bound on λ_c for the Gilbert's random disc model can be used for obtaining a lower bound on the λ_c for the path-loss plus fading model as follows.

Theorem 7.3.30 $\lambda_c(\rho_h) \geq \frac{1}{c\mathbb{E}\{\rho_h^2\}}$, *where c is a constant.*

Proof: For the sake of brevity, we just indicate how to modify the proof of Theorem 7.3.21, to get the proof of Theorem 7.3.30. Corresponding to Definitions 7.3.15, 7.3.18, and 7.3.19, we define the following quantities

Definition 7.3.31 *For any bounded set $\mathcal{B} \subset \mathbb{R}^2$, let $A_{\mathcal{B}}(r)$ be the event that the maximum length of any edge between any pair of nodes $x, y \in \Phi \cap \mathcal{B}$ is less than r, that is, $A_{\mathcal{B}}(r) = \{r_{xy} \leq r, \forall\, x, y \in \Phi \cap \mathcal{B}\}$.*

Definition 7.3.32 *For a point $q \in \mathbb{R}^2$, let $E(q, r)$ denote the event that there is a path from a node $x \in \Phi \cap q + B_r$ to a node $y \in \Phi \cap q + B_{9r} \backslash q + B_{8r}$ with all the nodes on the path between x and y lying inside $B_{10r} + q$.*

Note that due to stationarity $\mathbb{P}(E(q, r)) = \mathbb{P}(E(o, r))$.

Definition 7.3.33 *Let $G(r)$ be the event that there is a node $x \in \Phi \cap B_{9r}$ that has an edge to a node lying outside of B_{10r}, that is, there is at least one node $x \in B_{9r}$, such that $||x - y|| \leq r_{xy}$ for $y \in B_{10r}^c$. Note that edges are directional now. Formally,*

$$G(r) = \{(||x - y|| \leq r_{xy}, \text{ for some } y \in \Phi \backslash B_{10r} \text{ and some } x \in B_{9r}\}.$$

Following the main ideas of proof of Theorem 7.3.21, one can easily show that if the farthest node x_F of the connected component of the origin o lies outside B_{10r} and if $G(r)$ does not occur, then $E(o, r)$ occurs, and we get

Proposition 7.3.34

$$\mathbb{P}(x_F \in B_{10r}^c) \leq \mathbb{P}(E(o, r)) + \mathbb{P}(G(r)). \tag{7.26}$$

Moreover, we have

$$\mathbb{P}(|\mathsf{C}| = \infty) \leq \lim_{r \to \infty} \mathbb{P}(x_F \in B_{10r}^c).$$

It can be shown similar to Proposition 7.3.29 that $\mathbb{P}(G(r)) \to 0$ as $r \to \infty$. Hence, showing that $\mathbb{P}(E(o, r))$ goes to zero as $r \to \infty$ for $\lambda < \lambda_c$ is sufficient to lower bound λ_c. Next, we indicate how to show $\mathbb{P}(E(o, r))$ goes to zero as $r \to \infty$ for $\lambda < \lambda_c$.

Similar to Lemma 7.3.22, if event $E(o, 10r)$ occurs together with $A_{100r}(r)$, then we have that two independent events $\cup_{k \in K} E(rk, r)$ and $\cup_{\ell \in L} E(r\ell, r)$ happen simultaneously, since $r_{xy} < r$ for all $x, y \in B_{100r}$ because of $A_{100r}(r)$, where K and L be two finite subsets of the boundary of B_{10} and B_{80}, respectively, and $B_{10} \backslash B_9 \subset K + B_1, B_{81} \backslash B_{80} \subset L + B_1$. Each of these two events is a translation of $E(o, r)$ and are i.i.d. to $E(o, r)$, that is, $\mathbb{P}(E(r\ell, r)) = \mathbb{P}(E(o, r))$ for any $\ell \in L$. Thus, once again, we get,

Lemma 7.3.35 $\mathbb{P}(E(o, 10r)) \leq c_3 \mathbb{P}(E(o, r))^2 + \mathbb{P}(A_{B_{100r}}(r)^c)$, *where c_3 is a constant.*

Rest of the proof follows almost identically to the proof of Theorem 7.3.21.

□

7.4 Reference Notes

The basic percolation results presented in Sections 7.2 and 7.3.1 are derived from [1, 2, 7, 8], and lecture notes of Rahul Roy from Indian Statistical Institute, New Delhi. The connectivity results of Section 7.3.2 are derived from the seminal paper [9], which kick started the research in connectivity of wireless networks. The lower bound on the critical density for the Gilbert's random disc model presented in Section 7.3.3 is derived from [11], while the upper bound follows from [12].

Bibliography

[1] G. Grimmett. 1980. *Percolation*. New York, NY: Springer-Verlag.

[2] B. Bollobas and O. Riordan. 2006. *Percolation*. Cambridge, UK: Cambridge University Press.

[3] R. Meester and R. Roy. 1996. *Continuum Percolation*. Cambridge, UK: Cambridge University Press.

[4] M. Penrose. 2002. *Random Geometric Graphs*. Oxford: Oxford University Press.

[5] S. R. Broadbent and J. M. Hammersley. 1957. "Percolation processes i. crystals and mazes." In *Proc. Cambridge Philos. Soc* 53 (629): 41.

[6] H. Kesten. 1980. "The critical probability of bond percolation on the square lattice equals 1/2." *Commun. Math. Phy.* 74 (1): 41–59, 1980.

[7] M. Franceschetti and R. Meester. 2007. *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge, UK: Cambridge University Press, no. 24.

[8] E. N. Gilbert. 1971. "Random plane networks." *J. Soci. Ind. Appl. Math.* 9 (4): 533–43.

[9] P. Gupta and P. Kumar. 1998. "Critical power for asymptotic connectivity in wireless networks." In *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming, W. M. McEneaney, G. Yin, and Q. Zhang (Eds.)*. Birkhauser, Boston, 1998.

[10] R. Durrett. 2007. *Random graph dynamics* Cambridge, UK: Cambridge University Press.

[11] J. Gouéré. 2008. "Subcritical regimes in the Poisson Boolean model of continuum percolation." *Ann. Prob.* 36 (4): 1209–20.

[12] S. Iyer, "Personal communication."

Chapter 8

Percolation and Connectivity in Wireless Networks

8.1 Introduction

In this chapter, we exploit the basic ideas of percolation theory discussed in Chapter 7 to study percolation and connectivity in wireless networks under a signal-to-interference-plus-noise ratio (SINR) model. In the SINR model, two nodes are connected if the SINR between them is larger than a threshold. The SINR model allows multiple transmitters to share the same time/frequency and consequently interfere with each other. This is in contrast to the Gilbert's disc model, where all nodes transmit in orthogonal time/frequency slots, and consequently the spatial reuse/capacity of the SINR model is significantly larger than that of the Gilbert's disc model.

We assume that the locations of the nodes of the wireless network are drawn from a Poisson point process (PPP), and study the critical density of nodes required for percolation in the SINR model. The SINR model is characterized by the interference suppression parameter γ that measures the capacity of any node to suppress the interference received from non-intended transmitters. We show that for any density larger than the critical density of an appropriate Gilbert's disc model, the SINR model percolates for a small enough $\gamma > 0$. Thus, if each node can sufficiently mitigate the interference, percolation in the Gilbert's disc model guarantees the percolation in the SINR model. Conversely, we will show that if the density is smaller than the critical density of an appropriate Gilbert's disc model, the SINR graph cannot percolate.

Next, we look at ensuring connectivity in the SINR model, where we assume that there are n nodes that are independently and uniformly distributed in a unit square. We consider the case when $\mathsf{Col}(n)$ separate frequency bands/time slots (called colors) are used by the n nodes for transmission and reception, where only signals belonging to the same color interfere with each other. We show that $\mathsf{Col}(n) = \Theta(\ln n)$ is necessary and sufficient for ensuring connectivity in the SINR model with high probability.

Finally, we study a secure SINR model, where legitimate nodes want to communicate securely among themselves in the presence of eavesdropper nodes. In the secure SINR graph, two legitimate nodes x_i and x_j are defined to be connected if the SINR at x_j from x_i is larger than the maximum of the SINR between x_i and all eavesdropper nodes. Assuming that the location of legitimate and eavesdropper nodes is drawn from two independent PPPs, we derive percolation regime for the

secure SINR graph by mapping the continuum percolation of the secure SINR graph to an appropriate percolation over a discrete graph.

8.2 SINR Graph

Consider a wireless network with the set of nodes denoted by $\Phi = \{x_n\}$. For $x_i, x_j \in \Phi$, let d_{ij} denote the distance between x_i and x_j. We assume an unit power transmission from each node. Then the received signal power at node x_j from node x_i is $g(d_{ij})$, where $g(.)$ is the monotonically decreasing signal attenuation function with distance.

With concurrent transmissions from all nodes of Φ, the received signal at node x_j at any time is

$$y_j = \sqrt{g(d_{ij})}s_i + \sum_{k\in\Phi, k\neq i} \sqrt{\gamma g(d_{kj})}s_k + w_j, \tag{8.1}$$

where s_k is signal transmitted from node x_k, and w_j is the AWGN with $\mathcal{CN}(0,1)$ distribution. From (8.1), the SINR for the x_i to x_j communication is

$$\mathsf{SINR}_{ij} = \frac{g(d_{ij})}{1 + \gamma \sum_{k\in\Phi, k\neq i} g(d_{kj})}, \tag{8.2}$$

where $0 < \gamma \leq 1$ is the processing gain of the system that captures the interference suppression capability of each node. For example, in a CDMA system, γ depends on the orthogonality between codes used by different legitimate nodes during simultaneous transmissions. Under this model, we define the SINR graph as follows.

Definition 8.2.1 *SINR graph is a directed graph $SG(\beta) = \{\Phi, \mathcal{E}\}$, with vertex set Φ, and edge set $\mathcal{E} = \{(x_i, x_j) : \mathsf{SINR}_{ij} \geq \beta\}$, where β is the SINR threshold required for correct decoding required between any two nodes of Φ.*

Each edge of the SINR graph represents successful communication between the two nodes at transmission rate determined by the threshold β.

To study percolation in the SINR graph, we assume that the locations of Φ are distributed as a homogenous PPP Φ with density λ, and refer to the SINR graph as the Poisson SINR graph (PSG). The object of interest is the critical density required for percolation in the PSG, that is,

$$\lambda_c(\mathsf{SINR}) = \inf\{\lambda : \mathbb{P}(|\mathsf{C}| = \infty) > 0\},$$

where C is the connected component of the node at the origin in the PSG.

Remark 8.2.2 *Note that we have defined PSG to be a directed graph, and the component of the origin o is its out-component, that is, the set of nodes with which o can communicate. One can similarly define the bi directional component $\mathsf{C}^{bd} = \{x_k \in \Phi, x_k \to o \text{ and } o \to x_k\}$, and either one-directional component $\mathsf{C}^{ed} = \{x_k \in \Phi, x_k \to o \text{ or } o \to x_k\}$.*

8.3 Percolation on the PSG

We first look at the sub-critical regime of percolation in PSG, where we can directly reuse results from the sub critical regime of the Gilbert's disc model discussed in Section 7.3.1.

Consider the special case of $\gamma = 0$, where there is an edge in the PSG between nodes x_i and x_j if $g(d_{ij}) \geq \beta$, that is, if the distance d_{ij} between them is less than a fixed threshold $g^{-1}(\beta)$. Thus, the PSG with $\gamma = 0$ is equivalent to a Gilbert's disc model $G_P(\lambda, r_{\mathrm{SINR}})$, where the radius $r_{\mathrm{SINR}} = g^{-1}(\beta)$. This correspondence between the PSG for $\gamma = 0$ and Gilbert's disc model can be used directly to find the sub-critical regime for the PSG percolation as follows.

Let $\lambda_c(r_{\mathrm{SINR}})$ be the critical density of percolation in Gilbert's disc model with radius r_{SINR}, that is, for $\lambda < \lambda_c(r_{\mathrm{SINR}})$, there is no percolation in the Gilbert's disc model with radius r_{SINR}. Thus, for $\gamma = 0$, there is no percolation in the PSG for $\lambda < \lambda_c(r_{\mathrm{SINR}})$. Since PSG with $\gamma > 0$ is a subgraph of the PSG with $\gamma = 0$, we have that if $\lambda < \lambda_c(r_{\mathrm{SINR}})$, the PSG cannot percolate for any $\gamma > 0$. Therefore,

$$\lambda_c(\mathsf{SINR}) > \lambda_c(r_{\mathrm{SINR}}).$$

Finding the super critical regime is comparatively non trivial, where we will be able to show, that for any $\lambda > \lambda_c(r_{\mathrm{SINR}})$, there exists a small enough $\gamma > 0$ such that the PSG percolates. Note that percolation is monotonic in γ, that is, if there is percolation for $\gamma = \gamma_0$, then there is percolation for all $\gamma < \gamma_0$.

Theorem 8.3.1 *For the PSG, where the attenuation function $g(.)$ is monotonically decreasing and satisfies $g(.) \leq 1$ and $\int xg(x)\mathrm{d}x < \infty$, for any $\lambda > \lambda_c(r_{\mathrm{SINR}})$, and any β, there exists a small enough $\gamma > 0$ for which the probability of percolation on the PSG is greater than zero.*

The main idea of the proof is to consider a super-critical Gilbert's disc model $G_P(\lambda, r)$ and map it to an appropriate square lattice percolation, such that whenever the square lattice percolates, the connections/edges in the super critical Gilbert's disc model $G_P(\lambda, r)$ are also present in the PSG. Hence, making the PSG also super critical.

Proof: For the Gilbert's disc model $G_P(\lambda, r)$, let $r_{\mathsf{c}}(\lambda)$ be the critical radius for a fixed density λ, that is,

$$r_{\mathsf{c}}(\lambda) = \inf\{r : \mathbb{P}(|\mathsf{C}| = \infty) > 0\}.$$

Thus, for $r > r_{\mathsf{c}}(\lambda)$, the Gilbert's disc model percolates. By invariance property (Remark 7.3.1), we have

$$\lambda_c(r_{\mathrm{SINR}}) r_{\mathrm{SINR}}^2 = \lambda r_{\mathsf{c}}^2(\lambda).$$

We consider a super critical Gilbert's disc model $G_P(\lambda, r)$, for $\lambda > \lambda_c(r_{\mathrm{SINR}})$ and $r_{\mathsf{c}}(\lambda) < r < r_{\mathrm{SINR}}$.

We will map the continuum percolation of $G_P(\lambda, r)$ to a discrete edge percolation over a square lattice $\mathbf{S}$ with side s. Let the center of any horizontal edge $e \in \mathbf{S}$ be $c_e = (x_e, y_e) \in \mathbb{R}^2$. Then we define a binary event A_e which is 1 if ($A_e = 0$ otherwise)

- the rectangle $\mathbf{R}_e = [x_e - \frac{3s}{4}, x_e + \frac{3s}{4}] \times [y_e - \frac{s}{4}, y_e + \frac{s}{4}]$ containing the edge e has a left-right crossing of the connected component of the super critical Gilbert's disc model $G_P(\lambda, r)$, and
- both squares $[x_e - \frac{3s}{4}, x_e - \frac{s}{4}] \times [y_e - \frac{s}{4}, y_e + \frac{s}{4}]$ and $[x_e + \frac{s}{4}, x_e + \frac{3s}{4}] \times [y_e - \frac{s}{4}, y_e + \frac{s}{4}]$ containing the edge e have a top-bottom crossing of the connected component of the super-critical Gilbert's disc model $G_P(\lambda, r)$. The event A_e is similarly defined for a vertical edge.

Both these conditions are pictorially illustrated in Fig. 8.1.

Next, we define another event B_e corresponding to each edge e that represents that the interference received inside any point of the rectangle $\mathbf{R}_e = [x_e - \frac{3s}{4}, x_e + \frac{3s}{4}] \times [y_e - \frac{s}{4}, y_e + \frac{s}{4}]$

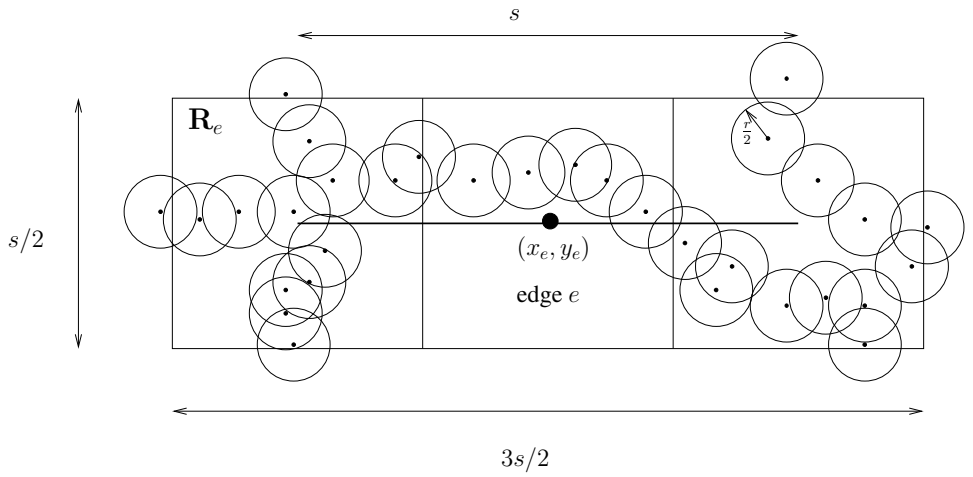

Figure 8.1: Definition of event A_e for any edge e in $\mathbf{S}$.

is less than a certain fixed threshold M. Towards that end, we define a new path-loss function $\tilde{g}$ as follows.

$$\tilde{g}(x) = \begin{cases} g(0), & x \leq \frac{\sqrt{10}s}{4}, \\ g\left(x - \frac{\sqrt{10}s}{4}\right), & x > \frac{\sqrt{10}s}{4}. \end{cases} \tag{8.3}$$

Then the interference $\tilde{I}$ received at any point $x \in \mathbb{R}^2$ is correspondingly defined as

$$\tilde{I}(x) = \sum_{x_k \in \Phi} \tilde{g}(|x - x_k|).$$

Recall that the interference seen at any point $x \in \mathbb{R}^2$ in the PSG without the interference suppression parameter γ is

$$I(x) = \sum_{x_k \in \Phi} g(|x - x_k|).$$

For an edge e, we define a binary event $B_e = 1$ if the interference defined by function $\tilde{I}(.)$ is less than M at the center of the edge, that is, if $\tilde{I}(c_e) < M$, and $B_e = 0$ otherwise. Since the distance between any point inside the rectangle $\mathbf{R}_e = [x_e - \frac{3s}{4}, x_e + \frac{3s}{4}] \times [y_e - \frac{s}{4}, y_e + \frac{s}{4}]$ and the center of the rectangle $\mathbf{R}_e$ is at most $\frac{\sqrt{10}s}{4}$, $\tilde{I}(c_e) < M$ implies $I(x) < M$ for any $x \in \mathbf{R}_e$. Thus, if the interference seen at the center is less than M with interference function $\tilde{I}$, then the interference seen at any point of the rectangle is less than M with interference function $I(.)$.

Then, we define an edge $e \in \mathbf{S}$ to be open (closed) if both $A_e = 1$ and $B_e = 1$, that is, $C_e = A_e B_e = 1$ ($C_e = 0$ otherwise). We will be interested in the regime when $\mathbf{S}$ percolates, that is, the connected component of $\mathbf{S}$ is of unbounded size with non-zero probability. Next, we show that if the square lattice $\mathbf{S}$ percolates, so does the PSG for some value of $\gamma > 0$.

First, we show that if $B_e = 1$, then any two nodes of $\mathbf{R}_e \cap \Phi$ connected in $G_P(\lambda, r)$ for $r_c(\lambda) < r < r_{\mathrm{SINR}}$ have an edge between them in the PSG as well. Note that the PSG is a directed graph, and what we claim is that edges exist in both directions in the PSG corresponding to the undirected edge

in the $G_P(\lambda, r)$. With $B_e = 1$, the interference received at any point in $\mathbf{R}_e$ is less than M. Thus, for any two nodes $x_i, x_j \in \mathbf{R}_e$ that are connected in $G_P(\lambda, r)$, that is, $|x_j - x_i| \leq r$, the SINR at node x_j from a node x_i is

$$\begin{aligned} \mathsf{SINR}_{ij} &= \frac{g(|x_i - x_j|)}{1 + \gamma \sum_{x \in \Phi \setminus \{x_i\}} g(|x - x_j|)}, \\ &\geq \frac{g(|x_i - x_j|)}{1 + \gamma M}, \text{ since } B_e = 1, I(x_j) \leq M, \\ &\geq \frac{g(r)}{1 + \gamma M}, \text{ since } |x_j - x_i| \leq r. \end{aligned}$$

As $r < r_{\mathrm{SINR}}$ and as $g(.)$ is strictly decreasing, we can choose

$$\gamma = \frac{1}{M} \left(\frac{g(r)}{g(r_{\mathrm{SINR}})} - 1 \right) > 0,$$

for which

$$\mathsf{SINR}_{ij} \geq g(r_{\mathrm{SINR}}) = \beta,$$

using the definition of $r_{\mathrm{SINR}} = g^{-1}(\beta)$. Thus, the nodes lying in $\mathbf{R}_e$ connected in the Gilbert's disc model are also connected in the PSG if $B_e = 1$.

If square lattice $\mathbf{S}$ percolates (for some choice of side length s and interference upper bounding parameter M in Lemma 8.3.2), then there is an infinite sequence of connected open edges of the square lattice. Using the definition of event A_e, consequently, we know that there is a connected component of $G_P(\lambda, r)$ crossing (both left-right and top-bottom) the rectangles $\mathbf{R}_e$'s corresponding to the connected open edges e's of $\mathbf{S}$, as shown in Fig. 8.2. We also know that an edge $e \in \mathbf{S}$ is open only if $B_e = 1$, hence from above, for small enough $\gamma > 0$, all the edges of the connected component of $G_P(\lambda, r)$ crossing the rectangles $\mathbf{R}_e$'s corresponding to the connected open edges e of $\mathbf{S}$ are also present in the PSG. Hence, PSG percolates whenever square lattice $\mathbf{S}$ and $G_P(\lambda, r)$ does for small enough $\gamma > 0$. □

The only thing left to be shown is that the square lattice percolates for a large enough choice of square lattice size s and interference threshold parameter M. This is accomplished as follows. We will use the idea of defining closed circuits in the dual lattice of $\mathbf{S}$ described in Definition 7.2.7 in Section 7.2.1 for this purpose and show that for large enough choice of square lattice side s and interference threshold parameter M, the probability of finding a closed circuit in the dual lattice around the origin in $\mathbf{S}$ is zero.

Lemma 8.3.2 *The square lattice* $\mathbf{S}$ *percolates for some choice of lattice side* s *and interference upper bounding parameter* M.

Proof: We want to bound the probability of having a closed circuit in the dual lattice of $\mathbf{S}$ surrounding the origin in $\mathbf{S}$. As before, the dual lattice $\mathbf{S}' = \mathbf{S} + \left(\frac{1}{2}, \frac{1}{2}\right)$ and any edge of the dual lattice is open/closed if the intersecting edge of $\mathbf{S}$ is open/closed. Recall that an edge $e \in \mathbf{S}$ is open (closed) if both $A_e = 1$ and $B_e = 1$, that is, $C_e = A_e B_e = 1$ ($C_e = 0$ otherwise).

Letting $A_i = A_{e_i}, B_i = B_{e_i}$, and $C_i = C_{e_i}$, we will first bound the probability of a closed circuit of length n, that is, $\mathbb{P}(C_1 = 0, C_2 = 0, \ldots, C_n = 0),\ \forall\ n \in \mathbb{N}$ considering n distinct edges. Let $p_A = \mathbb{P}(A_n = 0)$ for any n. Then we have the following intermediate results to upper bound $\mathbb{P}(C_1 = 0, C_2 = 0, \ldots, C_n = 0)$.

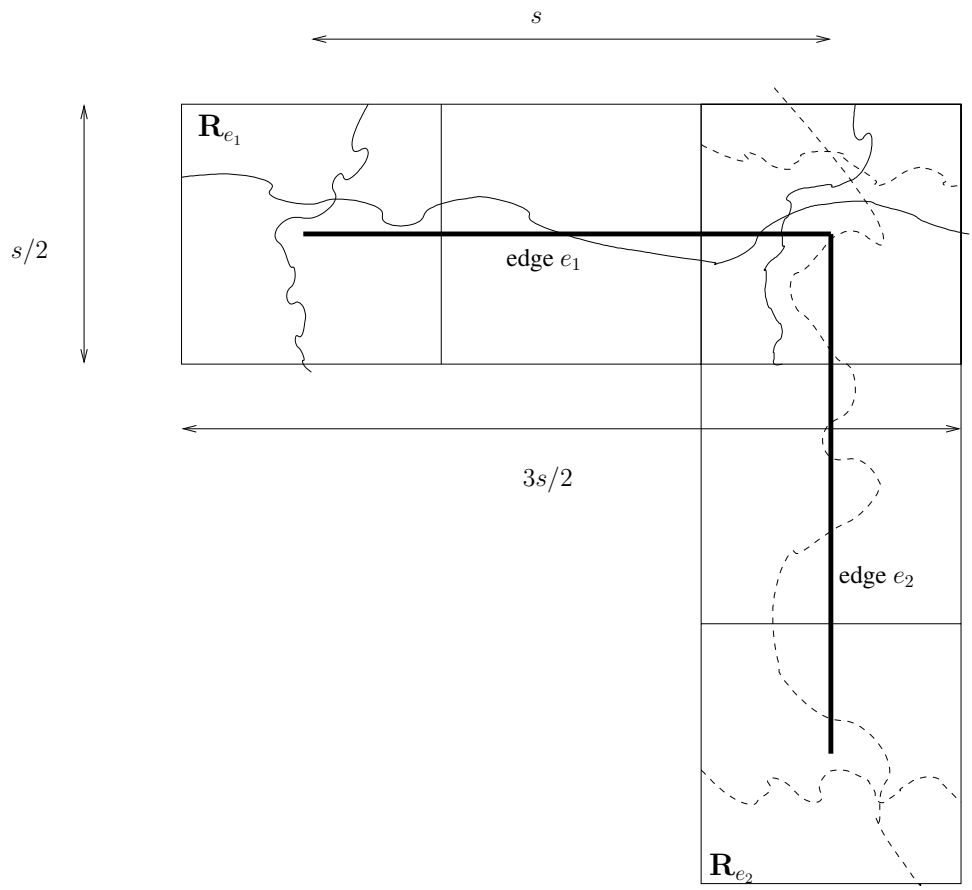

Figure 8.2: Two adjacent open edges of $\mathbf{S}$ imply a connected component of $G_P(\lambda, r)$ crossing rectangle $\mathbf{R}_e$'s corresponding to the open edges of $\mathbf{S}$. Solid lines are for rectangle $\mathbf{R}_{e_1}$ and dashed lines for $\mathbf{R}_{e_2}$.

Lemma 8.3.3 $\mathbb{P}(A_1 = 0, A_2 = 0, \ldots, A_n = 0) \leq p_1^n$, *where* $p_1 = p_A^{1/4}$.

Proof: Follows from the fact that in any sequence of n edges of $\mathbf{S}$, there are at least $n/4$ edges such that their adjacent rectangles $\mathbf{R}_e$'s do not overlap. Therefore, $\mathbb{P}(A_1 = 0, A_2 = 0, \ldots, A_n = 0) \leq \mathbb{P}(\cap_{e \in O} A_e = 0)$, where O is the set of edges for which their adjacent squares $\mathbf{R}_1(e) \cup \mathbf{R}_2(e)$ have no overlap, and $|O| \geq n/4$. Since $\mathbf{R}_e,\ e \in O$ have no overlap, the events $A_e = 0$ are independent for $e \in O$, and the result follows. □

Remark 8.3.4 *Since $G_P(\lambda, r)$ is super critical, by choosing a large enough side length s, there is always a connected component of $G_P(\lambda, r)$ satisfying the event A for any edge [1], and hence p_A can be made arbitrarily small by choosing a large enough side length s.*

Lemma 8.3.5 *[3, Proposition 2] For $\int_0^\infty xg(x)\mathrm{d}x < \infty$ and $g(.) \leq 1$,*

$$\mathbb{P}(B_1 = 0, B_2 = 0, \ldots, B_n = 0) \leq p_2^n,$$

where $p_2 = \exp\left(\frac{2\lambda}{K}\int \tilde{g}(x)\mathrm{d}x - \frac{M}{K}\right)$, and K is a constant.

Proof: With c_i as center of the edge e_i, by definition,

$$\begin{aligned}\mathbb{P}(B_1 = 0, B_2 = 0, \ldots, B_n = 0) &= \mathbb{P}(\tilde{I}(c_1) \geq M, \tilde{I}(c_2) \geq M, \ldots, \tilde{I}(c_3) \geq M),\\ &\leq \mathbb{P}\left(\sum_{i=1}^n \tilde{I}(c_i) \geq nM\right),\\ &\leq \exp(-tnM)\mathbb{E}\left\{\exp\left(t\sum_{i=1}^n \tilde{I}(c_i)\right)\right\},\end{aligned}$$

where in the last inequality we have used Markov inequality for any $t > 0$. Now we make use of Campbell's theorem (Theorem 2.3.7) to compute the expectation on the R.H.S. to get

$$\mathbb{E}\left\{\exp\left(t\sum_{i=1}^n \tilde{I}(c_i)\right)\right\} = \exp\left(\lambda\left(\int_{\mathbb{R}^2}\left(\exp\left(t\sum_{i=1}^n \tilde{g}(|x-c_i|)\right)\right) - 1\right)\mathrm{d}x\right). \tag{8.4}$$

The centers c_i of edges are located on a square lattice with side $\frac{s}{\sqrt{2}}$, thus the interference seen at any point $x \in \mathbb{R}^2$, $\sum_{i=1}^n \tilde{g}(|x-c_i|)$, is coming from vertices of a square lattice with side $\frac{s}{\sqrt{2}}$, as shown in Fig. 8.3.

For any $x \in \mathbb{R}^2$, the contribution in $\sum_{i=1}^n g(|x-c_i|)$ from the four nearest vertices c_i is 4, since $g(.) \leq 1$. Looking at the next nearest level in Fig. 8.3, there are at most 12 (solid) vertices c_i at a distance of at least $\frac{s}{\sqrt{2}}$. Continuing on, at the k^{th} nearest level, there are at most $4 + 8k$ vertices c_i at a distance of at least $\frac{ks}{\sqrt{2}}$. Thus,

$$\begin{aligned}\sum_{i=1}^n \tilde{g}(|x-c_i|) &\leq 4 + \sum_{k=1}^\infty (4+8k)\tilde{g}\left(\frac{ks}{\sqrt{2}}\right),\\ &= K.\end{aligned} \tag{8.5}$$

From assumption, $\int_y^\infty xg(x)\mathrm{d}x < \infty$ for some $y > 0$, it follows that for some $y > 0$, $\int_y^\infty x\tilde{g}(x)\mathrm{d}x < \infty$. Hence, K is bounded. Since t can be arbitrary, we let $t = \frac{1}{K}$ and get

$$t\sum_{i=1}^n \tilde{g}(|x-c_i|) \leq 1, \ \forall\, x. \tag{8.6}$$

Noting that $\exp(x) - 1 < 2x$ for all $x \leq 1$, we have

$$\exp\left(t\sum_{i=1}^n \tilde{g}(|x-c_i|)\right) - 1 \leq 2t\sum_{i=1}^n \tilde{g}(|x-c_i|) = \frac{2}{K}\sum_{i=1}^n \tilde{g}(|x-c_i|), \tag{8.7}$$

by substituting $t = 1/K$. Thus, using (8.4), we get

$$\mathbb{E}\left\{\exp\left(t\sum_{i=1}^n \tilde{I}(c_i)\right)\right\} \leq \exp\left(\lambda\int_{\mathbb{R}^2}\frac{2}{K}\sum_{i=1}^n \tilde{g}(|x-c_i|)\mathrm{d}x\right),$$

$$= \exp\left(\frac{2n\lambda}{K}\int_{\mathbb{R}^2}\tilde{g}(|x|)\mathrm{d}x\right),$$
$$= \left(\exp\left(\frac{2\lambda}{K}\int_{\mathbb{R}^2}\tilde{g}(|x|)\mathrm{d}x\right)\right)^n. \quad (8.8)$$

Thus, we get

$$\begin{aligned}\mathbb{P}(B_1=0,B_2=0,\ldots,B_n=0) &= \exp(-tnM)\mathbb{E}\left\{\exp\left(s\sum_{i=1}^{n}\tilde{I}(c_i)\right)\right\},\\ &\leq \exp\left(-\frac{nM}{K}\right)\left[\exp\left(\frac{2\lambda}{K}\int_{\mathbb{R}^2}\tilde{g}(|x|)\mathrm{d}x\right)\right]^n,\\ &= \left[\exp\left(\left(\frac{2\lambda}{K}\int_{\mathbb{R}^2}\tilde{g}(|x|)\mathrm{d}x-\frac{M}{K}\right)\right)\right]^n,\end{aligned}$$

where the second inequality is obtained from (8.8) for $t = 1/K$. Finally, we define $p_2 = \exp\left(\left(\frac{2\lambda}{K}\int_{\mathbb{R}^2}\tilde{g}(|x|)\mathrm{d}x-\frac{M}{K}\right)\right)$.

□

Next, we upper bound the probability of having n consecutive closed edges $\mathbb{P}(C_1=0,C_2=0,\ldots,C_n=0)$.

Lemma 8.3.6 *[3, Proposition 3]* $\mathbb{P}(C_1=0,C_2=0,\ldots,C_n=0)\leq(\sqrt{p_1}+\sqrt{p_2})^n$.

Proof:

$$\begin{aligned}\mathbb{P}(C_1=0,C_2=0,\ldots,C_n=0) &\overset{(a)}{=} \mathbb{E}\left\{(1-C_1)(1-C_2)\ldots(1-C_n)\right\},\\ &\overset{(b)}{\leq} \mathbb{E}\left\{(\bar{A}_1+\bar{B}_1)(\bar{A}_2+\bar{B}_2)\ldots(\bar{A}_n+\bar{B}_n)\right\},\\ &\overset{(c)}{\leq} \sum_{\mathbf{b}\in[2^n]}\mathbb{E}\left\{\left(\prod_{i:\mathbf{b}_i=0}\bar{A}_i\right)\left(\prod_{i:\mathbf{b}_i=1}\bar{B}_i\right)\right\},\end{aligned}$$

where (a) follows since $C_i\in\{0,1\}$, (b) follows by defining $\bar{A}_i=1-A_i$, and noting that

$$1-C_i=1-A_iB_i\leq(1-A_i)+(1-B_i)=\bar{A}_i+\bar{B}_i,$$

and (c) follows by taking the sum over all binary 2^n binary sequences $\mathbf{b}$. Next, we use the Cauchy–Schwarz inequality to upper bound the expectation of the product to get

$$\begin{aligned}\mathbb{P}(C_1=0,C_2=0,\ldots,C_n=0) &\overset{(d)}{\leq} \sum_{\mathbf{b}\in[2^n]}\sqrt{\mathbb{E}\left\{\prod_{i:\mathbf{b}_i=0}\bar{A}_i^2\right\}\mathbb{E}\left\{\prod_{i:\mathbf{b}_i=1}\bar{B}_i^2\right\}},\\ &\overset{(e)}{\leq} \sum_{\mathbf{b}\in[2^n]}\sqrt{\mathbb{E}\left\{\prod_{i:\mathbf{b}_i=0}\bar{A}_i\right\}\mathbb{E}\left\{\prod_{i:\mathbf{b}_i=1}\bar{B}_i\right\}},\\ &\overset{(f)}{\leq} \sum_{\mathbf{b}\in[2^n]}\sqrt{\prod_{i:\mathbf{b}_i=0}p_1\prod_{i:\mathbf{b}_i=1}p_2},\end{aligned}$$

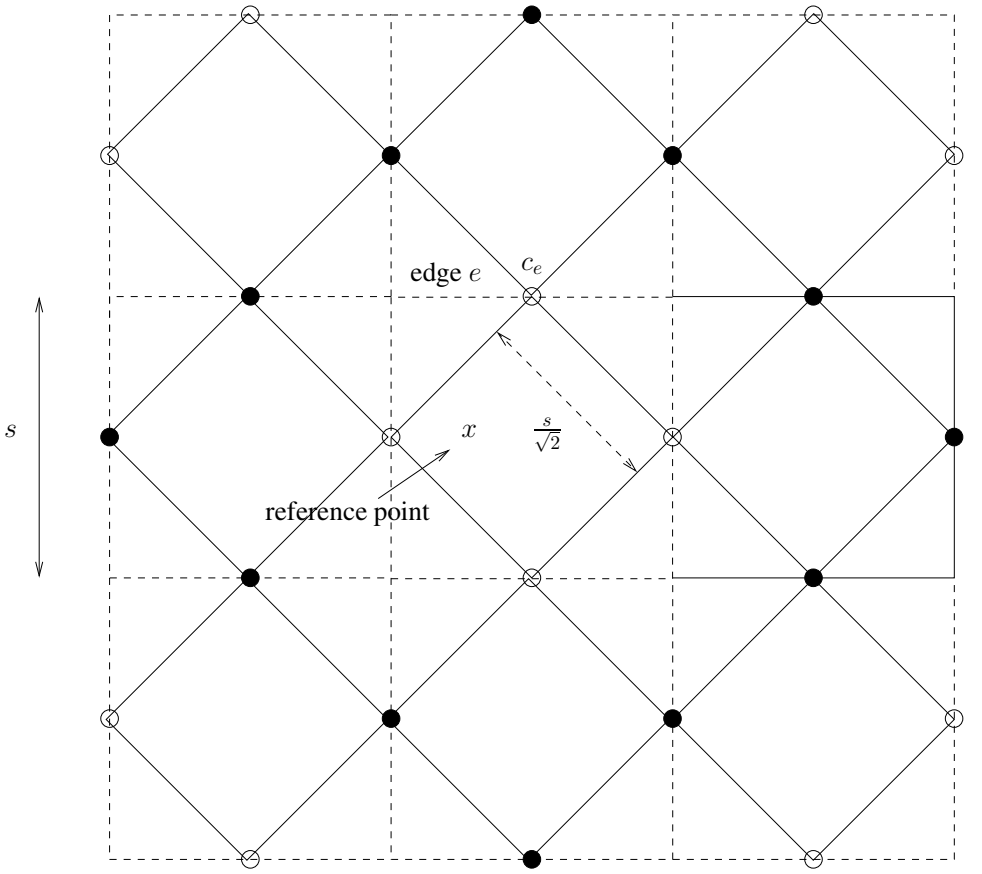

Figure 8.3: Square grid formed by centers (represented as dots) of edges of $\mathbf{S}$ with side $\frac{s}{\sqrt{2}}$.

$$\begin{aligned} &\leq \sum_{\mathbf{b}\in[2^n]} \left(\prod_{i:\mathbf{b}_i=0} \sqrt{p_1}\right)\left(\prod_{i:\mathbf{b}_i=1} \sqrt{p_2}\right), \\ &\leq (\sqrt{p_1}+\sqrt{p_2})^n, \end{aligned}$$

where (d) follows from the inequality of Cauchy–Schwarz, (e) follows since by definition $\bar{A}_i^2 = \bar{A}_i$, and (f) follows from the definition of p_1 and p_2 from Lemma 8.3.3 and Lemma 8.3.5. □

Let $q = (\sqrt{p_1} + \sqrt{p_2})$. Using the Peierls argument, the next Lemma characterizes an upper bound on q for which having a closed circuit in $\mathbf{S}$ surrounding the origin is less than 1.

Lemma 8.3.7 *If $q < \frac{11-2\sqrt{10}}{27}$, then the probability of having a closed circuit in the dual lattice $\mathbf{S}'$ surrounding the origin is less than 1.*

Proof: As shown in Proposition 7.2.6, the number of possible closed circuits of length n around the origin is less than or equal to $4n3^{n-2}$. From Lemma 8.3.6, we know that the probability of a closed

circuit of length n is upper bounded by q^n. Thus,

$$\begin{aligned}\mathbb{P}(\text{closed circuit around origin in } \mathbf{S}') &\leq \sum_{n=1}^{\infty} 4n3^{n-2}q^n, \\ &= \frac{4q}{3(1-3q)^2},\end{aligned}$$

which is less than 1 for $q < \frac{11-2\sqrt{10}}{27}$. □

Back to the proof of Lemma 8.3.2, for $q = (\sqrt{p_1} + \sqrt{p_2})$, using the definitions of p_1 and p_2 from Lemma 8.3.3 and Lemma 8.3.5, by appropriately choosing the side length s and the interference threshold M, for any density λ, we can have $q < \frac{11-2\sqrt{10}}{27}$. Hence, from Lemma 8.3.7, the probability of having a closed circuit surrounding the origin is less than 1 and therefore, from Lemma 7.2.8, we can conclude that the square lattice $\mathbf{S}$ percolates for appropriate choice of s and M. □

Thus, we have shown that for small enough interference suppression parameter γ, if the Gilbert's disc model percolates, so does the SINR graph, where the radio range used for the Gilbert's disc model is $g^{-1}(\beta)$. An alternate way of stating these results could be in terms of the SINR threshold β. Since the SINR graph percolates for small enough $\gamma > 0$, it is clear that if an edge exists between nodes x_i and x_j, that is,

$$\frac{g(d_{ij})}{1 + \gamma \sum_{k \neq i} g(d_{kj})} \geq \beta$$

for small enough γ, it will also exist even if we fix $\gamma > 0$ and choose a small enough β. Thus, alternatively, we can say that the SINR graph percolates for a small enough SINR threshold β for any fixed γ. The physical interpretation of choosing small enough β is to say that if the rate of communication between any nodes is small enough, then the connected component with the SINR graph is of unbounded size.

After establishing the percolation regime for the SINR graph, next we consider ensuring connectivity in wireless networks with the SINR connection model.

8.4 Connectivity on the SINR Graph

In this section, we now consider the problem of ensuring connectivity in the wireless network with the SINR graph. Similar to Section 7.3.2, for studying the SINR graph connectivity, we restrict ourselves to a bounded area, in particular, an unit square $\mathbf{S}_1$, where n nodes $\Phi_n = \{x_1, \ldots, x_n\}$ are located independently and uniformly randomly.

Following Section 8.2, the SINR graph on the unit square is defined as $SG(\beta, \mathbf{S}_1) = \{\Phi_n, \mathcal{E}_n\}$, where $\mathcal{E}_n = \{(x_i, x_j) \ : \mathsf{SINR}_{ij} \geq \beta\}$, and

$$\mathsf{SINR}_{ij} = \frac{g(d_{ij})}{1 + \gamma \sum_{k \in \Phi_n, k \neq i} g(d_{kj})}.$$

The presented results will not depend on γ, hence for the ease of exposition, we let $\gamma = 1$.

Definition 8.4.1 *The SINR graph $SG(\beta, \mathbf{S}_1)$ is defined to be connected if there is a path from $x_i \rightarrow x_j$ in $SG(\beta, \mathbf{S}_1)$, $\forall\, i, j = 1, 2, \ldots, n, i \neq j$.*

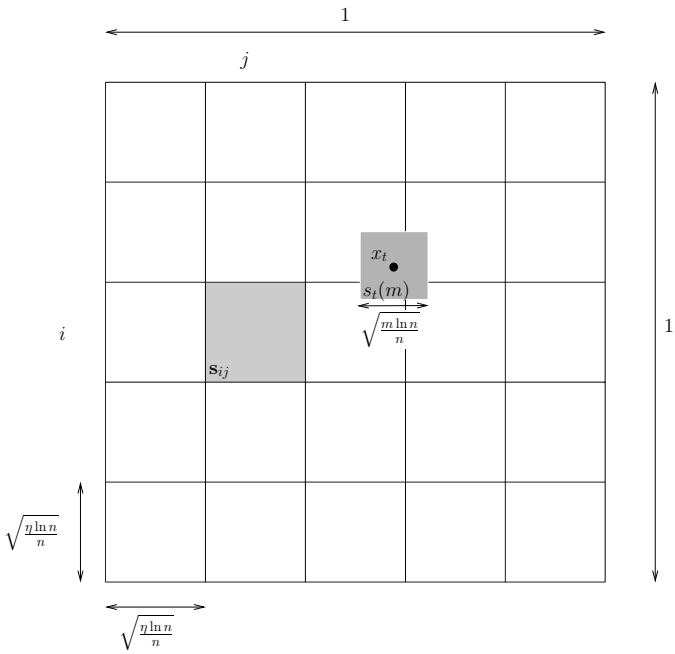

Figure 8.4: Square tiling of the unit square, and pictorial definition of square $s_t(m)$ for each node x_t.

To analyze the connectivity of the SINR graph, we color the nodes of network Φ with $\mathsf{Col}(n)$ different colors, where nodes assigned different colors correspond to having orthogonal signals in either time or frequency, which do not interfere with each other. $\mathsf{Col}(n)$ essentially corresponds to the spatial reuse in the network, smaller the value of $\mathsf{Col}(n)$, larger is the number of nodes allowed to transmit at same time/frequency and consequently larger the spatial capacity. Thus, we want to find the minimum $\mathsf{Col}(n)$ that ensures connectivity in the wireless network.

Let the color of node x_i be $\mathsf{c}(x_i)$. Then the colored SINR graph (CSG) is defined as

$$SG(\beta, \mathbf{S}_1, \mathsf{Col}(n)) = \{\Phi_n, \mathcal{E}_n\},$$

where $\mathcal{E}_n = \{(x_i, x_j) \ : \mathsf{SINR}_{ij} \geq \beta\}$, and

$$\mathsf{SINR}_{ij} = \frac{g(d_{ij})}{1 + \sum_{k \in \Phi_n, k \neq i, \mathsf{c}(x_k) = \mathsf{c}(x_i)} g(d_{kj})},$$

where only the same colored node contribute to the interference. $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is defined to be connected if there is a path from $x_i \rightarrow x_j$ in $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n)) \ \forall, i, j = 1, 2, \ldots, n, i \neq j$. Note that $SG(\beta, \mathbf{S}_1) = SG(\beta, \mathbf{S}_1, 1)$. In the next theorem, we find an upper bound on $\mathsf{Col}(n)$ for which $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is connected with high probability.

8.4.1 Upper bound on $\mathsf{Col}(n)$

We will prove the upper bound on $\mathsf{Col}(n)$ for the path-loss model $g(d_{ij}) = d_{ij}^{-\alpha}$, which can be easily extended to all other path-loss models with monotonically decreasing $g(.)$ and $\int xg(x)\mathrm{d}x < \infty$. The upper bound on the number of colors $\mathsf{Col}(n)$ that ensures connectivity in the CSG is as follows.

Theorem 8.4.2 *CSG $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is connected with high probability if $\mathsf{Col}(n) \geq 4(1+\delta)\eta \ln n$ colors are used, where η and δ are constants.*

Proof: Consider a 1×1 square $\mathbf{S}_1$, where n nodes are distributed uniformly in $\mathbf{S}_1$. We tile $\mathbf{S}_1$ into smaller squares $\mathbf{s}_{ij}$ with side $\sqrt{\frac{\eta \ln n}{n}}$ as shown in Fig. 8.4. Let the number of nodes lying in any small square $\mathbf{s}_{ij}$ be $N(\mathbf{s}_{ij})$. From Chernoff bound (Lemma 7.3.7),

$$\mathbb{P}\left(N(\mathbf{s}_{ij}) \geq (1+\delta)\eta \ln n\right) \leq n^{\frac{-\eta\delta^2}{3}}. \tag{8.9}$$

Thus, there are most a constant times $\ln n$ nodes in any small square $\mathbf{s}_{ij}$ with high probability.

For any node x_t, consider a square $\mathbf{s}_t(m)$ with side $\sqrt{\frac{m \ln n}{n}}$ centered at x_t as shown in Fig. 8.6, where $m < \eta$ is a constant. Define event

$$F_t(m) = \left\{\text{there are less than } \frac{m}{2} \ln n \text{ nodes in } \mathbf{s}_t(m)\right\}.$$

Again, using the Chernoff bound (Lemma 7.3.7), we have that

$$\mathbb{P}(F_t(m)) \leq n^{-2}$$

and taking the union bound over all n nodes x_t, we get

$$\mathbb{P}(\cup_{x_t} F_t(m)) \leq n^{-1}. \tag{8.10}$$

Thus, for large n, every square $\mathbf{s}_t(m)$ has at least a constant times $\ln n$ nodes in it.

Coloring strategy: Let $|\mathsf{Col}(n)| = 4(1+\delta)\eta \ln n$, and partition $\mathsf{Col}(n)$ into four parts $\mathsf{Col}(n) = \{\mathsf{Col}_1(n), \mathsf{Col}_2(n), \mathsf{Col}_3(n), \mathsf{Col}_4(n)\}$, where $|\mathsf{Col}_\ell(n)| = (1+\delta)\eta \ln n$, $\ell = 1, 2, 3, 4$ and $\mathsf{Col}_\ell(n) \cap \mathsf{Col}_k(n) = \phi, \forall\, \ell, k$. We drop the index n from $\mathsf{Col}_\ell(n)$ for ease of notation. Colors from set Col_1 and Col_2 are associated with alternate rows in odd numbered columns, while sets Col_3 and Col_4 are associated with alternate rows in even numbered columns in the tilting of $\mathbf{S}_1$ using $\mathbf{s}_{ij}$ as shown in Fig. 8.5. Nodes in each smaller square $\mathbf{s}_{ij}$ are colored as follows. Let the nodes lying in each $\mathbf{s}_{ij}$ be indexed using numbers 1 to $N(\mathbf{s}_{ij})$. Then we associate $(1+\delta)\eta \ln n$ colors to each $\mathbf{s}_{ij}$ in a regular fashion, that is, color of node x, $x = 1, \ldots, N(\mathbf{s}_{ij})$ is $\mathsf{c}(x) = x \mod (1+\delta)\eta \ln n$.

Define $E_{ij} = \{\text{two nodes with the same color lie in } \mathbf{s}_{ij}\}$. From (8.9), $N(\mathbf{s}_{ij}) < (1+\delta)\eta \ln n$ with high probability for any square $\mathbf{s}_{ij}$, hence with this coloring, the probability that there are two or more nodes with the same color in a given square $\mathbf{s}_{ij}$ is

$$\mathbb{P}(E_{ij}) \leq n^{\frac{-\eta\delta^2}{3}}. \tag{8.11}$$

Taking the union bound over all squares, we get

$$\mathbb{P}(\cup_{i,j} E_{ij}) \leq n^{1-\frac{\eta\delta^2}{3}}. \tag{8.12}$$

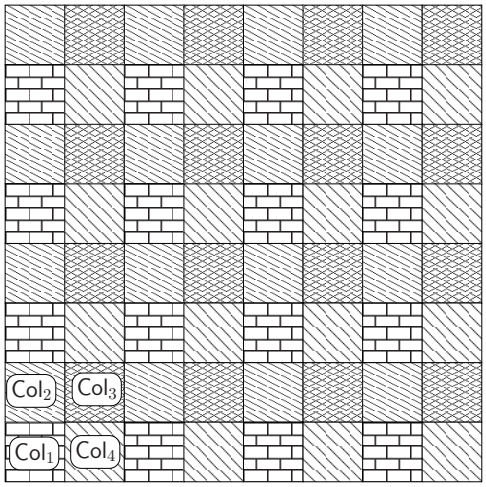

Figure 8.5: Coloring the square tiling of the unit square with four sets of colors.

Thus,

$$\lim_{n\to\infty} \mathbb{P}(\cup_{i,j} E_{ij}) = 0, \tag{8.13}$$

for $\eta\delta^2/3 \geq 1$. Using the union bound, $\mathbb{P}(\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\}) \leq \mathbb{P}(\cup_{ij} E_{ij}) + \mathbb{P}(\cup_{ij} E_{ij})$ and from (8.10) and (8.13), we have

$$\mathbb{P}(\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\}) \to 0, \tag{8.14}$$

as $n \to \infty$.

Thus, with high probability, each square $\mathbf{s}_t(m)$ has at least $\eta \ln n$ nodes, and no square $\mathbf{s}_{ij}$ contains two or more nodes with the same color.

With the definition of events E_{ij} and $F_t(m)$, the probability that the CSG is connected can be written as $\mathbb{P}(SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is connected)

$$\begin{aligned}
= \quad & \mathbb{P}(\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\}) \\
& \mathbb{P}\left(SG(\beta, \mathbf{S}_1, \mathsf{Col}(n)) \text{ is connected } | \{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\}\right) \\
& + \mathbb{P}((\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\})^c) \\
& \mathbb{P}\left(SG(\beta, \mathbf{S}_1, \mathsf{Col}(n)) \text{ is connected } | (\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\})^c\right), \\
\sim \quad & \mathbb{P}\left(SG(\beta, \mathbf{S}_1, \mathsf{Col}(n)) \text{ is connected} | (\{\cup_{ij} E_{ij}\} \cup \{\cup_t F_t(m)\})^c\right), \text{from (8.14).}
\end{aligned}$$

Hence, we can analyze the SINR connectivity while conditioning on the event that no square $\mathbf{s}_{ij}$ has more than two nodes with the same color, and each square $\mathbf{s}_t(m)$ has at least $\frac{m}{2} \ln n$ nodes.

Under this conditioning, to show that the CSG is connected, it is sufficient to show that for any $t = 1, \ldots, n$, x_t is connected to all nodes in $\mathbf{s}_t(m)$ (square of side $\sqrt{\frac{m \ln n}{n}}$ centered at node x_t) in the CSG. To see this, ignoring edge effects, if x_t is connected to all nodes in $\mathbf{s}_t(m)$ in the CSG, it is also connected to all nodes lying in four equal partitions of the square $\mathbf{s}_t(m)$. Since each of the four smaller squares also have a constant times $\ln n$ nodes from the Chernoff bound, thus x_t has a

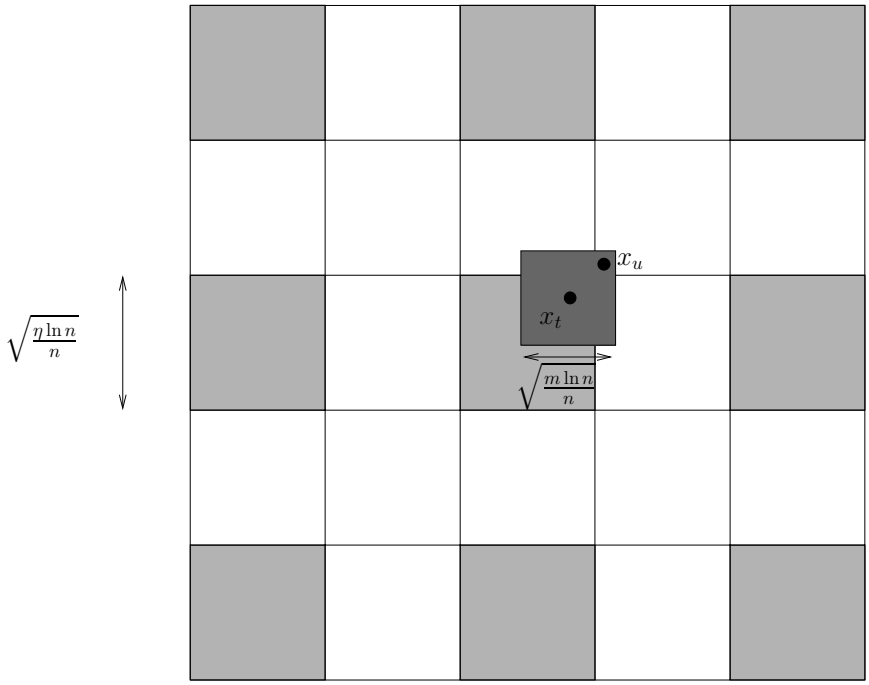

Figure 8.6: Interference for node x_u with respect to node x_t only comes from at most one node lying in the shaded squares, where distance from x_u to nodes in the shaded squares at level z is at least $\left(2z\left(\sqrt{\frac{\eta \ln n}{n}} - \sqrt{\frac{m \ln n}{n}}\right)\right)$.

connection to nodes in all four directions, and recursing this argument for each node x_t, we see that there is a connected path from each node to every other node in the network.

Consider a node x_t and its corresponding square $\mathbf{s}_t(m)$ of side $\sqrt{\frac{m \ln n}{n}}$ centered at x_t. Next, we show that x_t is connected to all nodes $x_u \in \mathbf{s}_t(m), u \neq t$ in the CSG. Since the side length of $\mathbf{s}_t(m)$ is $\sqrt{\frac{m \ln n}{n}}$, the distance between x_t and x_u, d_{tu}, is upper bounded by $\sqrt{\frac{2m \ln n}{n}}$, thus the signal power between x_t and x_u is lower bounded by

$$d_{tu}^{-\alpha} \geq \left(\frac{2m \ln n}{n}\right)^{-\alpha/2}. \tag{8.15}$$

Now consider Fig. 8.6 for analyzing the interference power seen at x_u. Without loss of generality, assume that x_t belongs to the square associated with color set Col_1, and is associated with color $\mathsf{c}(x_t)$. Note that there is no other node in $\mathbf{s}_{ij}$ that has the same color as x_t. So the interference received at x_u is attributed to nodes lying in square $\mathbf{s}_{i'j'}$ associated with color $\mathsf{c}(x_t) \in \mathsf{Col}_1$, where either $i' \neq i$ or $j' \neq j$.

From Fig. 8.6, it is clear that for any level k, $\mathsf{k} = 1, 2, \ldots, n$, there are maximum $8\mathsf{k}$ nodes using the same color as x_t, at a distance at least $\mathsf{k}\left(\sqrt{\frac{\eta \ln n}{n}} - \sqrt{\frac{m \ln n}{n}}\right)$ from x_u. Thus, the interference

power

$$\sum_{v\neq t,\mathsf{c}(x_v)=\mathsf{c}(x_t)} d_{vu}^{-\alpha} \leq \sum_{\mathsf{k}=1}^{n} \frac{8\mathsf{k}}{\left(\mathsf{k}\left(\sqrt{\frac{\eta \ln n}{n}}-\sqrt{\frac{m\ln n}{n}}\right)\right)^{\alpha}}.$$

Since $m<\eta$ is a constant by construction,

$$\left(\sqrt{\frac{\eta \ln n}{n}}-\sqrt{\frac{m\ln n}{n}}\right) \geq \kappa\left(\sqrt{\frac{\eta \ln n}{n}}\right)$$

for some $\kappa>0$. Hence, the interference seen at x_u is upper bounded by

$$\sum_{v\neq t,\mathsf{c}(x_v)=\mathsf{c}(x_t)} d_{vu}^{-\alpha} \leq \frac{8}{(\kappa)^{\alpha}\left(\frac{\eta\ln n}{n}\right)^{\alpha/2}} \sum_{\mathsf{k}=1}^{n} \mathsf{k}^{1-\alpha}. \tag{8.16}$$

Since $\sum_{\mathsf{k}=1}^{n}\mathsf{k}^{1-\alpha}$ converges for $\alpha\geq 2$, let $c_5=\sum_{\mathsf{k}=1}^{n}\mathsf{k}^{1-\alpha}$. Then, from (8.15) and (8.16) computing the SINR, we have

$$\mathsf{SINR}_{tu}=\frac{d_{tu}^{-\alpha}}{1+\sum_{v\neq t,\mathsf{c}(x_v)=\mathsf{c}(x_t)} d_{vu}^{-\alpha}} \geq \frac{\kappa^{\alpha}\left(\frac{2m}{\eta}\right)^{-\alpha/2}}{8c_5}.$$

Appropriately choosing constant m, SINR_{tu} be made larger than β as required to show that x_t is connected to every $x_u\in \mathbf{s}_t(m)$.

Thus, for an appropriate choice of m and η,

$$\mathbb{P}\left(SG(\beta,\mathbf{S}_1,\mathsf{Col}(n)) \text{ is connected}|(\{\cup_{ij}E_{ij}\}\cup\{\cup_t F_t(m)\})^c\right)=1,$$

and from (8.15),

$$\lim_{n\to\infty}\mathbb{P}(SG(\beta,\mathbf{S}_1,\mathsf{Col}(n)) \text{ is connected})=1.$$

□

Theorem 8.4.2 implies that $\mathcal{O}(\ln n)$ colors are sufficient for guaranteeing the connectivity of $SG(\beta,\mathbf{S}_1,\mathsf{Col}(n))$ with high probability. The intuition behind this result is that if only $n/\ln(n)$ nodes interfere with any node's transmission, then the total interference received at any node is bounded with high probability, and each node can connect to a large number of nodes. In the next subsection, we show that actually $\mathsf{Col}(n)=\Omega(\ln n)$ colors are also necessary for the $SG(\beta,\mathbf{S}_1,\mathsf{Col}(n))$ to be connected with high probability, and if $\mathsf{Col}(n)$ is less than order $\ln n$, then the interference power can be arbitrarily large and difficult to bound, making $SG(\beta,\mathbf{S}_1,\mathsf{Col}(n))$ disconnected with high probability.

8.4.2 Lower bound on $\mathsf{Col}(n)$

In this section, we show that if less than order $\ln(n)$ colors are used, then the CSG is disconnected with high probability. To show this, we actually show that any node is not connected to any other node with high probability if less than order $\ln(n)$ colors are used. For proving this lower bound, we will restrict ourselves to path-loss models with monotonically decreasing $g(.)$ and $\int xg(x)\mathrm{d}x<\infty$, since with singular path-loss models, $g(d_{ij})=d_{ij}^{-\alpha}$, the signal power between any two nodes cannot be bounded. Formally, our result is as follows.

Theorem 8.4.3 *If* $\mathsf{Col}(n) = o(\ln n)$*, that is,* $\mathsf{Col}(n)$ *is sub-logarithmic in* n*, then the CSG* $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ *is not connected with high probability for path-loss models with monotonically decreasing* $g(.)$ *and* $\int xg(x)\mathrm{d}x < \infty$.

Proof: Similar to the last subsection, we consider the tiling of the unit square $\mathbf{S}_1$ by squares $\mathbf{s}_{ij}$, but with side $\sqrt{\frac{\ln n}{n}}$, instead of $\sqrt{\frac{\eta \ln n}{n}}$.

With this tiling, the expected number of nodes in any square $\mathbb{E}\{N(\mathbf{s}_{ij})\} = \ln n$, and from Chernoff bound $\mathbb{P}(N(\mathbf{s}_{ij}) < (1-\delta)\ln n) \leq n^{-\delta^2/2}$, for any $0 < \delta < 1$. Thus, there are at least a constant times $\ln n$ nodes in each smaller square.

Let the number of colors $|\mathsf{Col}(n)| = \frac{\beta f(n)}{\kappa}$, where β is the SINR threshold and $\lim_{n\to\infty} \frac{f(n)}{\ln n} = 0$, that is, the number of distinct colors scales slower than $\ln n$. Since there are at least a constant times $\ln n$ nodes in each smaller square $\mathbf{s}_{ij}$ with high probability, with any coloring using $|\mathsf{Col}(n)| = \frac{\beta f(n)}{\kappa}$ colors, there are at least κ/β nodes in any square $\mathbf{s}_{ij}$ using any one particular color $\mathsf{col} \in \mathsf{Col}(n)$, with high probability.

Let $\Phi_{\mathsf{col}}(\mathbf{s}_{ij}) = \{x_m : \mathsf{c}(x_m) = \mathsf{col}, x_m \in \mathbf{s}_{ij}\}$ be the set of nodes in square $\mathbf{s}_{ij}$ that use the same color col. As noted $|\Phi_{\mathsf{col}}(\mathbf{s}_{ij})| \geq \kappa/\beta$ with high probability. Consider two nodes $x_k, x_m \in \Phi_{\mathsf{col}}(\mathbf{s}_{ij}) \cap \mathbf{s}_{ij}$, that is, both x_k and x_m lie in $\mathbf{s}_{ij}$ and share the same color. Let x_n be any other node lying in $\mathbf{s}_{ij}$. By the definition of smaller square $\mathbf{s}_{ij}$, the distance between nodes x_m and x_n, d_{mn} is no more than $d_{kn} + \sqrt{\frac{2\ln n}{n}}$, where d_{kn} is the distance between x_k and x_n. Therefore, the interference received at node x_n from nodes inside $\mathbf{s}_{ij}$ using color col is

$$\sum_{x_m \in \Phi_{\mathsf{col}}(\mathbf{s}_{ij}), m \neq k} g\left(d_{mn}\right)$$

which is greater than $(\kappa/\beta - 1)g\left(d_{kn} + \sqrt{\frac{2\ln n}{n}}\right)$ since $|\Phi_{\mathsf{col}}(\mathbf{s}_{ij})| \geq \kappa/\beta$. Thus, the SINR between x_k and x_n is

$$\begin{aligned} \mathsf{SINR}_{kn} &\leq \frac{g(d_{kn})}{1 + (\frac{\kappa}{\beta} - 1)g\left(d_{kn} + \sqrt{\frac{2\ln n}{n}},\right)}, \\ &\leq \frac{g(d_{kn})}{(\frac{\kappa}{\beta} - 1)g\left(d_{kn} + \sqrt{\frac{2\ln n}{n}},\right)}. \end{aligned}$$

Since $g(.)$ is bounded, for large enough n, choosing κ appropriately, we can make $\mathsf{SINR}_{kn} < \beta$. Thus, we have shown that node x_k is not connected to any node inside $\mathbf{s}_{ij}$. Similarly, it follows that $x_k \in \mathbf{s}_{ij}$ is not connected to any node outside of $\mathbf{s}_{ij}$, since for $x_q \notin \mathbf{s}_{ij}$, the signal power $g(d_{kq})$ is less compared to $g(d_{kn})$, the signal power at any node $x_n \in \mathbf{s}_{ij}$, while the interference powers at $x_q \notin \mathbf{s}_{ij}$ and $x_n \in \mathbf{s}_{ij}$ are identical. Thus, we conclude that if less than order $\ln n$ colors are used, then $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is not connected with high probability. □

We have shown that the CSG $SG(\beta, \mathbf{S}_1, \mathsf{Col}(n))$ is disconnected as long as the number of colors or orthogonal time/frequency slots is less than order $\ln n$ colors. Thus, for ensuring connectivity, the network can tolerate at most $n/\ln n$ interferers for any node. This result holds for any SINR threshold β and interference suppression parameter γ, and hence, even for small enough β and γ, the CSG cannot be connected by using less than $\ln n$ colors. This is in contrast to our percolation

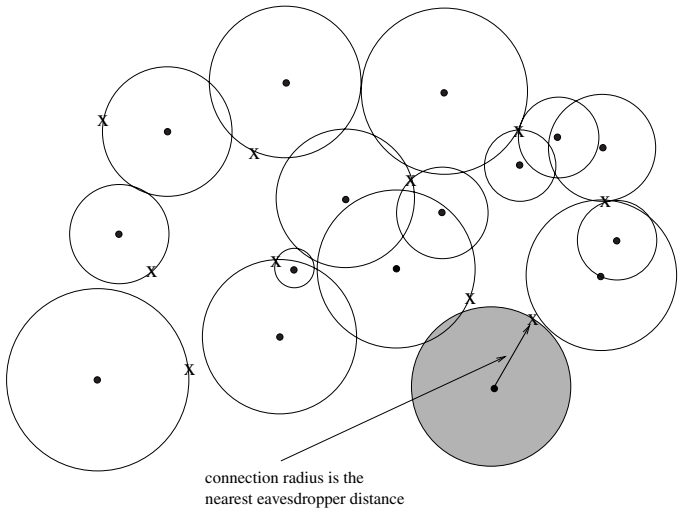

Figure 8.7: Distance-based secure graph model, where dots are legitimate nodes and crosses are eavesdropper nodes, and x_i is connected to x_j if x_j lies in the disc of x_i with radius equal to the nearest eavesdropper distance.

result over PSG, where we showed that for small enough β (γ), percolation happens for densities above the critical density for any fixed γ (β) while using a single color.

We next, consider a more general SINR graph model that allows secure communication between legitimate nodes in a wireless network in the presence of eavesdroppers.

8.5 Information Theoretic Secure SINR Graph

In this section, we consider the information theoretic secure model for communication in wireless networks [6, 7], where two legitimate nodes can communicate with each other at a non-zero rate even in the presence of eavesdroppers. Under this model, the security guarantee is absolute rather than dependent on the large complexity of problem to be solved at eavesdroppers as guaranteed by cryptography.

The simplest example of secure communication is a distance-based secure model, where two legitimate nodes i and j are connected securely, provided node j is closer to node i than its nearest eavesdropper, as shown in Fig. 8.7. These are the links over which secure communication can take place in the presence of eavesdroppers of arbitrary capability. A natural question from percolation point of view that arises is "when does the secure graph percolates as a function of the legitimate and the eavesdropper node locations?

When both the legitimate and eavesdropper nodes are distributed according to independent PPP in $\mathbb{R}^2$ of densities λ and λ_E, respectively, with the distance-based secure model, the expected number of legitimate neighbors of any legitimate node is λ/λ_E. Thus, using a branching process argument, one can show that if the expected number of legitimate neighbors $\lambda/\lambda_E < 1$, then almost surely, no unbounded connected component exits [8, 9]. Concrete results for the super critical regime are, however, harder to derive. Existential results for the super critical regime can be derived by an

appropriate mapping of the continuum percolation of the distance-based secure graph model to a square lattice [10], similar to the procedure used in Section 8.3.

The distance-based secrecy graph model has a limited scope since it assumes that the signals transmitted from different legitimate nodes do not interfere with each other. In reality, that is difficult to incorporate, since there are large number of legitimate nodes, and all nodes cannot transmit on orthogonal frequency or time slots. To generalize the distance-based secrecy graph and model a more realistic scenario, the secrecy graph is defined using the SINR model, where two legitimate nodes are connected if the SINR between them is more than the SINR at any other eavesdropper node.

The secure PSG (called SSG) allows all legitimate nodes to transmit at the same time/frequency and interfere with each other's communication. Let Φ be the set of legitimate nodes, and Φ_E be the set of eavesdropper nodes. We assume that the location of nodes in Φ and Φ_E are distributed according to independent PPPs with densities λ and λ_E, respectively. Let $x_i, x_j \in \Phi$, and $e \in \Phi_E$. Without loss of generality, we assume an average power constraint of unity $(P = 1)$ at each node in Φ, and unit noise variance at all nodes of Φ, Φ_E. Let $0 < \gamma \leq 1$ be the processing gain of the system (interference suppression parameter), which depends on the orthogonality between codes used by different legitimate nodes during simultaneous transmissions. Then the SINR between two legitimate nodes x_i and x_j is

$$\mathsf{SINR}_{ij} = \frac{g(|x_i - x_j|)}{\gamma \sum_{k \in \Phi k \neq i} g(|x_j - x_j|) + 1},$$

and between x_i and an eavesdropper e is

$$\mathsf{SINR}_{ie} = \frac{g(|x_i - e|)}{\sum_{k \in \Phi k \neq i} g(|x_j - x_j|) + 1}.$$

Note that the parameter γ is absent in the SINR expression at the eavesdropper e, since γ depends on the coding strategy between the legitimate nodes, for example, spreading code used in a CDMA system, that is not known to the eavesdroppers, hence no processing gain can be obtained at any of the eavesdroppers. Then the maximum rate of reliable communication between x_i and x_j such that an eavesdropper e gets no knowledge of communication between x_i and x_j is [6]

$$R_{ij}^{\mathsf{secure}}(e) = \left[\log_2\left(1 + \mathsf{SINR}_{ij}\right) - \log_2\left(1 + \mathsf{SINR}_{ie}\right)\right]^+,$$

and the maximum rate of communication between x_i and x_j that is secured from all the eavesdropper nodes of Φ_E,

$$R_{ij}^{\mathsf{secure}} = \min_{e \in \Phi_E} R_{ij}(e).$$

Definition 8.5.1 *SINR secrecy graph (SSG) is a directed graph $SSG(\tau) = \{\Phi, \mathcal{E}\}$, with vertex set Φ, and edge set $\mathcal{E} = \{(x_i, x_j) : R_{ij}^{\mathsf{secure}} > \tau\}$, where τ is the minimum rate of secure communication required between any two nodes of Φ.*

We will consider $\tau = 0$ and represent $SSG(0)$ as SSG, since the derived results can be generalized easily for $\tau > 0$. With $\tau = 0$, $SSG = \{\Phi, \mathcal{E}\}$, with edge set $\mathcal{E} = \{(x_i, x_j) : \mathsf{SINR}_{ij} > \mathsf{SINR}_{ie},\ \forall\, e \in \Phi_E\}$.

We want to find the critical density for percolation for the legitimate nodes for any eavesdropper nodes density, that is,

$$\lambda_c(\lambda_E) = \inf\{\lambda : \mathbb{P}(|\mathsf{C}| = \infty) > 0\},$$

where C is the connected component of the node located at origin in the SSG.

We will show that for any given density λ_E of the eavesdropper nodes, the secrecy graph SSG percolates for sufficiently large density λ of legitimate nodes and sufficiently small interference suppression parameter γ [11]. Conversely, we show that for a given density λ_E of the eavesdropper nodes, and $\gamma > 0$, if the density of legitimate nodes λ is below a threshold, then the SSG does not percolate [11].

Remark 8.5.2 *Instead of existential results, concrete bounds can be found for percolation in SSG, by following the proof of Theorem 7.3.21 for the Gilbert's random disc model, as long as the path-loss function $g(.)$ has a finite support, that is, $g(x) = 0$ for $x \geq \eta$ [12]. With a finite support path-loss function, the dependency of the SSG is restricted to distances of at most η, since $(x_i, x_j) \in \mathcal{E}$ only depends on $x_j \in \Phi \cap \mathbf{B}(x_i, \eta)$ and $e \in \Phi_E \cap \mathbf{B}(x_i, 2\eta)$. Hence, the proof idea of Theorem 7.3.21 can be used, by defining the maximum distance to which a node $x_i \in \Phi$ can have a connection in SSG as*

$$\mathsf{cov}(x_i) = \sup \left\{ d : \; g(d) > \max_{e \in \Phi_E} \mathsf{SINR}_{ie} \right\}.$$

Therefore, a node x_j can be connected to node x_i only if $d_{ij} < \mathsf{cov}(x_i)$. Note that $\mathsf{cov}(x_i) \leq \eta$, $\forall\, x_i$, since the signal power received at a distance x, $g(x)$, is zero for $x \geq \eta$. Now, $\mathsf{cov}(x_i)$ can play the role of random radius r_x of the Gilbert's random disc model, and similar to Theorem 7.3.21, we can obtain a lower bound on the critical density as $\lambda_c \geq \frac{1}{c\mathbb{E}\{\mathsf{cov}^2\}}$.

8.5.1 Super Critical Regime

In Theorem 8.5.3, we show that for small enough γ, there exists a large enough λ for which the SSG percolates with positive probability for any value of λ_E.

Theorem 8.5.3 *For any λ_E, there exists $\lambda' < \infty$ and a function $\gamma'(\lambda, \lambda_E) > 0$, such that $\mathbb{P}(|\mathsf{C}| = \infty) > 0$ in the SSG for $\lambda > \lambda'$, and $\gamma < \gamma'(\lambda, \lambda_E)$, for any signal attenuation function $g(x)$, such that $\int x g(x) \mathrm{d}x < \infty$ and $g(.) \leq 1$.*

This result is similar in spirit to Theorem 8.3.1, where percolation is shown to happen in the PSG, where two nodes are connected if the SINR between them is more than a fixed threshold β, (without the secrecy constraint due to eavesdroppers) for small enough interference suppression parameter γ. The major difference between the SSG and PSG is that with the SSG, the threshold for connection between two nodes (maximum of SINRs received at all eavesdroppers) is a random variable that depends on both the legitimate and eavesdropper density, while in contrast, the threshold in the PSG is a fixed constant.

Proof: To prove the result, we consider percolation on an enhanced graph SSG^e that is a subset of SSG obtained by removing the interference power at each of the eavesdropper nodes, thereby increasing the SINR seen at any eavesdropper node. The SSG^e is defined as follows, $SSG^e = \{\Phi, \mathcal{E}^e\}$, with edge set

$$\mathcal{E}^e = \{(x_i, x_j) : \mathsf{SINR}_{ij} \geq g(d_{ie}),\; \forall\, e \in \Phi_E\}.$$

Since $g(d_{ie}) \geq \mathsf{SINR}_{ie}$, $SSG^e \subseteq SSG$. Thus, if SSG^e percolates, so does SSG.

To find conditions for SSG^e to percolate, we map the continuum percolation of SSG^e to an appropriate discrete percolation on $\mathbb{Z}^2$, similar to the technique used for Gilbert's disc model in Section 7.3.1. Consider a square lattice $\mathbf{S}$ with side s. Let $\mathbf{S}' = \mathbf{S} + (\frac{s}{2}, \frac{s}{2})$ be the dual lattice of $\mathbf{S}$ obtained by translating each edge of $\mathbf{S}$ by $(\frac{s}{2}, \frac{s}{2})$. For any edge $\mathbf{a}$ of $\mathbf{S}$, let $S_1(\mathbf{a})$ and $S_2(\mathbf{a})$ be

the two adjacent squares to $\mathbf{a}$. Let $\{a_i\}_{i=1}^4$ denote the four vertices of the rectangle $S_1(\mathbf{a}) \cup S_2(\mathbf{a})$. Let $Y(\mathbf{a})$ be the smallest square containing $\cup_{i=1}^4 \mathbf{B}(a_i, t)$, where t is such that $g(t) < \frac{g(\sqrt{5}s)}{2}$. See Fig. 8.8 for a pictorial description.

Definition 8.5.4 *Any edge* $\mathbf{a}$ *of* $\mathbf{S}$ *is defined to be open if*

1. *there is at least one legitimate node of* Φ *in both the adjacent squares* $S_1(\mathbf{a})$ *and* $S_2(\mathbf{a})$,
2. *there are no eavesdropper nodes of* Φ_E *in* $Y(\mathbf{a})$,
3. *and for any legitimate node* $x_i \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$, *the interference received at any legitimate node* $x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$,

$$I_j^i = \sum_{k \in \Phi, k \neq i} g(|x_k - x_j|) \leq \frac{1}{\gamma}.$$

An open edge is pictorially described in Fig. 8.8 by edge $\mathbf{a}$, where the black dots represent a legitimate node while a cross represents an eavesdropper node.

The next lemma allows us to tie up the continuum percolation on SSG^e to the percolation on the square lattice, where we show that if an edge $\mathbf{a}$ is open, then all legitimate nodes lying in $S_1(\mathbf{a}) \cup S_2(\mathbf{a})$ can connect to each other in the SSG^e.

Lemma 8.5.5 *If an edge* $\mathbf{a}$ *of* $\mathbf{S}$ *is open, then any node* $x_i \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ *can connect to any node* $x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ *in* SSG^e.

Proof: For any $x_i, x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$, $\mathsf{SINR}_{ij} \geq \frac{g(\sqrt{5}s)}{2}$, since $I_j^i = \sum_{k \in \Phi, k \neq i} g(|x_k - x_j|) \leq \frac{1}{\gamma}$ for each $x_i, x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$. Moreover, since there are no eavesdropper nodes in $Y(\mathbf{a})$, the minimum distance between any eavesdropper node from any legitimate node in $\Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ is at least t. By choice, t is such that $g(t) < \frac{g(\sqrt{5}s)}{2}$. Therefore, $\mathsf{SINR}_{ij} \geq \max_e \mathsf{SINR}_{ie}$, and hence $x_i, x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ are connected in SSG^e. □

Recall from Lemma 7.2.8 that the connected component of $\mathbf{S}$ containing the origin is finite if and only if there is a closed circuit in dual lattice $\mathbf{S}'$ surrounding the origin. Hence, if we can show that the probability that there exists a closed circuit in $\mathbf{S}'$ surrounding the origin is less than 1, then it follows that an unbounded connected component exists in $\mathbf{S}$ with non-zero probability. Moreover, having an unbounded connected component in the square lattice $\mathbf{S}$ implies that there is an unbounded connected component in SSG^e from Lemma 8.5.5. Thus, to show that SSG^e percolates, it is sufficient to show that the probability of having a closed circuit in $\mathbf{S}'$ surrounding the origin is less than 1.

Next, we find a bound on λ as a function of λ_E such that the probability of having a closed circuit in $\mathbf{S}'$ surrounding the origin is less than 1.

For an edge $\mathbf{a}$, let $A(\mathbf{a}) = 1$ if $\Phi \cap S_i(\mathbf{a}) \neq \emptyset$, $i = 1, 2$, and zero otherwise. Similarly, let $B(\mathbf{a}) = 1\ (= 0)$ if $I_j^i = \sum_{k \in \Phi, k \neq i} g(|x_k - x_j|) \leq \frac{1}{\gamma}$ for $x_i, x_j \in \Phi \cap (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ (otherwise), and $C(\mathbf{a}) = 1\ (= 0)$ if there are no eavesdropper nodes in $Y(\mathbf{a})$ (otherwise). Then by definition, the edge $\mathbf{a}$ is open if $D(\mathbf{a}) = A(\mathbf{a})B(\mathbf{a})C(\mathbf{a}) = 1$.

Now we want to bound the probability of having a closed circuit surrounding the origin in $\mathbf{S}'$. Towards that end, we will first bound the probability of a closed circuit of length n, that is, $\mathbb{P}(D(\mathbf{a}_1) = 0, D(\mathbf{a}_2) = 0, \ldots, D(\mathbf{a}_n) = 0)$, $\forall\ n \in \mathbb{N}$ considering n distinct edges. Let

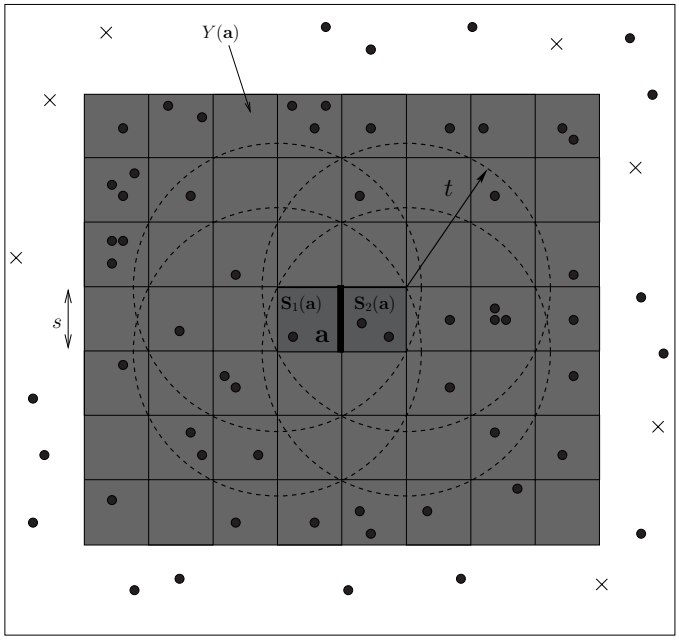

Figure 8.8: Open edge definition on a square lattice for super critical regime.

$p_A := \mathbb{P}(A(\mathbf{a}_i) = 0)$ for any i. Since Φ is a PPP with density λ, $p_A = 1 - (1 - \exp(-\lambda s^2))^2$. Then we have the following intermediate results to upper bound $\mathbb{P}(D(\mathbf{a}_1) = 0, D(\mathbf{a}_2) = 0, \ldots, D(\mathbf{a}_n) = 0)$.

Lemma 8.5.6 $\mathbb{P}(A(\mathbf{a}_1) = 0, A(\mathbf{a}_2) = 0, \ldots, A(\mathbf{a}_n) = 0) \leq p_1^n$, *where* $p_1 = p_A^{1/4}$.

Proof: Follows from the fact that in any sequence of n edges of $\mathbf{S}$, there are at least $n/4$ edges such that their adjacent rectangles $S_1(\mathbf{a}_e) \cup S_2(\mathbf{a}_e)$ do not overlap. Therefore, $\mathbb{P}(A(\mathbf{a}_1) = 0, A(\mathbf{a}_2) = 0, \ldots, A(\mathbf{a}_n) = 0) \leq \mathbb{P}(\cap_{e \in O} A(\mathbf{a}_e) = 0)$, where O is the set of edges for which their adjacent rectangles $S_1(\mathbf{a}_e) \cup S_2(\mathbf{a}_e)$ have no overlap and $|O| \geq n/4$. Since $S_1(\mathbf{a}_e) \cup S_2(\mathbf{a}_e),\ e \in O$ have no overlap, and events $A(\mathbf{a}_e) = 0$ are independent for $\mathbf{a}_e \in O$, the result follows. □

Recall from Lemma 8.3.5, that for $\int_0^\infty x g(x) \mathrm{d}x < \infty$ and $g(.) \leq 1$,

$$\mathbb{P}(B(\mathbf{a}_1) = 0, B(\mathbf{a}_2) = 0, \ldots, B(\mathbf{a}_n) = 0) \leq p_2^n,$$

where

$$p_2 = \exp\left(\frac{2\lambda}{K} \int g(x) \mathrm{d}x - \frac{1}{\gamma K}\right),$$

and K is a constant and $\frac{1}{\gamma}$ is the interference upper bound (Definition 8.5.10) corresponding to M in Lemma 8.3.5.

Lemma 8.5.7 $\mathbb{P}(C(\mathbf{a}_1) = 0, C(\mathbf{a}_2) = 0, \ldots, C(\mathbf{a}_n) = 0) \leq p_3^n$, *for some* p_3 *independent of* n.

Proof: By definition, events $C(\mathbf{a}_i)$ and $C(\mathbf{a}_j)$ are independent if $Y(\mathbf{a}_i) \cap Y(\mathbf{a}_j) = \phi$. Consider a circuit $\mathcal{P}_n$ in $\mathbf{S}$ of length n, with a subset $\mathcal{P}_n^s \subset \mathcal{P}_n$, where $\mathcal{P}_n^s = \{\mathbf{a}_i\}_{i \in \mathcal{I}}$, where for any $n, m \in$

$\mathcal{I}, Y(\mathbf{a}_n) \cap Y(\mathbf{a}_m) = \phi$. Since $Y(\mathbf{a})$ occupies at most $\left(L + \lceil \frac{2t}{s} \rceil\right) \times \left(L + 1 + \lceil \frac{2t}{s} \rceil\right)$ squares of lattice $\mathbf{S}$, where $L = 2\lceil\sqrt{5}\rceil$, it follows that $|\mathcal{I}| \geq \frac{n}{\psi}$, where $\psi = 8\left(L + \lceil \frac{2t}{s} \rceil\right)^2 - 1$. Thus, $\mathbb{P}(C(\mathbf{a}_1) = 0, C(\mathbf{a}_2) = 0, \ldots, C(\mathbf{a}_n) = 0) \leq p_3^n$, where $p_3 = \mathbb{P}(C(\mathbf{a}_i) = 0)^{\frac{1}{\psi}}$ and $\mathbb{P}(C(\mathbf{a}_i) = 0) = \exp\left(-\lambda_E \nu(Y(\mathbf{a}_i))\right)$. □

Lemma 8.5.8 $\mathbb{P}(D(\mathbf{a}_1) = 0, D(\mathbf{a}_2) = 0, \ldots, D(\mathbf{a}_n) = 0) \leq (\sqrt{p_1} + {p_2}^{1/4} + {p_3}^{1/4})^n$.

Proof: Follows from Lemma 8.3.6, where event $D(\mathbf{a}) = 1$ if $A(\mathbf{a})B(\mathbf{a})C(\mathbf{a}) = 1$. □

Let $q = (\sqrt{p_1} + {p_2}^{1/4} + {p_3}^{1/4})$. Then from Lemma 8.3.7, we have that if $q < \frac{11-2\sqrt{10}}{27}$, then the probability of having a closed circuit in $\mathbf{S}'$ surrounding the origin is less than 1.

Thus, for $q = (\sqrt{p_1} + {p_2}^{1/4} + {p_3}^{1/4}) < \frac{11-2\sqrt{10}}{27}$, the

$$\mathbb{P}(\text{closed circuit around origin}) < 1,$$

and hence, $\mathbb{P}(|\mathsf{C}| = \infty) > 0$ from Lemma 7.2.8. Now we show that for an appropriate choice of s, t, γ, we can make q as small as we like. For any eavesdropper density λ_E, p_3 can be made arbitrarily small by choosing small enough s (side length) and t (radius used to define $Y(\mathbf{a})$). Depending on the choice of s, p_1 can be made arbitrarily small for large enough legitimate node density λ, and finally depending on the choice of λ, choosing small enough γ, p_2 can be made arbitrarily small. In summary, we see that for any λ_E, there exists a small enough $\gamma'(\lambda, \lambda_E)$ and λ' such that for $\lambda > \lambda'$ and $\gamma < \gamma'(\lambda, \lambda_E)$, SSG^e percolates. □

8.5.2 Sub-Critical Regime

In this section, through Theorem 8.5.9, we lower bound the critical density λ_c and show that for any γ, and λ_E, there exists a λ_ℓ for which the SSG does not percolate with positive probability, that is, $\lambda_c > \lambda_\ell$.

Theorem 8.5.9 *For every $\lambda_E \geq 0$ and $\gamma \in (0, 1)$, there exists a $\lambda_c(\lambda_E, \gamma) > 0$ such that for all $\lambda < \lambda_c(\lambda_E, \gamma)$, $\mathbb{P}(|\mathsf{C}| = \infty) = 0$ in the SSG.*

To prove this result, we define a graph G_E over eavesdropper nodes, such that if the straight line in $\mathbb{R}^2$ between two eavesdropper nodes that have an edge in G_E intersects the straight line between any legitimate node pair $x_i, x_j \in \Phi$ in $\mathbb{R}^2$, then $(x_i, x_j) \notin SSG$, that is, x_i and x_j do not have an edge in the SSG. Assume that the eavesdropper graph G_E percolates. Then there are left-right crossings and top-bottom crossings of any square box of large size in $\mathbb{R}^2$ by connected edges of the eavesdropper graph G_E. Since edges of SSG and G_E cannot cross each other, percolation over G_E precludes the possibility of having left-right crossings and top-bottom crossings of any square box of large size in $\mathbb{R}^2$ by connected edges of SSG, which is necessary for percolation in the SSG. Thus, only one of the two graphs G_E and SSG can percolate. We derive conditions for percolation on G_E to characterize the sub-critical regime of percolation over the SSG. We show that for any given λ_E and $\gamma > 0$, if the density of legitimate nodes is below a threshold, then the eavesdropper nodes' graph G_E percolates.

Proof: Without loss of generality, we consider the case of $\gamma = 0$, where the SINR between any two

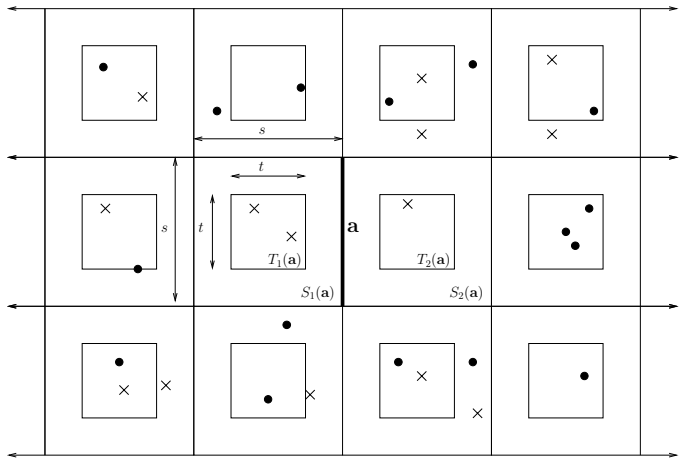

Figure 8.9: Open edge definition on a square lattice for sub critical regime.

legitimate nodes is $g(d_{ij})$, while the SINR at the eavesdropper nodes remains the same. Thus, two legitimate nodes x_i and x_j are connected in the SSG if

$$g(d_{ij}) \geq \mathsf{SINR}_{ie}, \ \forall \, e \in \Phi_E.$$

If we can show that $\lambda_c > \lambda_\ell$ for $\gamma = 0$, then since SSG with $\gamma > 0$ is contained in SSG with $\gamma = 0$, we have that for all $\gamma > 0$, $\lambda_c > \lambda_\ell$. So the lower bound λ_ℓ for λ_c obtained with $\gamma = 0$ serves as a universal lower bound on the critical density λ_c required for percolation.

For the case of $\gamma = 0$, we proceed as follows. Consider a square lattice $\mathbf{S}$ with side s. Let $\mathbf{S}' = \mathbf{S} + (\frac{s}{2}, \frac{s}{2})$ be the dual lattice of $\mathbf{S}$ obtained by translating each edge of $\mathbf{S}$ by $(\frac{s}{2}, \frac{s}{2})$. For any edge $\mathbf{a}$ of $\mathbf{S}$, let $S_1(\mathbf{a})$ and $S_2(\mathbf{a})$ be the two adjacent squares to $\mathbf{a}$. See Fig. 8.9 for a pictorial description. Let $T_1(\mathbf{a})$ and $T_2(\mathbf{a})$ be the smaller squares with side t contained inside $S_1(\mathbf{a})$ and $S_2(\mathbf{a})$, respectively, as shown in Fig. 8.9, with centers identical to that of $S_1(\mathbf{a})$ and $S_2(\mathbf{a})$.

Let the interference power received at any eavesdropper e with respect to signal from x_k is $I_e^k = \sum_{x_j \in \Phi, j \neq k} g(|x_j - e|)$.

Definition 8.5.10 *For any edge* $\mathbf{a}$ *of* $\mathbf{S}$*, we define three indicator variables* $\tilde{A}(\mathbf{a}), \tilde{B}(\mathbf{a})$*, and* $\tilde{C}(\mathbf{a})$ *as follows.*

1. $\tilde{A}(\mathbf{a}) = 1$ *if there is at least one eavesdropper node of* Φ_E *in both the adjacent squares* $T_1(\mathbf{a})$ *and* $T_2(\mathbf{a})$.

2. $\tilde{B}(\mathbf{a}) = 1$ *if there are no legitimate nodes of* Φ *in* $S_1(\mathbf{a})$ *and* $S_2(\mathbf{a})$.

3. $\tilde{C}(\mathbf{a}) = 1$ *if for any eavesdropper node* $e \in \Phi_E \cap (T_1(\mathbf{a}) \cup T_2(\mathbf{a}))$*, the interference received from all the legitimate nodes* $I_e = \sum_{x_k \in \Phi} g(x_k, e) \leq c$.

Then an edge $\mathbf{a}$ is defined to be open if $\tilde{D}(\mathbf{a}) = \tilde{A}(\mathbf{a})\tilde{B}(\mathbf{a})\tilde{C}(\mathbf{a}) = 1$. An open edge is pictorially described in Fig. 8.9 by a solid edge $\mathbf{a}$, where the black dots represent legitimate nodes while crosses are used to represent eavesdropper nodes.

Lemma 8.5.11 *For any c, for large enough s, the straight line between two nodes that have an edge $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$.*

Proof: Let two legitimate nodes $x_i, x_j \in \Phi$ be such that the straight line between x_i and x_j intersects an open edge $\mathbf{a}$ of $\mathbf{S}$. Then by definition of an open edge, $x_i, x_j \notin (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$. Thus, the distance between x_i and x_j is at least s, $d_{ij} \geq s$. Moreover, the SINR between x_i and any eavesdropper node $e \in (T_1(\mathbf{a}) \cup T_2(\mathbf{a}))$,

$$\mathsf{SINR}_{ie} \geq \frac{g(d_{ie})}{1+c},$$

since edge $\mathbf{a}$ is open and hence, $I_e \leq c$ for any $e \in (T_1(\mathbf{a}) \cup T_2(\mathbf{a})$. Thus, choosing s large enough, we can have $g(d_{ij}) < \mathsf{SINR}_{ie}$ for any $e \in (T_1(\mathbf{a}) \cup T_2(\mathbf{a}))$, and hence, x_i and x_j cannot be connected directly in SSG if the straight line between them happens to cross an open edge $\mathbf{a}$ of $\mathbf{S}$. □

Lemma 8.5.12 *If percolation happens on square lattice $\mathbf{S}$ for large enough s for which an edge $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$, then the connected component of SSG is finite.*

Proof: Consider a square box $B_{N/2}$ of side N centered at the origin. Let s be large enough such that the straight line between two nodes that have an edge $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$ (Lemma 8.5.11). Let $L \to R$ and $T \to D$ denote the left-right and top-bottom crossing. If percolation happens on square lattice $\mathbf{S}$, then

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \text{ a } L \to R \text{ and } T \to D \text{ crossing of } B_{N/2} \text{ by open edges of } \mathbf{S}\right) = 1. \tag{8.17}$$

We use contradiction to conclude the result. Let there be an infinite connected component in the SSG with probability 1. Then, necessarily

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \text{ a } L \to R \text{ and } T \to D \text{ crossing of } B_{N/2} \text{ by } SSG\right) = 1. \tag{8.18}$$

Since for large enough side length s of $\mathbf{S}$ an edge $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$, (8.17) and (8.18) cannot hold simultaneously. □

Next, we show that for any λ_E, square lattice $\mathbf{S}$ percolates for small enough density of legitimate nodes λ, as long as s is such that an edge $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$.

Theorem 8.5.13 *For large enough s that ensures that $(x_i, x_j) \in SSG$ cannot cross an open edge $\mathbf{a}$ of $\mathbf{S}$, for any λ_E, $\mathbf{S}$ percolates for a small enough density of legitimate nodes λ.*

Proof: Similar to the proof in the super-critical regime, we need to show that the probability of having a closed circuit surrounding the origin in $\mathbf{S}$ is less than 1. Towards that end, consider the probability of a closed circuit of length n,

$$\mathbb{P}(\tilde{D}(\mathbf{a}_1) = 0, \tilde{D}(\mathbf{a}_2) = 0, \ldots, \tilde{D}(\mathbf{a}_n) = 0),$$

where $\tilde{D}(\mathbf{a}_1) = \tilde{A}(\mathbf{a}_1)\tilde{B}(\mathbf{a}_1)\tilde{C}(\mathbf{a}_1)$. Similar to Lemma 8.5.6,

$$\mathbb{P}(\tilde{A}(\mathbf{a}_1) = 0, \tilde{A}(\mathbf{a}_2) = 0, \ldots, \tilde{A}(\mathbf{a}_n) = 0) \leq q_1^n,$$

where $q_1 = p_A^{1/4}$ and $p_A = 1 - \left(1 - \exp\left(-\lambda_E t^2\right)\right)^2$ is the probability that there is no eavesdropper in either $T_1(\mathbf{a})$ or $T_2(\mathbf{a})$. Similarly, following Lemma 8.5.6,

$$\mathbb{P}(\tilde{B}(\mathbf{a}_1) = 0, \tilde{B}(\mathbf{a}_2) = 0, \ldots, \tilde{B}(\mathbf{a}_n) = 0) \leq q_2^n,$$

where $q_2 = p_B^{1/4}$ and $p_B = 1 - \exp(-2\lambda s^2)$ is the probability that there is at least one legitimate node of Φ in $S_1(\mathbf{a})$ or $S_2(\mathbf{a})$,

$$\mathbb{P}(\tilde{C}(\mathbf{a}_1) = 0, \tilde{C}(\mathbf{a}_2) = 0, \ldots, \tilde{C}(\mathbf{a}_n) = 0) \leq q_3^n,$$

where $q_3 = \exp\left(\frac{2\lambda}{K} \int \tilde{g}(x)\mathrm{d}x - \frac{c}{K}\right)$ following Lemma 8.3.5, and finally

$$\mathbb{P}(\tilde{D}(\mathbf{a}_1) = 0, \tilde{D}(\mathbf{a}_2) = 0, \ldots, \tilde{D}(\mathbf{a}_n) = 0) \leq \left(\sqrt{q_1} + q_2^{1/4} + q_3^{1/4}\right)^n,$$

following Lemma 8.5.8. Let $q = \sqrt{q_1} + q_2^{1/4} + q_3^{1/4}$.

Using Lemma 8.3.7, percolation happens in $\mathbf{S}$ if we can get $q < \kappa$ for sufficiently small $\kappa > 0$. Let us fix such a $\kappa \geq 0$. Then, by choosing t large enough, we can have $\sqrt{q_1} < \frac{\kappa}{3}$. Moreover, for fixed interference upper bound c, let s be large enough such that for any pair of legitimate nodes $x_i, x_j \notin (S_1(\mathbf{a}) \cup S_2(\mathbf{a}))$ for which the straight line between them intersects an open edge $\mathbf{a}$ of $\mathbf{S}$, $g(d_{ij}) < \mathsf{SINR}_{ie}$ for any $e \in (T_1(\mathbf{a}) \cup T_2(\mathbf{a}))$. Now, given c, t, and s, we can choose λ small enough so that $q_2^{1/4} < \frac{\kappa}{3}$ and $q_3^{1/4} < \frac{\kappa}{3}$. Thus, we have that $q < \kappa$ as required for an appropriate choice of c, t, s and λ, for any λ_E. □

Thus, from Theorem 8.5.13, we have that square lattice $\mathbf{S}$ percolates. In addition, from Lemma 8.5.12, we know that for small enough legitimate node density λ, SIG with $\gamma = 0$ does not percolate as long as $\mathbf{S}$ percolates, completing the proof. □

8.6 Reference Notes

Percolation on the SINR graph was first studied in [2] with a finite support path-loss function, and later generalized in [3] for all path-loss functions. Connectivity on the SINR graph over a one-dimensional network was studied in [5] that was generalized for the two-dimensional network in [4]. The distance-based secure model was introduced in [8, 9] and studied in [8–10]. Percolation results for the more general secure SINR graph with a finite support path-loss function can be found in [11], and for all path-loss functions in [12].

Bibliography

[1] R. Meester and R. Roy. 1996. *Continuum Percolation*. Cambridge University Press.

[2] O. Dousse, F. Baccelli, and P. Thiran "Impact of Interferences on Connectivity in Ad Hoc Networks," In *IEEE INFOCOM 2003*, San Francisco.

[3] O. Dousse, M. Franceschetti, N. Macris, R. Meester, and P. P. Thiran. 2006. "Percolation in the signal to interference ratio graph." *J. Appl. Prob.* 43 (2): 552–62.

[4] R. Vaze. 2012. "Percolation and connectivity on the signal to interference ratio graph." In *Proceedings IEEE INFOCOM, 2012*. 513 –21.

[5] C. Avin, Z. Lotker, F. Pasquale, and Y.-A. Pignolet. 2010. "A note on uniform power connectivity in the SINR model." Availabe at *http://arxiv.org/abs/0906.2311*.

[6] A. Wyner. 1975. "The wire-tap channel." *Bell Syst. Tech. J.* 54 (8): 1355–67.

[7] S. Leung-Yan-Cheong and M. Hellman. 1978. "The Gaussian wire-tap channel." *IEEE Trans. Inf. Theory*. 24 (4): 451–56.

[8] M. Haenggi. 2008. "The secrecy graph and some of its properties." In *Proc. IEEE Int. Symp. Information Theory ISIT 2008*. 539–43.

[9] A. Sarkar and M. Haenggi. 2011. "Percolation in the Secrecy Graph." In *2011 Information Theory and Applications Workshop (ITA'11)*. Available at http://www.nd.edu/mhaenggi/pubs/ita11.pdf.

[10] P. Pinto and M. Win. 2010. "Percolation and connectivity in the intrinsically secure communications graph." submitted. Available at http://arxiv.org/abs/1008.4161.

[11] R. Vaze and S. Iyer. 2013. "Percolation on the information-theoretically secure signal to interference ratio graph." *J. Appl. Prob. Accepted Dec 2013.*

[12] ——. 2014. "Percolation on the information theoretic secure SINR graph: Upper and lower bounds." In *IEEE Spaswin 2014*.

Chapter 9

Throughput Capacity

9.1 Introduction

We are at the final frontier of this book, where we study the concept of transport capacity of a random wireless network. The transport capacity captures the maximum number of bits that can be transported across the network multiplied by the distance they (bits) travel. Neglecting the effect of distance, transport capacity becomes the throughput capacity, which measures the successful sum-rate of the transmission between all pairs of source–destination pairs at the same time. For a random wireless network located in a bounded area, the average source–destination pair distance is a constant, and hence, it is sufficient to analyze the throughput capacity rather than the transport capacity. Following this, in this chapter, we analyze the throughput capacity of a random wireless network and present its fundamental characterization.

We begin by deriving an upper bound on the throughput capacity of a random wireless network under the SINR model, where communication between two nodes is successful if the signal-to-interference-plus-noise ratio (SINR) between them is larger than a threshold. We show that the throughput capacity of a random wireless network with n nodes cannot scale faster than order $\sqrt{n}$ under the SINR model. Next, using results from percolation theory, we construct a three-phase achievable strategy, where nearest neighbor routing is used, which achieves the upper bound of order $\sqrt{n}$ on the throughput capacity scaling under the SINR model.

To identify whether the order $\sqrt{n}$ scaling is fundamental or a result of the considered SINR model, we next turn toward an information theoretic upper bound on the throughput capacity and show that no matter what strategy one uses, the throughput capacity of a random wireless network cannot scale faster than order $n \ln n$. Next, we present the most remarkable result of this chapter, which shows that almost linear scaling of the throughput capacity can be achieved, that is, each source can communicate at a constant rate with its destination simultaneously, leaving a gap of only order $\ln n$ between the upper and the lower bound. This in turn shows that the nearest neighbor routing strategy is not optimal in general.

Almost linear scaling of the throughput capacity is achieved by a hierarchical cooperation strategy, where we progressively combine weak strategies with small throughput capacity scalings to get an overall strategy with better throughput capacity scaling. In particular, the network is divided into smaller clusters, and the given weak strategy is used within each cluster for intra-cluster communication. Inter-cluster communication is accomplished via a *virtual* multiple

antenna channel. By reducing the number of hops between source–destination pairs and exploiting the multiplexing gain of multiple antennas, the new strategy provides a better throughput capacity scaling. Recursing this procedure with the new strategy, we can build better and better strategies with throughput capacity approaching linear scaling.

For most part of this book, and this chapter, we consider a *dense* wireless network, where we fix the area of operation and increase the density of nodes. An alternate model is what is called as the *extended* network, where we fix the density of nodes and increase the area of operation. Dense and extended networks are fundamentally different, the former being interference-limited, while the latter is power-limited since ensuring minimum received SNR with increasing distances is a challenge. In the last section of this chapter, we present results on the throughput capacity of the random extended wireless networks that are fairly different than the results for dense networks. For example, we show that for path-loss exponents larger than 3, the nearest neighbor routing strategy is in fact optimal, and the throughput capacity cannot scale faster than order $\sqrt{n}$.

9.2 Throughtput Capacity Formulation

Consider the unit square $\mathbf{S}_1$ and assume that n nodes $\Phi_n = \{x_1, \ldots, x_n\}$ are distributed uniformly randomly in $\mathbf{S}_1$. Each node x_i has a destination that is defined to be the node in $\Phi \backslash \{x_i\}$ that is closest to a randomly chosen point in $\mathbf{S}_1$. Ignoring edge effects, the average distance W between any source–destination pair is a constant. Packets between source–destination pairs travel over multiple hops, where in any hop, a receiving node decodes an incoming packet, and then either re-transmits it to the next node or buffers it up for a later retransmission.. Finding routing paths for packet forwarding between different source–destination pairs is also part of the problem.

We let each node to have a single antenna each and consider the path-loss model. For ease of exposition, we assume that each node transmits unit power, and the SINR between nodes x_i and x_j with distance $d_{ij} = |x_i - x_j|$ at time t is

$$\mathsf{SINR}_{ij}(t) = \frac{d_{ij}^{-\alpha}}{1 + \sum_{x_k \in \Phi_n \backslash \{x_i\}} \mathbf{1}_k(t) d_{kj}^{-\alpha}}, \tag{9.1}$$

where $\mathbf{1}_k(t)$ is the indicator variable denoting whether node x_k is active at time t or not.

SINR model: With the SINR model, the transmission between nodes x_i and x_j is deemed to be successful at time t if the SINR between them at time t is larger than a threshold β, that is, $\mathsf{SINR}_{ij}(t) > \beta$. Thus, the rate of transmission between x_i and x_j, if successful, is $B(\beta)$ bits/sec/Hz.

Remark 9.2.1 *In this chapter, we will primarily deal with the SINR model, but another model of interest is the protocol model, where communication between nodes x_i and x_j is defined to be successful if there is no other transmitter in a disc of radius $(1+\Delta)|x_i - x_j|$ around the receiver x_j. The quantity Δ defines a guard zone around each receiver.*

Definition 9.2.2 *We define a per-node throughput of $\mathtt{t}(n)$ bits/sec/Hz to be feasible if there is a spatial and temporal scheduling strategy, such that each node can send $\mathtt{t}(n)$ bits/sec/Hz on average to its randomly chosen destination. The per-node* **throughput capacity** *is defined to be of order $\Theta(f(n))$ bits/sec/Hz if there exists constants c_1 and c_2 such that*

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\mathtt{t}(n) = c_1 f(n) \textit{ is feasible}) &= 1, \\ \liminf_{n\to\infty} \mathbb{P}(\mathtt{t}(n) = c_2 f(n) \textit{ is feasible}) &< 1, \end{aligned}$$

where the probability is over random node locations.

Summing over all n nodes, we define **throughput capacity** *of the network to be $\mathsf{T}(n) = n\mathtt{t}(n)$, or $\mathsf{T}(n) = \Theta(nf(n))$.*

Definition 9.2.3 *For any wireless network, we define that the network transport one bit-meter if one bit is successfully transmitted to a distance of one meter toward its destination. Then the transport capacity is defined to be the sum of all such bit-meters across the network.*

Remark 9.2.4 *For a random wireless network, where n nodes are distributed uniformly randomly in an unit area, the average distance between source-destination pair is $W = \Theta(1)$. Therefore, assuming a symmetric traffic demand of $\mathtt{t}(n)$ bits/sec/Hz/node, the transport capacity of a random wireless network is $W\mathsf{T}(n) = \Theta(T(n))$. Thus, to characterize the transport capacity of a random wireless network, it is sufficient to study the throughput capacity of a random wireless network as done in the rest of this chapter.*

It is important to compare the definition of the throughput capacity under the SINR model and the transmission capacity definition (Definition 2.2.4) of Chapter 2 that we have used in the earlier chapters of this book. The philosophy behind the throughput capacity definition is that it tries to capture how much data can be "transported" across the network with n nodes per unit time, using multi-hops between randomly placed source-destination pairs. In contrast, transmission capacity counts the number of successful transmissions satisfying the outage probability constraint $\mathbb{P}(\mathsf{SINR} < \beta) \leq \epsilon$, over a single hop and fixed source-destination distances. Throughput capacity takes a network-wide view of data dissemination and critically depends on routing protocols, while in contrast, the transmission capacity tries to express the physical layer aspects of the network, for example, effect of signal processing algorithms and multiple antennas.

Another important distinction between the two capacity definitions is in the definition of successful transmission event. The transmission capacity uses an outage probability constraint $\mathbb{P}(\mathsf{SINR} \leq \beta) \leq \epsilon$ and counts the number of successful transmissions satisfying the outage probability constraint. In contrast, in the throughput capacity definition, we are counting the number of nodes that can simultaneously transmit, so that the SINR (realization and not the probability) for each pair of transmissions is above a threshold.

In Sections 9.2.1 and 9.2.2, we show that under the SINR model, the per-node throughput capacity $\mathtt{t}(n)$ of the random wireless network is $\Theta\left(\frac{1}{\sqrt{n}}\right)$. This is indeed a negative result, since it shows that as the number of nodes increase, the per-node throughput goes down to zero.

9.2.1 Upper Bound on the Per-Node Throughput Capacity t*(n)*

Theorem 9.2.5 *The per-node throughput capacity $\mathtt{t}(n)$ of random wireless network under the SINR model is $\mathtt{t}(n) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.*

The main idea of the proof is to show that if a transmission between a tagged transmitter–receiver pair is successful in the SINR model, it precludes the possibility of other transmitters lying *physically* close to the tagged receiver. Thus, each successful transmission consumes a minimum area (bounded from below) of the unit square $\mathbf{S}_1$, and since the total area of $\mathbf{S}_1$ is bounded, the total number of simultaneously successful transmissions are also bounded. Since the rate of transmission between any transmitter-receiver pair is B (fixed), this gives us an upper bound on the throughput capacity.

Proof: To upper bound the per-node throughput capacity, we map the SINR constraint on successful transmissions to a spatial separation constraint between any transmitter–receiver pairs that are active at the same time.

Let the transmission from node x_i be successfully received at node x_j at time t. Moreover, let transmission from node x_k be also successfully received at node x_ℓ at time t. Thus, both $\mathsf{SINR}_{ij}(t) > \beta$ and $\mathsf{SINR}_{k\ell}(t) > \beta$.

Therefore, by the definition of $\mathsf{SINR}_{ij}(t)$ (9.1), we have

$$\frac{d_{ij}^{-\alpha}}{1+\sum_{x_m\in\Phi_n\setminus\{x_i\}}\mathbf{1}_m(t)d_{mj}^{-\alpha}} \quad > \quad \beta. \tag{9.2}$$

Since node x_k is also transmitting, that is, $\mathbf{1}_k(t)=1$, from (9.2), we get

$$\frac{d_{ij}^{-\alpha}}{d_{kj}^{-\alpha}} \quad > \quad \beta. \tag{9.3}$$

which implies that

$$|x_k - x_j| > (1+\Delta)|x_i - x_j|, \tag{9.4}$$

where $\Delta = \beta^{\frac{1}{\alpha}} - 1$, that is, for communication to be successful between a tagged source-destination pair, the distance between them has to be less than a scaled multiple of the distance between any other active transmitter and the tagged receiver. Hence, with the SINR model, if two transmitter-receiver pairs can simultaneously communicate with each other successfully, they have to be sufficiently separated in space, as shown in Fig. 9.1. The quantity Δ defines a guard-zone/exclusion region around each receiver, as shown in Fig. 9.1, where an interfering transmitter cannot be present.

The spatial separation constraint (9.4) imposed by the SINR model is identical to the protocol model (Remark 9.2.1), where any node is allowed to transmit if it satisfies (9.4) for all receivers in the network, except the intended receiver.

Since the SINR constraint implies a spatial separation constraint, that is, SINR model implies a protocol model with appropriate Δ, the transport capacity with the SINR model is bounded from above by the transport capacity of the protocol model. Next, we derive an upper bound on the transport capacity of the protocol model to get an upper bound on the per-node throughput capacity of the random wireless network.

As before, let communication between transmitter-receiver pairs (x_i, x_j) and (x_k, x_ℓ) be simultaneously successful. From the triangle inequality, we have

$$\begin{aligned} |x_j - x_\ell| &\geq |x_j - x_k| - |x_\ell - x_k|, \\ &> (1+\Delta)|x_i - x_j| - |x_\ell - x_k|, \end{aligned} \tag{9.5}$$

where the second inequality follows from (9.4).

Similarly, we can get

$$|x_\ell - x_j| > (1+\Delta)|x_k - x_\ell| - |x_j - x_i|. \tag{9.6}$$

Combining (9.5) and (9.6), we get

$$|x_\ell - x_j| > \frac{\Delta}{2}\left(|x_k - x_\ell| + |x_i - x_j|\right). \tag{9.7}$$

Condition (9.7) implies that the two discs $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$ and $\mathbf{B}\left(x_\ell, \frac{\Delta}{2}|x_k - x_\ell|\right)$ of radius $\frac{\Delta}{2}|x_i - x_j|$ and $\frac{\Delta}{2}|x_k - x_\ell|$ centered at receivers x_j and x_ℓ do not overlap as shown in Fig. 9.2. Thus,

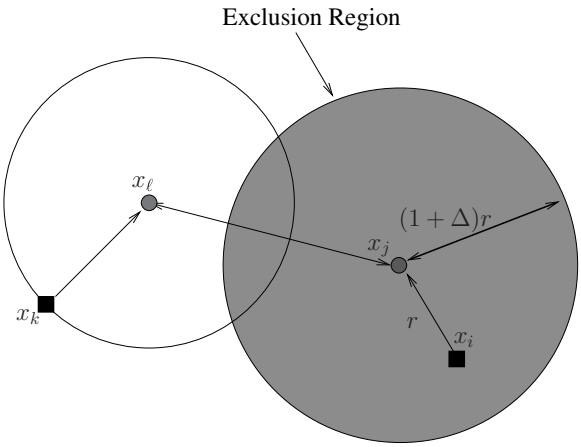

Figure 9.1: Shaded region is the guard-zone based exclusion region around the receiver x_j, where no transmitter (squares) other than the intended transmitter x_i is allowed to lie.

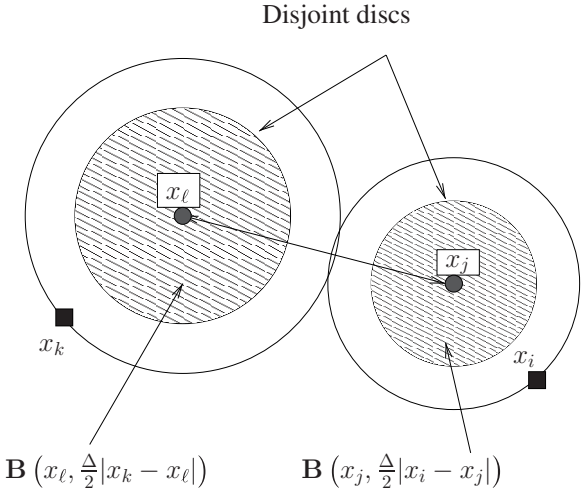

Figure 9.2: Shaded discs around the two receivers (dots) x_j and x_ℓ are not allowed to overlap for successful reception at both x_j and x_ℓ from x_i and x_k, respectively.

if we consider the set of all simultaneously active successful transmissions $\mathcal{A}(t)$ at any time t, we have that all discs of radius $\frac{\Delta}{2}$ times the transmission distance (distance between the transmitter and its receiver at time t) centered at the successful receivers have to be non-overlapping. Equivalently, what we have shown is that each successful transmission requires a certain exclusive area of $\mathbf{S}_1$, and since the total area of $\mathbf{S}_1$ is 1, the total number of successful transmissions can be bounded as follows.

It can be the case that nodes $x_i, x_j \in \mathbf{S}_1$, but $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right) \not\subset \mathbf{S}_1$. However, even in this case, one can easily argue that the area of the disc $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$ that lies inside $\mathbf{S}_1$ is at least

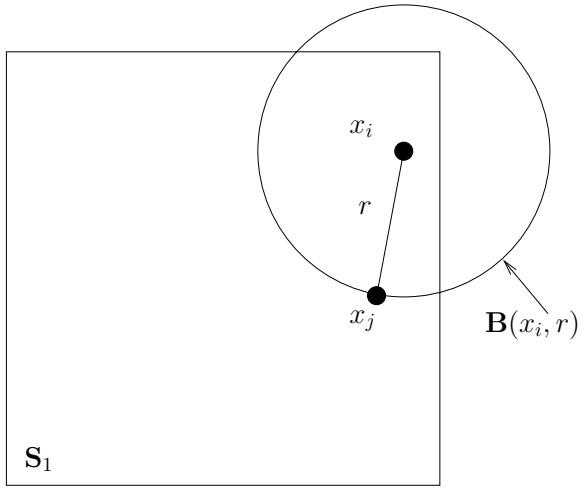

Figure 9.3: Overlap of $\mathbf{B}(x_i, |x_i - x_j| = r)$ with $\mathbf{S}_1$ when $x_i, x_j \in \mathbf{S}_1$.

$\frac{1}{2\pi}$ of the area of disc $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$, whenever $x_i, x_j \in \mathbf{S}_1$. To see this, consider Fig. 9.3, where for $x_i, x_j \in \mathbf{S}_1$, the minimum overlap of disc $\mathbf{B}(x_i, |x_i - x_j| = r)$ with $\mathbf{S}_1$ is achieved if x_i and x_j lie on the diagonally opposite corners of $\mathbf{S}_1$, in which case $\mathbf{S}_1 \subset \mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$ and exactly $\frac{1}{2\pi}$ fraction of the area of $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$ is inside $\mathbf{S}_1$.

Thus, only accounting for the $\frac{1}{2\pi}$ fraction of the exclusive area of $\mathbf{B}\left(x_j, \frac{\Delta}{2}|x_i - x_j|\right)$ occupied by each successful transmission, and summing over all areas corresponding to the set of simultaneously successful transmissions $\mathcal{A}(t)$, we have that

$$\sum_{(i,j)\in\mathcal{A}(t)} \frac{\pi}{2\pi}\left(\frac{\Delta}{2}|x_i - x_j|\right)^2 \leq 1. \tag{9.8}$$

Using Cauchy–Schwarz inequality, we get

$$\begin{aligned}
\sum_{(i,j)\in\mathcal{A}(t)} |x_i - x_j| &\leq \sqrt{\sum_{(i,j)\in\mathcal{A}(t)} (|x_i - x_j|)^2 \sum_{(i,j)\in\mathcal{A}(t)} 1^2}, \\
&\leq \sqrt{\sum_{(i,j)\in\mathcal{A}(t)} (|x_i - x_j|)^2 |\mathcal{A}(t)|}, \\
&\leq \sqrt{\sum_{(i,j)\in\mathcal{A}(t)} (|x_i - x_j|)^2 \frac{n}{2}}, \\
&\leq \sqrt{\frac{8}{\Delta^2}\frac{n}{2}} \qquad \text{from (9.8)},
\end{aligned}$$

where the third inequality follows since at most half the nodes transmit at any time t, that is, $|\mathcal{A}(t)| \leq \frac{n}{2}$.

Thus, summing over all the successful bit-meters, the transport capacity of the protocol model at any time t is upper bounded by

$$B(\beta) \sum_{(i,j)\in\mathcal{A}(t)} (|x_i - x_j|) \leq \frac{2B(\beta)}{\Delta}\sqrt{n}, \tag{9.9}$$

where $B(\beta)$ is the rate of transmission between any two nodes with SINR constraint of β. Since, we know that the SINR model is equivalent to a protocol model with $\Delta = \beta^{1/\alpha} - 1$, therefore, the transport capacity with the SINR constraint is also $\mathcal{O}(\sqrt{n})$.

Recall that the transport capacity of the random wireless network is $W\mathsf{T}(n) = Wn\mathsf{t}(n)$, where $W = \Theta(1)$ is the average source-destination distance. Thus, $Wn\mathsf{t}(n) = \mathcal{O}(\sqrt{n})$, and we get that the per-node throughput capacity of the wireless network with SINR constraint scales as

$$\mathsf{t}(n) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

□

There is one fine point to be noted in this discussion. We have used a binary rate of transmission, that is, rate is $B(\beta)$ if $\mathsf{SINR} > \beta$, and zero otherwise. One can argue that if the minimum separation between successful transmitter-receiver pairs is limiting the throughput capacity, then to be fair, the rate obtained between each transmitter-receiver pair should be made dependent on the SINR between them, that is, rate should be $\log(1 + \mathsf{SINR})$ from the Shannon's formula. It turns out that even with the SINR-dependent rate definition, the throughput capacity cannot scale faster than $\sqrt{n}$ as shown in [1].

Remark 9.2.6 *In the process of proving Theoerem 9.3.1, we have shown another result, that the transport capacity of any arbitrary wireless network, where n nodes are located arbitrarily (not randomly) in a unit area, and source-destination choices are also arbitrary can at best scale as $\mathcal{O}(\sqrt{n})$ under the protocol model. A matching lower bound can also be shown when source-destination pairs are located on a grid [2].*

9.2.2 Lower Bound on the Per-Node Throughput Capacity

In this section, we show that the upper bound of $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ derived in Theorem 9.2.5 is actually achievable. Toward that end, we will use results from percolation theory and ideas presented in Section 7.2.

Theorem 9.2.7 *A per-node throughput of $\mathsf{t}(n) = \Omega\left(\frac{1}{\sqrt{n}}\right)$ is achievable with the SINR model.*

The main idea behind the proof is as follows. Assume that there exists equally spaced horizontal and vertical connected paths (called highways) crossing the unit square $\mathbf{S}_1$, where each highway can support a unit capacity simultaneously. Such a construction will be guaranteed by results from percolation theory (Theorem 7.2.12 and Theorem 7.2.13), even though highways will not be perfect straight lines, however, no two horizontal or vertical highways will have any intersection. Then think of a simple driving strategy to reach location B from location A in $\mathbf{S}_1$ (A corresponds to a source and B to a destination), where we first enter the horizontal highway at a point that is nearest to location A, and then switch to the vertical highway that is closest to location B, and exit the vertical

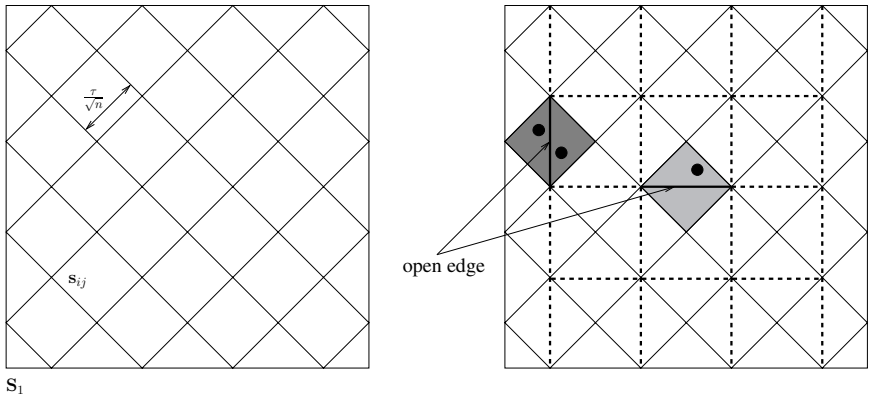

Figure 9.4: Left figure defines a tiling of $\mathbf{S}_1$ by smaller squares of side $\frac{\tau}{\sqrt{n}}$. On the right figure we join the opposite sides of square by an edge (dashed line) and define it to be open (solid line) if the corresponding square contains at least one node in it.

highway at the point closest to location B. We show that with randomly located locations A and B for different users, each of the horizontal and vertical highways serve at most $\sqrt{n}$ users with high probability, and thus a per-user throughput of order $1/\sqrt{n}$ is achievable, since each highway has unit capacity.

Proof: As before, n nodes of Φ_n are distributed uniformly randomly in $\mathbf{S}_1$. Consider a tiling of $\mathbf{S}_1$ by smaller squares $\mathbf{s}_{ij}$ of side $\frac{\tau}{\sqrt{n}}$ with $45\deg$ orientation as shown in Fig. 9.4 (a). Any square $\mathbf{s}_{ij}$ is defined to be *open* if at least one node of Φ_n lies in $\mathbf{s}_{ij}$, and defined to be *closed* otherwise. Since nodes are distributed uniformly randomly, squares $\mathbf{s}_{ij}$ are open and closed independently, and

$$\mathbb{P}(\mathbf{s}_{ij} \text{ is open}) = 1 - (1 - \tau^2/n)^n \approx 1 - \exp(-\tau^2),$$

for large n.

Now we connect an edge (dashed line) between opposite corners of each small square as shown in Fig. 9.4 (b), and define the edge to be open (solid line) or closed (dashed line) depending on whether the square containing it is open or closed. Thus, we get a square grid, where each edge is open/closed independently, and for which we can use results from Section 7.2 to study formation of connected paths over the square grid.

Let us partition $\mathbf{S}_1$ into horizontal rectangles $\mathbf{R}_i$'s of size $\frac{\sqrt{2}\tau}{\sqrt{n}} \ln \frac{\sqrt{n}}{\sqrt{2}\tau} \times 1$, as shown in Fig. 9.5. Since $\mathbf{S}_1$ is a scaled version of box $B_{n/2}$ a square with side n, from Theorem 7.2.13, we know that for large enough $\mathbb{P}(\mathbf{s}_{ij} \text{ is open})$, that is, for large enough τ, there are $\delta \ln \frac{\sqrt{n}}{\sqrt{2}\tau}$ ($\delta = \delta(\tau) > 0$) disjoint left-right crossings (called highway hereafter) of each rectangle $\mathbf{R}_i \in \mathbf{S}_1$ by the connected edges of the square grid with probability 1 as n goes to infinity. Thus, for each of the $\frac{\sqrt{n}}{\sqrt{2}\tau} / \ln \frac{\sqrt{n}}{\sqrt{2}\tau}$ rectangles, there are at least $\delta \ln \frac{\sqrt{n}}{\sqrt{2}\tau}$ disjoint left-right crossings. Similarly, defining vertical rectangles, we can get the same conclusion about top-bottom crossings of $\mathbf{S}_1$.

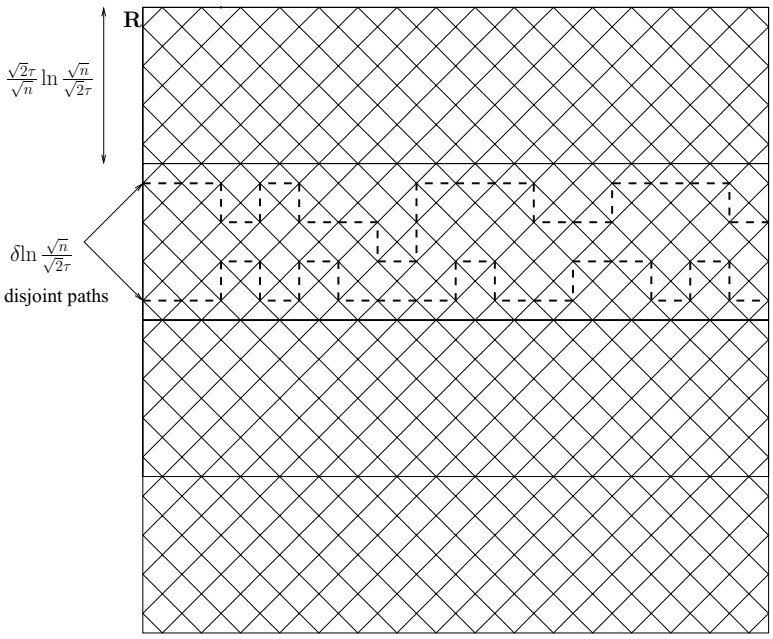

Figure 9.5: Partitioning $\mathbf{S}_1$ into rectangles of size $\frac{\sqrt{2}\tau}{\sqrt{n}} \ln \frac{\sqrt{n}}{\sqrt{2}\tau} \times 1$, where each rectangle contains at least $\delta \ln \frac{\sqrt{n}}{\sqrt{2}\tau}$ disjoint left-right crossings of the square grid defined over $\mathbf{S}_1$.

For any two consecutive connected open edges of the square grid, there is at least one node in each of the two corresponding smaller squares, thus, one can think of these left-right and top-bottom crossings of $\mathbf{S}_1$ as horizontal and vertical highways, where inter-hop distance between any two connected nodes on the highways is less than $2\sqrt{2}\frac{\tau}{\sqrt{n}}$, since the edge length is $\sqrt{2}\frac{\tau}{\sqrt{n}}$.

Moreover, we divide each rectangle horizontally into slices of width $\frac{\mathsf{w}}{\sqrt{n}} \times 1$, where w is chosen so that the number of slices within any rectangle is less than or equal to the number of left-right crossings (highways) $\delta \ln \frac{\sqrt{n}}{\sqrt{2}\tau}$ of each rectangle. Then we associate each slice with one of the highways.

We first compute the total achievable throughput on these highways. An edge being open means the corresponding square contains at least one node. Designate any one node of the square as *relay* that acts as the sender and the receiver for the square corresponding to the open edge of the horizontal highway. Note that the distance between any two consecutive relays lying on any horizontal or vertical highway is less than equal to $2\sqrt{2}\frac{\tau}{\sqrt{n}}$, since the inter-hop distance on any highway is $2\sqrt{2}\frac{\tau}{\sqrt{n}}$.

We next show that if relay nodes employ a time-sharing strategy, then we can achieve a constant total throughput on each highway.

Lemma 9.2.8 *A time-sharing schedule in which each relay on any highway transmits once every K^2 time slots achieves a constant throughput along each highway, where K is a constant that depends on the SINR threshold β and the path-loss exponent α.*

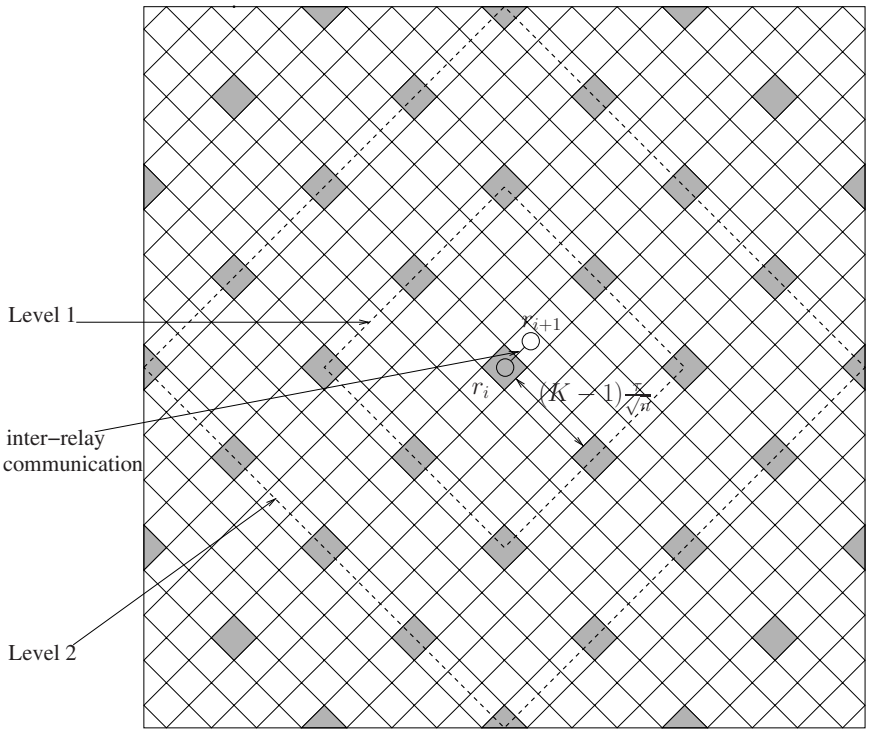

Figure 9.6: Time-sharing by relay nodes using K^2 different time slots, where at any time relays lying in shaded squares transmit.

Proof: We divide the set of relays into K^2 equal-sized groups $G_m, m = 1, \ldots, K^2$, where any two relays in a single group lie in squares that are at least at a distance of $K\frac{\tau}{\sqrt{n}}$ from each other as shown in Fig. 9.6, and only one group is active at a particular time. We consider the SINR between two consecutive relays r_i and r_{i+1} on any horizontal or vertical path, where without loss of generality $r_i \in G_1$. As noted before, the distance between any two consecutive relays is at most $2\sqrt{2}\frac{\tau}{\sqrt{n}}$, then with unit power transmission, the signal power received at r_{i+1} from r_i is at least

$$\left(2\sqrt{2}\frac{\tau}{\sqrt{n}}\right)^{-\alpha}. \tag{9.10}$$

Next, we look at interference power seen at relay r_{i+1}. Because of time-sharing, the interference seen at relay r_{i+1} is coming from multiple levels of relays, where in level ℓ, there are 8ℓ relays that are at a distance of $(\ell K - 1)\frac{\tau}{\sqrt{n}}$ from relay r_{i+1} as shown in Fig. 9.6. Thus, the total interference power seen at r_{i+1} is upper bounded by

$$\sum_{j\in G_1, j\neq i} |r_j - r_{i+1}|^{-\alpha} \leq \sum_{\ell=1}^{\infty} \frac{8\ell}{\left((\ell K-1)\frac{\tau}{\sqrt{n}}\right)^{\alpha}}. \tag{9.11}$$

Hence, the SINR between r_i and r_{i+1} is bounded as follows.

$$\mathsf{SINR}_{i,i+1} \quad = \quad \frac{|r_i - r_{i+1}|^{-\alpha}}{1 + \sum_{j\in G_1, j\neq i} |r_j - r_{i+1}|^{-\alpha}},$$

$$\geq \frac{\left(2\sqrt{2}\frac{\tau}{\sqrt{n}}\right)^{-\alpha}}{1+\sum_{\ell=1}^{\infty}\frac{8\ell}{\left((\ell K-1)\frac{\tau}{\sqrt{n}}\right)^{\alpha}}}, \text{ from (9.10) and (9.11),}$$

$$\geq \frac{\left(2\sqrt{2}\right)^{-\alpha}}{\left(\frac{\tau}{\sqrt{n}}\right)^{\alpha}+\frac{8}{K^{\alpha}}\sum_{\ell=1}^{\infty}\frac{\ell}{\left((\ell-\frac{1}{K})\right)^{\alpha}}},$$

It is easy to check that $\exists$ a constant $c>0$, such that $\sum_{\ell=1}^{\infty}\frac{\ell}{\left(\ell-\frac{1}{K}\right)^{\alpha}} \leq c$ for $\alpha>2$ and $K>1$. Thus, we get

$$\mathsf{SINR}_{i,i+1} > \frac{\left(\frac{K}{2\sqrt{2}}\right)^{\alpha}}{8c}(1+o(1)). \tag{9.12}$$

Choosing $K = \left\lceil 2\sqrt{2}(8c\beta)^{\frac{1}{\alpha}}\right\rceil$, we have that $\mathsf{SINR}_{i,i+1} > \beta$ as required for successful communication between relay r_i and r_{i+1}. Most importantly, note that the time-sharing parameter K is a constant, hence, we get that each relay on the highway can get a constant throughput per unit time.

□

Next, we define a three-phase protocol to show that the throughput of order $\frac{1}{\sqrt{n}}$ is achievable for each node.

Phase 1: Draining Recall that each rectangle is horizontally divided into slices of width $\frac{\mathsf{w}}{\sqrt{n}}\times 1$, where the number of slices within any rectangle is less than or equal to the number of left-right crossings (highways) $\delta\ln\frac{\sqrt{n}}{\sqrt{2}\tau}$ of each rectangle.

In the draining phase, each node lying in the ith slice transmits to its nearest relay on the i^{th} horizontal highway of its own rectangle. From the Chernoff bound (Lemma 7.3.7), we know that each smaller square $\mathbf{s}_{ij}$ with side $\frac{\tau}{\sqrt{n}}$ contains at most $\tau^2\ln n$ nodes with high probability. We will let each of the $\tau^2\ln n$ cohabitants of any smaller square $\mathbf{s}_{ij}$ time share equally to access their nearest relay.

Next, we find the throughput achievable from any one node of a square to its nearest relay assuming that there is only one node in each square.

Since there are $\frac{\sqrt{n}}{\sqrt{2}\tau}/\ln\frac{\sqrt{n}}{\sqrt{2}\tau}$ horizontal rectangles in $\mathbf{S}_1$, each containing at least $\delta\ln\frac{\sqrt{n}}{\sqrt{2}\tau}$ disjoint left-right crossings of $\mathbf{S}_1$, the distance between any node and its nearest relay is no more than $\frac{\tau\sqrt{2}\ln\frac{\sqrt{n}}{\sqrt{2}\tau}}{\sqrt{n}}$ as shown in Fig. 9.7.

Similar to Lemma 9.2.8, we can show that since any node and its nearest relay are at most distance $\frac{\tau\sqrt{2}\ln\frac{\sqrt{n}}{\sqrt{2}\tau}}{\sqrt{n}}$ away, using a $K^2(n)$ phased time-sharing protocol across different squares, where

$$K(n)=2\sqrt{2}(8c\beta)^{\frac{1}{\alpha}}\tau\sqrt{2}\ln\left(\frac{\sqrt{n}}{\sqrt{2}\tau}\right),$$

each node can get a throughput of order $\frac{1}{K(n)^2}$ to their nearest relay. Since all nodes (at most $\tau^2\ln n$) within each square equally time-share, the per-node achievable throughput is order $\frac{1}{K(n)^2\ln n} = \frac{1}{(\ln n)^3}$.

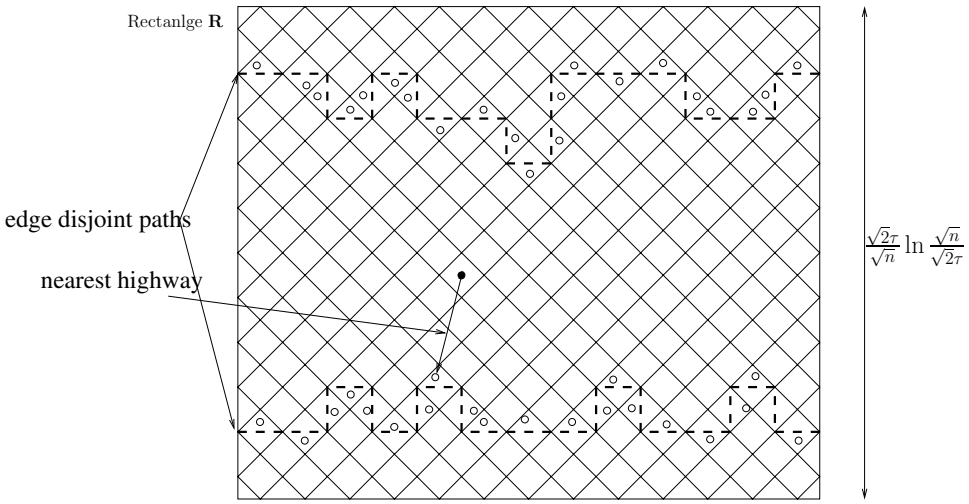

Figure 9.7: Each node (black dots) connects to its nearest relay (hollow circle) and the distance between any node and its nearest relay is no more than the width of the rectangle $\frac{c\sqrt{2}\ln\frac{\sqrt{n}}{\sqrt{2}\tau}}{\sqrt{n}}$.

Phase 2: Highway In the highway phase, the packet is first transmitted horizontally and then vertically toward its destination. The packet is switched over to the vertical highway that is nearest to the destination in the vertical highway.

Any relay on the left-right crossing or highway path lying in slice i of any rectangle serves only those nodes that lie within slice i. From the Chernoff bound, we know that each slice of width $\frac{w}{\sqrt{n}}$ contains at most $w\sqrt{n}$ nodes. Thus, each relay has to route data for at most $\Theta(\sqrt{n})$ nodes, and since each highway can support a constant throughput (Lemma 9.2.8), a per-node throughput of $\Theta\left(\frac{1}{\sqrt{n}}\right)$ is achievable. Vertical highway analysis follows identically.

Phase 3: Delivery Delivery phase is identical to the draining phase, where the packet is delivered to the destination from the nearest relay on the vertical highway. Thus, it follows that the delivery from the nearest relay to the destination can also be done at a per-node throughput of $\Theta\left(\frac{1}{(\ln n)^3}\right)$.

Hence, we have shown that among the three phases, Phase 2 (highway-phase) is the bottleneck, where a per-node throughput of $\Theta\left(\frac{1}{\sqrt{n}}\right)$ can be achieved, completing the proof.

□

In the initial paper by Gupta and Kumar [2], a similar achievable strategy of partitioning the square $\mathbf{S}_1$ into smaller squares $s_{i,j}$ of side $\sqrt{\frac{\tau \ln n}{n}}$ was proposed but it achieved only a per-node throughput of $\Omega\left(\frac{1}{\sqrt{n \ln n}}\right)$ compared to the three phase strategy [5] presented above to prove Theorem 9.2.7 that achieves a per-node throughput of $\Omega\left(\frac{1}{\sqrt{n}}\right)$. The limiting factor of the strategy [2] was its routing protocol. Let line L_i be the straight line between the source-destination pair i. Then hop-by-hop, the packet for destination i is forwarded to a node lying in a square that intersects L_i, where any one node acts as a forwarder (receiver and transmitter). Thus, the overall throughput is limited by the maximum number of lines L_i that cross any one particular square,

since nodes in that square can only transmit at rate proportional to the reciprocal of the number of lines crossing it.

The best bound one can show is that

$$\mathbb{P}\left(\sup_{i,j}\{\text{number of lines intersecting square } s_{i,j}\} \leq \kappa_1\sqrt{n\ln n}\right) \to 1. \tag{9.13}$$

Hence, the per-node throughput capacity with this strategy is $\Omega\left(\frac{1}{\sqrt{n\ln n}}\right)$.

By now, we have shown that under the SINR model, the per-node throughput capacity is $\Theta\left(\frac{1}{\sqrt{n}}\right)$, which is essentially a negative result, since the throughput of each node goes down to zero as we add more and more nodes/users in the system. One natural question to ask at this stage is "whether this scaling result is fundamental or a manifestation of the considered assumptions, such as the SINR model, no cooperation between different nodes, and nearest-neighbor multi-hop routing?. To answer this question, in the next section, we first consider an information theoretic upper bound on the throughput capacity, and then present a hierarchical cooperation strategy that achieves the information theoretic upper bound upto an order $\ln n$ factor.

9.3 Information Theoretic Upper Bound on the Throughput Capacity

In Theorem 9.2.5, we derived an upper bound of $\mathcal{O}(\sqrt{n})$ on the throughput capacity of a wireless network under the SINR model of communication. Ideally, we would like to know a protocol/model independent upper bound in the spirit of Shannon capacity, where no matter what the model is, or how smart the signal processing is, that upper bound cannot be breached. For this purpose, in this section, we work toward an information theoretic upper bound and show that the throughput capacity of a wireless network cannot scale faster than order $n\ln n$.

In this section, we consider the same assumptions as in Section 9.2, except now we consider the fading-plus-path-loss model of wireless signal propagation, where the received signal at node x_j at time t is

$$y_j(t) = \sum_{k\in[1:n],k\neq j} d_{ij}^{-\alpha}\exp\left(\jmath\,\theta_{kj}(t)\right)s_k(t) + w_j(t), \tag{9.14}$$

where $s_k(t)$ is the signal transmitted by node x_k at time k, $w_j(t)$ is AWGN with zero mean and unit variance, and $\theta_{kj}(t)$ is the random phase between nodes x_k and x_j at time t, that is i.i.d. across time slots. The main idea behind considering the fading-plus-path-loss model is to obtain multiplexing gain similar to the point-to-point multiple input multiple output (MIMO) channel, by allowing multiple distributed nodes to collaborate for transmission and reception, forming a virtual multiple antenna channel.

Note that this model is a slight deviation from our usual fading-plus-path-loss model, where we typically have

$$y_j(t) = \sum_{k\in[1:n],k\neq j} d_{ij}^{-\alpha}h_{kj}(t)s_k(t) + w_j(t), \tag{9.15}$$

where $h_{kj}(t) = |h_{kj}(t)|\exp\left(\jmath\,\theta_{kj}(t)\right)$ consists of attenuation by multi-path and the random phase. In (9.14), we have ignored the multi path attenuation, that is, $|h_{kj}(t)| = 1$. This simplification allows us simpler analytical exposition, though the results presented next extend to the general case as well.

9.3.1 Upper Bound on Throughput Capacity $\mathsf{T}(n) = \mathcal{O}(n \ln n)$

Theorem 9.3.1 *The throughput capacity of a wireless network is* $\mathsf{T}(n) = \mathcal{O}(n \ln n)$.

Proof: Consider a source-destination pair x_1, x_2. Then the rate of communication between x_1 and x_2 is clearly upper bounded by the rate of communication between x_1 and $\Phi_n \backslash \{x_1\}$, where we assume that all nodes of $\Phi_n \backslash \{x_1\}$ are collocated and have access to each other's received signals without any additional noise. This is equivalent to a point-to-point communication setting between a single transmit antenna and $n-1$ receive antennas, for which the capacity is upper bounded by $\log\left(1 + \left(\sum_{j=2}^{n} d_{1j}^{-\alpha}\right)\right)$. Thus, the throughput of any source-destination pair is upper bounded by

$$R(n) \leq \log\left(1 + \left(\sum_{j=2}^{n} d_{1j}^{-\alpha}\right)\right). \tag{9.16}$$

We can bound the R.H.S. as long as $d_{1j}^{-\alpha}$ is bounded. Recall that n nodes are uniformly distributed in a unit square $\mathbf{S}_1$, hence the probability that d_{ij} is greater than $\frac{1}{n^{1+\kappa}}$ for any $\kappa > 0$ is equal to the case when none of the $n-1$ nodes lie inside a disc of radius $\frac{1}{n^{1+\kappa}}$ with center x_i. Thus,

$$\mathbb{P}\left(\min_j d_{ij} < \frac{1}{n^{1+\kappa}}\right) = 1 - \left(1 - \frac{1}{n^{2+2\kappa}}\right)^{n-1}. \tag{9.17}$$

Taking the union bound over all nodes,

$$\mathbb{P}(\min_j d_{ij} < \frac{1}{n^{1+\kappa}} \text{ for any } i) \leq n\left(1 - \left(1 - \frac{1}{n^{2+2\kappa}}\right)^{n-1}\right). \tag{9.18}$$

Thus, $\lim_{n\to\infty} \mathbb{P}\left(\min_{i,j} d_{ij} < \frac{1}{n^{1+\kappa}}\right) = 0$, for any $\kappa > 0$, and hence there is a minimum separation of $\frac{1}{n^{1+\kappa}}$ between any two nodes of the network with high probability. Therefore, from (9.16), we have

$$\begin{aligned} R(n) &\leq \log\left(1 + \left(\sum_{j=2}^{n} n^{1+\kappa}\right)\right), &(9.19)\\ &\leq c \ln n. &(9.20)\end{aligned}$$

Thus, per-node throughput cannot scale faster than order $\ln n$ and counting for all n nodes, we get

$$\mathsf{T}(n) \leq \mathcal{O}(n \ln n).$$

□

To derive this upper bound, we have assumed that for each source, all the $n-1$ nodes can collaborate (noiselessly) to receive the signal for the corresponding destination, and all source transmissions can be scheduled simultaneously without any interference, which is clearly a gross simplification. Remarkably, we will show next that a hierarchical cooperation strategy can achieve an almost linear scaling of the throughput capacity, leaving only a gap of order $\ln n$ between the upper and the lower bound.

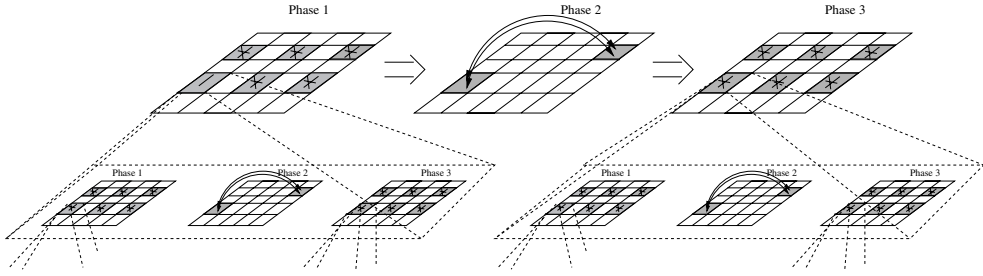

Figure 9.8: Hierarchical layered strategy for achieving almost linear scaling of the throughput capacity.

9.3.2 Achieving Throughput Capacity of $\mathsf{T}(n) = \Omega\left(n^{1-\delta}\right)$

In this section, we describe a hierarchical cooperation strategy proposed in [3] to achieve almost linear scaling of throughput capacity with the number of nodes n. Assuming that there is a strategy $\mathcal{P}$ with $\mathsf{T}(n) = n^b, 0 \leq b < 1$ (a simple time-sharing strategy has $\mathsf{T}(n) = \Theta(1)$, i.e., $b = 0$), the two key ideas behind the hierarchical cooperation strategy are

1. Partition the unit square $\mathbf{S}_1$ into small clusters and use $\mathcal{P}$ simultaneously in sufficiently separated clusters to exchange all bits of all nodes within the cluster. One can show that a constant fraction of clusters can do this simultaneously.

2. Then for each source-destination pair, consider the corresponding clusters, called the source and destination clusters. For transmitting any one source's bits, schedule long-range communication between all the nodes of the source and the destination clusters, which resembles a multiple antenna channel. Long-range communication is done for the n sources serially in n time slots. Within each cluster, each receiver quantizes the received signal to Q bits and sends it to the destination, with intra-cluster communication using $\mathcal{P}$.

The throughput capacity of this new strategy critically depends on the cluster size, since smaller the cluster size, less multiplexing gain is available for the long-range communication, while large cluster sizes require larger intra-cluster communication overhead. Choosing the cluster size such that each cluster has $n^{\frac{1}{2-b}}$ expected number of nodes, one can show that the new strategy has $\mathsf{T}(n) = n^{\frac{1}{2-b}}$. Since for $0 \leq b < 1$, $\frac{1}{2-b} > b$, the new strategy has a higher throughput capacity. Recursing this procedure multiple times, which resembles a hierarchical structure as shown in Fig. 9.8, one can achieve $\mathsf{T}(n) = n^{1-\delta}$, for any $\delta > 0$. This idea of combining short-range and long-range communication was first presented in [4], which showed a strategy with $\mathsf{T}(n) = n^{2/3}$, which was later refined in [3] to get an almost linear scaling of throughput capacity as stated next.

Theorem 9.3.2 *For any $\delta > 0$, there exists constants $K(\delta) > 0$ such that with high probability, with hierarchical cooperation, the throughput capacity*

$$\mathsf{T}(n) \geq K(\delta) n^{1-\delta}.$$

The proof of Theorem 9.3.2 critically depends on the following Lemma that shows that if there exists a strategy with throughput capacity of $\mathsf{T}(n) \geq K_1 n^b$ for any $0 \leq b < 1$, then one can use that to build another strategy (hierarchical cooperation) with throughput capacity of $\mathsf{T}(n) \geq K_2 n^{\frac{1}{2-b}}$. Since for $0 \leq b < 1$, $\frac{1}{2-b} > b$, the new strategy provides a better scaling of the throughput capacity.

Lemma 9.3.3 *For $\alpha > 2$, let there exists a strategy with*

$$\mathsf{T}(n) \geq K_1 n^b$$

for $0 \leq b < 1$ with probability $1 - \exp(n^{c_1})$ for $c_1 > 0$, where each node uses power $\Theta\left(\frac{1}{n}\right)$. Then one can construct another strategy with a higher throughput capacity

$$\mathsf{T}(n) \geq K_2 n^{\frac{1}{2-b}}$$

with probability $1 - \exp(n^{c_2})$ for $c_2 > 0$, where each node again uses power $\Theta\left(\frac{1}{n}\right)$.

Since we are going to recursively use older strategies to produce better ones, it is necessary to ensure that per-node power of each new strategy is $\Theta\left(\frac{1}{n}\right)$ so that the per-node power used by the final strategy remains bounded.

Remark 9.3.4 *Lemma 9.3.3 holds with a scaling by $\frac{1}{\ln n}$ for $\alpha = 2$, that is, if there is a strategy with*

$$\mathsf{T}(n) \geq K_1 \frac{n^b}{\ln n}$$

for $0 \leq b < 1$ with probability $1 - \exp(n^{c_1})$, then one can construct another strategy with

$$\mathsf{T}(n) \geq K_2 \frac{n^{\frac{1}{2-b}}}{(\ln n)^2}$$

with probability $1 - \exp(n^{c_2})$. The scaling with $\ln n$ appears because the interference received at any cluster with a time-sharing protocol scales as $\ln n$ for $\alpha = 2$.

We first present the proof of Theorem 9.3.2 using Lemma 9.3.3, and thereafter describe the proof of Lemma 9.3.3.

Proof: (Theorem 9.3.2) To make use of Lemma 9.3.3, we need a strategy to begin with. Let that be the simplest time-sharing strategy, where there is direct transmission between each source-destination pair, which take turns and only one source-destination pair is active at any time slot. Thus, using $\Theta(n)$ slots, all source-destination pairs can communicate at a constant rate. Therefore, the throughput capacity of the time-sharing strategy is $\mathsf{T}(n) = \Theta(1)$, thus satisfying $\mathsf{T}(n) = n^b$ for $b = 0$. Moreover, the probability of achieving $\mathsf{T}(n) = \Theta(1)$ is 1, since all communication happens in orthogonal time slots, and since each node transmits once in n slots, the transmit power is $\Theta\left(\frac{1}{n}\right)$ for each node. This is our base starting point.

Starting with $b = 0$, applying Lemma 9.3.3 recursively m times, we get that there exists a strategy with $\mathsf{T}(n) \geq \Theta\left(n^{\frac{m}{m+1}}\right)$. Thus, given any $\delta > 0$, we can choose m such that $\frac{m}{m+1} > 1 - \delta$, proving the desired existence of a strategy with

$$\mathsf{T}(n) \geq \Theta(n^{1-\delta}).$$

□

Now we present the proof of Lemma 9.3.3, which shows that there is a way of hierarchically building a better strategy using a weaker strategy in terms of the throughput capacity.

Proof: (Lemma 9.3.3) We divide the unit square $\mathbf{S}_1$ into smaller squares/clusters $\mathbf{s}_{ij}$ with area $\nu(s)$, where each cluster contains $M = \nu(s)n$ nodes on average, as shown in Fig. 9.9. Using Chernoff bound, we can show that each cluster contains $\Theta(M)$ nodes with high probability as follows.

Lemma 9.3.5 *For any $\kappa > 0$, the number of nodes lying in any smaller square/cluster s with area $\nu(s)$ lies in the interval $[(1-\kappa)\nu(s)n, (1+\kappa)\nu(s)n]$ with probability at least*

$$1 - \frac{2}{\nu(s)} \exp\left(-c(\kappa)\nu(s)n\right),$$

where $c(\kappa) > 0$ for $\kappa > 0$ and is independent of n and $\nu(s)$.

In particular, we will assume that each cluster contains M nodes. We restrict our attention to a particular source-destination pair $\mathcal{S} - \mathcal{D}$ that lie in non-adjacent clusters, where $\mathcal{S}$ intends to send M bits to $\mathcal{D}$ using the following three phase strategy. Remark 9.3.6 points out how to handle the case when $\mathcal{S} - \mathcal{D}$ lie in the same cluster or in adjacent clusters.

1. Short-range cluster cooperation: $\mathcal{S}$ distributes its M bits to the M nodes in its cluster, one bit to each node. Within a cluster, there are M nodes, and each wants to send M bits to its other M nodes in the cluster. Thus, the total traffic demand is for M^2 bits. One can think of this as small-scale counterpart of the original problem, where M^2 bits need to be communicated in a network of M nodes spread over an area of size $\nu(s)$. We divide this communication over M sessions, where in each session a designated set of M source-destination pairs exchange their 1 bit. M sessions are held in series one after another, completing the traffic demand. Phase 1 is illustrated in Fig. 9.9.

   For this intra-cluster communication, we make use of the strategy given by the hypothesis that promises a throughput capacity of $\mathsf{T}(M) = \Theta(M^b)$. Therefore, each of the M sessions can be

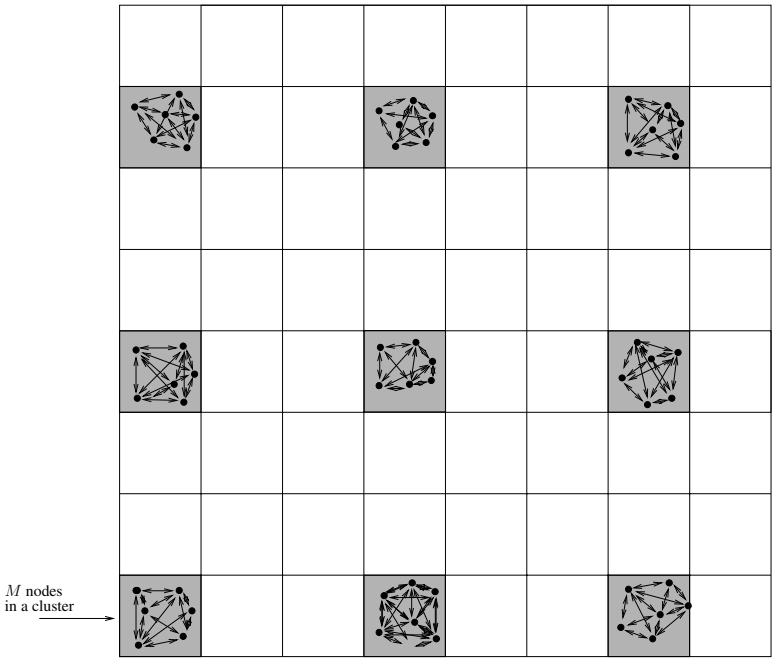

Figure 9.9: In phase 1 all nodes in a cluster exchange their bits, where only the clusters in shaded squares are active at any time.

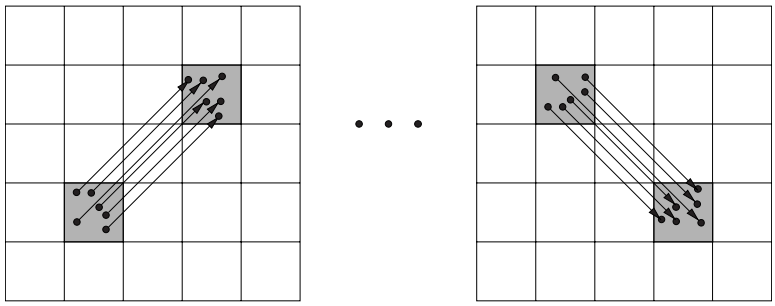

Figure 9.10: Sequential transmission of information between all M nodes of two clusters over long-range multiple antenna communication.

completed in $\Theta\left(M^{1-b}\right)$ time slots, and all M sessions can be completed in $\Theta\left(M^{2-b}\right)$ time slots. We need to show this rigorously (Lemma 9.A.1), because we are now working over a smaller area $\nu(s)$ rather than $\mathbf{S}_1$ as promised by the hypothesis.

We note an important point that multiple clusters can be active at the same time as long as they are sufficiently separated by a constant distance as shown in Lemma 9.2.8, since interference seen at any receiving node can be bounded by a constant for $\alpha > 2$. Therefore, using a constant time-sharing parameter, all clusters can finish their M sessions in $\Theta\left(M^{2-b}\right)$ time slots.

2. Long-range virtual MIMO transmission: Consider a single source-destination pair $\mathcal{S} - \mathcal{D}$. At the end of phase 1, out of total M bits of $\mathcal{S}$, 1 bit is available with each of the M nodes of its cluster (source cluster). In phase 2, at any time slot, all the M bits of a source $\mathcal{S}$ are transmitted together by the M nodes of the source cluster to the M nodes of the cluster of $\mathcal{D}$ (destination cluster). This communication is equivalent to virtual multiple antenna transmission, since the M transmit and M receive antennas are not co located but cooperate distributively for transmission and reception. This process is repeated for all source-destination pairs, and hence takes $\Theta(n)$ time slots to complete. Fig. 9.10 depicts the long-range multiple antenna transmission in phase 2.

3. Short-range cluster dissemination: In phase 2, each node of any cluster receives M signals corresponding to the M nodes in its cluster. In phase 3, each node quantizes each of its M received signals using Q bits and sends them to their respective destinations, similar to phase 1. In total, the traffic demand is QM^2 bits, and similar to phase 1, this can be satisfied in $\Theta(M^{2-b})$ time slots using the strategy described in the hypothesis (Lemma 9.A.1).

Assuming that the aggregate rate of $M \times M$ virtual multiple antenna link scales linearly in M (Lemma 9.B.1 and Lemma 9.B.4) with quantization used at each node of the destination clusters in phases 2 and 3, we calculate the throughput capacity of this three-phase strategy as follows.

Each source transmits M bits to its destination, thus overall, nM bits are transported through the network in total $\Theta(M^{2-b} + n + M^{2-b})$ time slots (accounting for the three phases), yielding a throughput capacity of

$$\mathsf{T}(n) \geq \frac{nM}{M^{2-b} + n + M^{2-b}}, \text{ bits/time slot}$$

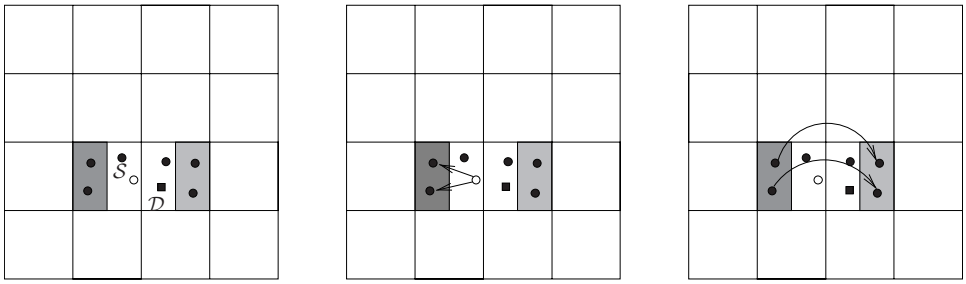

Figure 9.11: Three-phase protocol for source (hollow dot)-destination (black square) pairs lying in adjacent clusters.

Recall that we can choose M (or the size of smaller square $\nu(s)$, where $M = n\nu(s)$). Fixing $M = n^{1/(2-b)}$, we get

$$\mathsf{T}(n) \geq n^{\frac{1}{2-b}}.$$

Note that we also need to show that $\mathsf{T}(n) \geq n^{\frac{1}{2-b}}$ with probability $1 - \exp(n^{c_2})$ and that the per-node power used in the new strategy scales as $\Theta\left(\frac{1}{n}\right)$. We next show that the success probability constraint is satisfied, while defer proving the required per-node powers scaling to Lemma 9.B.1.

Note that the new strategy fails to achieve the promised throughput if each cluster does not contain M nodes or when the old (hypothesis) strategy fails to achieve the promised throughput in any of the clusters. From the Chernoff bound, we know that each cluster contains M nodes with high probability, so that takes care of first "bad" event. Moreover, the old strategy is used in each of the n/M cluster M times, which is polynomial in n and M. Thus, taking the union bound, we conclude that the new strategy has $\mathsf{T}(n) \geq n^{\frac{1}{2-b}}$ with probability $1 - \exp(n^{c_2})$ whenever the old strategy achieves $\mathsf{T}(n) \geq n^b$ with probability $1 - \exp(n^{c_1})$. □

All the intermediate lemmas used in the proof of Lemma 9.3.3 are presented in Appendix 9.A and 9.B.

Remark 9.3.6 *If a particular source-destination pair lies in the same cluster, then phase* 2 *and phase* 3 *are not required. Otherwise, if the source-destination pair lie in adjacent clusters, then we divide both the smaller squares/clusters in two halves, and the source transmits its* $M/2$ *bits to the* $M/2$ *nodes of its cluster that lie in the partition farther away from the destination node. Then in phase* 2*, all the* $M/2$ *nodes of the source cluster transmit their bits to the* $M/2$ *nodes of the destination cluster that lie in the partition that is farther from the source node. Thereafter, these* $M/2$ *nodes communicate their quantized signals to the destination in phase* 3*. Pictorially this is depicted in Fig. 9.11.*

In summary, this hierarchical cooperation strategy, where nodes cooperate at multiple levels, is able to beat the bottleneck of $\Theta\left(\frac{1}{\sqrt{n}}\right)$ per-node throughput capacity of the nearest neighbor multi-hop routing. Recall that the key limitation of the nearest neighbor multi-hop routing is that each node has to relay traffic for roughly $\Theta(\sqrt{n})$ nodes. By allowing cooperation between nodes, we can break this ceiling, and with hierarchical cooperation strategy each node needs to route traffic for far fewer nodes, and each source can reach its destination in significantly less number of hops. Even though the hierarchical cooperation strategy provides better per-node throughput capacity, it also requires larger signaling overhead required for cooperation compared to the nearest neighbor

multi-hop routing. Thus, there is an inherent tradeoff between the per-node throughput capacity and the signaling overhead.

In Section 2.4.1, we compared the transmission capacity and the throughput capacity with multi-hop routing under the SINR model and showed that they have identical scaling, that is, the transmission capacity scales as $\Theta(\sqrt{n})$. The poor scaling performance of the transmission capacity in comparison to the throughput capacity of the hierarchical cooperation strategy is attributed to node cooperation, and the ensuing large signaling overhead, since in the transmission capacity framework, all transmissions are over a single hop with no cooperation.

9.3.3 Multiple Antennas

In addition to using virtual multiple antenna channel, one can ask what if we equip each node with N antennas, how will the scaling of throughput capacity change with n or N? Not surprisingly, in terms of the number of nodes n, there is no change, while in terms of N, the throughput capacity scales linearly with N. Both of these claims can be easily established as follows. For the upper bound on the throughput capacity, (following Theorem 9.3.1), one can easily see that the per-node throughput capacity is bounded by order $N\log(nN)$, since the per-node throughput capacity is bounded from above by a point-to-point multiple antenna channel with N transmit antennas and $N(n-1)$ receive antennas. Similarly, for the lower bound, following Theorem 9.3.2, assume that each node has MN bits to transmit to its destination. Assuming that there exists a strategy with throughput capacity of Nn^b, in phase 1, each source transmits MN bits to its $M-1$ neighbors in clusters, which takes M^{2-b} time slots. Similarly, for phase 2, using $NM \times NM$ virtual multiple antenna channel, long-range communication of MN bits takes n time slots, and phase 3 is similar to phase 1. Thus, overall, nMN bits are transported through the network in total $\Theta(M^{2-b}+n+M^{2-b})$, yielding a throughput capacity of

$$\mathsf{T}(n) \geq \frac{nMN}{M^{2-b}+n+M^{2-b}} \text{ bits/time slot.}$$

Recall that we can choose M (or the size of smaller square $\nu(s)$, where $M = n\nu(s)$), fixing $M = n^{1/(2-b)}$ and we get

$$\mathsf{T}(n) \geq Nn^{\frac{1}{2-b}}.$$

Note that the linear scaling of the throughput capacity with multiple antennas is similar to the case of transmission capacity (Theorem 3.3.5); however, finding the role of multiple antennas and the analysis in the transmission capacity framework is much more complicated as seen in Chapter 3. The main reason being that in the transmission capacity framework, there is no cooperation and the system is interference limited, and the optimal role of multiple antennas is to balance the increase in signal power with respect to mitigating the interference. In comparison, with hierarchical cooperation strategy, nodes cooperate with each other and essentially avoid all interference, allowing full multiplexing gain of the multiple antennas to be realized.

9.4 Extended Networks

In this chapter, we have essentially considered what is called as the *dense* wireless network, where we have fixed a unit area and increased the density of the nodes n. An alternate way to model a wireless network is the *extended* network, where the density of nodes is fixed, while increasing the area, say a square of side $\sqrt{n}$. Dense network model nodes locations in urban areas, where more and

more basestations and users are added in a fixed area, while extended networks are more suited for rural areas, where the density of nodes is fixed and we want to cover a large piece of geographical area.

It is easy to notice that dense networks are inherently interference limited, while extended networks are power/coverage limited, since with large transmitter-receiver distances in extended networks, maintaining the minimum received SNRs is a challenge. An extended network with unit density of nodes over a square of side $\sqrt{n}$ is equivalent to a dense network with distance between any pair of nodes scaled by $\sqrt{n}$. Thus, equivalently, in terms of power, an extended network with per-node power constraint of P is identical to a dense network with power constraint of $\frac{P}{n^{\alpha/2}}$ at each node. This observation allows us to construct the hierarchical cooperation strategy for the extended network case as follows.

We noted in Theorem 9.3.2 that for the hierarchical cooperation strategy, each node uses power of $\Theta\left(\frac{1}{n}\right)$. Thus, for $\alpha = 2$, the hierarchical cooperation strategy satisfies the power constraint of $\frac{P}{n^{\alpha/2}}$ imposed by the extended networks. Thus, an almost linear scaling of the throughput capacity can also be achieved in an extended network with the hierarchical cooperation strategy. For $\alpha > 2$, the power constraint of $\frac{P}{n^{\alpha/2}}$ is violated if we use the hierarchical cooperation strategy since that uses power $\Theta\left(\frac{1}{n}\right)$. To fix this, one can operate the hierarchical cooperation strategy for a fraction $\frac{1}{n^{2/\alpha-1}}$ of time with power $\Theta\left(\frac{1}{n}\right)$ and remain silent for the rest of the time. Thus, satisfying the power constraint $\frac{P}{n^{\alpha/2}}$ imposed by extended networks. Thus, using the hierarchical cooperation strategy, the throughput capacity of

$$n^{1-\delta}\frac{1}{n^{2/\alpha}-1} = n^{2-\alpha/2-\delta}$$

is achievable in an extended network, where the throughput capacity of $n^{1-\delta}$ follows from Theorem 9.3.2.

So for $2 < \alpha < 3$, we have $\mathsf{T}(n) \geq n^{2-\alpha/2-\delta}$, which is better than the order $\sqrt{n}$ scaling of the nearest neighbor multi-hop strategy (Section 9.2.2), while for $\alpha \geq 3$, $n^{2-\alpha/2-\delta} < \sqrt{n}$, and the nearest neighbor multi-hop strategy starts outperforming the hierarchical cooperation strategy. Thus, we get that the following throughput is achievable in an extended network.

Theorem 9.4.1 *For an extended network with unit density over a square of side $\sqrt{n}$, the achievable throughput scales as*

1. *For $2 \leq \alpha < 3$, $\mathsf{T}(n) \geq K(\delta)n^{2-\alpha/2-\delta}$,*
2. *For $\alpha \geq 3$, $\mathsf{T}(n) \geq K(\delta)\sqrt{n}$,*

with high probability for any $\delta > 0$.

Matching upper bounds can also be derived using the cut-set bound that are stated as follows.

Theorem 9.4.2 *For an extended network with unit density over a square of side $\sqrt{n}$, the throughput capacity is bounded by*

1. *For $2 \leq \alpha \leq 3$, $\mathsf{T}(n) \leq K'(\delta)n^{2-\alpha/2+\delta}$,*
2. *For $\alpha > 3$, $\mathsf{T}(n) \geq K'(\delta)n^{1/2+\delta}$,*

with high probability for any $\delta > 0$.

Theorem 9.4.2 is obtained by considering a cut across the extended network that divides the square with side $\sqrt{n}$ in two equal halves. With random source-destination pairs, with high probability, one-fourth of the source-destination pairs lie on opposite sides of the cut. Thus, the throughput across the cut is at least one-fourth of the total throughput. Then, the throughput across the cut is bounded by the capacity of the multiple antenna channel, assuming that all nodes on either side of the cut cooperate with each other under suitable power constraints.

9.5 Reference Notes

This chapter is based on work started by the seminal paper of Gupta and Kumar [2]. The upper bound on the throughput capacity of the SINR model follows from [2], while the achievability of $\sqrt{n}$ scaling of the throughput capacity is based on [5]. The information theoretic study of throughput capacity was initiated by [6], while the idea of using multi phase transmission to beat the order $\sqrt{n}$ scaling was first presented in [4] and later refined to show that linear scaling of throughput capacity is possible in [3]. The results presented in Section 9.3 are based on [3]. The hierarchical cooperation strategy for arbitrary networks has been analyzed in [7]. Several other important results in this area include an upper bound on the throughput capacity by [8–10], analyzing effect of mobility [11], and characterizing the effect of fading channels [12]. Results on the transport capacity of extended networks can be found in [13, 14]. Some results on the physics of signal transmission showing that throughput capacity cannot scale better than $\sqrt{n}$ under realistic assumptions appear in [15].

9.A Hierarchical Cooperation

Lemma 9.A.1 *If any strategy has a throughput capacity of* $\mathsf{T}(M) = \Theta(M^b)$ *for* M *nodes in* $\mathbf{S}_1$. *Then* $\mathsf{T}(M) = \theta(M^b)$ *can be achieved even if* M *nodes lie in a smaller square* $\mathbf{s}_{ij}$ *of area* $\nu(s)$.

Proof: Recall that to exchange M^2 bits in phase 1, M sessions are organized within each cluster, where in each session, one bit is exchanged between each of the M source-destination pairs inside the cluster. Changing the choice of source-destination pairs over the M sessions ensures that each node in the clusters gets one bit from every other node. We next show that in each session, we can achieve a throughput capacity of M^b given the hypothesis strategy.

For the hypothesis strategy, let power P is transmitted by each of the M nodes lying in $\mathbf{S}_1$ for which $\mathsf{T}(M) = \Theta(M^b)$. Compared to $\mathbf{S}_1$, in any smaller square $\mathbf{s}_{ij}$, the distances are scaled down by $1/\nu(s)$. Hence, the received power at any node within $\mathbf{s}_{ij}$ is increased by a factor $\frac{1}{\nu(s)^{\alpha/2}}$ compared to the case when the node was lying in $\mathbf{S}_1$. Thus, if originally each node lying in $\mathbf{S}_1$ used transmit power P, transmit power of only $P\nu(s)^{\alpha/2}$ is needed by each node lying in any smaller square/cluster $\mathbf{s}_{ij}$ with area $\nu(s)$.

Consider any one session in a given cluster. With the reduced transmit power of $P\nu(s)^{\alpha/2}$ used by each node of each cluster, similar to the proof of Lemma 9.2.8, if simultaneously active clusters are separated by a constant distance, the interference received at any node in a cluster is bounded by a constant. Hence, the rate of transmission supported between nodes lying in $\mathbf{s}_{ij}$ with power $P\nu(s)^{\alpha/2}$ is order wise similar to when nodes lie in $\mathbf{S}_1$ with power P. Thus, using the hypothesis strategy, we can get an order wise rate of M^b for each session within each cluster. Moreover, note that, at any time a constant fraction of clusters are active, and hence all clusters finish their session in constant time. Thus, we conclude that the throughput capacity of M^b with M nodes can be achieved even on a smaller square $\mathbf{s}_{ij}$ of $\mathbf{S}_1$ in Phases 1 and 3. □

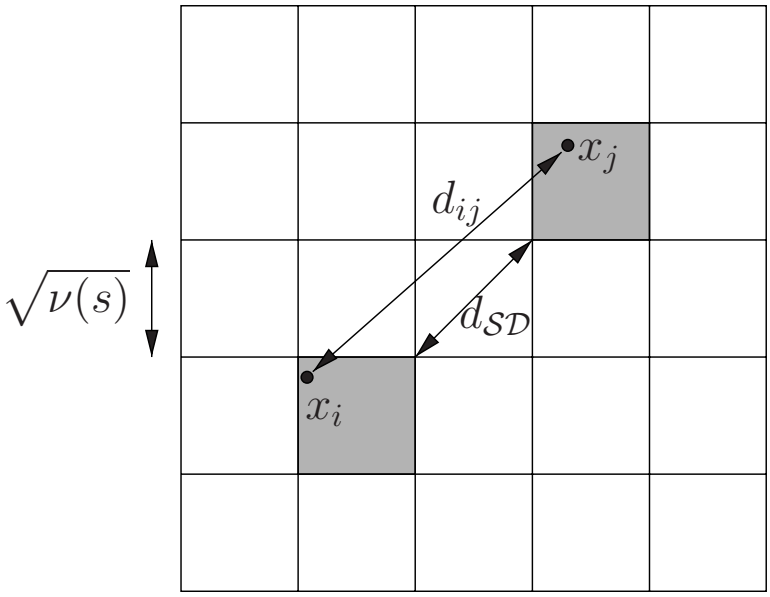

Figure 9.12: Shaded squares are active source-destination clusters in Phase 2.

9.B Mutual Information of Multiple Antenna Channel with Quantization

Note that to conclude the proof of Lemma 9.3.3, we need to show that the mutual information between the $M \times M$ virtual multiple antenna communication with quantization increases linearly in the number of nodes M. In particular, we are interested in the spatial multiplexing gain r that is defined as

$$r = \lim_{N\to\infty} \frac{\mathsf{I}(\mathbf{x};\mathbf{y})}{N},$$

where $\mathsf{I}(\mathbf{x};\mathbf{y})$ is the mutual information between the input $\mathbf{x}$ and output $\mathbf{y}$, and N is the number of antennas at the input and the output. For our case $N = M$ and we want $r = 1$.

Toward, that end, we first show in Lemma 9.B.1 that without quantization, $r = 1$ for the $M \times M$ virtual multiple antenna communication, that is, the mutual information increases linearly in the number of nodes M. Thereafter by arguing that the scaling of the mutual information with quantization cannot be less than without quantization, we prove the required result in Lemma 9.B.4.

Lemma 9.B.1 *The mutual information achieved by the $M \times M$ virtual multiple antenna transmission in Phase 2 between any two clusters grows at least linearly in M.*

Proof: Following (9.14), the $M \times M$ multiple antenna channel between the M receivers with single antenna of the destination cluster and the M transmitters with single antenna of the source cluster is

$$\mathbf{y}_d(t) = \sqrt{\frac{1}{M}}\mathbf{F}_{sd}(t)\mathbf{s}(t) + \mathbf{w}(t), \tag{9.21}$$

where normalization $\sqrt{\frac{1}{M}}$ is to ensure that total unit power is transmitted by the cluster, $\mathbf{s}(t)$ is the $M \times 1$ signal sent from the M nodes of a cluster, $\mathbf{w}(t)$ is the AWGN vector with each entry independent with zero mean and unit variance, and the matrix $\mathbf{F}_{sd}$ is such that its $(i,j)\text{th}$ entry is the path-loss times the random phase between i^{th} receive antenna of the destination cluster and the j^{th} transmit antenna of the source cluster, that is, $\mathbf{F}_{sd}(i,j) = d_{ij}^{-\alpha/2} \exp(\jmath\theta_{ij})$.

In phase 2, any cluster is active for only $\frac{M}{n}$ time slots and remains silent for the rest of the slots. Therefore, the average energy used by any node in phase 2 is $\Theta\left(\frac{1}{n}\right)$ even if it transmits with power $\frac{1}{M}$ whenever active. For phases 1 and 3, the hypothesis strategy is used within each cluster; hence, the power used in phases 1 and 3 is also $\theta\left(\frac{1}{n}\right)$. Thus, counting for all the 3 phases, we conclude that the per-node power with the new strategy is also $\theta\left(\frac{1}{n}\right)$ as required in Lemma 9.3.3.

Let the distance between the closest points (not nodes) of the source-destination clusters be $d_{\mathcal{SD}}$. Then, we have that $d_{\mathcal{SD}} \leq d_{ij} \leq d_{\mathcal{SD}} + 2\sqrt{2\nu(s)}$ as shown in Fig. 9.12, where d_{ij} is the distance between the i^{th} node of the source cluster and j^{th} node of the destination cluster. Moreover, when source-destination clusters are not adjacent, then $d_{\mathcal{SD}} > \sqrt{\nu(s)}$. Thus, we get

$$a = \left(\frac{1}{1+2\sqrt{2}}\right)^{\alpha} \leq \left(\frac{d_{\mathcal{SD}}}{d_{ij}}\right)^{\alpha} \leq 1.$$

For adjacent source-destination clusters, using Remark 9.3.6, we know that the minimum separation between source-destination clusters is at least $\sqrt{\nu(s)}$.

Thus, there exists $b > a > 0$ with a and b independent of n, such that $d_{ij}^{-\alpha/2} = d_{\mathcal{SD}}^{-\alpha/2}\mathsf{z}_{ij}$, where $\mathsf{z}_{ij} \in [a,b]$, both in cases when $\mathcal{S}$ and $\mathcal{D}$ are in neighboring clusters or not. Let $\mathbf{G}_{sd}$ be a matrix with entries $\mathbf{G}_{sd}(i,j) = \mathsf{z}_{ij}\exp(\jmath\theta_{ij})$, that is, $\mathbf{F}_{sd} = d_{\mathcal{SD}}^{-\alpha/2}\mathbf{G}_{sd}$. Thus, we can write the received signal (9.21) as

$$\mathbf{y}_d(t) = \sqrt{\frac{1}{M}} d_{\mathcal{SD}}^{-\alpha/2}\mathbf{G}_{sd}(t)\mathbf{s}(t) + \mathbf{w}(t). \tag{9.22}$$

Assuming perfect knowledge of channel state information (CSI) at all the receivers in the destination cluster, the mutual information of this channel is

$$I(\mathbf{s};\mathbf{y},\mathbf{G}_{sd}) = \mathbb{E}\left\{\log\det\left(\mathbf{I} + \frac{d_{\mathcal{SD}}^{-\alpha}}{M}\mathbf{G}\mathbf{G}^{\dagger}\right)\right\}. \tag{9.23}$$

Let $\sigma_i(\mathbf{G}), i = 1,\ldots,M$ be the eigenvalues of $\frac{1}{M}\mathbf{G}\mathbf{G}^{\dagger}$. Then

$$I(\mathbf{s};\mathbf{y},\mathbf{G}_{sd}) = \mathbb{E}\left\{\sum_{i=1}^{M}\log\left(1 + d_{\mathcal{SD}}^{-\alpha}\sigma_i(\mathbf{G})\right)\right\}. \tag{9.24}$$

Let $\sigma(\mathbf{G})$ be an eigenvalue picked uniformly randomly from $\sigma_1(\mathbf{G}),\ldots,\sigma_M(\mathbf{G})$, then

$$\begin{aligned} I(\mathbf{s};\mathbf{y},\mathbf{G}_{sd}) &\overset{(a)}{=} M\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left\{\log\left(1 + d_{\mathcal{SD}}^{-\alpha}\sigma_i(\mathbf{G})\right)\right\}, \\ &\overset{(b)}{=} M\mathbb{E}\left\{\log\left(1 + d_{\mathcal{SD}}^{-\alpha}\sigma(\mathbf{G})\right)\right\}, \end{aligned}$$

$$\overset{(c)}{\geq} \quad M \log(1 + d_{\mathcal{SD}}^{-\alpha} t) \mathbb{P}(\sigma(\mathbf{G}) > t),$$

for any $t \geq 0$, where (a) follows from the linearity of expectation, (b) follows by definition of expectation, and (c) follows by only considering $\sigma(\mathbf{G}) > t$. From Lemma 9.B.2, we have

$$I(\mathbf{s}; \mathbf{y}, \mathbf{G}_{sd}) = M \log(1 + d_{\mathcal{SD}}^{-\alpha} t) \frac{(\mathbb{E}\{\sigma(\mathbf{G})\} - t)^2}{\mathbb{E}\{\sigma(\mathbf{G})^2\}}. \tag{9.25}$$

It is easy to check that

$$\mathbb{E}\{\sigma(\mathbf{G})\} = \frac{1}{M^2} \sum_{i,j=1}^{M} \mathsf{z}_{ij}^2 \geq a^2.$$

Moreover,

$$\mathbb{E}\{\sigma(\mathbf{G})^2\} = \frac{2}{M^3} \sum_{i,j,k=1}^{M} \mathsf{z}_{ij}^2 \mathsf{z}_{ik}^2 \leq 2b^4.$$

Thus, we have for any $t < a$

$$I(\mathbf{s}; \mathbf{y}, \mathbf{G}_{sd}) = M \log(1 + d_{\mathcal{SD}}^{-\alpha} t) \frac{(a^2 - t)^2}{2b^4}. \tag{9.26}$$

Choosing $t = a/2$ shows that $(\mathbf{s}; \mathbf{y}, \mathbf{G}_{sd})$ grows linearly with M, that is, $r = 1$. □

Lemma 9.B.2 *(Paley–Zygmund inequality) For any non-negative random variable X such that $\mathbb{E}\{X^2\} < \infty$, for any $t \geq 0$*

$$\mathbb{P}(X > t) \geq \frac{(\mathbb{E}\{X\} - t)^2}{\mathbb{E}\{X^2\}}$$

for any $t < \mathbb{E}\{X\}$.

Now we come to the final stage, where we show that even with quantized signals being sent at a constant rate from each of the nodes of the destination cluster, the spatial multiplexing gain $\mathsf{r} = 1$.

Consider a discrete memoryless channel with single input $\mathcal{X}$ and M outputs $\mathcal{Y}_1, \ldots, \mathcal{Y}_M$ defined by $p(y_1, \ldots, y_M|x)$. The outputs of the channel are quantized independently to get $\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_M$ that are available at the final destination. Note that in our setting, $\mathcal{X}$ corresponds to the vector transmitted by the source cluster, $\mathcal{Y}_1, \ldots, \mathcal{Y}_M$ are the outputs at each of the nodes in the destination cluster, which are quantized independently by these nodes to send $\tilde{\mathcal{Y}}_i$ to the destination. Hence, the goal is to recover the transmitted information x by observing the outputs of $\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_M$ at the destination. The following result from information theory [16] using [[17], 13.6] is useful to characterize the transmission rates that ensure decoding of x with arbitrarily small probability at the destination.

Proposition 9.B.3 *Given a probability distribution $q(x)$ (input), $q_M(y_1, \ldots, y_M|x)$ (channel), and $q_j(\tilde{y}_j|y_j)$ (quantizers) for $j = 1, \ldots, M$, all rates $(R; R_1, \ldots, R_m)$ are achievable that satisfy*

$$R < \mathsf{I}(X; \tilde{Y}_1, \ldots, \tilde{Y}_M),$$

and

$$R_j > \mathsf{I}(Y_j; \tilde{Y}_j),$$

where R is rate of transmission from source X to the destination using $\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_M$, and R_j's are quantization rates from the j^{th} receiver to the destination.

Lemma 9.B.4 *There exists a strategy to encode the observations at a fixed rate Q bits per observation to get $\mathsf{r} = 1$ for the $M \times M$ quantized virtual multiple antenna channel.*

Proof: Proof is based on Proposition 9.B.3. Consider the conditional probability densities

$$q_j(\tilde{y}_j|y_j) \sim \mathcal{N}(y_j, \mathsf{d}^2) \tag{9.27}$$

for the quantization process, where d is the distortion corresponding to the quantization. From Proposition 9.B.3, we know that all rate pairs $(R; R_1 \dots R_M)$ are achievable if

$$R < \mathsf{I}(X; \tilde{Y}_1, \dots, \tilde{Y}_M)$$

and

$$R_j > \mathsf{I}(Y_j; \tilde{Y}_j), j = 1, \dots, M,$$

where R is the transmission rate from the source to the destination, and R_j's are the quantized encoding rate of the j^{th} stream from the j^{th} receiver in the destination cluster to the destination.

We note the following observation that the received power at any node in the destination cluster in phase 2 is bounded by a constant, say P_2, since the distance between any node in the destination cluster and any node in the source cluster is bounded from below. Thus, the received signal at the j^{th} node of the destination cluster Y_j has an average power constraint of P_2. Hence, from (9.27), and using the fact that Gaussian distribution maximizes mutual information, we have

$$\mathsf{I}(Y_j; \tilde{Y}_j) \leq \log\left(1 + \frac{P_2}{\mathsf{d}^2}\right)$$

for any probability distribution on the input space $p(x)$.

So if we choose the constant quantization rate (as it does not depend on M),

$$R_j = \log\left(1 + \frac{P_2}{\mathsf{d}^2}\right) + \psi, \ \forall \ j,$$

for some $\psi > 0$, all rates

$$R \leq \mathsf{I}(X; \tilde{Y}_1, \dots, \tilde{Y}_M),$$

are achievable between the source and the destination for any input distribution $p(x)$, where the equivalent channel between the source and the destination using the conditional distribution (9.27), can be written as

$$\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{x} + \mathbf{w} + \mathsf{D}, \tag{9.28}$$

where $\tilde{\mathbf{Y}} = [\tilde{Y}_1, \dots, \tilde{Y}_M]^T$, and D represents the distortion vector whose each entry is independent and Gaussian distributed with zero mean and d^2 variance. R_j plays the role of Q quantized bits being sent from each receiver in the destination cluster to the destination.

Comparing (9.28) to (9.21), we note that (9.28) has been degraded by an additive independent noise compared to (9.21), thus the spatial multiplexing gain r of the channel with quantization (9.28) is no less than the original virtual multiple antenna channel (9.21). Thus, we conclude that $\mathsf{r} = 1$ even for the $M \times M$ quantized virtual multiple antenna channel. □

Bibliography

[1] A. Agarwal and P. Kumar. 2004. "Capacity bounds for ad hoc and hybrid wireless networks." *ACM SIGCOMM Comp. Commun. Rev.* 34 (3): 71–81.

[2] P. Gupta and P. Kumar. 2000. "The capacity of wireless networks." *IEEE Trans. Inf. Theory* 46 (2): 388–04.

[3] A. Ozgur, O. Leveque, and D. Tse. 2007. "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks." *IEEE Trans. Inf. Theory* 53 (10): 3549–72.

[4] S. Aeron and V. Saligrama. 2007. "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime." *IEEE Trans. Inf. Theory* 53 (6): 2044–59.

[5] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran. 2007. "Closing the gap in the capacity of wireless networks via percolation theory." *IEEE Trans. Inf. Theory* 53 (3): 1009–18.

[6] L. Xie and P. Kumar. 2006. "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks." *IEEE Trans. Inf. Theory* 52 (6): 2313–28.

[7] U. Niesen, P. Gupta, and D. Shah. 2009. "On capacity scaling in arbitrary wireless networks." *IEEE Trans. Inf. Theory* 55 (9): 3959–82.

[8] S. H. A. Ahmad, A. Jovičić, and P. Viswanath. 2006. "On outer bounds to the capacity region of wireless networks." *IEEE/ACM Trans. Netw.* 14 2770–76.

[9] A. Jovičić, P. Viswanath, and S. Kulkarni. 2004. "Upper bounds to transport capacity of wireless networks." *IEEE Trans. Inf. Theory* 50 (11): 2555–65.

[10] E. Telatar and O. Lévêque. 2004. "Information theoretic upper bounds on the capacity of large extended ad-hoc wireless networks." In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, no. LTHI-CONF-2006-010, 2004.

[11] M. Grossglauser and D. Tse. 2002. "Mobility increases the capacity of ad hoc wireless networks." *IEEE/ACM Trans. Netw.* 10 (4): 477–86.

[12] F. Xue, L.-L. Xie, and P. Kumar. 2005. "The transport capacity of wireless networks over fading channels." *IEEE Trans. Inf. Theory* 51 (3): 834–47.

[13] E. Telatar and O. Lévêque. 2004. "Information theoretic upper bounds on the capacity of large extended ad-hoc wireless networks." In *Proc. of the 2004 IEEE International Symposium on Information Theory*.

[14] A. Ozgur, O. Lévêque, and E. Preissmann. 2007. "Scaling laws for one-and two-dimensional random wireless networks in the low-attenuation regime." *IEEE Trans. Inf. Theory* 53 (10): 3573–85.

[15] M. Franceschetti, D. Migliore, and P. Minero. 2009. "The capacity of wireless networks: Information-theoretic and physical limits." *IEEE Trans. Inf. Theory* 55 (8): 3413–24.

[16] G. Kramer, M. Gastpar, and P. Gupta. 2005. "Cooperative strategies and capacity theorems for relay networks." *IEEE Trans. Inf. Theory* 51 (9): 3037–63.

[17] T. Cover and J. Thomas. 2004. *Elements of Information Theory*. John Wiley and Sons.

Index