Capacity Scaling and Optimal Operation of Wireless Networks

by

Javad Ghaderi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

How much information can be transferred over a wireless network and what is the optimal strategy for the operation of such network? This thesis tries to answer some of these questions from an information theoretic approach.

A model of wireless network is formulated to capture the main features of the wireless medium as well as topology of the network. The performance metrics are throughput and transport capacity. The throughput is the summation of all reliable communication rates for all source-destination pairs in the network. The transport capacity is a sum rate where each rate is weighted by the distance over which it is transported. Based on the network model, we study the scaling laws for the performance measures as the number of users in the network grows.

First, we analyze the performance of multihop wireless network under different criteria for successful reception of packets at the receiver. Then, we consider the problem of information transfer without arbitrary assumptions on the operation of the network. We observe that there is a dichotomy between the cases of relatively high signal attenuation and low attenuation. Moreover, a fundamental relationship between the performance metrics and the total transmitted power of users is discovered. As a result, the optimality of multihop is demonstrated for some scenarios in high attenuation regime, and better strategies than multihop are proposed for the operation in the low attenuation regime. Then, we study the performance of a special class of networks, random networks, where the traffic is uniformly distributed inside the networks. For this special class, the upperbounds on the throughput are presented for both low and high attenuation cases. To achieve the presented upperbounds, a hierarchical cooperation scheme is analyzed and optimized by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. In addition, to apply the hierarchical cooperation scheme to random networks, a clustering algorithm is developed, which divides the whole network into quadrilateral clusters, each with exactly the number of nodes required.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Sherman Shen for his support and encouragement in the course of this thesis. Sherman was more than just a supervisor for me; he gave me freedom to choose the topic of my interest, and taught me invaluable lessons about research and life. I was also fortunate to work with Professor Liang-Liang Xie. I am grateful to him for his insightful discussions and comments about my work. Without his guidance, preparing this thesis would have been impossible. I will miss our numerous meetings and discussions that shaped my ideas. I would like to thank the other member of my seminar committee, Professor Patrick Mitran, for his time and thoughtful comments.

My sincere thanks go to my parents for their support and dedication through my life. There is no way that I can express my gratitude to them. Besides, I am really grateful to Majid, my brother, for his inspiration and guidance.

I would also like to thank my colleagues at the Broadband Communications Research (BBCR) group for their support. They provided an excellent environment for me during my studies at UW. Thanks to all my friends for their advice and support: Amir, Mohammad Hossein, Mohammad and Xiaou.

Dedication

To my Parents

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation and Background

Wireless networks formed by radio nodes are a subject of much topical interest, and they are found in various applications such as ad hoc networks, mesh networks, sensor networks, etc. These wireless networks without infrastructure (see figure 1.1) have been supposed to work in multi-hop mode: packets are relayed from node to node in several hops until they reach their destinations and interference is essentially regarded as noise, i.e., only point-to-point coding is considered. While this may model how current technology operates, this model does not tell what are the ultimate limits to the information transfer in future wireless networks. The reason is that interference can carry information; so, one wishes to study wireless networks without making arbitrary assumptions about how they operate.

For the optimal design and operation of such networks, it is of fundamental importance to determine the information-theoretic capacity of such networks, which, however, is a formidable task, since even for the simple three-node scenario [2], the exact capacity is still undetermined after several decades' effort. In fact, networks with a few nodes have very complex cooperation possibilities. The situation becomes more complicated for networks with several source-destination pairs among a large number of nodes, all cooperating in whatever ways are imaginable to maximize information transfer. As observed in [22], the union between information theory and networking is not consummated.

Although the exact capacity is extremely difficult to determine, a lot of insightful upper and lower bounds on the capacity of large wireless networks have been obtained in recent years [3]-[18]. The progress has been made by asking for less.

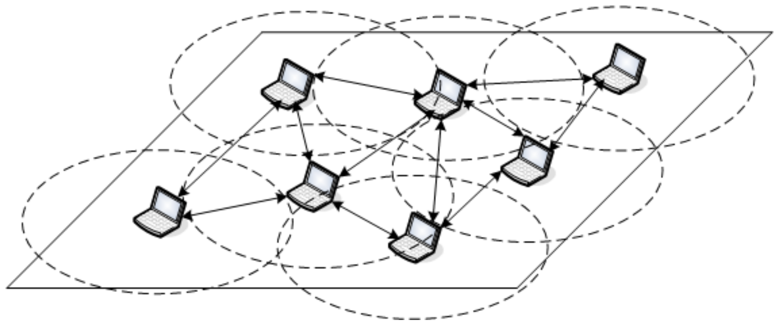

Figure 1.1: A wireless Ad Hoc network without infrastructure.

Instead of studying the capacity region, which is the closure of all feasible rate vectors, we study two scalars, transport capacity and throughput. Transport capacity is defined as

$$C_T = \max_{\substack{All\ R_{ij} \\ in\ the\ capacity\ region}} \sum_{i,j=1}^{n} R_{ij} r_{ij}$$

where R_{ij} is the rate of reliable communication between users i and j, and r_{ij} denotes the Euclidean distance between them. Throughput is another performance metric defined as the sum of all transmission rates in the network. Again we ask for less, and instead of the exact quantity, we study the scaling laws for it as the number n of nodes in the network grows.

The seminal work of Gupta-Kumar [3] initiated the study of scaling laws for multihop wireless networks under several communication models. They essentially demonstrated a trade off between the effects of the number of hops and the amount of traffic routed across the network on the throughput of multihop networks. They proved the former wins this tradeoff. In other words, a nearest neighbor multihop scheme can achieve a higher throughput than what can be achieved by a single hop network. Although this result justified the current layered or crosslayer activities on optimizing the network performance based on multihop, it did not justify the priority of multihop over other possible strategies.

Gupta-Kumar strategy for the constructive lower bound was based on a load balancing routing scheme across the network and discovered a throughput scaling of $\Theta(\sqrt{n/\log n})$ for the class of random networks. Franceschetti et al. [10] later presented a better scheme using percolation theory, to achieve $\Theta(\sqrt{n})$ which meets the upperbound for multihop wireless networks.

Subsequently, a purely information-theoretic approach without any restrictions on the communication schemes was taken in [4], where a more fundamental connection between the total network transmit power and the transport capacity was discovered. As a consequence, when fixing the minimum separation distance and letting the number of nodes increase, the scaling law of $\Theta(\sqrt{n})$ was confirmed in the high signal attenuation regime. However, when the signal attenuation was low, higher scaling laws were shown to be possible for some special relay networks.

Therefore, an interesting question was raised as to what exactly the scaling laws are in the low signal attenuation regime. By incorporating long-range MIMO communications with local cooperations as proposed in [11], a recent work [1] developed a hierarchical architecture which was able to continually increase the scaling by adding more hierarchical stages. Specifically, for a network model where all the nodes are confined in a unit area but still with the far-field signal attenuation, the scaling with h hierarchical stages was claimed to be $\Theta(n^{\frac{h}{h+1}})$. Thus, by letting $h \to \infty$, any scaling of $\Theta(n^{1-\epsilon})$ is achievable, where $\epsilon > 0$ can be arbitrarily small. However, there is a fundamentally important issue that needed to be addressed, i.e., the pre-constant of the scaling. The pre-constants of the scalings for different h are different, and they are not even lower bounded from zero.

1.2 Thesis Objectives

The focus of this thesis is on scaling laws for the capacity of wireless networks. The two fundamental questions of interest are as follows.

1. How much information can wireless networks transport?
2. How should one operate wireless networks?

We consider the classical general setting of figure 1.2 where n users have been distributed within a network of area A. There are multiple source-destination pairs. In general, we allow a source node to generate traffic for several destinations. The goal is to identify the scaling of transport capacity and throughput of such networks as the number of nodes in the network grows.

To answer the first question, we try to find some upperbounds on scaling laws. To do this, we resort to the *max-flow min-cut theorem* and *sum rate bounds based on MIMO techniques*. To answer the second question, we try to find some new strategies for the operation of network that are able to achieve the upperbounds, at least in the sense of scaling growth. The design of such strategies necessitates consideration of important features such as spatial distribution of nodes, the medium

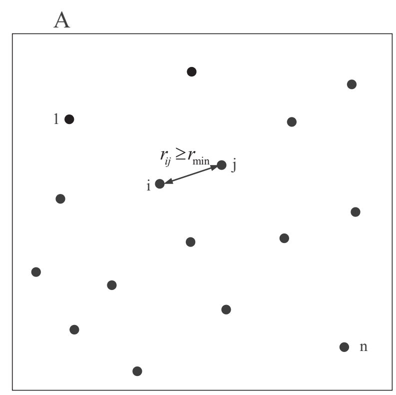

Figure 1.2: A planer network of area A with n nodes.

access control, the attenuation of signals with distance, interference management by spatio-temporal scheduling of transmissions, and different choices of cooperation between nodes.

It will be observed that choosing the appropriate strategy for the operation of the network, significantly depends on the wireless channel properties such as fading and signal attenuation. The topology and the traffic distribution of the network are important factors as well. Such results can help in understanding the complicated wireless interactions, and shed light on efficient design and operation of future wireless networks.

1.3 Thesis Organization

The remainder of this thesis is organized as follows:

Chapter 2 studies the performance of wireless networks under the multihop operation. The concept of transport capacity is defined, and its scaling behavior is analyzed under two models, namely, *protocol model* and *physical model*. The criterion for successful transmission in these models is based on collision and signal-to-noise-plus-interference ratio, which well models how current technology operates.

Chapter 3 is devoted to study the transport capacity from an information theoretic perspective. An interesting dichotomy is observed between low and high attenuation cases. It is shown that multihop is an order optimal strategy for some scenarios under the high attenuation regime. For the low attenuation regime, there are better scalings. Actually, we can achieve unbounded transport capacity in some networks for a limited total power. The strategy is coherent multistage relaying with interference subtraction (CRIS), where users profitably cooperate over long distance by using coherent and multiuser estimation instead of multihop.

In Chapter 4, we study the throughput of random networks, where traffic is uniformly distributed in the network. We consider two cases, dense networks and extended networks, and present upperbounds on the throughput scaling in each case. In particular, we are interested to discover whether a linear scaling is possible.

Chapter 5 develops a hierarchical cooperation scheme to achieve the scaling results of Chapter 4. The scheme is analyzed and optimized by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. In addition, to apply the hierarchical cooperation scheme to random networks, a clustering algorithm is developed, which divides the whole network into quadrilateral clusters, each with exactly the number of nodes required.

Finally, a summary of the thesis and some future research directions are presented in Chapter 6.

Chapter 2

The Capacity of Multihop Wireless Networks

2.1 Transport Capacity of Arbitrary Wireless Networks

This section considers arbitrary networks operating under a multi-hop mode of information transfer, where the locations of the nodes, the choice of source-destination pairs, rates along each hop, transmission time slots, and routing, can be jointly optimized. How will the transport capacity scale with the network size, the number of nodes in the network? This is the question we try to answer in this section. We present two models for successful transmissions: *protocol model* and *physical model*. Given the protocol model, assuming that the nodes are located in a region of area A, we show that the transport capacity cannot grow faster than $O(\sqrt{An})$[1] for a network of n nodes. On the other hand, networks with grid distribution and neighbor-only transmissions are shown to achieve $\Omega(\sqrt{An})$ bit-meters per second. This shows transport capacity scales as $\Theta(\sqrt{An})$ for arbitrary networks under protocol model. To understand the importance of this result, suppose all nodes share equally in this transport capacity. Then each obtains only $\Theta(\frac{\sqrt{A}}{\sqrt{n}})$ bit-meters/second, when the area is held fixed. To interpret this result note that if each node communicates to a distant node at a distance of $\Omega(\sqrt{A})$ meters away, then it can only obtain a rate of $O(\frac{1}{\sqrt{n}})$. On the other hand, if each node only wishes to communicate with its

[1] Given two functions f and g, we say f=$O(g)$ if $\sup_n |f(n)/g(n)| < \infty$. We say $f = \Omega(g)$ if $g = O(f)$. If both f=$O(g)$ and $f = \Omega(g)$, then we say that $f = \Theta(g)$.

nearest neighbor that is to a distance of $O(\frac{\sqrt{A}}{\sqrt{n}})$ meters away, then it can do so at a non-vanishing rate.

Similar results also hold under an alternate physical model, where a required signal-to-noise-plus-interference ratio (SINR) is specified for successful reception.

2.1.1 The protocol model

Consider the following model for an *arbitrary network*. A set of n nodes $\mathcal{N} = \{1, 2, ..., n\}$ is arbitrarily located in a disk of area A on the plane[2]. The scenario when nodes locations are random will be discussed later. Let X_i denote the location of node i and $r_{ij} = |X_i - X_j|$ be the Euclidian distance between node i and node j, for all $i, j \in \mathcal{N}$. We assume that each node can transmit at W bits per second over a common wireless channel shared by all nodes[3]. It will be shown that it will not change the ensuring capacity results if the channel is broken up into several sub-channels of capacities $W_1, W_2, \cdots, W_M$ such that $W = W_1 + \cdots + W_M$. Note that the choice of the sequence of nodes along which a packet is sent from its origin to its final destination is the routing problem. Suppose node i transmits over the m-th sub-channel to the node j. Then this transmission at rate W_m bits/sec is assumed to be successfully received by node j if

$$r_{kj} \geq (1 + \Delta) r_{ij} \tag{2.1}$$

for every other node k simultaneously transmitting over the same sub-channel. See figure (2.1); The quantity $\Delta > 0$, or more properly a circle of radius $(1 + \Delta)r_{ij}$ quantifies a guard zone required around the receiver to ensure that there is no destructive interference from neighboring nodes transmitting on the same m-th sub-channel at the same time.

Definition 2.1.1. *Transport capacity of a network is defined as*

$$C_T = \sup_{\{all\ feasible\ R_{ij}\}} \sum_{\substack{i,j \in \mathcal{N} \\ i \neq j}} R_{ij} r_{ij} \tag{2.2}$$

where R_{ij} is the rate of reliable communication from the source node i to the destination node j, for all $i, j \in \mathcal{N}$.

[2]Choosing a disk does not degrade the generality. The results hold true for square or, in fact, any domain which is the closure of its interior

[3]Note that the choice of the transmission range is achieved by power control

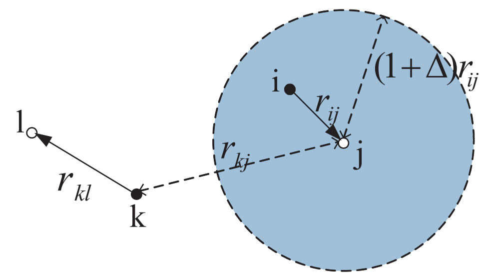

Figure 2.1: Successful transmission from i to j according to the protocol model

Theorem 2.1.1. *The transport capacity of an arbitrary network of* n *nodes in area* A *under the protocol model is* $\Theta(W\sqrt{An})$ *bit-meters/sec.*

This is achievable when the locations of the nodes and the source-destination pairs are chosen optimally, and the network is optimally operated. By the phrase "optimally operated" we mean optimized over the choice of a route, or a multiple set of routes to be used for each source-destination pair, as well as optimal timing of all transmissions.

Specifically, the upper bound is $\sqrt{\frac{8}{\pi}\frac{W}{\Delta}}\sqrt{An}$ bit-meters/sec for every arbitrary network for all scheduling strategies, while $\frac{W\sqrt{A}}{1+2\Delta}\frac{n}{\sqrt{n}+\sqrt{8\pi}}$ bit-meters/sec is an achievable lower bound, when the node locations and the transmissions are chosen appropriately.

Proof of Theorem 2.1.1. Here, we only mention the main ideas of the proof. The rigorous proof can be found in the references mentioned in the section "Note" at the end of this chapter.

The essential idea to upper-bound the transport capacity is that successful transmissions consume an area as they happen. The radius of such a consumed area is proportional to the transmission range. To observe this, let $\mathcal{T}_m(t)$ denote the set of all concurrent transmissions over the m-th subchannel at time slot t. Consider figure (2.1). Assume both transmissions, from i to j and from k to l, are successful. Then, by the triangle inequality and (2.1), we have

$$r_{jl} \geq r_{jk} - r_{kl} \tag{2.3}$$

$$\geq (1+\Delta)r_{ij} - r_{kl} \tag{2.4}$$

and similarly

$$r_{jl} \geq (1+\Delta)r_{kl} - r_{ij} \tag{2.5}$$

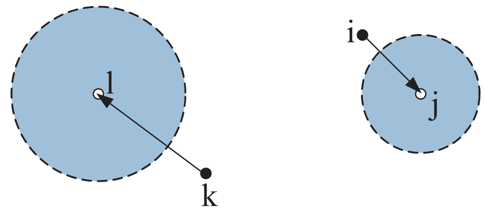

Figure 2.2: The exclusion disks around receivers are disjoint.

Adding the two inequalities yields

$$r_{jl} \geq \frac{\Delta}{2}(r_{kl} + r_{ij}) \tag{2.6}$$

This is equivalent to say that two disks, one of radius $\frac{\Delta}{2}r_{ij}$ catered at i and the second one of radius $\frac{\Delta}{2}r_{kl}$ centered at l, are disjoint. In other words, *exclusion disks* of radius $\frac{\Delta}{2}$ times the length of hops centered at the receivers, over the same subchannel in the same slot, are disjoint; as shown in figure (2.2). Obviously, the sum of such areas is upper-bounded by the limited total area A. Considering the edge effect where a node is near the boundary of the domain, and noting that a transmission range greater than the diameter of the domain is unnecessary, we see that at least a quarter of each exclusion disk is inside the domain. Hence,

$$\sum_{(i,j)\in\mathcal{T}_m(t)} \frac{1}{4}\pi(\frac{\Delta}{2}r_{ij})^2 \leq A \tag{2.7}$$

Since at most half of the nodes can transmit at each time, $|\mathcal{T}_m(t)| \leq n/2$ and furthermore, by (2.7), we have

$$\sum_{(i,j)\in\mathcal{T}_m(t)} r_{ij} \leq \sqrt{\frac{n}{2}\sum_{(i,j)\in\mathcal{T}_m(t)} r_{ij}^2} \tag{2.8}$$

$$\leq \sqrt{\frac{8An}{\pi\Delta^2}} \tag{2.9}$$

Hence, for any time t, the total bit-meters/sec for subchannel m is given by

$$W_m \sum_{(i,j)\in\mathcal{T}_m(t)} r_{ij} \leq W_m\sqrt{\frac{8An}{\pi\Delta^2}} \tag{2.10}$$

Summing over all subchannels, we get the upperbound.

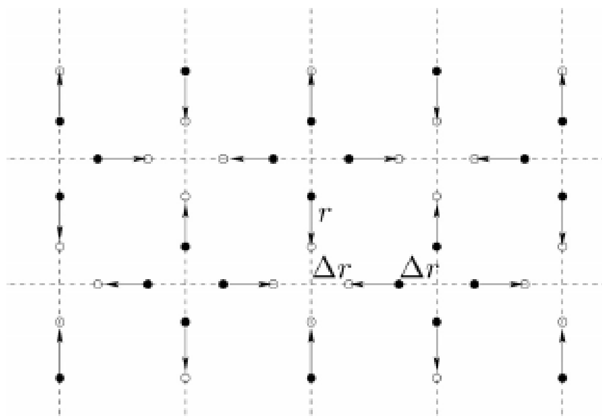

Figure 2.3: The grid-like node arrangement for the lowerbound.

To show the achievability of $\Theta(\sqrt{An})$ bit-meters/second, we can first arrange the n nodes in grid-like positions, as shown in figure (2.3), then choose $n/2$ nodes as senders with each of them transmitting only to one of its nearest neighbors. Note that all the transmitters can transmit simultaneously according to the protocol model. Therefore, $\frac{n}{2}Wr$ bit-meters/second is achievable. To make sure that there are exactly $n/2$ transmitter nodes inside the disk of area A, by a simple calculation, we choose $r = \frac{1}{1+2\Delta}\frac{\sqrt{A}}{\sqrt{n}+\sqrt{8\pi}}$. It is worth pointing out that, to achieve the upperbound, there is no need to divide the original bandwidth to several sub-channels. □

2.1.2 The physical model

In the previous section, we considered a geometric model for successful reception. In this section, we analyze a more realistic model. In the *physical model*, the criterion for the successful reception is that the SINR at the receiver must be greater than a threshold. The threshold must be large enough to enable the receiver to decode a packet, which is a better modeling of current radio technology decoding.

Let $\mathcal{T}(t)$ be the subset of nodes simultaneously transmitting at some time instant t. Let P_k be the power level chosen by node k, for $k \in \mathcal{T}(t)$. Then the transmission between nodes $i, j \in \mathcal{T}(t)$ is successful if

$$\frac{\frac{P_i}{r_{ij}^{\alpha}}}{N + \sum_{\substack{k\in\mathcal{T}(t)\\ k\neq i}} \frac{P_k}{r_{kj}^{\alpha}}} \geq \beta \tag{2.11}$$

where N is the noise power level. The signal power is assumed to decay with distance r as $1/r^{\alpha}$. The constant α is called path loss exponent for power. We assume $\alpha \geq 2$. If the required SINR is satisfied, we assume a data rate of W bits/sec over the link.

Theorem 2.1.2. *For an arbitrary network of unit area*[4], $\Omega(W\sqrt{n})$ *bit-meters per second is feasible under the physical model, when the network appropriately designed, while an upper bound is* $O(Wn^{\frac{\alpha-1}{\alpha}})$.

It is worth noting that there is a gap between these two bounds.

Proof of Theorem 2.1.2. The main idea to obtain the upper bound is using (2.11) to find an upper bound for r_{ij}^{α}. By using (2.11), and the fact that $r_{kj} \leq 2/\sqrt{\pi}$, we get

$$r_{ij}^{\alpha} \leq \frac{\beta+1}{\beta} \frac{P_i}{N + (\pi/4)^{\alpha/2} \sum_{k \in \mathcal{T}(t)} P_k} \tag{2.12}$$

and consequently

$$\sum_{i \in \mathcal{T}(t)} r_{ij}^{\alpha} \leq \frac{\beta+1}{\beta} \frac{\sum_{i \in \mathcal{T}(t)} P_i}{N + (\pi/4)^{\alpha/2} \sum_{k \in \mathcal{T}(t)} P_k} \tag{2.13}$$

$$\leq \frac{\beta+1}{\beta} (\frac{4}{\pi})^{\alpha/2} \tag{2.14}$$

Then the rest of proof proceeds similar to the proof of upperbound for the protocol model, using (2.13) instead of (2.10), and invoking the convexity of r^{α} instead of r^2 in (2.8). The exact upperbound is given by

$$\frac{1}{\sqrt{\pi}} \left(\frac{2\beta+2}{\beta} \right)^{\frac{1}{\alpha}} Wn^{\frac{\alpha-1}{\alpha}}$$

For proving the feasibility, we find a correspondence between the protocol model and the physical model such that for every β there exists a Δ determining the total number of simultaneous transmissions for each time instant and sub-channel. In fact, a simple calculation for the SINR of the grid-like node arrangement of figure (2.3) shows it is lowerbounded by $\frac{(1+2\Delta)^{\alpha}}{16(2^{\alpha/2}+\frac{6\alpha-2}{\alpha-2})}$. So, it is sufficient to choose Δ to make this lowerbound equal to β. Then there exists a power assignment allowing

[4]The results can be easily extended to any area A, by multiplying by a factor of $\sqrt{A}$

the same set of transmissions under the physical model with threshold β. More specifically,

$$\frac{1}{(16(2^{\alpha/2}+\frac{6\alpha-2}{\alpha-2}))^{1/\alpha}}\frac{Wn}{\sqrt{n}+\sqrt{8\pi}}$$

bit-meters/sec is achievable. □

Remark 2.1.1. *As compared to $\Theta(W\sqrt{n})$ of the protocol model, the physical model yields a larger bound for $\alpha > 2$. In a special case where the ratio between the maximum and the minimum powers that transmitters can employ satisfies $\frac{P_{max}}{P_{min}} < \beta$, the upperbound is in fact*

$$\sqrt{\frac{8}{\pi}}\frac{1}{(\frac{\beta P_{min}}{P_{max}})^{\frac{1}{\alpha}}-1}W\sqrt{n}$$

bit-meters/second.

2.1.3 The generalized physical model

The physical model assumes that a transmission can only occur at two rates: W bits/sec if SINR exceeds β, and 0 bits/sec otherwise. This model can be generalized to be continuous in SINR, based on Shannon's capacity formula for AWGN channel. In this case, the data rate from node i to its receiver j is assumed to be

$$W_i = H_m \log\left(1+\frac{\frac{P_i}{r_{ij}^\alpha}}{NH_m+\sum_{\substack{k\in\mathcal{T}(t)\\k\neq i}}\frac{P_k}{r_{kj}^\alpha}}\right) \tag{2.15}$$

where H_m is the bandwidth of channel m in hertz, such that the total bandwidth is finite, $\sum_m H_m \le H_0$, and $N/2$ is the noise spectral density in watts/hertz.

Theorem 2.1.3. *For the Generalized Physical Model with $N > 0$ and $\alpha > 2$ and available total bandwidth H_0, if the maximum power that a node can employ on the m^{th} sub-channel is $P_{max} = H_m N n^{\alpha/2}$, then the transport capacity of an n node network located in a unit square is upper bounded by $O(H_0\sqrt{n})$. The lower bound for feasibility is $\Omega(\sqrt{n})$, so $\Theta(\sqrt{n})$ is indeed the scaling law for the transport capacity.*

Proof of Theorem 2.1.3. The feasibility is similar to physical model because any SINR larger than a threshold enables a constant rate between two nodes. To prove the upper bound, one needs to carefully examine the interference experienced by each receiver. □

Remark 2.1.2. *For a square of area A, by shrinking it to a unit area and scaling the powers to $P_i A^{-\alpha/2}$, it follows that for a power constraint of $P_{max} = H_m N(nA)^{\alpha/2}$ the transport capacity scales as $\Theta(\sqrt{An})$. Hence, this closes the gap between the lower bound and the upper bound in the physical model.*

Remark 2.1.3. *Note that the above attenuation model is based on far-field assumption [29]. Let r_f denote the far-field distance of a transmitter antenna. r_f is defined as*

$$r_f = 2D^2/\lambda_c$$

where D is the largest physical linear dimension of the antenna and λ_c is the carrier wavelength. Moreover, r_f should satisfy

$$r_f \gg D \text{ and } r_f \gg \lambda_c$$

which imposes the following constraint on the minimum separation distance r_{min} between nodes.

$$r_{min} \geq r_f \tag{2.16}$$

When increasing the number of nodes in a fixed area A, the physical model does not seem reasonable, since the model fails to satisfy the far-field requirement after some point. Therefore, nodes have to be separated by a positive distance. This is the model we consider in Chapter 3, and as we see, it yields different results.

2.2 Throughput of Random Wireless Networks

The locations of source-destination pairs are not often known *a priori*; so, one is interested in how random settign will influence the performance. In this section, we consider randomly distributed networks. For simplicity, we assume that each source has one randomly chosen destination. The results are extendable to a more general case, where a source node can generate traffic for more several destination nodes.

Consider the following model of random networks. n nodes are uniformly and independently distributed in a unit square. Each node has a random destination. The destination is chosen as follows. A position is first picked uniformly from within the unit square, then the node nearest to it is chosen as the destination. We present the results on the average achievable throughput of each node under the protocol model and the physical model.

2.2.1 Throughput under the protocol model

The model is essentially the same as the protocol model considered for the arbitrary network. For the random network, we impose another constraint on the range of transmissions. We assume that all nodes employ a common transmission range r. Therefore, node i can successfully transmit to node j if

(i) The distance between the transmitter and the receiver is no more than r, i.e., $r_{ij} \leq r$.

(ii) For every other node k, $k \neq i$, transmitting at the same time, $r_{kj} \geq (1+\Delta)r$.

(iii) The data rate for such successful transmitter-receiver pair is W bits/second.

Theorem 2.2.1. *The order of the per-node throughput of random wireless networks, $\lambda(n)$, under the protocol model is $\Theta(\frac{W}{\sqrt{n \log n}})$ bits/sec.*

Proof of Theorem 2.2.1. First, let us focus on the achievability part. The key ideas are the following:

1. **A constructive lower bound**

   We need to present a scheme that achieves $\lambda(n) = \Theta(\frac{W}{\sqrt{n \log n}})$.

   a) Tessellating the units quare by small squares

   We tessellate the unit square by square cells of side $s_n = \sqrt{\frac{K \log n}{n}}$. It is easy to prove, using the binomial distribution and Chernoff bound, that for any $K \geq 1$, with probability going to one, each cell holds at least one but no more than $Ke \log n$ nodes[5].

   b) Transmission schedule

   Index the cells as S_{ij}, with i denoting the column number and j the row number. For a positive integer M, let $C(k_1, k_2)$, for $0 \leq k_1, k_2 \leq M-1$, be the group of all cells $\{S_{ij} : i \bmod M = k_1, j \bmod M = k_2\}$. All nodes choose a common transmission range $r_n = 2\sqrt{2}s_n$, so that every node can cover all its neighboring cells. If M is large enough, then all cells in one of the M^2 groups can transmit simultaneously in a time slot.

   c) The routing

[5] A general form of clustering of random networks, based on Vapnik-Chervonekis theorem, will be presented in Chapter 5

Each node i, located in X_i, for $1 \leq i \leq n$, generates data packets at rate $\lambda(n)$ with an end destination chosen as the node nearest to a randomly chosen location Y_i. Denote by $X_{dest}(i)$ the node nearest to Y_i, and by L_i the straight-line segment connecting X_i and Y_i. The packets generated by node i will be forwarded toward $X_{dest}(i)$ in a multi-hop manner, from cell to cell in the order that they are intersected by L_i. Any node in the cell can be chosen as a receiver. Then there exists a constant $c \geq 0$ such that

$$Prob\left(\sup_{(k,j)}\{\text{Number of lines } L_i \text{ intersecting } S_{kj}\} \leq c\sqrt{n\log n}\right) \rightarrow 1.$$

This gives a bound on the maximum amount of traffic which must be relayed by each cell.

d) Lower bound on per-node throughput

In every M^2 slots, each cell gets one slot to transmit at rate W bits/sec. So the rate of each call is $\frac{W}{M^2}$ bits/sec. Each cell needs to handle the traffic assigned to it and according to the routing, this traffic is less than $\lambda(n)c\sqrt{n\log n}$. This can be therefore accommodated by all cells if $\lambda(n)c\sqrt{n\log n} \leq \frac{W}{M^2}$ and consequently, any $\lambda(n) \leq (\frac{c'W}{\sqrt{n\log n}})$ is achievable.

2. **Upper bound**

The proof is analogous to the proof of the upperbound for arbitrary networks. Exclusion disks of radius $\frac{\Delta}{2}r$ around every receiver are disjoint. So, we can find the maximum total number of simultaneous transmissions, that is $\frac{16}{\Delta^2 r^2}$. Let $\overline{L}$ be the expected distance between source and destination. Then each packet needs $\frac{\overline{L}}{r}$ hops to reach its destination, this means the bits/sec being transmitted by the network is at least $n\lambda(n)\frac{\overline{L}}{r_n}$ and this value must be less than the maximum number of simultaneous transmissions times W. Using this inequality, and choosing $r = \sqrt{\frac{\log n}{\pi n}}$ to guarantee the connectivity of the network [15] yield the result.

□

As observed in the above argument for the upperbound, an interesting tradeoff is involved for random networks: The desire to reduce the multihop burden is in contradiction with the desire to increase the number of concurrent transmissions and spacial reuse. In other words, by increasing the transmission range of each node, packets can reach the destinations by less number of hops but this, in turn, results in a decrease in the total number of simultaneous transmissions. Considering the both issues, we found out that we need to decrease r as much as possible. But this small transmission range must guarantee the connectivity of the network. Hence, among all the multihop schemes, the nearest-neighbor multihop has the best performance. The critical transmission range for a random network on a unit area is $\sqrt{\frac{\log n}{\pi n}}$ [15]. This critical value is the transmission range we chose for the constructive lower bound. As a result of this argument, we mention the following remark:

Remark 2.2.1. *To achieve a better scaling, one must be able to perform many simultaneous long-range communications. A technique which achieves this is MIMO (Multi-Input Multi-Output). In this way, mutually interfering signals between source-destination pairs can be tuned to useful ones to realize spatial multiplexing gain. This will be the topic of Chapter 5.*

2.2.2 Throughput under the physical model

The model is similar to the physical model of the arbitrary networks, with an additional constraint that all nodes employ a common transmission power P. The required SINR must satisfy (2.11).

Theorem 2.2.2. *There exist positive constants c and c' such that a per-node throughput of $\lambda(n) = \frac{cW}{\sqrt{n \log n}}$ is feasible, while $\lambda(n) = \frac{c'W}{\sqrt{n}}$ is not, both with probability approaching one as $n \to \infty$.*

Proof of Theorem 2.2.2. The proof is very similar to the proof under the protocol model. For feasibility we can use the constructive lower bound of the protocol model and adjust M to make SINR$\geq \beta$. For upper bound, we find a correspondence with the protocol model. Assume i is successfully transmitting to j, and k is to l. From (2.11), we have

$$\frac{P/r_{ij}^{\alpha}}{P/r_{kl}^{\alpha}} \geq \beta$$

which is equivalent to

$$r_{kl} \geq (1 + \Delta) r_{ij}$$

for $\Delta = (\beta^{\frac{1}{\alpha}} - 1)$. Thus, the upper bound on the transport capacity of arbitrary networks under the protocol model, Theorem 2.1.1, also holds for our physical model. Using this transport capacity, and noting that the average distance between a source node and its destination is a constant $\overline{L}$, prove the stated result. □

Remark 2.2.2. *As observed for arbitrary or random networks, the results are relatively robust to the choice of geometric model or more physically realistic model. In the next chapters, we will see a similar correspondence with the other modeling approaches as well.*

2.3 Notes

This section was mainly based on [3], [12] and [13]. The transport capacity was first studied in [3] where there was a gap between the lower bound and the upper bound on the transport capacity of arbitrary networks under the physical model. This gap was closed in [12] by using the generalized physical model. For the throughput of random networks under the physical model, as we saw, the constructive lower bound shows $\Theta(\frac{W}{\sqrt{n \log n}})$ is feasible, while $\Theta(\frac{W}{\sqrt{n}})$ is an upperbound. This gap was later closed in [10], by showing a constructive scheme to achieve $\Theta(\frac{W}{\sqrt{n}})$. The scheme is using techniques from percolation theory. In addition, nodes can use different transmission ranges instead of the common range used here. As a result, the factor $\sqrt{\log n}$ present in the earlier result can be vanished, however at the expense of a more complicated architecture.

Chapter 3

An Information Theory for Transport Capacity

In this chapter, we try to shed some information theoretical light on the appropriate architecture for two separate cases, namely high attenuation and low attenuation. We consider a model where nodes are located on a plane with a minimum distance separation between them, with each node having an individual power constraint or a total power constraint on all nodes. We model the signal attenuation as $\frac{\sqrt{G}e^{-\gamma r}}{r^{\alpha/2}}$ where α is the path loss exponent for power and γ is the absorption coefficient. For transmission in vacuum absorption is zero. G is another constant depending on the transmitter and receiver antenna gains and the carrier wavelength.

The main results are as follows:

i) The transport capacity grows like $O(n)$ when $\gamma \geq 0$ or $\alpha \geq 6$. This is established by showing that transport capacity is upper bounded by the multiple of the total transmission powers of all the nodes. As the result, multi-hop is an order-optimal strategy under some scenarios when $\gamma \geq 0$ or $\alpha \geq 6$ [1] . For example, if the nodes are located at integer lattice sites in a square and randomly choose their destinations, multihop transport is differing at most by a factor $\frac{1}{\sqrt{\log n}}$ from optimal order. If traffic can be load-balanced across the network by multipath routing with bounded distance at each hop, then the scaling law is sharp, i.e., the transport capacity order is $\Theta(n)$, and is achieved by multihop transport. In overall, this provide justification for using multihop when the traffic is balanced or can be balanced by using multipath routing if necessary.

[1] Note that the linear scaling of transport capacity, in this case, agrees with the square-root scaling $O(\sqrt{An})$ in Chapter 2. This is because the area A scales as $\Theta(n)$ due to the minimum seperation distance constraint.

ii) There is a dichotomy between the low and high attenuation cases. When $\gamma = 0$ and $\alpha \leq 3$, we can achieve unbounded transport capacity in some networks for a limited total power. In particular, for nodes on a line when $\alpha \leq 2$, there are networks where their transport capacity scales super linearly like $\Theta(n^{\theta})$ for $1 \leq \theta \leq 2$. The strategy is coherent multistage relaying with interference subtraction (CRIS) where the nodes profitably cooperate over long distance by using coherent and multiuser estimation instead of multihop.

3.1 Model of Wireless Networks

A set of n nodes is located on a plane. There is a minimum distance r_{min} between nodes. At time instants $t = 1, 2, ...$ node i sends $X_i(t)$ and receives $Y_i(t)$ such that

$$Y_i(t) = \sum_{j \neq i} \frac{\sqrt{G} e^{-\gamma r_{ij}} X_j(t)}{r_{ij}^{\alpha/2}} + Z_i(t)$$

where $Z_i(t)$, for $0 \leq i \leq n$ and $t = 1, 2, ...$, are i.i.d random variables with zero mean and variance σ^2. We consider constraint on the total power of the nodes (P_{total}) or individual power constraint on each node (P_{ind}). The network can have several source-destination pairs $(s_l, d_l), l = 1, ..., m$. A regular planer network is a special case of planer networks where the nodes are located at the points $(i, j), 0 \leq i, j \leq \sqrt{n}$. In a linear network all the nodes are on a line, and in regular linear network the nodes are located on the points $1, 2, ..., n$.

3.2 The High Attenuation Regime

3.2.1 Main results under high attenuation

For the case of total power constraint, we have the following bound on the transport capacity.

Theorem 3.2.1. *If $\gamma \geq 0$ or $\alpha \geq 6$, then for every planer network we have*

$$C_T(n) \leq \frac{c_1(\gamma, \alpha, r_{min})}{\sigma^2} P_{total}$$

where $c_1(\gamma, \alpha, r_{min})$ is a constant.

The exact value of c_1 can be found in the references mentioned in the section "Note" at the end of this chapter.

It follows immediately from this theorem that under the individual power constraint, transport capacity can not grow faster than linear in n. A corollary of the above theorem is as follows.

Corollary 3.2.1. *If $\gamma \geq 0$ or $\alpha \geq 6$, then for any planer network*

$$C_T(n) \leq \frac{c_1(\gamma, \alpha, r_{min})}{\sigma^2} \ P_{ind} \ n$$

On the other hand, according to the following theorem, the linear growth is indeed achievable for regular networks.

Theorem 3.2.2. *Suppose $\gamma \geq 0$ or $\alpha \geq 2$, and each node is subject to an individual power constraint P_{ind}. A regular planer network of n nodes can achieve*

$$C_T(n) \geq S(\frac{e^{-2\gamma} P_{ind}}{c_3(\gamma, \alpha/2) P_{ind}} + \sigma^2) \ n$$

where $S(x)$ is the Shannon function $\frac{1}{2}\log(1+x)$.

Again, we do not mention the exact value of c_3 here. The same results, as above theorems, can be derived for linear networks as well, when dichotomy occurs for $\alpha = 4$, and c_3 is replaced by another constant c_2.

3.2.2 Main ideas behind the proof under high attenuation

Proof of Theorem 3.2.1. The proof is based on a *Max-flow min-cut lemma* relating rates with received power. The lemma is similar in spirit to Theorem 14.10.1 in [27].

Lemma 3.2.1. *Let $\mathcal{N}_1$ be any subset of all the nodes $\mathcal{N}$. If $(R_1, ..., R_m)$ is a feasible rate vector, then*

$$\sum_{\left\{l: d_l \in \mathcal{N}_1, s_l \in \mathcal{N}_1^c\right\}} R_l \leq \frac{\log e}{\sigma^2} \liminf_{T \to \infty} P_{\mathcal{N}_1}^{rec}(T)$$

where $P_{\mathcal{N}_1}^{rec}(T)$ is the average power received by $\mathcal{N}_1$, from outside of $\mathcal{N}_1$, $\mathcal{N}_1^c$, i.e.,

$$P_{\mathcal{N}_1}^{rec}(T) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in \mathcal{N}_1} E\left(\sum_{j \in \mathcal{N}_1^c} \frac{e^{-\gamma r_{ij}} X_j(t)}{r_{ij}^{\alpha/2}}\right)^2$$

For simplicity, let us focus on linear networks. The idea is the same for planer networks and the proof of their upperbound proceeds similarly. Let $a_i r_{min}$ denote the coordinate of the node i. Let R_S denote the sum of the rates with source nodes in set S and destinations in the rest of the network. For every integer q, define the following subsets.

$$\mathcal{N}_q^- = \{i \in \mathcal{N} : a_i \leq q\}$$

and

$$\mathcal{N}_q^+ = \{i \in \mathcal{N} : a_i > q\}$$

Then by using Lemma 3.2.1, we can bound $R_{\mathcal{N}_q^-}$ and $R_{\mathcal{N}_q^+}$. The key idea in the proof is that every source-destination pair (s_l, d_l) with distance r_l cuts at least $\lfloor \frac{r_l}{r_{min}} \rfloor$ subsets among all $\mathcal{N}_q^-$s (d_l on the right side of s_l) or $\mathcal{N}_q^+$s (d_l on the left side of s_l). Hence, we can upperbound the transport capacity as

$$\begin{aligned} \sum_{i,j \in \mathcal{N}} R_{ij} d_{ij} &\leq 2r_{min} \sum_{i,j \in \mathcal{N}} R_{ij} \lfloor \frac{d_{ij}}{r_{min}} \rfloor \\ &\leq 2r_{min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-} + 2r_{min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+} \end{aligned} \tag{3.1}$$

So, it is sufficient to calculate $\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-}$ or $\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+}$ (they are equal) where $R_{\mathcal{N}_q^-}$ and $R_{\mathcal{N}_q^+}$ are given by Lemma 3.2.1. The series will converge for $\alpha > 4$ or $\gamma > 0$. This concludes the proof and yields the stated result. □

Proof of Theorem 3.2.2. Consider a regular planer network with n nodes such that two neighboring nodes are one meter apart from each other, and each source chooses one of its 4 nearest neighbors as its destination. Let us only focus on the case where $\gamma = 0$ and $\alpha > 2$. Each node generates its codebook according to a Gaussian distribution with variance P_{ind}. After the block of T transmissions, each destination decodes its intended message treating all the other transmissions as noise (This is, in fact, the grid-like node arrangement of Figure 2.3 that we used to show the achievablity results for the multihop strategy). It is easy to verify that the interference due to other transmissions can be bounded by $c_3(\gamma, \alpha) P_{ind}$, and consequently the following rate is achievable for each source-destination pair:

$$R < \frac{1}{2} \log \left(1 + \frac{P_{ind}}{c_3(\gamma, \alpha) P_{ind} + \sigma^2} \right). \tag{3.2}$$

Therefore, a linear growth for transport capacity is achievable, which concludes the proof. □

3.3 Low Attenuation Regime

We continue with the scaling behavior of transport capacity when there is no absorbtion and the attenuation is low. In this case, by *coherent relaying with interference subtraction* (CRIS) strategy, super linear growth in transport capacity can be achieved for networks under the individual power constraint. The ratio of transport capacity to total power can also be unbounded. These results reveal that there must be a fundamental relationship between the properties of the medium, the information transfer capacity and appropriate architecture for different attenuations. At the present, however, the results presented are under unrealistically low attenuations and further research is required to bridge the gap with high attenuation. Next, we present the main features of CRIS scheme.

3.3.1 The Gaussian multiple-relay channel and CRIS strategy

The basic idea of this new coding scheme is the same as the *block markov coding* scheme for simple relay channel of figure (3.1).

Let s, r, d respectively denote source, relay and destination nodes, and $\alpha_{sr}, \alpha_{sd}, \alpha_{rd}$ denote the signal attenuation factors. The transmission time is divided into a sequence of blocks. In each block, the source divides its power into two parts: θP_s for informing the relay from its intention for the next block and $(1-\theta)P_s$ for coherent cooperation with the relay to transmit to d. So, we can achieve any rate $R \leq S(\frac{\alpha_{sr}^2 \theta P_s}{\sigma^2})$ for informing the relay. The received signal of the destination has three parts:

1. The signal consisting of coherent transmission of r and s with power
$$(\alpha_{sd}\sqrt{(1-\theta)P_s} + \alpha_{rd}\sqrt{P_r})^2.$$
2. The signal with power $\alpha_{sd}^2 \theta P_s$.
3. The noise.

For decoding, node d treats the second part in the previous block and the first part in the current block as signal since they represent the same information. Note that the destination can recover the second part in the previous block, since it has already decoded the previous block, and can deducts the first part from the total

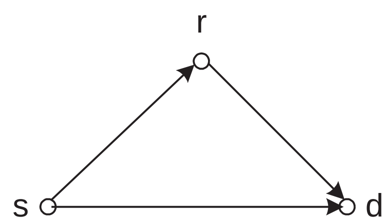

Figure 3.1: The single relay channel

received signal. Hence, the destination can do decoding in the present block with rate

$$\begin{aligned} R &< S\left(\frac{(\alpha_{sd}\sqrt{(1-\theta)P_s}+\alpha_{rd}\sqrt{P_r})^2}{\alpha_{sd}^2\theta P_s+\sigma^2}\right)+S\left(\frac{\alpha_{sd}^2\theta P}{\sigma^2}\right) \\ &= S\left(\frac{\alpha_{sd}^2 P_s+\alpha_{rd}^2 P_r+2\alpha_{sd}\alpha_{sd}\sqrt{(1-\theta)P_sP_r}}{\sigma^2}\right). \end{aligned} \tag{3.3}$$

So, the achievable rate, for reliable decoding at both the relay and the destination, is

$$R < \max_{0\le\theta\le1}\min\left\{S(\frac{\alpha_{sr}^2\theta P_s}{\sigma^2}) \ , \ S(\frac{\alpha_{sd}^2 P_s+\alpha_{rd}^2 P_r+2\alpha_{sd}\alpha_{sd}\sqrt{(1-\theta)P_sP_r}}{\sigma^2})\right\}. \tag{3.4}$$

The above method can be extended as CRIS for multiple-relay. Consider one source, $M-1$ relay nodes, and a destination. Index them respectively by $0, 1, ..., M$. At the beginning of each block b, every relay node i has estimates of the previous source messages up to i blocks before the present block, i.e., blocks $b-M-i, \cdots$, up to $b-i$. Each node transmits coherently to all its downstream nodes (with larger node number). The allocated power of node i to transmit to node k is $P_{ik}, k \ge i+1$. Then any rate satisfying the following inequality is achievable:

$$R < \min_{1\le j\le M} S(\frac{1}{\sigma^2}\sum_{k=1}^{j}(\sum_{i=0}^{k-1}\alpha_{ij}\sqrt{P_{ik}})^2) \tag{3.5}$$

where P_{ik} satisfies $\sum_{k=i+1}^{M} P_{ik} \le P_i$.

3.3.2 Main results under low attenuation

Theorem 3.3.1. *For a regular planer network*

(i) If $\gamma = 0$ and $\alpha < 3$, then even with a fixed total power, we can achieve unbounded transport capacity when n grows.

(ii) If $\gamma = 0$ and $\alpha < 2$, then even with a fixed total power, we can achieve a fixed minimum rate between every source-destination pair, irrespective of their distance.

We have a similar theorem for regular linear networks, for $\alpha < 2$ in the first part and $\alpha < 1$ in the second part of Theorem 3.3.1.

Theorem 3.3.2. *If $\gamma = 0$ and $1 < \alpha < 2$, under an individual power constraint, a superlinear $\Theta(n^\theta)$ scaling law with $1 < \theta < 2/\alpha$ is feasible for some linear networks.*

3.3.3 Main ideas behind the proofs in low attenuation regime

In this subsection, we show how to use the CRIS strategy to achieve a superlinear growth for the transport capacity. The key steps in proving Theorem 3.3.1 are presented here. We refer the interested reader to [4] for complete proofs.

Proof of Theorem 3.3.1. For simplicity, consider a linear network. The proof for planer case proceeds similarly. We consider one source-destination pair where the source is located at 0 and the destination is located at n. So, there are $n-1$ nodes, located at $1, 2, ..., n-1$ acting as relay nodes. Then the following rate is achievable by CRIS:

$$R < \min_{1 \le j \le n} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}}{(j-i)^{\alpha/2}}\right)^2\right) \tag{3.6}$$

with the total power constraint

$$\sum_{k=1}^{n}\sum_{i=0}^{k-1} P_{ik} \le P_{total}. \tag{3.7}$$

Recall that P_{ik} is the part of the power used by node i for direct transmission to node k. Let

$$P_{ik} = \frac{P}{(k-i)^{\beta_1}k^{\beta_2}}, \quad 0 \le i < k \le n \tag{3.8}$$

and

$$P = \frac{(\beta_1 - 1)(\beta_2 - 1)}{\beta_1 \beta_2} P_{total}. \tag{3.9}$$

The above power assignment satisfies the total power constraint for $\beta_1, \beta_2 > 1$. For $3 - \beta_1 - \beta_2 > 0$, the following lower bound can be established:

$$\sum_{k=1}^{j} \left(\sum_{i=0}^{k-1} \frac{\sqrt{P_{ik}}}{(j-i)^{\alpha/2}} \right)^2 = \Omega(j^{3-\beta_1-\beta_2-\alpha}) \tag{3.10}$$

As the result, for $\alpha < 1$ and $3 - \beta_1 - \beta_2 - \alpha > 0$, (3.10) is lower bounded by a positive constant, say P^*, from zero, and consequently any rate $R < S(\frac{P^*}{\sigma^2})$ is achievable. Note that the transport capacity in this case is Rn. Therefore, an unbounded transport capacity is achievable as n grows.

For the case where $1 \leq \alpha < 2$, $3 - \beta_1 - \beta_2 - \alpha < 0$; hence, the minimum of (3.10) is attained at $j = n$. So, the transport capacity is lowerbounded by $\Omega(n^{4-\beta_1-\beta_2-\alpha})$. By choosing β_1 and β_2 such that $4 - \beta_1 - \beta_2 - \alpha > 0$, arbitrary large transport capacity is achievable with a fixed total power constraint. This concludes the proof. Similarly, an unbounded transport capacity can be obtained for a fixed total power constraint in planer networks. □

3.4 Notes

This section is mainly based on [4]. In [8], it has been shown that, for upper bounding the transport capacity by a multiple of the total power, $\alpha > 3$ suffices for linear networks and $\alpha > 5$ for planar networks. The effect of random phases of channels is also discussed in the same paper. The linear growth rate of transport capacity for an improved bound on α, and a generalized transport capacity are also given in [14].

Chapter 4

Upper Bounds on the Throughput of Random Networks

In the previous chapter, we presented some results on the transport capacity of wireless networks from an information theoretic point of view. The upperbounds were quite general and applicable to arbitrary wireless networks as well as some special classes such as regular networks and random networks. Since for the special cases, the topology of the network is known, one wishes to analyze the throughput of such networks. Some interesting questions can be raised here that whether a constant per-node throughput, irrespective of the number of nodes of the networks, is possible or how fast the rate tends to zero as the number of users gets large.

In this chapter, we characterize the fundamental scaling behavior of the throughput of random/regular wireless networks. We show how to derive performance bounds by using *multi-input multi-output (MIMO)* techniques and by analysis of random matrices. This will provide us with an impetus to build a hierarchical cooperation scheme, in chapter 5, to achieve the performance bounds.

4.1 Model of Random Wireless Networks

The model is similar to the standard additive white Gaussian noise channel model considered in Chapter 3.

1. There are a set of n nodes $\mathcal{N}$ located on a plane.

2. Each node uses a common average power P to transmit.

3. At any time t, each node i transmits the signal $X_i(t) \in \mathbb{C}$, and receives the signal $Y_i(t) \in \mathbb{C}$. The received signal depends on the transmitted signals of all the other nodes as

$$Y_i(t) = \sum_{k \neq i} H_{ik} X_k(t) + Z_i(t) \tag{4.1}$$

where $Z_i(t)$ is white circularly symmetric Gaussian noise of variance N, and the gain

$$H_{ik} = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}) \tag{4.2}$$

where r_{ik} is the distance between nodes i and k, and θ_{ik} is the phase uniformly distributed in $[0, 2\pi]$. The parameter $\alpha \geq 2$ is the power path-loss exponent and G is another constant depending on the transmitter and receiver antenna gains and the carrier wavelength.

In order to obtain the upperbound on the throughput, we make a series of optimistic assumptions: We assume that the nodes are distributed randomly (uniformly and independently) on a planer network with area A or they are located regularly on the grid sites inside the area A. We further assume that source-destination pairs are chosen i.i.d. Each node can be a source and also a destination for another source node.

4.2 The Sum-Rate Upperbound For Random Dense Networks

For dense networks, the network area is fixed and does not change with n. Let us assume a unit area for the network. The following Theorem gives an upperbound on the throughput of such networks.

Theorem 4.2.1. *The aggregate throughput of a dense network is bounded by*

$$T(n) \leq K_1 n \log n$$

with high probability[1] *for some constant $K_1 > 0$ independent of n.*

Proof of Theorem 4.2.1. Consider a source node s and its randomly selected destination d. The transmission rate from s to d, is upper bounded by the capacity

[1] i.e., probability goes to one as number of nodes grows.

of the single-input multiple-output (SIMO) channel between s and the rest of the network. This capacity (see [28]) is given by

$$R(n) \leq \log\left(1+\frac{P}{N}\sum_{i\neq s}|H_{is}|^2\right) \quad (4.3)$$

$$= \log\left(1+\frac{P}{N}\sum_{i\neq s}\frac{G}{r_{is}^{\alpha}}\right) \quad (4.4)$$

It is easy to verify that in a dense random network, the minimum distance between any two nodes is larger than $\frac{1}{n^{1+\delta}}$ with high probability, for any $\delta > 0$. Hence, by using this fact, we obtain

$$R(n) \leq \log\left(1+\frac{GP}{N}n^{\alpha(1+\delta)+1}\right) \quad (4.5)$$

$$\leq K_1 \log n \quad (4.6)$$

for some constant K_1 independent of n. Since there are n such source-destination pairs, $T(n) = nR(n)$, and the theorem follows. □

So, the upperbound does not deprive us from the hope for a linear throughput scaling. However, to achieve this scaling, the removal of interference between simultaneous transmissions from different sources is necessary.

4.3 The Sum-Rate Upperbound for Random Extended Networks

Another natural scaling is the extended case, where the density of nodes is fixed and the area increases linearly with the number of nodes. Let us model the geographical area by a $\sqrt{n}\times\sqrt{n}$ square. As compared to the dense network, the distance between the nodes is increased by a factor of $\sqrt{n}$, and therefore the received powers, for the same transmit power, are reduced by a factor of $n^{\alpha/2}$. The following Theorem gives an upperbound on the capacity scaling of extended networks.

Theorem 4.3.1. *For a random extended network, for any $\epsilon > 0$, the aggregate throughput is bounded by*

$$T(n) \leq \begin{cases} K_2 n^{2-\alpha/2+\epsilon} & 2 \leq \alpha < 3 \\ K_2 n^{1/2+\epsilon} & \alpha \geq 3 \end{cases}$$

with high probability for a constant $K_2 > 0$ independent of n.

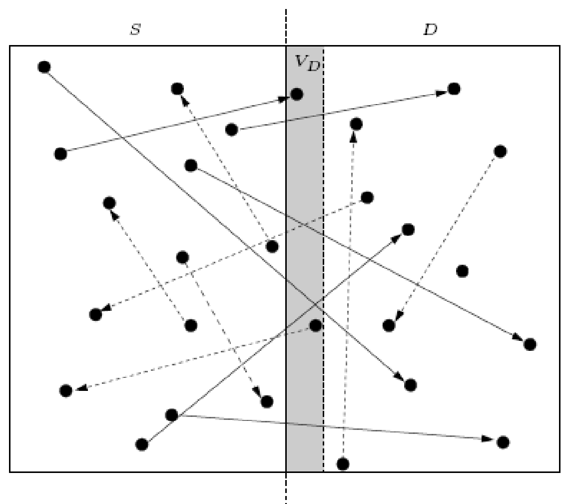

Figure 4.1: The cut set considered in the proof of Theorem 4.3.1.

Proof of Theorem 4.3.1. The proof is based on careful analyzing the power transfer through a cut-set bound. Consider a cut dividing the network into two equal halves (See figure 4.1. The source-destination pairs that pass the cut from left to right are depicted in bold lines). Because of the random source-destination pairing, the sum rate of communication requests passing through the cut from left to right, is equal to 1/4 of the total throughput of the network with high probability. This sum rate is bounded by the capacity of the MIMO channel between the nodes S located on the left of the cut and the nodes D located to the right. By an standard formula [28], and under fast fading assumption, the capacity is given by

$$\sum_{k\in S, i\in D} R_{ik} \leq \max_{\substack{Q(H)\succeq 0 \\ \mathbb{E}(Q_{kk}(H))\leq P, \forall k\in S}} \mathbb{E}(\log\det(I + HQ(H)H^*)) \tag{4.7}$$

where elements of H are the channel gains $H_{ik},\ k \in S, i \in D$. $Q(H) \succeq 0$ denotes a positive semi-definite transmit covariance matrix corresponding to the channel realization H. The diagonal element $Q_{kk}(H)$ denotes the power allocated to the node k at channel state H. Let V_D be the set of nodes on the $1 \times \sqrt{n}$ rectangular area to the right of the cut. By chernoff bound argument, there are no more than

$\sqrt{n}\log n$ in V_D. By Hadamards's inequality, we have

$$\sum_{k\in S, i\in D} R_{ik} \leq \max_{\substack{Q(H^{(1)})\succeq 0 \\ \mathbb{E}(Q_{kk}(H^{(1)}))\leq P, \forall k\in S}} \mathbb{E}(\log\det(I + H^{(1)}Q(H^{(1)})H^{(1)*})) + \max_{\substack{Q(H^{(2)})\succeq 0 \\ \mathbb{E}(Q_{kk}(H^{(2)}))\leq P, \forall k\in S}} \mathbb{E}(\log\det(I + H^{(2)}Q(H^{(2)})H^{(2)*})) \quad (4.8)$$

where $H^{(1)}$ is the channel matrix between S and V_D, and $H^{(2)}$ is between S and $D - V_D$. The first term in 4.8 can be easily upperbounded by the sum of the capacities of MISO channels between nodes in S and each node in V_D. The capacity of each individual channel is of the order of $\log n$. Since there are at most an order of $n\log n$ such channels, the first term is bounded by $K_3\sqrt{n}(\log n)^2$.

The hard part is the second term, i.e., the capacity of MIMO channel between nodes in S and nodes in $D - V_D$. Since, the channel matrix $H^{(2)}$ is random, computing the second term invokes analysis of the distribution of the largest eigenvalue of the channel matrix. Here, the channel gains H_{ik}, are independent and the distribution of real and imaginary parts are symmetric around the origin. In this case, the maximum of the second term is attained with a diagonal $Q(H^{(2)})$. In other words, independent signaling can achieve the capacity. This result has been shown in different ways in the literature [9], [24], [25], and [26]. Evaluating the second term for diagonal Q, and using the Hadamard inequality yields:

$$\text{The second term of (4.8)} \leq \mathbb{E}\left(\sum_{i\in D-V_D} \log\left(1 + \sum_{k\in S} P_k|H_{ik}|^2)\right)\right) \quad (4.9)$$

$$\leq \sum_{i\in D-V_D} \log\left(1 + \sum_{k\in S}\frac{GP_k}{r_{ik}^{\alpha}})\right) \quad (4.10)$$

$$\leq \sum_{i\in D-V_D}\sum_{k\in S}\frac{GP_k}{r_{ik}^{\alpha}} = P_{tot}. \quad (4.11)$$

where we have used the Jensen's inequality in the second step. It is worth noting that (4.11) is equivalent to the max-flow min-cut lemma developed in the chapter 3, both relating the sum rate across a cut-set to the total power transfer. The last thing needs therefore to be calculated is P_{tot}.

Lemma 4.3.1. *The scaling of the total received power is bounded by*

$$P_{tot} \leq \begin{cases} K_4 n(\log n)^3 & \alpha = 2 \\ K_4 n^{2-\alpha/2}(\log n)^2 & 2 < \alpha < 3 \\ K_4\sqrt{n}(\log n)^3 & \alpha = 3 \\ K_4\sqrt{n}(\log n)^2 & \alpha > 3 \end{cases}$$

with high probability for a constant $K_4 > 0$ independent of n.

We have mentioned the sketch of the proof of the above lemma at the end of the chapter for completeness. Putting every thing together, the result stated in Theorem is concluded. □

4.4 Notes

This chapter was mainly based on [1], along with a simplified method for calculating (4.8). In driving the upperbounds based on the capacity of MIMO channel, we assumed a fast fading environment. In a recent work [18], it has been shown that the same results can be extended to the slow fading setting as well. The throughput of random extended networks has been also studied in [6] under the channel model considered in chapter 3, i.e., attenuation of the transmitted signal over distance without fading. The result was a throughput scaling of $Kn^{1/2+1/\alpha}\log n$ for all $\alpha \geq 2$. It is interesting that this bound is looser than what was presented under the fading assumption, especially for $2 < \alpha < 3$.

4.5 Some Proofs

Proof of Lemma 4.3.1. Divide the network into n squares of area 1. By Chernoff bound argument, there are no more than $\log n$ nodes in each square with high probability. Consider figure 4.5; Each rectangle S_m of height 1 and width $\sqrt{n}$ consists of $\sqrt{n}$ squares. We aim to bound

$$P_{tot} = PG \sum_{m=1}^{\sqrt{n}} \sum_{k \in S_m} \sum_{i \in D-V_D} r_{ik}^{-\alpha}$$

Let us first focus on $\sum_{i \in D-V_D} r_{ik}^{-\alpha}$. Note that if we move the nodes that lie in each square to the square vertex, as indicated in figure 4.5, all terms in the summation only increase. Considering the same node displacement for all rectangular slabs S_m, $m = 1, ..., \sqrt{n}$ results in a regular network with at most $\log n$ nodes at each vertex on the left and $2\log n$ nodes at each vertex on the right. So, the problem reduces to computing P_{tot} for the resulting regular network. Let us index the left-hand side nodes by $(-k_x + 1, k_y)$ and those on the right by (i_x, i_y) where

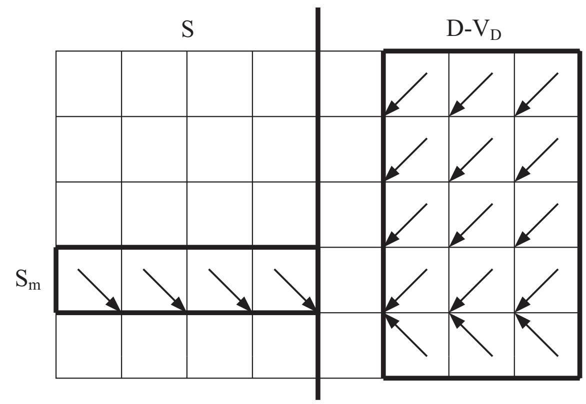

Figure 4.2: The displacement of nodes to the sqaure vertices, indicated by arrows.

$k_x, k_y, i_x, i_y = 1, ..., \sqrt{n}$. Then

$$P_{tot} \leq 2(\log n)^2 PG \sum_{k_x,k_y=1}^{\sqrt{n}} d_{k_x,k_y} \tag{4.12}$$

where

$$d_{k_x,k_y} = \sum_{i_x,i_y=1}^{\sqrt{n}} \frac{1}{((i_x + k_x - 1)^2 + (i_y - k_y)^2)^{\alpha/2}} \tag{4.13}$$

The upper bound of P_{tot} is obtained by manipulations of the above series. $\square$

Chapter 5

Hierarchical Cooperation in Ad Hoc Networks

In Chapter 4, we presented the upperbounds on the throughput scaling of the extended random networks. The results demonstrated the optimality of multihop for the high attenuation. An important question remained unanswered that what the optimal operation scheme is for the low attenuation case. In this chapter, we try to answer this question by introducing the hierarchical cooperation scheme, which is proposed in [1] for the operation of random wireless networks with n source-destination pairs communicating with each other at some common rate. We analyze the scheme and optimize it by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. In addition, to apply the hierarchical cooperation scheme to random networks, a clustering algorithm is developed, which divides the whole network into quadrilateral clusters, each with exactly the number of nodes required.

As a new result, we will show that the complete expression for the scaling with h hierarchical stages should be $c(h)n^{\frac{h}{h+1}}$. Since the pre-constant $c(h)$ affects the scaling behavior, we will present what can be achieved with the hierarchical scheme by providing an explicit expression of the pre-constant. Actually, for each n, the optimal number of stages to choose is $\sqrt{\log_\beta(n/2)}$, where β is a constant to be defined later, and the corresponding maximum achievable throughput is

$$\frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \tag{5.1}$$

where R is another constant. Therefore, as shown in (5.1), the hierarchical scheme actually achieves a scaling with the exponent depending on n.

Generally, a network with area A is distinguished into two categories based on whether $A^{\alpha/2} \leq n$, where $\alpha \geq 2$ is the power path loss exponent. In the case where $A^{\alpha/2} \leq n$, (5.1) is achievable. In the other case where $A^{\alpha/2} > n$, (5.1) has to be multiplied by $n/A^{\alpha/2}$ in order to meet the power constraint. It is worth pointing out that our results such as (5.1) apply to finite n. When trying to draw conclusions on scaling laws by taking $n \to \infty$, however, it should be noted that the results for the first case cannot remain valid if $\alpha > 2$, since the far-field model would fail to apply after some point.

For clarity, we will first present the results for regular networks. Then the extension to random networks is trivial after we introduce a clustering algorithm that divides the whole network into quadrilateral clusters, each with exactly the number of nodes required for carrying out the hierarchical cooperation scheme.

The remainder of this Chapter is organized as follows. In section 5.1, the wireless network model is described. Section 5.2 is devoted to the hierarchical cooperation scheme in regular networks, where we present the optimal throughput-delay results for the scheme with different stages. In section 5.3, a clustering algorithm is developed to extend the results to general random networks. In section 5.4, we study the appropriate strategy for the operation of random networks. In sections 5.5 and 5.6, we mention some notes and proofs.

5.1 Wireless Network Model

Recall the standard additive white Gaussian noise channel model of wireless networks presented in Chapter 4. Briefly,

1. There are a set of n nodes $\mathcal{N}$ located on a plane.
2. Each node uses a common average power P to transmit.
3. At any time t, each node i transmits the signal $X_i(t) \in \mathbb{C}$, and receives the signal $Y_i(t) \in \mathbb{C}$. The received signal depends on the transmitted signals of all the other nodes as

$$Y_i(t) = \sum_{k \neq i} H_{ik} X_k(t) + Z_i(t) \tag{5.2}$$

where $Z_i(t)$ is white circularly symmetric Gaussian noise of variance N, and the gain

$$H_{ik} = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}) \tag{5.3}$$

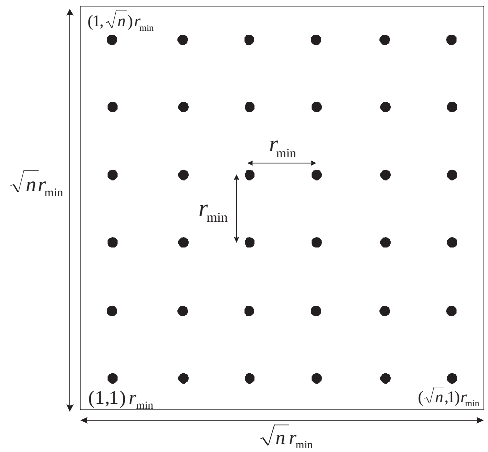

Figure 5.1: A regular network with n nodes and a minimum distance $r_{\min}$.

Consider the problem of n source-destination pairs in the network, where each node is a source, with its destination node arbitrarily chosen from the other nodes. For simplicity, assume that each node chooses a different node as its destination, although this requirement can be relaxed to some extend as we can see from the coding strategy described later. Therefore, each node is a source and also a destination for another source. We only consider the case where all pairs communicate at the same rate.

For the simplicity of presentation, and in order to expose the key features of the coding strategy, we will first consider a regular network as depicted in figure 5.1, where nodes are located at the grid points $(xr_{\min}, yr_{\min})$ for $1 \leq x, y \leq \sqrt{n}$ in an area $A = nr_{\min}^2$. Then the results can be easily extended to general random networks with high probability, where n nodes are randomly and uniformly distributed inside a square of area A.

5.2 Hierarchical Cooperation in Regular Networks

5.2.1 Double stage cooperation scheme

As a prelude, consider only two stages for the scheme and assume $A = 1$ unit. We will follow [1], but show what is really achievable by presenting a more transparent description. Divide the regular network into clusters of size M nodes (See figure 5.2; A source node s and its cluster S, and the destination d within cluster D have been depicted.). The *double stage scheme* is based on local transmit and receive

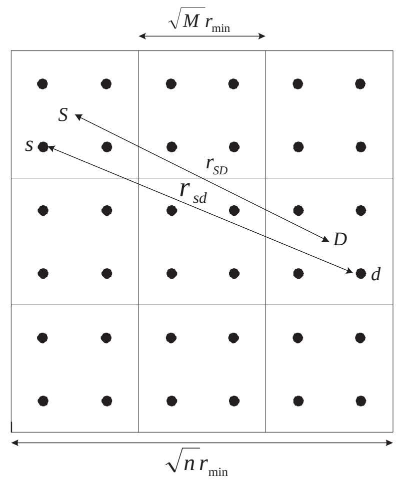

Figure 5.2: Dividing the network of n nodes into clusters of size M nodes.

cooperation in clusters and MIMO transmissions between clusters. Consider one source node s and its destination node d. The goal of s is to send M sub-blocks of length L bits (in overal, ML bits) to d. Let these bits be arranged in a *data matrix* $B^{M\times L}$

$$B = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1L} \\ B_{21} & B_{22} & \dots & B_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ B_{M1} & B_{M2} & \dots & B_{ML} \end{pmatrix}$$

which corresponds to choose one message W from 2^{LM} possible messages $\{1, \cdots, 2^{LM}\}$. Denote the i-th row by b^i (i-th *data sub-block*) and the m-th column by b_m (m-th *data vector*). The node s sends its data matrix to the node d in three steps:

1. s distributes its sub-blocks among the M nodes in its cluster by using TDMA. For this purpose, for each node k in the source cluster, s encodes the data sub-block b^k to a codeword of length C_0 chosen from a randomly generated Gaussian codebook $\mathcal{C}_0 \overset{i.i.d}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_0^2)$ where $\sigma_0^2 = Pr_{sk}^{\alpha}$. Sending one sub-block requires C_0 time slots and distributing all sub-blocks needs MC_0 time slots. At the end, each node in the cluster obtains one data sub-block of s.

2. The nodes of the source cluster form a distributed array antenna and send the LM bits of information to the destination cluster by MIMO transmissions. To accomplish this step, each node i encodes its sub-block b^i to a codeword $X_i^{C_1} = (X_{i1}X_{i2}\cdots X_{iC_1})$ of C_1 symbols by using a randomly generated Gaus-

sian codebook $\mathcal{C}_1 \overset{i.i.d}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_1^2)$ where $\sigma_1^2 = P\frac{r_{SD}^{\alpha}}{M}$ and r_{SD} is the distance between the centers of two clusters. Then nodes of the source cluster send their codewords simultaneously to the destination cluster. Therefore this step needs C_1 time slots to complete. Each node k in destination cluster receives an observation Y_{kt} from the MIMO transmission at time t for $1 \leq t \leq C_1$ according to (5.2) or the following vector form.

$$Y^M(t) = HX^M(t) + Z(t) \tag{5.4}$$

where $X^M(t) = (X_{1t}X_{2t}\cdots X_{Mt})^T$ and $Y^M(t) = (Y_{1t}Y_{2t}\cdots Y_{Mt})^T$ is the *observation vector* at time t. $Z(t) = (Z_{1t}Z_{2t}\cdots Z_{Mt})^T$ is uncorrelated noise at the receiver nodes, and H_{ik} are given by (5.3). The nodes simply store their observations. At the end of this step, each node k in destination cluster has accumulated an *observation sub-block* $Y_k^{C_1} = (Y_{k1}Y_{k2}\cdots Y_{kC_1})$ of C_1 observations.

3. Each node k in the destination cluster quantizes its observations with Q bits per observation to obtain a *quantized observation sub-block* $V_k^{C_1Q} = (V_{k1}V_{k2}\cdots V_{k(C_1Q)})$ of length C_1Q bits. From now on, the step is similar to step 1 but in reverse order. The cluster nodes send their quantized observation sub-blocks V_k to d by using the codewords of length C_0C_1Q/L chosen from a randomly generated Gaussian codebook with power σ_0^2 where $\sigma_0^2 = Pr_{kd}^{\alpha}$. The destination d can decode the quantized observations and estimate the observation sub-blocks and consequently, the observation vector $Y^M(t)$ by an *estimated observation vector* $\widehat{Y}^M(t)$. Then d can decode the transmitted data vectors $\widehat{b}_t$. The required number of slots for this step is MC_0C_1Q/L.

In the double stage cooperation strategy, the power of each observation must be upper bounded independent of cluster size which leads to quantization with a fixed number of bits for an average distortion Δ^2. When two clusters are neighbor, using the power assignment of $\sigma^2 = Pr_{SD}^{\alpha}/M$ yields an unbounded received power when the cluster size increases. A simple solution is to divide these clusters into two equal halves, each with $M/2$ nodes. The source node s distributes its sub-blocks among $M/2$ nodes of the half located farther to the border. Then these $M/2$ nodes form a distributed antenna and perform MIMO between the halves located farther away. Now, the required time for the step 2 is twice the time needed for disjoint clusters, i.e. the required time is $2C_1$ slots. In step 3, $M/2$ nodes take part in delivering the observations to the destination. For source and destination nodes located in the same cluster, we can simply ignore the second step. According to Lemma 4.5

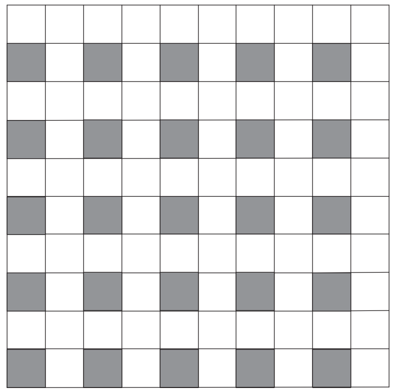

Figure 5.3: Parallel operating clusters according to 4-TDMA

of [1], the power received by each node in destination cluster P^{rec} in the step 2 is lower and upper bounded independent of cluster size such that

$$GPa^2 + N \leq P^{rec} \leq GPb^2 + N \tag{5.5}$$

where

$$a = (2 - \sqrt{2})^{\alpha/2} \tag{5.6}$$
$$b = (2 + \sqrt{2})^{\alpha/2} \tag{5.7}$$

Each source-destination pair must accomplish the three steps. Clustering also enables spatial reuse in the sense that clusters can work in parallel for local cooperations (step 1 and step 3) provided they locate far enough from each other. This leads to three phases in the operation of the network:

Phase 1: Setting up transmit cooperation. Clusters work in parallel according to the 4-TDMA scheme in figure 5.3 (as opposed to 9-TDMA scheme in [1][1]) where each cluster is active a fraction 1/4 of the total time of this phase. When a cluster becomes active, its source nodes must perform the first step, i.e. distributing their sub-blocks to the other nodes of the cluster by a simple TDMA. Each source node needs MC_0 slots, hence the required time for source nodes of one cluster to exchange their bits is at most M^2C_0 slots and due to 4-TDMA, the whole phase needs $4M^2C_0$ slots to complete. Each node transmits with power σ_0^2 in at most fraction $\frac{1}{4M}$ of the total time of the phase. It can be shown that this power assignment satisfies an overal average power consumption less than P/n. Using the 4-TDMA ensures us that the interference power each node received from all simultaneously transmitting nodes is bounded according to the following Lemma.

[1]4-TDMA actually saves time compared to 9-TDMA. However, the scaling won't be changed.

Lemma 5.2.1. *The interference signals received by different nodes, due to parallel operating clusters using 4-TDMA, are independent and for $\alpha > 2$ the interference power that each node is received is given by*

$$P_I \leq GP(8 + \frac{2}{\alpha - 2} + \frac{2}{\alpha - 1})$$

Phase 2: MIMO transmissions. We perform successive MIMO transmissions according to the step 2, one MIMO for each source-destination pair from source cluster to destination cluster in one time slot, hence we need at most $2nC_1$ slots. Each node encodes the sub-blocks by using a Gaussian code of power σ_1^2 as defined earlier. Since at most M MIMO transmissions are originated from each cluster, each node is active at most a fraction M/n of the total time of this phase and remains silent during the rest of the phase which yields an average power consumption less than P/n.

Phase 3: Cooperate to decode. After the first two phases, each source-destination pair has completed the steps 1 and 2.Each cluster should accomplish the step 3 by conveying the quantized observations to the corresponding destination nodes located in the cluster. This phase is identical to the first phase, except that each node has C_1Q bits to transmit to each node in the same cluster instead of L bits. Therefore, this phase needs $4M^2C_0C_1Q/L$ slots to complete.

In summary, the required time D_2 for the double stage scheme is

$$\begin{aligned} D_2 &= D(phase1) + D(phase2) + D(phase3) \\ &= 4M^2C_0 + 2nC_1 + 4M^2C_0C_1Q/L \\ &= 4M^2C_0(1 + C_1Q/L) + 2nC_1 \end{aligned}$$

Assume the channel gains are known at all nodes. All communication links in the first phase can operate at any rate less than the following:

$$R_0 \leq \log\left(1 + \frac{GP}{P_I + N}\right) \tag{5.8}$$

Communications in the second phase are performed over the quantized MIMO channel of figure 5.4 where the notation $(\cdots)_L$ is used for an i.i.d sequence of L random variables. The following lemma asserts that an spatial multiplexing gain of M is achievable for this channel.

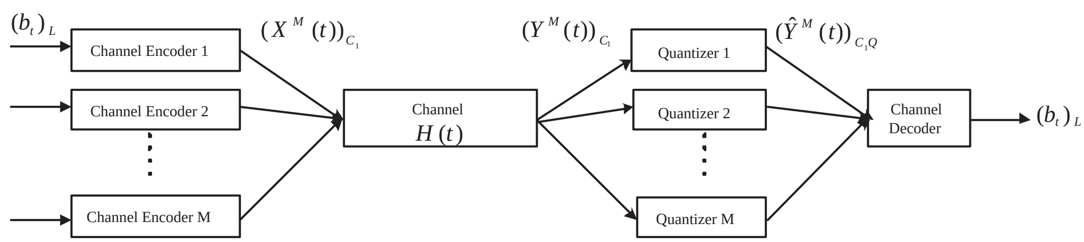

Figure 5.4: A model for the quantized MIMO channel

Lemma 5.2.2. *Define the average probability of error for the quantized MIMO channel by*

$$P_e^L = \frac{1}{2^{LM}} \sum_{k=1}^{2^{LM}} \mathbb{P}(\widehat{W} \neq k | W = k)$$

Then there exists a strategy to quantize the observations with Q bits per observation and a codebook $\mathcal{C}_1$ satisfying power constraint $\sigma_1^2 = Pr_{SD}^{\alpha}/M$ to encode the data sub-blocks such that arbitrary low P_e^L is feasible. Moreover, the minimum quantization rate Q and the maximum achievable rate R_1 of the codebook satisfy

$$Q > \log\left(1 + \frac{GPb^2 + N}{\Delta^2}\right)$$

and

$$\log\left(1 + t\frac{GP}{N + \Delta^2}\right)\frac{(a^2 - t)^2}{2b^4} \leq R_1 < \log\left(1 + b^2\frac{GP}{N + \Delta^2}\right) \tag{5.9}$$

for any $0 \leq t \leq a^2$.

For simplicity, all nodes use the same rate for their codewords $\mathcal{C}_0$ and $\mathcal{C}_1$, i.e., $\frac{L}{C_0} = \frac{L}{C_1} = R$, where

$$R \leq \min\{\text{LHS of (5.9), RHS of (5.8)}\}$$

Hence, the required time D_2 can be written as

$$D_2(n, M, L) \;\; = \;\; 4M^2\frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n\frac{L}{R}$$

We call this quantity delay because each destination can decode its intended bits only after receiving all the corresponding observations , i.e., after the step 3. At the end of this time, each node has delivered ML bits to its destination which yields a total throughput of

$$T_2 = \frac{nML}{D_2}$$

which is maximized by choosing $M = \sqrt{\frac{R}{2(R+Q)}}\sqrt{n}$:

$$T_2^{opt}(n) = \frac{R}{4\sqrt{2(1+Q/R)}}n^{1/2} \tag{5.10}$$

and the corresponding delay is

$$D_2^{opt}(n, L) = 4\frac{L}{R}n. \tag{5.11}$$

Obviously, by repeating n times, the double stage scheme can also be used for the problem where each node needs to send different information to all the other nodes. The achievable rate is as the following.

Lemma 5.2.3. *For a regular network of size n, by the double-stage cooperation scheme with clusters of size M, each node can deliver ML different bits to each of the other nodes in a time block of*

$$nD_2(n, M, L) = 4nM^2\frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n^2\frac{L}{R}.$$

Remark 5.2.1. *Note that L denotes the number of bits to be transmitted in a basic time block, and is proportional to the block length for any fixed communication rate. Although for the interest of delay, it is better to choose smaller L as shown in Lemma 5.2.3, shorter block length leads to higher decoding errors. Hence, there is always a minimum L required to ensure enough reliability.*

5.2.2 Triple stage cooperation scheme

Is it possible to achieve a better throughput by local cooperation and MIMO transmissions? Recall that in Phase 1 and Phase 3 of the double stage scheme, TDMA was used in each cluster to deliver the bits. Since each cluster itself is a network similar to the original network only with a smaller number of nodes, this implies that one can use the double stage scheme in each cluster to exchange the bits as well. Next, we analyze the throughput and delay of this new *triple stage scheme* when the double stage scheme is used in Phase 1 and Phase 3.

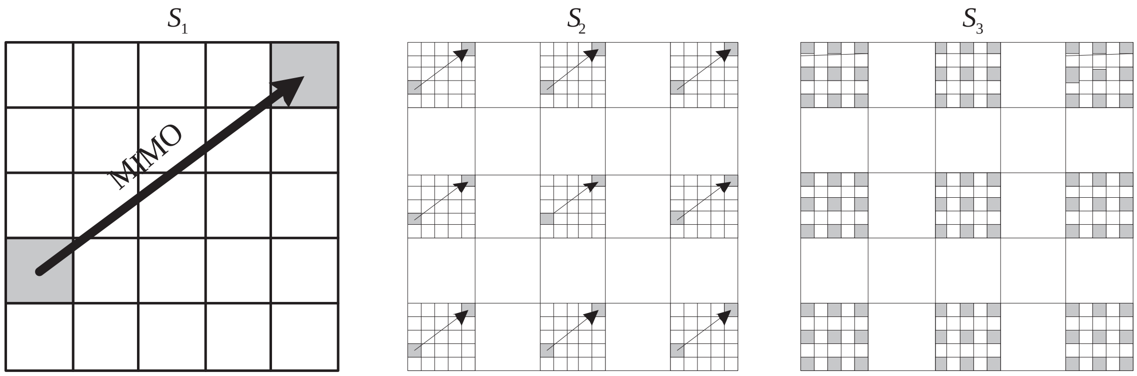

Figure 5.5: The three stages of the triple stage cooperation scheme.

First, divide the whole network into clusters of size M_1, and then divide each cluster of size M_1 into sub-clusters of size M_2. Apply the double stage scheme to each cluster of size M_1. To avoid the interference from neighboring clusters, use 4-TDMA as before. Hence, according to Lemma 5.2.3, it takes $M_1 D_2(M_1, M_2, L)$ time slots for each node to deliver $M_2 L$ bits to each node in the same cluster and this phase needs $4M_1 D_2(M_1, M_2, L)$ time slots to complete.

In Phase 2, as before, it takes $2n\frac{M_2 L}{R}$ time slots to complete.

In Phase 3, same as phase 1 except that there are $\frac{Q}{R}$ times as many bits to transmit, it takes $4M_1 D_2(M_1, M_2, L)\frac{Q}{R}$ time slots to complete.

Totally, with the triple stage scheme, it takes

$$D_3(n, M_1, M_2) = 4M_1 D_2(M_1, M_2, L)\left(1 + \frac{Q}{R}\right) + 2n\frac{M_2 L}{R}$$

time slots to communicate $M_1 M_2 L$ bits for each source-destination pair. This yields a throughput of

$$T = \frac{nM_1 M_2 L}{D_3(n, M_1, M_2)}. \tag{5.12}$$

The three stages of the scheme, namely S_1, S_2 and S_3 have been depicted in figure 5.5. In stage S_1, global MIMO transmissions are performed between clusters of size M_1. In stage S_2, clusters M_1 work in parallel and local MIMO transmissions are performed between sub-clusters of size M_2. S_3 is the bottom stage where point-to-point communications take place between nodes of sub-clusters.

Using the partial derivatives with respect to M_1 and M_2 to maximize the throughput in (5.12), the optimal cluster sizes are given by

$$M_2 = \frac{1}{2(1+Q/R)}\left(\frac{n}{2}\right)^{1/3}$$

$$M_1 = \frac{1}{2(1+Q/R)}\left(\frac{n}{2}\right)^{2/3}$$

and consequently the optimal throughput and the delay of the triple stage scheme are given by

$$T_3^{opt}(n) = \frac{2^{1/3}}{24(1+Q/R)}Rn^{2/3}$$

$$D_3^{opt}(n) = \frac{3}{2^{1/3}(1+Q/R)}\frac{L}{R}n^{4/3}$$

Remark 5.2.2. *It is easy to prove as an extension of Lemma 5.2.1 that for the triple stage cooperation scheme, the received interference signals by different nodes of the network are uncorrelated in all the stages. Moreover the stage* S_3 *has the largest interference power which can be bounded by* P_I [2]. *Hence, the following coding rate* R *and quantization rate* Q *can be used in all the stages.*

$$R \leq \max_{0\leq t\leq a^2} \log\left(1 + t\frac{GP}{N+P_I+\Delta^2}\right)\frac{(a^2-t)^2}{2b^4} \tag{5.13}$$

$$Q > \log\left(1+\frac{GPb^2+N+P_I}{\Delta^2}\right) \tag{5.14}$$

Compared to the double stage scheme, the triple stage scheme can achieve a higher order of n for throughput (an order of $n^{2/3}$ for the triple stage scheme in contrast with an order of $n^{1/2}$ for the double stage scheme), but the pre-constant of throughput decreases by increasing the number of stages. The desirable and adverse effects of increasing the number of stages can be explained as follows.

- Increasing the number of stages results in a better use of the degrees of freedom as the network transports more portion of the traffic by MIMO transmissions and less by TDMA. This in turn leads to an increase in order of n in the throughput.

[2] Since the scheme runs 4-TDMA in both network and clusters, the exact interference power is less than P_I, nevertheless this bound is sufficient to verify that a universal coding rate R is feasible.

- For a higher stage scheme, one should be able to bound the interference power due to parallel operating clusters which invokes running 4-TDMA in the network and at the same time inside the clusters. This yields an increase in the delay and consequently a reduction in the throughput. Another overhead arises from quantizing and re-encoding the observations at different stages which further increases the delay and reduces the throughput.

5.2.3 h-stage hierarchical cooperation scheme

Generally, suppose that with the $(h-1)$-stage hierarchical cooperation scheme with cluster sizes $M_1, M_2, \ldots, M_{h-2}$, it takes $D_{h-1}(n, M_1, M_2, \ldots, M_{h-2})$ time slots to communicate $M_1 M_2 \cdots M_{h-2} L$ bits for each source-destination pair.

Replacing phase 1 and phase 3 of the double stage scheme with the $(h-1)$-stage scheme, we have the h-stage scheme. Obviously, for the h-stage scheme with cluster sizes $M_1, M_2, \ldots, M_{h-1}$, it takes

$$\begin{aligned} & D_h(n, M_1, M_2, \ldots, M_{h-1}) \\ & = 4M_1 D_{h-1}(M_1, M_2, \ldots, M_{h-1}) \left(1 + \frac{Q}{R}\right) \\ & \quad + 2n \frac{M_2 \cdots M_{h-1} L}{R} \end{aligned}$$

time slots to communicate $M_1 M_2 \cdots M_{h-1} L$ bits for each source-destination pair.

It can be verified that the general formula is

$$D_h(n, M_1, M_2, \ldots, M_{h-1}) = M_1 M_2 \cdots M_{h-1} \frac{L}{R} \times \left\{ [4(1+Q/R)]^{h-1} M_{h-1} + 2 \sum_{i=0}^{h-2} [4(1+Q/R)]^i \frac{M_i}{M_{i+1}} \right\}$$

Consequently, the throughput is given by

$$T_h(n, M_1, M_2, \ldots, M_{h-1}) = \frac{n M_1 M_2 \cdots M_{h-1} L}{D_h(n, M_1, M_2, \ldots, M_{h-1})}$$

which in general is a function of all the cluster sizes.

We maximize the throughput by using the partial derivatives. Solving $\partial T_h / \partial M_i = 0$ for $1 \leq i \leq h-1$ yields

$$M_i^2 = \frac{M_{i-1} M_{i+1}}{4(1+Q/R)}$$

where let $M_0 = n$ and $M_h = 2$. Therefore, the optimal choices of the cluster sizes are

$$M_i = \frac{2(n/2)^{(h-i)/h}}{[4(1+Q/R)]^{i(h-i)/2}} \quad \text{for } 1 \le i \le h-1 \tag{5.15}$$

Next we present one of our main results.

Theorem 5.2.1. *For a regular network of n nodes in a unit area, by the h-stage hierarchical cooperation scheme with the optimal cluster sizes (5.15), the throughput is given by*

$$T_h^{opt}(n) = \frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}} \tag{5.16}$$

and the corresponding delay is

$$D_h^{opt}(n, L) = \frac{h2^{(h+2)(h-1)/(2h)}}{(2\sqrt{1+Q/R})^{(h+3)(h-2)(h-1)/6}} \frac{L}{R} n^{\frac{h-1}{2}+\frac{1}{h}}.$$

For any fixed n, we can find the optimal h to maximize $T_h^{opt}(n)$. Let

$$\frac{dT_h^{opt}(n)}{dh} = 0$$

which leads to

$$h^2 \ln(2\sqrt{1+Q/R}) + h - \ln(n/2) = 0.$$

Hence, the optimal number of stages to choose is

$$h^* = \frac{\sqrt{1 + 4\ln(2\sqrt{1+Q/R})\ln(n/2)} - 1}{2\ln(2\sqrt{1+Q/R})}. \tag{5.17}$$

In order to obtain a simple formula, let

$$\begin{aligned} h^* &= \sqrt{\frac{\ln(n/2)}{\ln(2\sqrt{1+Q/R})}} \\ &= \sqrt{\log_\beta(n/2)} \end{aligned} \tag{5.18}$$

where $\beta := 2\sqrt{1+Q/R}$. Note that

$$\beta^h = \beta^{\log_\beta(n/2)\frac{h}{\log_\beta(n/2)}} = (n/2)^{\frac{h}{\log_\beta(n/2)}}.$$

Therefore,

$$\begin{aligned} T_h^{opt}(n) &= \frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}} \\ &= \frac{\beta R}{h\beta^h}(n/2)^{1-\frac{1}{h}} \\ &= \frac{\beta R}{h}(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}} \end{aligned} \tag{5.19}$$

where choosing h as in (5.18), we have

$$T^{opt}(n) = \frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \tag{5.20}$$

Obviously (5.20) is a very accurate estimation, although we made some approximation in (5.18) and h^* should always be an integer.

Theorem 5.2.2. *For a regular network of n nodes in the unit area, by the hierarchical cooperation scheme with the optimal number of stages (5.17) and the optimal cluster sizes (5.15), the maximum throughput is approximately given by (5.20).*

Actually, we can provide an exact upper bound of $T^{opt}(n)$. It follows from (5.19) that

$$\begin{aligned} T_h^{opt}(n) &\leq \beta R(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}} \\ &\leq \beta R(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \end{aligned} \tag{5.21}$$

where, in the last inequality, "=" holds if $h = \sqrt{\log_\beta(n/2)}$.

To check how much different (5.21) is from the linear scaling law $\Theta(n)$, we take the ratio:

$$\begin{aligned} \frac{n/2}{(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}}} &= (n/2)^{\frac{2}{\sqrt{\log_\beta(n/2)}}} \\ &= \left(\beta^{\log_\beta(n/2)}\right)^{\frac{2}{\sqrt{\log_\beta(n/2)}}} \\ &= \beta^{2\sqrt{\log_\beta(n/2)}} \\ &\to \infty. \end{aligned}$$

Hence, the hierarchical cooperation scheme cannot achieve arbitrarily close to linear scaling as claimed in [1]. Instead, the difference grows to infinity as n increases.

5.2.4 Hierarchical cooperation for networks with area A

Generally, consider a regular network with area A. Note that distance affects the power loss. We can scale down the general regular network with area A to a regular network with unit area, but with the power constraint $\frac{P}{(\sqrt{A})^\alpha}$ at each node, since the distance between nodes is reduced by a factor of $\sqrt{A}$. Recall that when $A = 1$ unit, running the hierarchy does not need the whole power budget P and the average power consumption is less than P/n per node. Thus, a general network can be dichotomized based on the relation between its area and the number of nodes into two cases:

- Dense network: The network is called dense when $A^{\alpha/2} \leq n$. Then the nodes have enough power to run the hierarchical scheme and get the throughput-delay results as discussed above.
- Sparse network: The network is called sparse when $A^{\alpha/2} > n$. Then the nodes do not have sufficient power to run the hierarchical scheme all the time. Instead, they run the scheme in a fraction $n/A^{\alpha/2}$ of the time with power $PA^{\alpha/2}/n$ and remain silent during the rest of the time. Obviously this bursty modification satisfies the original average power constraint P, and correspondingly, the achieved throughput is modified by a factor of $n/A^{\alpha/2}$, e.g., in (5.16) and (5.20).

5.3 Extension to Random Networks

In this section, we extend the results of regular networks to random networks. We first review the extension method of [1]: Consider a random network of unit area with n nodes. Since the average number of nodes in a cluster of area $A_c = \frac{M}{n}$ is M, the hierarchical scheme was applied to this random network by dividing the network into the clusters of area $\frac{M_1}{n}$ and proceeding to clusters of area $\frac{M_{h-1}}{n}$, for the h-stage scheme, and get the throughput-delay of the regular network but with a failure probability. Failure arises from the deviation of number of nodes in each cluster from its average. By a simple Chernoff bound argument, the probability of having large deviations from the average can be bounded (see Lemma 4.1 of [1]). As $n \to \infty$, this probability goes to zero.

The above *clustering* method is not sufficient for the following reasons:

1. The clusters of area $A_c = \frac{M}{n}$ are required to contain *exactly* M nodes to perform the hierarchical scheme. A deviation from the average number of

nodes M, even very small, results in failure of the scheme. However, [1] only bounded the probability of large deviation.

2. The probability of having exactly M nodes in a cluster of area $A_c = \frac{M}{n}$ is given by the binomial distribution $\mathfrak{p}(M; n, M/n) = \binom{n}{M}(\frac{M}{n})^M \ (1 - \frac{M}{n})^{n-M}$. Using the Stirling's formula to approximate the factorial terms, as $n \to \infty$, yields
$$\mathfrak{p}(M; n, M/n) \approx \frac{M^M}{e^M M!}$$
Recall that for the optimal operation of the scheme, the cluster sizes M are chosen proportional to n^γ where $0 < \gamma < 1$. Hence, the probability of having M nodes is proportional to $\frac{1}{\sqrt{2\pi M}}$ which, in fact, goes to zero.

To resolve the issue of making clusters of exactly M nodes, we will develop a clustering algorithm in this paper. To achieve high probability, we need to consider simultaneously the probabilities of events of the entire class of clusters, which invokes a sort of uniform convergence (in probability) of law of large numbers over the entire class. To resolve this, we will resort to the *Vapnik-Chervonekis* theorem.

5.3.1 Choosing an appropriate cluster shape

We use the Vapnik-Chervonekis theorem [19], [20] to find the appropriate cluster shape. Let $\mathcal{F}$ be a set of subsets and A a finite set of points. First, we recall some definitions:

Definition 1: $Proj_{\mathcal{F}}(A)$ is the projection of $\mathcal{F}$ on A which is defined as $\{F \cap A \ : \ F \in \mathcal{F}\}$.

Definition 2: A is *shattered* by $\mathcal{F}$ if $Proj_{\mathcal{F}}(A) = 2^A$, i.e., if the projection of $\mathcal{F}$ on A includes all possible subsets of A.

Definition 3: The *VC-dimension of* $\mathcal{F}$, denoted by VC-d($\mathcal{F}$) is the cardinality of the largest set A that $\mathcal{F}$ shatters. It may be infinite.

The Vapnik-Chervonekis Theorem: If $\mathcal{F}$ is a set of finite VC-dimension and $\{X_i\}$ is a sequence of n i.i.d random variables with common probability distribution P, then for every $\epsilon, \ \delta > 0$

$$\text{Prob}\left\{\sup_{F\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} I(X_i \in F) - P(F)\right| \leq \epsilon\right\} > 1 - \delta \tag{5.22}$$

whenever

$$n > \max\left\{\frac{8\text{VC-d}(\mathcal{F})}{\epsilon}\log\frac{16e}{\epsilon}, \frac{4}{\epsilon}\log\frac{2}{\delta}\right\} \tag{5.23}$$

An application of this theorem has been already presented in [3] for the set of disks on the plane. In this section, we consider a more general case; we apply the Vapnik-Chervonekis theorem to the set of all the clusters that partition the given random network with n nodes in the unit area. Note that a finite VC-dimension, for the set of clusters $\mathcal{F}$, is a sufficient condition for the uniform convergence in the weak law of large numbers. Assume that this condition is satisfied and the set of clusters has a finite VC-dimension (We will later derive a sufficient condition for the cluster shapes to make the VC-dimension finite). Denote the area of each cluster $c \in \mathcal{F}$ by A_c and its number of nodes with N_c, then we have the following lemma:

Lemma 5.3.1. *For every cluster $c \in \mathcal{F}$ that contains exactly M nodes,*

$$\frac{M - \xi \log n}{n} < A_c < \frac{M + \xi \log n}{n} \tag{5.24}$$

with probability larger than $1 - \frac{\xi \log n}{n}$ where $\xi = \max\{8\, VC\text{-}d(\mathcal{F}), 16e\}$.

Proof. Let $\mathcal{F}$ denote the class of clusters with finite VC-dimension VC-d($\mathcal{F}$). To satisfy (5.23), let $\epsilon = \delta = \frac{\xi \log n}{n}$. Then the Vapnik-Chervonekis theorem states that

$$\text{Prob}\left\{\sup_{c \in \mathcal{F}}\left(\left|\frac{N_c}{n} - A_c\right| \leq \frac{\xi \log n}{n}\right)\right\} > 1 - \frac{\xi \log n}{n} \tag{5.25}$$

Therefore, if a cluster c contains exactly M nodes, i.e., $N_c = M$, then its area must satisfy (5.24) with high probability. □

Note that if a cluster has an area less than $\frac{M - \xi \log n}{n}$, then with high probability it contains less than M nodes. Similarly, if its area is greater than $\frac{M + \xi \log n}{n}$, with high probability, it contains more than M nodes. Next, we need to choose a right shape for clusters to make the VC-dimension finite. We will make use of the following lemma, due to [21], in finding the appropriate shape. We have presented the sketch of the proof at the end of the chapter for completeness.

Lemma 5.3.2. *Let $\mathcal{F}$ be a set of subsets with VC dimension d. Consider another set $\mathcal{F}_{\cap r}$ which consists of r-wise intersections of subsets in $\mathcal{F}$. The VC-dimension of the new set is at most $2rd \log(3r)$.*

Corollary 5.3.1. *The VC-dimension of the set of convex r-laterals is finite and upper bounded by $6r \log(3r)$ where r is the number of sides.*

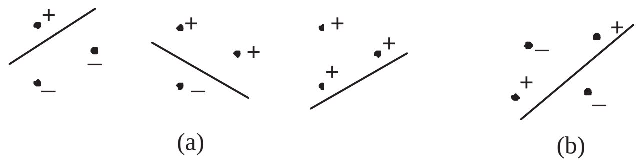

Figure 5.6: VC-dimension for the set of half-spaces is 3. (a): A set of 3 points is shattered, (b): No set of 4 points can be shattered.

Proof. Consider a line in the plane. It divides the plane into two half-spaces. Choose one of the half-spaces as subset. Define $\mathcal{F}'$ as the set of all half-spaces produced by considering different lines in the plane. It is easy to prove that VC-d($\mathcal{F}'$) = 3 since a set of 3 nodes that are not collinear can be shattered (see figure 5.6(a)) but it is impossible to find a set of 4 nodes that are shattered by $\mathcal{F}'$ (see figure 5.6(b)). The labels "+" and "-" in figure 5.6 have been used to specify different subsets of points. The key observation is that any convex r-lateral is an intersection of r half-spaces. In the light of this observation and by using Lemma 5.3.2, it is concluded that the VC-dimension of the set of convex r-laterals is at most $6r \log(3r)$. □

We will use a set of quadrilaterals as the clusters. Since the VC-dimension is at most $24 \log 12$, we can apply Lemma 5.3.1 with $\xi = 800$ to these clusters. Next, we develop an algorithm to make clusters of exactly M nodes.

5.3.2 Clustering algorithm

Divide the network into squares of area $\frac{M}{n}$, and start from the square located on the top left corner. Depending on how many nodes are within this square, three situations may arise:

1. if the number of nodes in the square is exactly M, ignore this square and go to the next one.

2. if the number of nodes in the square is less than M, make a quadrilateral cluster by expanding the square: Move the top right vertex of the square to the right such that the created quadrilateral cluster contains exactly M nodes.

3. if the number of nodes in the square is more than M, make a quadrilateral by shrinking the square: Move the top right vertex of the square to the left such that the resultant quadrilateral cluster contains exactly M nodes.

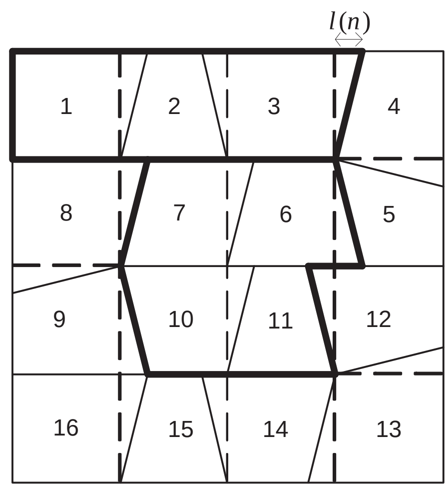

Figure 5.7: Clustering of a random network with exactly M nodes in each quadrilateral cluster.

After making the first cluster, go to the second cluster on the right side and make it a quadrilateral with exactly M nodes by expanding or shrinking as discussed above. Repeat the procedure for all the squares in the first row. For the top right square, use its bottom right vertex to do expanding/shrinking. For the second row, starting from the right square, move to the left side, and make the quadrilateral clusters of M nodes by expanding-shrinking. Perform the same procedure for all the rows, and we will have a set of quadrilateral clusters; each one contains exactly M nodes. One instance of such a clustering algorithm has been depicted in figure 5.7. Note that according to Lemma 5.3.1, the amount of expanding/shrinking in the areas of the squares is less than $\frac{\xi \log n}{n}$ with high probability.

5.3.3 Network operation

The operation of random networks is similar to the operation of the regular networks. The centers of the quadrilateral clusters are defined as the centers of the original squares. Note that the new quadrilateral cluster will include the center of its original square with high probability. To observe this property of our clustering algorithm, consider the combination of the clusters 1, 2, and 3 in figure 5.7. This combination gives a larger quadrilateral cluster with $N_c = 3M$, hence according to (5.25) the deviation of the area of this cluster from its average $(3M/n)$ must be less than $\frac{\xi \log n}{n}$ and consequently $l(n) \leq \frac{2\xi \log n}{\sqrt{nM}}$. Therefore $l(n)$ is much smaller than the square side $\sqrt{\frac{M}{n}}$ (recall that $M = n^\gamma$ for $0 < \gamma < 1$) and the quadrilaterals are concentrated on the squares. In other words, each quadrilateral corresponds only to one square, and *vice versa*. Hence, the hierarchical scheme can be applied to the random networks by using the corresponding quadrilateral of each square instead of original square cluster. By making clusters of M_{h-1} nodes for the bottom stage of the hierarchy using the clustering algorithm, these clusters can be combined to make larger clusters of M_{h-2} nodes for the upper stage. Following the same procedure, make clusters of exactly M_1 nodes for the top stage. It is worth noting that for combined clusters, for example, combination of clusters 6, 7, 10, and 11 in figure 5.7, we can define the same deviation factor $l(n)$ as defined for the clusters of the bottom stage. As the result, the received power of each MIMO transmission will be lower-bounded and upper-bounded by (5.5). The only difference is that the coefficients a and b in (5.6)-(5.7) should be replaced by

$$a' = (1 + \frac{\sqrt{2}}{2} \frac{\sqrt{\frac{M}{n}} + l(n)}{\sqrt{\frac{M}{n}} - l(n)})^{-\alpha/2}$$

$$b' = (1 - \frac{\sqrt{2}}{2} \frac{\sqrt{\frac{M}{n}} + l(n)}{\sqrt{\frac{M}{n}} - l(n)})^{-\alpha/2}$$

or equivalently

$$a' = \left(1 + \frac{\sqrt{2}}{2} \frac{M + 2\xi \log n}{M - 2\xi \log n}\right)^{-\alpha/2}$$
$$b' = \left(1 - \frac{\sqrt{2}}{2} \frac{M + 2\xi \log n}{M - 2\xi \log n}\right)^{-\alpha/2}$$

M can be chosen as the optimal cluster size for different stages, i.e., M_i for $1 \leq i \leq h-1$. But it holds for any i that $\frac{M_i + 2\xi \log n}{M_i - 2\xi \log n} \leq \frac{M_{h-1} + 2\xi \log n}{M_{h-1} - 2\xi \log n}$. Moreover, since $M_{h-1} \propto n^{1/h}$, the right hand side (RHS) of this inequality is a decreasing function of n for large values of n and approaches to one. In fact, for any given $\eta > 1$, there exists a n_0 such that for all $n > n_0$, the RHS is less than η. Hence, for all $n > n_0$

the coefficients can be chosen as

$$a' = \left(1 + \frac{\sqrt{2}}{2}\eta\right)^{-\alpha/2}$$
$$b' = \left(1 - \frac{\sqrt{2}}{2}\eta\right)^{-\alpha/2} \tag{5.26}$$

Consequently, the required quantization rate Q and the channel coding rate R can be defined based on the above coefficients. Obviously as $n \to \infty$, $a' \to a$ and $b' \to b$ and we can use the same quantization and coding rates as the rates already used for regular networks.

5.4 The Appropriate Operation Strategy

In previous sections of this chapter, we derived the exact achievable throughput of the hierarchical scheme with any number of stages. As observed, for dense networks, there is a gap between the maximum achieved throughput and the linear-scaling upperbound of Theorem 4.2.1. As we increases the number of stages of the hierarchy to achieve a scaling closer to the linear one, the overhead due to using the 4-TDMA for parallel operating clusters and quantizing and re-encoding the observations at different stages, reduces the performance significantly.

For the extended network model of Chapter 4, the area $A = n$, and as an immediate application of presented results in section 5.2.4, a throughput of

$$O\left(n^{2-\frac{\alpha}{2}-\frac{2}{\sqrt{\log_{\beta}(n/2)}}}\right) \tag{5.27}$$

is achievable.

Comparing the above scaling law with the upperbound of Theorem 4.3.1 for the extended networks, we observe the sub-optimality of the hierarchical cooperation scheme for $2 \leq \alpha \leq 3$. Furthermore, the multihop outperforms the hierarchical scheme for $\alpha > 3$.

A natural question is why attenuation has a fundamental effect on choosing the appropriate strategy. A natural place to look for the answer is the cutset of figure (4.1) in Chapter 4. Recall that the total throughput between the $S-D$ pairs with sources on the left half and destinations on the right half of the cut, which with

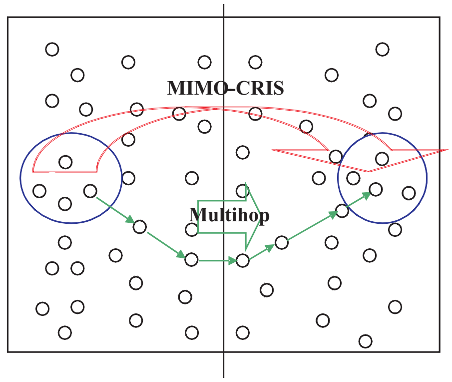

Figure 5.8: Power transfer is dominated by long distance transmissions under high attenuation and nearest-neighbor multi-hop under low attenuation

high probability is 1/4 of the total aggregate throughput, is bounded by a multiple of the total received energy by the nodes on the right from the nodes located on the left side. The total received power is dominated either by power transfer between the nodes near the cut or by power transferred by nodes far away rom the cut. There are roughly an order of $\sqrt{n}$ nodes near the cut and order of n^2 nodes away from the cut but the channels between the nodes near the cut are much stronger. When $\alpha \leq 3$, the received power is dominated by transfer between nodes far away from the cut because there are fewer nodes near the cut than far away from the cut. In this case, hierarchical scheme approximately achieves the required power by long distance MIMO transmissions (and also the CRIS strategy by coherent long distance transmissions). When $\alpha > 3$, the received power is dominated by power transferred by the nodes near the cut. This can be achieved by nearest-neighbor multihop and therefore, multi hop is optimal in this regime.

5.5 Notes

The idea of clustering and MIMO transmission between clusters was first introduced in [11]. The idea was later developed in [1] to the hierarchical architecture. The results presented in this chapter have been mainly derived from the papers published based on this thesis, [16] and [17], where the exact achievable throughput for any number of users has been found. The clustering algorithm for random networks is another contribution.

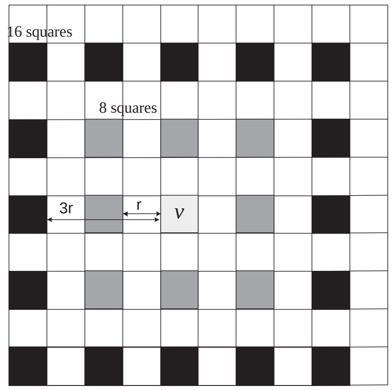

Figure 5.9: Grouping of interfering clusters in 4-TDMA.

5.6 Some Proofs

Here, we mention the proofs for some of the stated Lemmas for completeness.

Proof of Lemma 5.2.1. The proof follows in parallel with Lemma 4.2 of [1] for 9-TDMA. Consider figure 5.9 for the regular network in the unit area. The interference signal received by each node v is given by

$$I_v = \sum_{u \in \mathcal{U}_v} H_{vu} X_u$$

where H_{vu} is given by (5.3) and X_u is the signal transmitted by an active node located in a simultaneously operating cluster u with power σ_0^2. $\mathcal{U}_v$ is the set of simultaneously operating clusters of node v which can be grouped such that a group $\mathcal{U}_v(k)$ contains $8k$ clusters which are separated from v by a distance larger than $(2k-1)r$ where $r = \sqrt{\frac{M}{n}}$. The number of such groups can be easily bounded by $k \leq 1/4\sqrt{f}$ where $f = n/M$ is the number of clusters.

$$P_I < \sum_{k=1}^{1/4\sqrt{f}} \sum_{U \in \mathcal{U}_v(k)} \frac{G\sigma_0^2}{((2k-1)r)^\alpha}$$

where we used the assumption that channel gains are independent. Substituting the value of σ_0^2 yields

$$P_I < 8GP \sum_{k=1}^{1/4\sqrt{f}} \frac{k}{(2k-1)^\alpha}$$

When $\alpha > 2$, the above summation can be bounded as follows

$$\sum_{k=1}^{\infty} \frac{k}{(2k-1)^{\alpha}} = \frac{1}{2^{\alpha}} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^{\alpha-1}} + \frac{1}{2^{\alpha+1}} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^{\alpha}} \tag{5.28}$$

$$\leq \frac{1}{2^{\alpha}} \left(\frac{1}{(1/2)^{\alpha-1}} + \int_{1/2}^{\infty} \frac{dx}{x^{\alpha-1}}\right) + \frac{1}{2^{\alpha+1}} \left(\frac{1}{(1/2)^{\alpha}} + \int_{1/2}^{\infty} \frac{dx}{x^{\alpha}}\right) \tag{5.29}$$

$$= 1 + \frac{1}{4}\left(\frac{1}{\alpha-2} + \frac{1}{\alpha-1}\right) \tag{5.30}$$

which concludes the proof. □

Proof of Lemma 5.2.2. Consider figure 5.4. In the simple strategy of [1], each node simply quantizes the observations with rate Q bits per observation. Let the observations be encoded independently with a distortion constraint Δ^2. Since each observation is $\mathcal{N}_{\mathbb{C}}(0, P^{rec})$, Q must satisfy

$$Q > \log\left(1 + \frac{GPb^2 + N}{\Delta^2}\right) \tag{5.31}$$

Now consider the quantized MIMO channel which can be written as

$$\widehat{Y}^M = HX^M + Z + D \tag{5.32}$$

and $D \sim \mathcal{N}_{\mathbb{C}}(0, \Delta^2 I)$. The mutual information of this channel with CSI at receiver is given by $\mathcal{I}(X^M; Y^M, \mathbf{H})$ which can be written as $\log\det\left(I + \frac{1}{N'} H Q_X H^*\right)$ where $N' = N + \Delta^2$ (Noise and distortion are assumed to be uncorrelated). When H varies in a stationary ergodic manner, in general Q_X is chosen to maximize the expectation. Recall that in our model, H varies according to a stationary ergodic process, and elements of H are independent with mean zero, and different variances, such that the distributions of real and imaginary parts of the elements of H are symmetric around the origin. In this case, this is a well known result that the optimal Q_X must be diagonal. In other words, independent signaling can achieve the capacity. Now, consider the strategy of [1] when the elements of transmitted vector X^M are i.i.d $\sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, i.e., nodes use the same power $\sigma^2 = P\frac{r_{SD}^{\alpha}}{M}$. In this case the mutual information is given by

$$\mathcal{I}(X^M; Y^M, \mathbf{H}) = \mathbb{E}\left[\log\det\left(I + \frac{\sigma^2}{N'} HH^*\right)\right]$$

Define ρ_{ik} as $(\frac{r_{SD}}{r_{ik}})^{\alpha/2}$, then the above mutual information can be written as

$$\mathbb{E}\left[\log\det\left(I + \frac{\text{SNR}}{M} FF^*\right)\right]$$

where SNR $=\frac{GP}{N}$ and $F_{ik} = \rho_{ik}\exp(j\theta_{ik})$. Let $\lambda_1,\ \lambda_2,\ \cdots\lambda_M$ be the M eigenvalues of $\frac{1}{M}FF^*$, then the mutual information is given by

$$\begin{aligned}
\mathcal{I}(X^M;Y^M,\mathbf{H}) &= \sum_{k=1}^{M}\mathbb{E}\left[\log\left(1+\text{SNR}\lambda_k\right)\right] \\
&\leq M\mathbb{E}\left[\log\left(1+\frac{\text{SNR}}{M}\sum_{k=1}^{M}\lambda_k\right)\right] \\
&\leq M\log\left(1+\frac{\text{SNR}}{M}\mathbb{E}\left[\sum_{k=1}^{M}\lambda_k\right]\right) \\
&= M\log\left(1+\frac{\text{SNR}}{M^2}\mathbb{E}\left[\text{Tr}\left(FF^*\right)\right]\right)
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}\left[\text{Tr}\left(FF^*\right)\right] &= \sum_{i,k=1}^{M}\mathbb{E}\left[|F_{ik}|^2\right] \\
&= \sum_{i,k=1}^{M}\rho_{ik}^2 \leq M^2b^2
\end{aligned}$$

Here, we have used the fact that $a \leq \rho_{ik} \leq b$, and a and b are given by (5.7). Therefore the mutual information can be upper-bounded by

$$\mathcal{I}(X^M;Y^M,\mathbf{H}) \leq M\log\left(1+\text{SNR}b^2\right)$$

This upper bound, along with the lower bound of Lemma 4.3 of [1], yields the given bounds on the maximum achievable rate $R_1 = L/C_1$. The destination can decode the message with an average error probability arbitrary close to zero if $2^{LM+\epsilon} = 2^{C_1\mathcal{I}(X^M;Y^M,\mathbf{H})}$ for any $\epsilon > 0$ as $C_1 \to \infty$. □

Proof of Lemma 5.3.2. The proof is based on the following Lemma:

Lemma 5.6.1. *Consider a set A of cardinality m and define $\Phi_d(m) = \sum_{i=0}^{d}\binom{m}{d}$ for $d \geq 0$. If the VC dimension of $\mathcal{F}$ is d, then*

(i) $|Proj_A(\mathcal{F})| \leq \Phi_d(m)$

(ii) $\Phi_d(m) \leq 2\left(m^d/d!\right) \leq (em/d)^d$ *for all* $m \geq d \geq 1$.

Proof of part (i): We show that for any $\mathcal{H} \subset 2^A$ that has VC dimension d, $|\mathcal{H}| \leq \Phi_d(m)$. Letting $\mathcal{H} = Proj_A(\mathcal{F})$, we get the result. The proof of the latter is based on induction. Consider any point $x \in A$. Define the following sets:

$$\mathcal{H} - x = \{H - \{x\} : H \in \mathcal{H}\}$$

$$\mathcal{H}^{(x)} = \{H \in \mathcal{H} : x \notin H,\ H \cup \{x\} \in \mathcal{H}\}$$

Note that $\mathcal{H}-x$ and $\mathcal{H}^{(x)}$ are families of subsets of A and that $|\mathcal{H}| = |\mathcal{H} - x| + \left|\mathcal{H}^{(x)}\right|$. Obviously $\mathcal{H} - x$ has VC dimension at most d and therefore $|\mathcal{H} - x| \leq \Phi_d(m-1)$. If we prove $\mathcal{H}^{(x)}$ has VC-dimension less than $d-1$, then the lemma follows since $\Phi_d(m-1) + \Phi_{d-1}(m-1) = \Phi_d(m)$. The VC-dimension of $\mathcal{H}^{(x)}$ is at most $d-1$ since if its VC-dimension is d, there exists a set $B \subset A - \{x\}$ such that it is shattered by $\mathcal{H}^{(x)}$ and $|B| = d$. But in this case, $B \cup \{x\} \subset A$ can be shattered by $\mathcal{H}$; it means that VC-dimension of $\mathcal{H}$ is $d+1$ which is impossible.

Proof of part (ii): The second inequality of part (ii) is based on Stirling's approximation for $d!$ and the proof of the first inequality is by induction on d and m.

To prove Lemma 5.3.2, suppose A has m elements. Note that, according to the above lemma, $|Proj_A(\mathcal{F})| \leq \Phi_d(m)$. Every set in $Proj_A(\mathcal{F}_{\cap r})$ is of the form $\bigcap_{i=1}^{r} A_i$ where $A_i \in Proj_A(\mathcal{F})$. This shows that $|Proj_A(\mathcal{F}_{\cap r})| \leq \binom{|Proj_A(\mathcal{F})|}{r}$. If $|Proj_A(\mathcal{F}_{\cap r})| < 2^m$, A can not be shattered by $\mathcal{F}_{\cap r}$. Therefore, by part (ii) of the lemma, it suffices to choose m such that

$$\binom{\Phi_d(m)}{r} \leq \left(\frac{em}{d}\right)^{dr} < 2^m$$

which is satisfied when $m = 2dr\log(3r)$. This concludes the proof. □

Chapter 6

Summary and Future Research

6.1 Summary

With the recent advances in wireless networks, it has become of fundamental importance to quantify the limitations of information transfer in these networks. Moreover, there have been much interest in designing appropriate architectures for the efficient operation of wireless networks. In this thesis, we have presented a theory to address those issues by using tools from information theory and networking. We have studied the performance of the multihop strategy by using measures such as transport capacity, which is a distance weighted capacity, and throughput. We have provided sharp estimates on the bast achievable transport capacity in terms of its growth rate as a function of number of nodes in the network. In addition, we have studied the maximum throughput which can be supported by random networks. The constructive lower bounds presented to achieve the scaling orders, can shed insight on the design of multihop networks, as well as the role of protocols.

However, the multihop analysis imposes a restriction of the operation of network, ignoring other possible opportunities such as multiuser coding and estimation, cooperative transmission, interference subtraction, and etc. Therefore, one is forced to turn to the Shannon's fundamental work for the study of communication. But, despite all the development and significant applications of information theory in point-to-point communication, progress in network information theory is hindered by intractability of a general solution for some simple networks like three-node relay channel or four-node interference channel.

To overcome this difficulty, we have considered the scaling of the performance measures as the number of nodes grows. The scaling results revealed an interesting connection between the physical properties of wireless medium, like signal attenuation with distance, and the network's capacity measures. Consequently, an interesting dichotomy was observed between two separate cases of high attenuation and low attenuation. For the high attenuation regime, the sub-optimality of multihop was shown for some load-balanced scenarios. Actually, there is a energy cost in joules per one bit-meter information transfer, which can be achieved by using multihop. For the low attenuation regime, a better energy transfer is possible via long-distance transmissions such as MIMO and CRIS.

To bridge the gap between the two regimes, we have confined the study and focused on the throughput of random networks in more details. For this special class of networks, the upperbounds on the throughput were obtained for the whole range of attenuation coefficients. The results identified an expected dichotomy between high and low attenuation cases, along with a sharp threshold to distinguish these two cases. The sub-optimality of multihop was re-established for the high attenuation case while predicting a better throughput for the low attenuation. To achieve the upperbound for the low attenuation case, a hierarchical cooperation scheme was studied which was able to continually increase the scaling by adding more hierarchical stages and incorporating long-range MIMO communications with local cooperations. The scheme was analyzed and optimized by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. In addition, to apply the hierarchical cooperation scheme to random networks, a clustering algorithm was developed. Based on the expression of the maximum achievable throughput, it was found that, actually, there is a gap between the throughout scaling of the hierarchical scheme and the presented upper bound, and, clearly, much remains to be done.

6.2 Future Research

Obviously, the problem is still open for a general network with n nodes in area A. Despite the given upperbound on the throughput of the extended networks, further research is needed to bridge the gap between the low and high attenuation cases for general networks. Moreover, an interesting problem is that of designing more efficient strategies for operation in the low attenuation. The bursty modification of the hierarchical scheme for general networks is not necessarily optimal. In fact, it is possible to do better by exploiting the available silent period to reduce the mutual interference between parallel operating clusters.

There is an important unaddressed issue that whether the scaling laws are just the artifacts of the unrealistic channel modelings, or they indeed have physical meanings. In a recent work [23], it has been claimed that the capacity scaling of wireless networks is subject to a fundamental limitation which is independent of power attenuation and fading models; It is a degrees of freedom limitation which is due to maxwell's equations. The key part of the argument is as follows. Consider the cut-set of Figure 4.1 for the extended networks. The number of independent channels from S to $D - V_D$ is at most an order of $\sqrt{n} \log n$ dictated by physics of wave propagation, contrary to Theorem 4.3.1. We believe that the argument is premature, and much more research is needed to be done for a comprehensive electromagnetic-information theoretic approach.

As a final remark, we underline that some of the results presented in this thesis, do not apply to fixed size networks. Furthermore, some of the preconstants in scaling order results need to be sharpened and explicitly expressed as a function of network parameters. Apparently, we are far from reaching a network information theory for networks with any number of nodes.

References

[1] A. Ozgur, O. Leveque, and D. Tse, Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks, *IEEE Trans. on Information Theory*, vol 53, no. 10, pp. 3549-3572, October 2007. 3, 31, 33, 35, 38, 46, 47, 54, 55, 56, 57

[2] E. C. van der Meulen, Three-terminal communication channels, *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971. 1

[3] P. Gupta and P. R. Kumar, The capacity of wireless networks, *IEEE Trans. on Information Theory*, vol. 46, pp. 388- 404, March 2000. 1, 2, 17, 49

[4] L.-L. Xie and P. R. Kumar, A network information theory for wireless communication: Scaling laws and optimal operation, *IEEE Trans. on Information Theory*, vol. 50, no. 5, pp. 748-767, February 2004. 2, 24, 25

[5] A. Jovicic, P. Viswanath, and S. R. Kulkarni, Upper bounds to transport capacity of wireless networks, *IEEE Transactions on Information Theory*, vol. 50, pp. 2555–2565, November 2004.

[6] O. Leveque and E. Telatar, Information theoretic upper bounds on the capacity of large extended ad hoc wireless networks, *IEEE Trans. Inform. Theory*, pp. 858–865, vol. 51, no. 3, March 2005. 31

[7] F. Xue, L.-L. Xie, and P. R. Kumar, The transport capacity of wireless networks over fading channels, *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 834–847, March 2005.

[8] L.-L. Xie and P. R. Kumar, On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks, *Joint Special Issue of IEEE Trans. on Information Theory and IEEE/ACM Trans. on Networking and Information Theory*, vol. 52, no. 6, pp. 2313-2328, June 2006. 25

[9] S. Ahmad, A. Jovicic, and P. Viswanath, Outer bounds to the capacity region of wireless networks, *IEEE Trans. on Information Theory*, vol. 52, pp. 2770-2776, June 2006. 30

[10] M. Franceschetti, O. Dousse, D. Tse and P. Thiran, Closing the gap in the capacity of wireless networks via percolation theory, *IEEE Trans. on Information Theory*, vol. 53, no. 3, pp. 1009-1018, March 2007. 2, 17

[11] S. Aeron and V. Saligrama, Wireless ad hoc networks: strategies and scaling laws for the fixed SNR regime, *IEEE Trans. on Information Theory*, vol. 53, no. 6, pp. 2044-2059, June 2007. 3, 54

[12] A. Agarwal and P. R. Kumar, Capacity bounds for ad hoc and hybrid wireless networks, *ACM SIGCOMM Computer Communications Reveiew*, vol. 34, no. 3, pp. 71-81, July 2004. 17

[13] F. Xue and P. R. Kumar, Scaling laws for Ad Hoc wireless networks: An Information Theortic Approach, *Foundation and Trends in Networking*, vol. 1, No. 2, pp. 145-270, 2006. 17

[14] S. H. A. Ahmad, A. Jovicic, and P. Viswanath, Upper bounds to generalized transport capacity of wireless networks, *IEEE Trans. on Information Theory*, vol. 52, no. 6, June 2006. 25

[15] P. Gupta and P. R. Kumar, Critical power for asymptotic connectivity in wireless networks, *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, pp. 547-560, Boston, MA: Birkhauser, March 1998. 15, 16

[16] J. Ghaderi, L.-L. Xie, and X. Shen, Throughput optimization for hierarchical cooperation in Ad Hoc networks, *Proc. 2008 IEEE International Conference on Communications*, Beijing, China, May 19-23, 2008. 54

[17] J. Ghaderi, L.-L. Xie and X. Shen, Hierarchical cooperation in ad hoc networks: optimal clustering and achievable throughput, submitted to *IEEE Trans. on Information Theory*, Feb. 2008. 54

[18] U. Niesen, P. Gupta, and D. Shah, On capacity scaling in arbitrary wireless networks, *submitted to IEEE Trans. on Information Theory*, November 2007. 1, 31

[19] V. N. Vapnik and A. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264280, 1971. 48

[20] V. N. Vapnik, Estimation of dependences based on empirical data. New York: Springer-Verlag, 1982. 48

[21] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, *Journal of ACM*, vol. 36, no. 4, pp. 929-965, 1989. 49

[22] A. Ephremides and B. Hajek, Information theory and communication networks: an unconsummated union, *IEEE Trans. on Information Theory,*vol. 44, pp. 2416-2434, Oct. 1998. 1

[23] M. Franceschetti, M.D. Migliore, P. Minero, The capacity of wireless networks: information-theoretic and physical limits *Allerton Conference on communication, control and computing*, 2007. 61

[24] S. A Jafar and A. Goldsmith, Multiple-antenna capacity in correlated rayleigh fading with channel covariance information, *IEEE Trans. on Wireless Communications*, vol. 4, no. 3, pp. 990-997, May 2005. 30

[25] A. M. Tulino, A. Lozano and S. Verdu, Capacity-achieving input covariance for correlated multi-antenna channels, *Forty-first Annual Allerton Conference on communication, control and computing*, 2003. 30

[26] V. V. Veeravalli, Y. Liang and A. M. Sayeed, Correlated MIMO rayleigh fading fhannels: capacity, optimal signaling, and scaling laws, *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 2058-2072, June 2005. 30

[27] Elements of Information Theory, T. M. Cover and J. A. Thomas, New York: Wiley, 1991. 20

[28] Fundamentals of wireless communications, Cambridge University Press, 2005. 28, 29

[29] C. A. Balanis, Antenna Theory: Analysis and Design, John Wiley and Sons, 2005. 13