MAXIMIZING DEGREES OF FREEDOM IN WIRELESS NETWORKS

by

Shashibhushan Prataprao Borade

B.Tech. Indian Institute of Technology, Bombay

(2002)

Submitted in partial fulfillment
of the requirements for the
Degree of

MASTER OF SCIENCE
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2004

Author ....
Department of Electrical Engineering and Computer Science
May 21, 2004

Certified by ....

Lizhong Zheng

Thesis Supervisor

Certified by ....

Robert Gallager

Thesis Co-supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

MAXIMIZING DEGREES OF FREEDOM IN WIRELESS NETWORKS

by

Shashibhushan Prataprao Borade

Submitted on May 21, 2004
in Partial Fulfillment of the Requirements for the Degree of
Master of Science
in Electrical Engineering and Computer Science

Abstract

We consider communication from a single source to a single destination in a wireless network with fading. Both source and destination have multiple antennas. The information reaches the destination through a sequence of layers of single-antenna relays. A *non-separation-based* strategy is proposed and shown to achieve a rate equal to the capacity of a point-to-point multiantenna system in the high SNR regime. This implies that lack of coordination between relay nodes does not reduce the achievable rate at high SNR. We then derive the tradeoffs between network size and rate. We also derive the rate-diversity tradeoff for this network and study how it is affected by the network size. This shows that increasing network size is much more difficult when the codelength does not span a large number of fading realizations. Finally some implications to ad-hoc networks are discussed.

Thesis Supervisor: Lizhong Zheng
Title: Assistant Professor, Electrical Engineering and Computer Science

Thesis Co-supervisor: Robert Gallager
Title: Professor, Electrical Engineering and Computer Science

Acknowledgments

I am indebted to my advisors Professor Lizhong Zheng and Professor Robert Gallager for their insightful guidance and care. It is an honor and a great pleasure to be advised by them. I am looking forward to more of this bliss. I wish to acknowledge the infinite patience shown by Lizhong towards my pace of understanding and my frequency of knocking on his office. Merely watching Bob's mind in action has been a tremendous learning and inspiring experience.

I am deeply indebted to Professors Emre Telatar and Rüdiger Urbanke for their support and help throughout. I thank Professors Muriel Médard and Gregory Wornell for their valuable time and care–especially in my initial days at MIT. I thank Professors David Forney, Sanjoy Mitter and Moe Win for their encouragement and guidance.

I am indebted to all my friends in LIDS and MIT for their camaraderie and the numerous tea-sessions. A partial list includes Emmanuel Abbe, Cemal Akçaba, Sreekar Bhaviripudi, Li-Wei Chen, Amit Deshpande, Shan-Yuan Ho, Masha Ishutkina, Amir Khandani, Ashish Khisti, Desmond Lun, Baris Nakiboglu, Prof. Mike Neely, Siddharth Ray, Anand Srivavas, Jun Sun, Jay-Kumar Sundararajan, Watcharapan Suwansantisuk, Guy Wichenberg, and Murtaza Zafer. Longtime buddies like Amit Chandak, Rohan Kekatpure, Swapnil Pawar, Vishwambhar Rathi, Kranthikumar Reddy, Sumitkumar Singh and Yogesh Wakchaure are a treasure indeed.

Words can never express my feelings towards my parents, Mrs. Shashikala Borade and Professor Pratap Borade, for their love, encouragement and sacrifice. This thesis is dedicated to them. I express a feeling of bliss for having a loving sister like Dr. Mrunmayee Vikhe. I thank my little niece Tanishka for being so cute and playful. I am grateful to my brother-in-law Dr. Vikram Vikhe for his affection. I owe a lot to Bhau-Babai, all my Kaka-Kakus (specially Dattakaka-Mangaltai), Mama-Mamis, Atya-Mawashis and cousins (and Uttam Pawar) for their love and caring.

This work was supported by an MIT Presidential Fellowship. This funding is greatly appreciated.

Contents

List of Figures

Chapter 1

Introduction

The seminal result by Gupta and Kumar [3] implies that the sum rate of communication in an ad-hoc wireless network with uniform traffic cannot grow linearly with the number of nodes in the network. The same result was proved in a more general communication framework in [4, 5].

Gupta and Kumar also analyzed a strategy of passing the message from source to destination through a sequence of relay nodes. Each relay node in the sequence first decodes the message fully and then transmits it to the next relay node in the sequence (Fig 1-1). This strategy of multi-hop through relay nodes was shown to be essentially almost optimal. We call such a strategy *separation-based*, where every relay node decodes the complete message before transmitting it. Equivalently in a *separation-based* strategy, reliable communication is possible to each relay node so that each relay node can fully decode its received signal. Thus in a *separation-based* strategy, the joint problem of communication in the network is separated into two less complex problems. The first problem is how to create reliable links to the relay nodes and the second problem is what should be sent over those reliable links. This is the reason why such a strategy is called *separation-based*. This definition of *separation* is also known as channel-network separation and it should not be confused with other common connotations of separation, e.g. source-channel separation.

We also know that multiple antennas provide enormous performance gains in

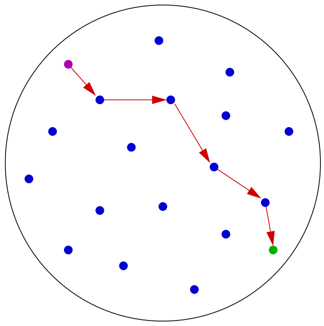

Figure 1-1: Multi-hop through a sequence of relay nodes

wireless systems. In a point-to-point multi-input multi-output (MIMO[1]) system with n transmit and n receive antennas and i.i.d. Rayleigh[2] flat fading, if the receiver knows the channel, then the capacity is approximately $n \log \mathsf{SNR}$ at high SNR [1, 2]. This result by Telatar and Foschini shows that the capacity of this MIMO channel is equal to the sum of the capacities of n parallel single antenna channels. Thus we say that this MIMO channel provides n degrees of freedom to communicate and hence improves the performance significantly. In general, we say that r degrees of freedom are achieved if a rate of $r \log \mathsf{SNR}$ is achieved.

Applying this MIMO idea in a wireless network with fading has the potential to achieve performance gains. A particular example for this claim using a single source-destination pair was given in [6]. There both source and destination use half the relay nodes. These relays are so close to the source or the destination that they act as multiple antennas of a single user (see Fig. 1-2). Now this network is like a point-to-point MIMO system and its capacity grows linearly with the number of antennas. The idea of making the relay nodes act like multiple antennas was also developed in [11].

We would like to explore this type of MIMO performance gain in more general wireless networks. Consider a layered relay network having a single source-destination pair. Let the message be passed from one layer to the next till it reaches the destination (Fig. 1-3). In this thesis, we will assume that this layering has been already

[1]We reserve the word "MIMO" for point-to-point channels with multiple transmit and multiple receive antennas.

[2]This means that all entries of the channel matrix are i.i.d. and circular symmetric complex Gaussian. We call such a random matrix as a "random Gaussian matrix", unless specified otherwise.

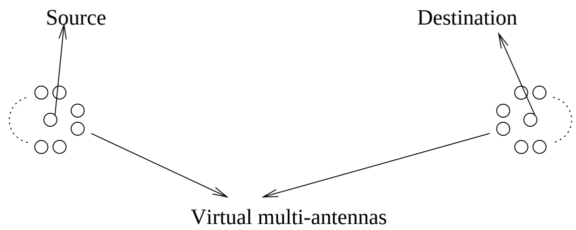

Figure 1-2: Relays acting as multi-antenna

done. We will not be concerned about how to divide a general wireless network into multiple layers and other such issues. Note that the message is now passed through a sequence of relay layers, as opposed to a sequence of relay nodes (Fig. 1-1). We will assume that each layer has n single-antenna relay nodes and the source and destination have n antennas each. Every hop of the message from one layer to its next layer *looks like* a MIMO system.

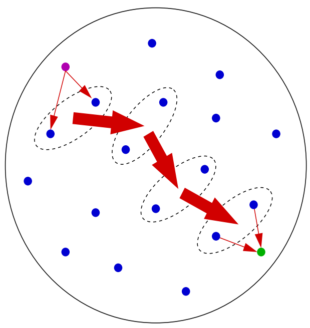

Figure 1-3: Multi-hop through sequence of relay layers

Nevertheless, as we will see later, this transmission between relay layers differs significantly from a point-to-point MIMO system because relay nodes in a layer are at different locations. They cannot coordinate with each other to act like a multiantenna node. This thesis is an attempt to answer the following questions:

How valuable is this coordination between relay nodes?

What is the performance loss if this coordination is lacking?

The next chapter reviews some basic concepts in information theory, which are used in later chapters. Some aspects of coordination between relay nodes are discussed in chapter 3. After describing the exact network model, it discusses the case where only one layer of relay nodes exists between the source and destination. We propose a strategy which achieves the same rate as the capacity of a point-to-point MIMO system in the high SNR regime. In other words, in spite of the lack of coordination between relay nodes, this strategy achieves all the degrees of freedom achievable in a point-to-point MIMO system.

In chapter 4, we study the general case where the number of relay layers can be one or more. The strategy for the single layer case fails to achieve all the degrees of freedom here. A different and surprisingly simple *non-separation-based* strategy is able to achieve all the degrees of freedom in this case.

All these results imply that at high SNR, lack of coordination does not reduce the achievable rate. Nonetheless, the precise meaning of "high SNR" is not made clear by these results. Chapter 5 addresses this issue. It studies how the achievable performance is affected if SNR is not high enough. It also sketches the tradeoff of the achievable rate with an increasing number of relay layers i.e. with increasing network size.

The results till this point were proved in the ergodic formulation. This formulation assumes that a transmitted codeword spans a large number of fading blocks i.e. faces a large number of channel realizations. If it is not possible, we study the outage formulation in chapter 6. In the outage formulation, we sketch the rate-diversity tradeoff for this network (like the rate-diversity tradeoff for the point-to-point MIMO channel found by Zheng and Tse [15]). We also show how this tradeoff gets worse with increasing network size. In other words, we plot the three dimensional tradeoff between rate, diversity and network size. It turns out that increasing the network size is much costlier in the outage formulation than in the ergodic formulation. In other words, increasing network size is much costlier when the codelength does not span a large number of fading realizations. Finally in chapter 7, we discuss some extensions of the derived results. The significance of these results for ad-hoc networks is also discussed.

Chapter 2

Background

This chapter reviews some pertinent results in information theory and multiantenna systems.

2.1 Coding Theorem for Point-to-Point Channels

Consider a discrete time communication channel whose input at any given time $i \in \mathbb{Z}$ is a continuous-valued random variable x_i and whose output is another continuous-valued random variable y_i. We denote the vector of channel inputs $[x_1, \cdots, x_N]^T$ by x^N; y^N is defined similarly. The channel behavior is completely described by the conditional probability density function $p(y^N = \tilde{y}^N | x^N = \tilde{x}^N)$ for all sample values $\tilde{x}^N$, $\tilde{y}^N$ and all N. If the channel is assumed to be memoryless and stationary, this decomposes into the product form $\Pi_{i=1}^N \, p_{y|x}(\tilde{y}_i|\tilde{x}_i)$, where $p_{y|x}(\tilde{y}|\tilde{x})$ denotes the stationary conditional pdf of the channel evaluated at $\tilde{x}$ and $\tilde{y}$, which are sample values or realizations of x and y. The channel is completely described by $p_{y|x}(\tilde{y}|\tilde{x})$ in this case. Let the set of all possible $\tilde{y}$ and $\tilde{x}$ be denoted by Ω. For example, Ω can be the real line or the complex plane.

The mutual information for this channel for a given input pdf $p_x(.)$ is defined as[1]

$$\mathcal{I}(y;x) = \int_\Omega \int_\Omega p_{yx}(\tilde{y},\tilde{x}) \log \frac{p_{y|x}(\tilde{y}|\tilde{x})}{p_y(\tilde{y})} \, d\tilde{y} \, d\tilde{x} \tag{2.1}$$

[1]The log function is assumed to have base 2 unless stated otherwise.

This can be also written in terms of differential entropies. The differential entropies are defined as follows:

$$
\begin{aligned}
h(y) &= \int_\Omega p_y(\tilde{y}) \log \frac{1}{p_y(\tilde{y})}\ d\tilde{y} \\
h(x) &= \int_\Omega p_x(\tilde{x}) \log \frac{1}{p_x(\tilde{x})}\ d\tilde{x} \\
h(y,x) &= \int_\Omega \int_\Omega p_{yx}(\tilde{y},\tilde{x}) \log \frac{1}{p_{yx}(\tilde{y},\tilde{x})}\ d\tilde{x}\ d\tilde{y}
\end{aligned}
$$

In a similar manner, the conditional differential entropies are defined as

$$
\begin{aligned}
h(y|x) &= \int_\Omega \int_\Omega p_{yx}(\tilde{y},\tilde{x}) \log \frac{1}{p_{y|x}(\tilde{y}|\tilde{x})}\ d\tilde{y}\ d\tilde{x} \\
&= \int_\Omega p_x(\tilde{x}) \left[\int_\Omega p_{y|x}(\tilde{y}|\tilde{x}) \log \frac{1}{p_{y|x}(\tilde{y}|\tilde{x})}\ d\tilde{y} \right]\ d\tilde{x} \\
&= \mathcal{E}_x \left[h\left(y|(x=\tilde{x}) \right) \right]
\end{aligned}
$$

Similarly, $h(x|y)$ is defined. Then, $\mathcal{I}(x;y) = \mathcal{I}(y;x) = h(y) - h(y|x)$. We now define the mutual information between two random variables x and y conditioned on a third random variable z as

$$
\begin{aligned}
\mathcal{I}(x;y|z) = h(y|z) - h(y|x,z) &= \mathcal{E}_z \left[h(y|(z=\tilde{z})) - h(y|(z=\tilde{z}),x) \right] && (2.2) \\
&= \mathcal{E}_z \left[\mathcal{I}(x;y|(z=\tilde{z})) \right] && (2.3)
\end{aligned}
$$

Note that for a given channel $p_{y|x}(.|.)$, the mutual information $\mathcal{I}(x;y)$ is fixed for each input pdf $p_x(.)$. Now define the channel capacity in bits per channel use as follows

$$
C = \max_{p_x(.)}\ \mathcal{I}(y;x) = \max_{p_x(.)}\ \left[h(y) - h(y|x) \right] \qquad (2.4)
$$

In his monumental work [12], Shannon showed that channel capacity defined as above is the long term maximum number of bits per channel use that can be transmitted reliably in the sense that the probability of decoding error can be made arbitrarily small. However the codewords may need to be very long to achieve an arbitrarily small probability of decoding error. Reliable communication is not possible at any rate $R > C$. To be precise, for any $R > C$, there is some $\epsilon(R) > 0$ such that the probability of decoding error cannot be made smaller than $\epsilon(R)$ for any code of any length of rate R.

We explain the above results with the example of additive white gaussian noise (AWGN) channel. The destination receives

$$y = x + w$$

where x is the channel input and w is Gaussian noise with variance σ^2, i.i.d. over time. The transmitter has an average power constraint $\mathcal{E}\left[x^2\right] \leq P$. Thus $p_{y|x}(y|x)$ is given by $\frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-(y-x)^2}{2\sigma^2})$. We can choose any $p_x(x)$ satisfying the power constraint and generate a valid random code. For any input pdf $p_x(x)$, the conditional entropy $h(y|x)$ is simply the (differential) entropy of the Gaussian noise given by,

$$h(y|x) = h(w) = \frac{1}{2}\log(2\pi e\sigma^2)$$

Thus maximizing $\mathcal{I}(x;y)$ reduces to maximizing $h(y)$. As the noise w is independent of the input x, the power in y is

$$\mathcal{E}\left[y^2\right] = \mathcal{E}\left[x^2\right] + \mathcal{E}\left[w^2\right] \leq P + \sigma^2$$

If x is chosen to be a zero mean Gaussian random variable with variance P, then y is a zero mean Gaussian random variable with variance $P + \sigma^2$. The Gaussian pdf maximizes the differential entropy under a power constraint. Hence the maximum possible $h(y)$ is equal to $h(y) = \frac{1}{2}\log\left(2\pi e(P + \sigma^2)\right)$. Thus capacity of this AWGN channel is achieved by the Gaussian input distribution with variance P. It is given by

$$C_{\text{AWGN}} = \frac{1}{2}\log(1 + P/\sigma^2) = \frac{1}{2}\log(1 + \mathsf{SNR})$$

where SNR is defined as the signal to noise power ratio P/σ^2. Similarly we can prove that the capacity of a complex AWGN channel is equal to $\log(1+\mathsf{SNR})$ since a complex AWGN channel is equivalent to two real AWGN channels.

2.2 Multiantenna Systems

2.2.1 Capacity Results

Multiple antennas at source and/or destination can provide significant performance gains. Consider the following MIMO channel in a fading environment, where the transmitted vector $\mathbf{x} \in \mathbb{C}^m$ is related to the received vector $\mathbf{y} \in \mathbb{C}^n$ as follows:

$$\mathbf{y} = \mathsf{H}\mathbf{x} + \mathbf{w} \tag{2.5}$$

where H, which denotes the fading state is an $n \times m$ matrix with i.i.d. complex Gaussian entries with unit variance. It is independent of the input and the noise $\mathbf{w}$. Each entry of the additive noise $\mathbf{w}$ is a zero mean complex Gaussian random variable with variance σ^2, i.e. its real and imaginary parts are real Gaussian variables of variance $\sigma^2/2$ independent of each other. Moreover, we assume that entries of this vector are independent of each other so that[2] $\mathcal{E}\left[\mathbf{w}\mathbf{w}^\dagger\right] = \sigma^2 I$. The input has a total power constraint given by $\mathcal{E}\left[\text{trace}(\mathbf{x}\mathbf{x}^\dagger)\right] \leq P$.

We assume an i.i.d. block fading model for H. This fading model means that the channel state H remains constant during a block of T_c symbols and changes to a new independent realization in the next block. Each such block of length T_c is called the fading block and T_c is usually called the coherence interval. The receiver is assumed to know H (i.e. the realization of this random matrix), but the transmitter does not know the channel. Thus in effect the output of this system at the receiver is $(\mathbf{y}, \mathsf{H})$.

[2] The symbol I is reserved for identity matrices, whose size will be clear by the context.

The mutual information between the input and output is:

$$\begin{aligned}
\mathcal{I}(\mathbf{x};(\mathbf{y},\mathsf{H})) &= \mathcal{I}(\mathbf{x};\mathsf{H})+\mathcal{I}(\mathbf{x};\mathbf{y}|\mathsf{H}) \\
&= 0+\mathcal{I}(\mathbf{x};\mathbf{y}|\mathsf{H}) \\
&= \mathcal{E}_{\mathsf{H}}\left[\mathcal{I}(\mathbf{x};\mathbf{y}|(\mathsf{H}=\tilde{\mathsf{H}}))\right] \quad \text{from Eq. 2.2} \qquad (2.6) \\
&= \mathcal{E}_{\mathsf{H}}\left[h(\mathbf{y}|(\mathsf{H}=\tilde{\mathsf{H}}))-\mathbf{h}(\mathbf{y}|(\mathsf{H}=\tilde{\mathsf{H}}),\mathbf{x})\right] \\
&= \mathcal{E}_{\mathsf{H}}\left[h(\mathbf{y}|(\mathsf{H}=\tilde{\mathsf{H}}))-\mathbf{h}(\mathbf{w})\right]
\end{aligned}$$

As in the AWGN case, $h(\mathbf{w})$ is independent of the input distribution. It equals $n\log(\pi e\sigma^2)$ for any H. Hence to maximize the mutual information, we simply maximize $\mathcal{E}\left[h(\mathbf{y}|\mathsf{H}=\tilde{\mathsf{H}})\right]=h(\mathbf{y}|\mathsf{H})$.

Now we prove that it is maximized by the complex (jointly) Gaussian distribution of $\mathbf{x}$ satisfying $\mathcal{E}\left[\mathbf{x}\mathbf{x}^\dagger\right]=\frac{P}{m}I$. Let the covariance matrix of the input $\mathbf{x}$ be denoted by K_{xx}. Then the covariance matrix of the output is given by

$$\mathcal{E}\left[\mathbf{y}\mathbf{y}^\dagger|\mathsf{H}=\tilde{\mathsf{H}}\right]=\tilde{\mathsf{H}}K_{xx}\tilde{\mathsf{H}}^\dagger+\sigma^2 I$$

For a given value of the covariance matrix above, the entropy $h(\mathbf{y}|\mathsf{H}=\tilde{\mathsf{H}})$ is maximized when $\mathbf{y}$ is a jointly Gaussian random vector with that covariance matrix. Choosing $\mathbf{x}$ to be a jointly Gaussian will cause $\mathbf{y}$ also to be jointly Gaussian (conditioned on the channel realization). Now remains the search of the optimal K_{xx} that maximizes $h(\mathbf{y}|\mathsf{H})$. Let the eigenvalue decomposition of K_{xx} be $UDU^\dagger$. The differential entropy of a complex Gaussian vector with covariance matrix Q is given by $\log\det(\pi eQ)$, hence

$$\begin{aligned}
h(\mathbf{y}|\mathsf{H}) &= \mathcal{E}_{\mathsf{H}}\left[\log\det(\pi e(\mathsf{H}K_{xx}\mathsf{H}^\dagger+\sigma^2 I))\right] \\
&= \mathcal{E}_{\mathsf{H}}\left[\log\det\left(\pi e\cdot((\mathsf{H}U)D(\mathsf{H}U)^\dagger+\sigma^2 I)\right)\right]
\end{aligned}$$

Note that distribution of $\mathsf{H}U$ is same as that of H for any unitary matrix U. Hence our search for optimal K_{xx} can be limited to non-negative diagonal matrices satisfying the power constraint i.e. whose trace equals P. Given any diagonal input covariance matrix D_1 and a permutation matrix Π, $h(\mathbf{y})$ is unchanged if the input covariance matrix is changed to $\Pi D_1 \Pi^\dagger$. It is because $\mathsf{H}\Pi$ has the same distribution as that of H. Now applying the concavity of $\log\det(\cdot)$ function on the set of positive definite matrices, we get the optimal K_{xx} by averaging all the permutations of a diagonal matrix satisfying the power constraint. Hence the optimal input distribution is jointly Gaussian with covariance matrix $\frac{P}{m}I$.

Intuitively, distributing the power uniformly in all directions is optimal because the transmitter does not know the channel H, so all the directions are equally good. Now by Eq. 2.6, we get the following expression for capacity of this MIMO channel [1]

$$C_{\mathrm{MIMO}} = \mathcal{E}_{\mathsf{H}}\left[\log\det\left(I + \frac{P}{m\sigma^2}\mathsf{H}\mathsf{H}^\dagger\right)\right] \tag{2.7}$$

$$= \sum_{i=1}^{n} \mathcal{E}_{\mathsf{H}}\left[\log\left(1 + \mathsf{SNR}\lambda_i^{\mathsf{H}}\right)\right] \tag{2.8}$$

where λ_i^{H} is the ith smallest eigenvalue of $\mathsf{H}\mathsf{H}^\dagger$ and $\mathsf{SNR} \triangleq P/m\sigma^2$ is the ratio of aggregate power transmitted by the transmit antennas to the aggregate noise power. Note that many outputs $(\mathbf{y}, \mathsf{H})$ and thus many channel realizations are needed for achieving this capacity reliably. More precisely, for any $R < C_{\mathrm{MIMO}}$ and any given $\epsilon > 0$, the error probability can be made smaller than ϵ if the codelength spans a large enough number of channel realizations.

Compare this with the AWGN case, where one needed to code over many noise realizations. Essentially, it ensured correct decoding even if the noise in some part of the codeword is large. Intuitively, a codeword could be decoded correctly by virtue of its part where the distortion by the channel (noise) is small. In the case of fading channels, there are two possible causes of distortion by the channel: fading and noise. By having long enough codelength, the effects of the additive noise are averaged out as in the case of AWGN channel. Moreover, coding over many fading realizations essentially ensures that some part of the codeword sees good channel gains. This enables correct decoding even if some parts of the codeword see bad channel gains. This capacity is called the ergodic capacity to emphasize the need to code over many fading realizations.

The case of $\mathsf{SNR} \gg 1$, where signal power is much larger than noise power, is called the high SNR scenario. Consider the case $m = n$, where the transmitter has the same number of antennas as the receiver.

$$\begin{aligned}
\frac{C_{\mathrm{MIMO}}}{\log \mathsf{SNR}} &\geq \sum_{i=1}^{n} \mathcal{E}_{\mathsf{H}}\left[\frac{\log(\mathsf{SNR}\cdot\lambda_i^{\mathsf{H}})}{\log \mathsf{SNR}}\right] \\
&= \sum_{i=1}^{n} \mathcal{E}_{\mathsf{H}}\left[1 + \frac{\log \lambda_i^{\mathsf{H}}}{\log \mathsf{SNR}}\right] \\
&= n + \frac{\mathcal{E}_{\mathsf{H}}\left[\log\det(\mathsf{H}\mathsf{H}^{\dagger})\right]}{\log \mathsf{SNR}}
\end{aligned}$$

Note that $\det(\mathsf{H}\mathsf{H}^{\dagger})$ equals $-\infty$ when $\mathsf{H}\mathsf{H}^{\dagger}$ is a singular matrix. However, this is a zero probability event. It turns out that $\mathcal{E}_{\mathsf{H}}\left[\log\det(\mathsf{H}\mathsf{H}^{\dagger})\right]$ equals a finite number [22]. This is easily verified by noting that $\det(\mathsf{H}\mathsf{H}^{\dagger}) = \prod_{\mathbf{i}} \lambda_{\mathbf{i}}$ is lower bounded by $\lambda_{\min}^{n}$. Given that $\lambda_{\min}$ is exponentially distributed, $\mathcal{E}\left[\log\lambda_{\min}\right]$ is a finite number, which explains the finiteness of $\mathcal{E}\left[\log\det(\mathsf{H}\mathsf{H}^{\dagger})\right]$. Hence the second term in the above equation becomes negligible when SNR is very large and the above lower bound for C_{MIMO} equals n when SNR tends to infinity. Similarly, we upper bound C_{MIMO} when SNR is large.

$$\begin{aligned}
\frac{C_{\mathrm{MIMO}}}{\log \mathsf{SNR}} &\leq \sum_{i=1}^{n} \mathcal{E}_{\mathsf{H}}\left[\frac{\log(\mathsf{SNR} + \mathsf{SNR}\lambda_i^{\mathsf{H}})}{\log \mathsf{SNR}}\right] \\
&= \sum_{i=1}^{n} \mathcal{E}_{\mathsf{H}}\left[1 + \frac{\log\left(1 + \lambda_i^{\mathsf{H}}\right)}{\log \mathsf{SNR}}\right] \\
&\leq n + \frac{n\mathcal{E}_{\mathsf{H}}\left[\log\left(1 + \mathrm{trace}(\mathsf{H}\mathsf{H}^{\dagger})\right)\right]}{\log \mathsf{SNR}} \qquad \text{as any } \lambda_i \leq \sum_{j} \lambda_j = \mathrm{trace}(\mathsf{H}\mathsf{H}^{\dagger}) \\
&\leq n + \frac{n\log\left(\mathcal{E}_{\mathsf{H}}\left[1 + \mathrm{trace}(\mathsf{H}\mathsf{H}^{\dagger})\right]\right)}{\log \mathsf{SNR}} \qquad \text{(Jensen's inequality)} \\
&= n + \frac{n\log(1+n^2)}{\log \mathsf{SNR}} \qquad \text{as } \mathcal{E}\left[\mathrm{trace}(\mathsf{H}\mathsf{H}^{\dagger})\right] \text{ equals } n^2
\end{aligned}$$

The numerator of the second term is a finite number, so the second term becomes negligible when SNR is very large. Thus the upper bound also equals n when SNR tends to infinity. Hence the ratio of C_{MIMO} to $\log \mathsf{SNR}$ is almost n when SNR is large. Thus the ergodic capacity at high SNR equals $n \log \mathsf{SNR} + \zeta(\mathsf{SNR})$, where $\zeta(\mathsf{SNR}) \ll n \log \mathsf{SNR}$. The $\ll$ sign here denotes that the ratio of the two sides goes to zero when SNR tends to infinity. We will also use the notation $C_{\text{MIMO}} \approx n \log \mathsf{SNR}$, where the $\approx$ sign means that the ratio of the sides goes to 1 as SNR goes to infinity. In words, we will say that the capacity of a MIMO channel is *approximately* $n \log \mathsf{SNR}$ *at high SNR.*

Notice that the capacity of a scalar fading channel (i.e. single transmit and receive antenna) is $C_{\text{SISO}} \approx \log \mathsf{SNR}$. The same is true for the capacity of a complex AWGN channel. Thus at high SNR, having n transmit and n receive antennas is equivalent to having n separate single antenna channels. Thus a MIMO system provides n dimensions in space for communication. Hence in MIMO, n degrees of freedom are said to be achieved. In general with m transmit and n receive antennas, $\min(m, n)$ degrees of freedom can be achieved. This is because $n - \min(m, n)$ (out of n) eigenvalues of $\mathsf{H}\mathsf{H}^\dagger$ (in Eq. 2.8) are zero when H is a rectangular matrix.

$$C_{\text{MIMO}} \approx \min(m, n) \log \mathsf{SNR} \tag{2.9}$$

One should note that a code achieving all the $\min(m, n)$ degrees of freedom need not be a capacity achieving code. For example, consider a scalar AWGN channel with $\mathsf{SNR} \gg 1$. Consider a code which uses only half the available power. The maximum achievable rate of this code is $\log(1+\mathsf{SNR}/2)$. Due to the $3dB$ power loss, this code is clearly not a capacity achieving code. However, it achieves all the degrees of freedom because $\log(1 + \mathsf{SNR}/2) \approx \log \mathsf{SNR}$. This example illustrates the coarseness in the measure of degrees of freedom.

Remark: Fading is caused when the signal from the transmitter reaches to the receiver via multiple paths. The constructive and destructive interference between those paths causes the fading. It is worth mentioning that only 1 degree of freedom could be achieved if fading was absent. That is the case when only a single line of sight path exists between each transmitter and receiver antenna. In that case, $\mathsf{HH}^\dagger$ is always singular. All but one of its eigenvalues are zero. Somewhat surprisingly, this means that randomness or fading in the channel can be advantageous as it provides multiple degrees of freedom.

A capacity achieving code spanning L fading blocks of length T_c each (i.e. codelength LT_c) has $2^{LT_c C_{\mathrm{MIMO}}}$ different codewords. For large LT_c, this may lead to impractically large computational complexity. Even when the computational complexity is not an issue, coding over large number of fading blocks increases the delay. This is especially prominent when T_c is large. Thus achieving ergodic capacity is difficult in practice.

As a small digression, let us see another way to achieve ergodic capacity when the transmitter has some information about the channel state. Assume that the transmitter knows the rate to be transmitted in each fading block, which is equal to $C(\mathsf{H}) = \log(\det(\mathbf{I} + \mathsf{SNR} \cdot \mathsf{HH}^\dagger))$. However it does not know the exact realization of the channel state H.

The communication is done in following manner. Suppose that the transmitter has a different code-book for each rate $R > 0$ and these code-books are known to the receiver[3]. At the beginning of every fading block, the transmitter has a codeword ready for each rate. After knowing $C(\mathsf{H})$, it starts transmitting the codeword of that rate. At the receiver, the channel state is known. Hence the rate of the transmitted codeword and its corresponding code-book is also known. The receiver selects one codeword from this code-book by performing ML decoding on the received signal. Note that the codelength for each rate need not be limited within a fading block. A codeword corresponding to a given rate R can be transmitted in multiple pieces, one piece at each interval when the channel state satisfies $C(\mathsf{H}) = R$.

In this scheme, an average rate of $\mathcal{E}_{\mathsf{H}}\left[C(\mathsf{H})\right] = C_{\mathrm{MIMO}}$ is achieved by choosing long enough codewords and by adjusting the transmission rate in each fading block. Sur-

[3]This requires an infinite number of such code-books. In practice, we use code-books for a large number of rates and essentially achieve the same performance as having code-books for all possible rates.

prisingly, the capacity when the transmitter has some information about the channel state (in terms of $C(\mathsf{H})$) is the same as the capacity when the transmitter has no information about the channel state.

2.2.2 Outage Formulation

We have seen how achieving ergodic capacity is difficult as the codelength needs to span many fading blocks to achieve the ergodic capacity. It may cause large delay and might become infeasible in practice. Hence we need to study the performance of a code which only spans a finite number of fading blocks. In particular, we analyze the performance of codes which span only one fading block. These are defined as "short codes". As shown later, this analysis easily extends to the analysis of codes which span L fading blocks.

To analyze the error probability of these short codes, consider the following non-ergodic channel model for a moment. The channel H is a random Gaussian matrix but it is fixed for all time once chosen. As before, the receiver knows the realization of H but the transmitter does not. Any codeword (of arbitrary length) will face only one channel realization in this model. Conditioned on the channel realization H, the capacity is

$$C(\mathsf{H}) = \log(\det(I + \mathsf{SNR} \cdot \mathsf{H}\mathsf{H}^\dagger))$$

Let us see what is the channel capacity under this non-ergodic fading model. In other words, let us find the maximum rate R for which the error probability can be made arbitrarily small. For any fixed rate $R > 0$, consider the set of all channel realizations which satisfy $C(\mathsf{H}) < R$. An *outage event* is said to occur when the channel realization belongs to this set. The probability of this event is called the outage probability at rate R.

$$\begin{aligned} P_{\text{out}}(R) &= \Pr\left(C(\mathsf{H}) < R\right) && (2.10) \\ &= \Pr\left(\log\left(\det\left(I + \mathsf{SNR} \cdot \mathsf{H}\mathsf{H}^\dagger\right)\right) < R\right) && (2.11) \end{aligned}$$

For any $R > 0$, this probability P_{out} is a strictly positive number. Recall Shannon's coding theorem, which says that the error probability for any code of any length with rate R greater than the channel capacity $C(\mathsf{H})$ cannot be smaller than some $\epsilon(R, \mathsf{H}) > 0$. For any code of any length having rate $R > 0$, we can lower bound the probability of decoding error as follows:

$$\begin{aligned} P_e &= P_{\text{out}} \cdot P(\text{error}|\text{outage}) + P(\text{no outage}) \cdot P(\text{decoding error}|\text{no outage}) \\ &\geq P_{\text{out}} \cdot P(\text{error}|\text{outage}) \\ &\geq P_{\text{out}} \cdot \min_{C(\mathsf{H})<R} \epsilon(R, \mathsf{H}) \\ &\triangleq P_{\text{out}} \cdot \epsilon_{\min} \end{aligned}$$

Thus the probability of error cannot be made arbitrarily small for any $R > 0$ and hence the capacity of this channel is 0. This lower bound for non-ergodic channel also gives an lower bound to the error probability of short codes in our block fading model.

Now we upper bound the error probability of short codes in our block fading model assuming high SNR. We assume that each codeword is within one fading block and length of each fading block $T_c > 2n-1$. An upper bound on the minimum probability of error is given by the probability of error for a random Gaussian code.

$$\begin{aligned} P_e &= P_{\text{out}} \cdot P(\text{error}|\text{outage}) + P(\text{no outage}) \cdot P(\text{decoding error}|\text{no outage}) \\ &\leq P_{\text{out}} \cdot 1 + 1 \cdot P(\text{decoding error}|\text{no outage}) \\ &\leq (1 + \delta')P_{\text{out}} \end{aligned}$$

$$\text{Thus } P_e \text{ is bounded as} \quad \epsilon_{\min} P_{\text{out}} \leq P_e \leq (1 + \delta')P_{\text{out}} \tag{2.12}$$

The last step for the upper bound followed from [15]. They showed that if $T_c > 2n-1$ and the SNR is high enough, $P(\text{decoding error}|\text{no outage})$ of a randomly generated Gaussian code is essentially upper bounded by $\delta' P_{\text{out}}$ for some $\delta' > 0$. From the lower and upper bounds above, we see that P_{out} is a good approximation to the error probability of short codes up to a fixed multiplicative factor. Moreover, P_{out} is usually easier to calculate than the exact error probability.

It is easy to see that the outage probability goes to zero as the SNR goes to infinity. In this case, using Eq. 2.12 we can say that,

$$\lim_{\mathsf{SNR}\to\infty} -\frac{\log P_e(\mathsf{SNR})}{\log \mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty} -\frac{\log P_{\text{out}}(\mathsf{SNR})}{\log \mathsf{SNR}}$$

The above limit always exists and is defined as the diversity d. Now following the same notation in [15], we use $f(\mathsf{SNR}) \doteq g(\mathsf{SNR})$ as a shorthand for

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log(f(\mathsf{SNR}))}{\log \mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty} \frac{\log(g(\mathsf{SNR}))}{\log \mathsf{SNR}}$$

Thus in this notation, for short codes $P_e(\mathsf{SNR}) \doteq P_{\text{out}}(\mathsf{SNR}) \doteq \mathsf{SNR}^{-d}$. The outage and error probability are not exactly equal to each other, but their SNR exponents are the same. Thus essentially all of the decoding errors are caused by outage events, which have a probability on the order of SNR^{-d}.

We saw how at high SNR, the capacity of an $n \times n$ MIMO system can be approximated as $n \log \mathsf{SNR}$ due to multiple spatial channels. Thus the MIMO system can be used to provide higher data rate as compared to a single antenna system. In addition to high rate, the MIMO system available can be used to improve reliability i.e. error probability or outage probability. Tarokh *et al* showed that for any fixed rate $R > 0$, the error probability P_e of a short code as a function of SNR is given by,

$$P_e(\mathsf{SNR}) \doteq \mathsf{SNR}^{-n^2}$$

Comparing this with the error probability for single antenna channel ($P_e \doteq \mathsf{SNR}^{-1}$) demonstrates the improvement in reliability due to multiple antennas. Thus a MIMO system can provide a diversity gain of $d = n^2$. One may wonder whether a MIMO system can simultaneously provide gains in reliability and data rate.

The difficulty in studying this is that for any fixed (and arbitrarily small) $\epsilon > 0$ error probability, all the n degrees of freedom i.e. the MIMO capacity can be achieved.

On the other hand, n^2 diversity can be achieved for any fixed (and however large) rate $R > 0$. Given these two extremes, it is difficult to see how the data rate and reliability interact with each other.

To get the complete picture of the interaction between reliability and rate, we should let the rate R grow to infinity with SNR instead of fixing it. How fast the probability of error decays with SNR when the rate is also growing with SNR? What is the achievable diversity if a fraction of the capacity is to be achieved i.e. a fraction of the n available degrees of freedom are to be achieved? Hence let the rate R grow with SNR as $R(\mathsf{SNR}) = r \log \mathsf{SNR}$ for some fixed number $0 \leq r \leq n$. This is a fraction $\frac{r}{n}$ of the capacity at high SNR. This rate is equivalent to having r separate single antenna channels, so we say that r degrees of freedom are to be achieved. We can define the degrees of freedom achieved as

$$r = \lim_{\mathsf{SNR} \to \infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} \tag{2.13}$$

Observe that for a fixed value of R (however large), only 0 degrees of freedom are achieved because R becomes a negligible fraction of the total capacity for large enough SNR.

Zheng and Tse [15] found the tradeoff between the error probability and rate for short codes. In other words, they found the maximum achievable diversity if r degrees of freedom are to be used. This optimal diversity $d^*(r)$ at r degrees of freedom is given by the piecewise linear function connecting the points $(k, (n-k)^2), k = 0, \cdots n$. For example, fig. 2-1 shows this tradeoff when $n = 3$. In our notation, the optimal rate-diversity tradeoff means that the achievable error probability of a short code of rate $R \approx r \log \mathsf{SNR}$ is equal to $P_e \doteq \mathsf{SNR}^{-d^*(r)}$. Note that for any $r < n$ degrees of freedom, strictly positive diversity is achieved. Thus although each codeword spans only a single channel realization, essentially all the n degrees of freedom can still be achieved.

The rate-diversity tradeoff shows that gains in rate and reliability can be achieved simultaneously in a MIMO system but there is a certain tradeoff between these two

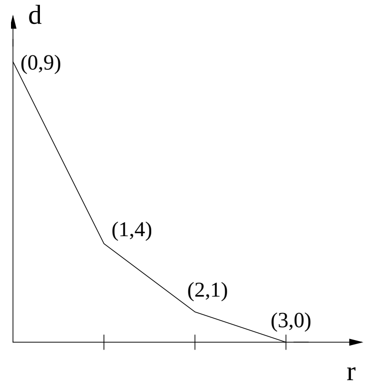

Figure 2-1: Optimal rate-diversity tradeoff for $n = 3$

gains. It also shows that any $r < n$ degrees of freedom (which were achievable in the ergodic formulation) can be *achieved* with positive diversity even though the codewords are limited to one fading block. It means that the error probability can be driven to zero by pushing the SNR to infinity. It appears as if there is no performance loss due to constraining the codewords within a single fading block. However, the caveat is that the error probability cannot be made arbitrarily small for a given SNR when the codewords span a single fading block. This is in contrast with the ergodic case.

Now consider the case when the codelength spans L fading blocks. If the coherence interval T_c satisfies $T_c > 2n - 1$, the error probability when r degrees of freedom are achieved is given in [15] as $P_e \doteq \mathsf{SNR}^{-Ld^*(r)} = (\mathsf{SNR}^{-d^*(r)})^L$. This is because an error happens essentially when all the L fading blocks are in outage. Thus it is sufficient to study the rate-diversity tradeoff for short codes ($L = 1$ case).

To summarize, we discussed the outage formulation and rate-diversity tradeoff in this section. More importantly, we discussed a type of asymptotic analysis which is useful later in this thesis.

Chapter 3

Issue of Coordination

The performance gains achievable in a point-to-point MIMO channel were discussed in the previous chapter. Can we also achieve these gains in the network scenario? We explore the problem of a single source communicating to a single destination in a network over multiple hops, where each hop starts from and/or reaches to a group of relay nodes. This will shed some light on the effect of lack of coordination between the relay nodes. We consider the following simplified model to gain some insight into this general problem.

3.1 Network Model

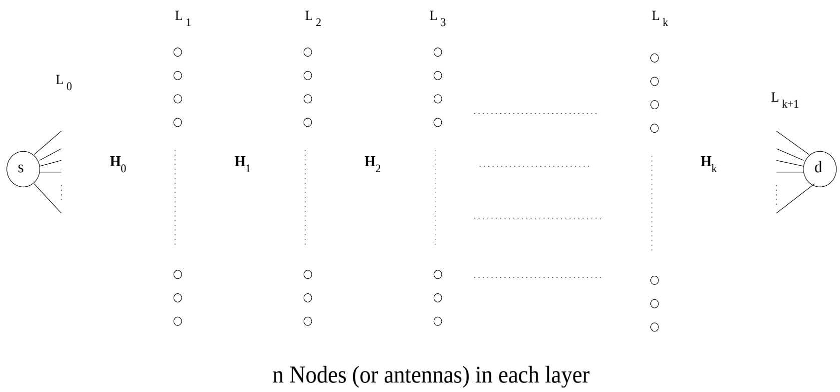

Figure 3-1: Network structure

The network under consideration has a single information source s with n transmit antennas and its final destination d also has n receive antennas (see Fig. 3-1). There are k layers of relay nodes between s and d, with each layer having n relay nodes[1]. Each relay node has a single antenna which can transmit and receive simultaneously. We denote the m'th layer of relay nodes by L_m and let L_0 denote the layer of the transmit-antennas of s. Similarly, let L_{k+1} denote the layer of the receive antennas at d. Each layer can only receive transmissions from its previous layer. Interference from all other layers is ignored which is in contrast with the existing network models (e.g. Gupta-Kumar model in [3],[4]). In our model, transmission from every layer to the next looks very much like a MIMO system, with the only difference being the lack of coordination between the relay nodes in a layer. As the relay nodes in a layer are at different locations, they cannot coordinate fully with each other to act like a single multiantenna node. Thus this oversimplified assumption (of ignoring the interference from other layers) allows us to study in isolation the value of coordination between relay nodes. Thus the complete network state is fully characterized by $k+1$ channel matrices denoting the channels between adjacent layers. Let matrix $\mathbf{H}_m$ denote the channel between layer m and $m+1$, i.e. $\mathbf{H}_m(i,j)$ is the value of the channel gain between the i'th node (or antenna) in L_{m+1} and the j'th relay node (or antenna) in L_m.

We assume an i.i.d. block fading model for each $\mathbf{H}_m$. The block fading model means that the channel state remains constant during a block of T_c symbols and changes to a new independent realization after that. We also assume that the coherence interval $T_c > 2n-1$, which ensures that the error probability and outage probability have the same SNR exponent i.e. $P_e(\mathsf{SNR}) \doteq P_{\text{out}}(\mathsf{SNR})$ [15]. Each $\mathbf{H}_m$ is assumed to be known at the final destination d. The source need not know the channel realizations. Each $\mathbf{H}_m$ is assumed to be independent of all others. All entries of each $\mathbf{H}_m$ are i.i.d. complex Gaussian variables with variance 1.

In our discrete-time model, $\mathbf{x}_{mi}$, $\mathbf{y}_{mi}$ and $\mathbf{w}_{mi}$ denote (respectively) the transmitted signal, the received signal and the noise at the i'th relay node (or antenna) in the m'th layer. Let $\mathbf{x}_m = [\mathbf{x}_{m1}, \mathbf{x}_{m2}, ..\mathbf{x}_{mN}]^T$ denote the transmitted vector by the m'th

[1]We assume that this layering of the network has been already done.

layer; $\mathbf{y}_m$ and $\mathbf{w}_m$ are defined similarly. For $0 \le m \le k$ we have,

$$\mathbf{y}_{m+1} = \mathbf{H}_m \mathbf{x}_m + \mathbf{w}_{m+1} \tag{3.1}$$

Each $\mathbf{w}_{mi}$ is a stationary white complex Gaussian process with variance σ^2. Moreover, each $\mathbf{w}_{mi}$ is independent of the input signal and of each other. Each of the relay nodes or antennas[2] has an average power constraint,

$$\mathcal{E}\left[||\mathbf{x}_{mi}||^2\right] \le P \qquad 1 \le i \le n, 0 \le m \le k \tag{3.2}$$

The $\mathsf{SNR} \triangleq P/\sigma^2$ is assumed to be very large. This is the ratio of the transmitted power of an individual relay node (or antenna) to the noise power, as opposed to the usual definition, which is the ratio of received signal power to noise power.

3.2 Single Layer Case

We saw that although the transmission from one relay layer to the next (in Fig. 1-3 or 3-1) *looks like* a MIMO system, it has an important difference with respect to a point-to-point MIMO system. The n transmit antennas can fully coordinate in a multiantenna transmitter. Similarly, a multiantenna receiver can decode based on the signals received at all of its n antennas. This is not the case in this wireless network (in Fig. 1-3 or 3-1) because the n relays in a layer are at different locations. How crucial is this coordination amongst relays for achieving all the degrees of freedom? We first review the problems of multiple access and broadcast to shed some light on this issue. For the multiple access to a multiantenna receiver, the coordination between the transmit antennas of a MIMO system is removed but the receive antennas can fully coordinate. On the other hand, for the broadcast from a multiantenna transmitter, the coordination between the receive antennas of a MIMO system is removed but the transmit antennas can fully coordinate.

[2]This simplifying assumption means the total available power at the source node s is nP. The achievable degrees of freedom are unchanged even if the total power at s is just P because we are assuming the SNR is very high.

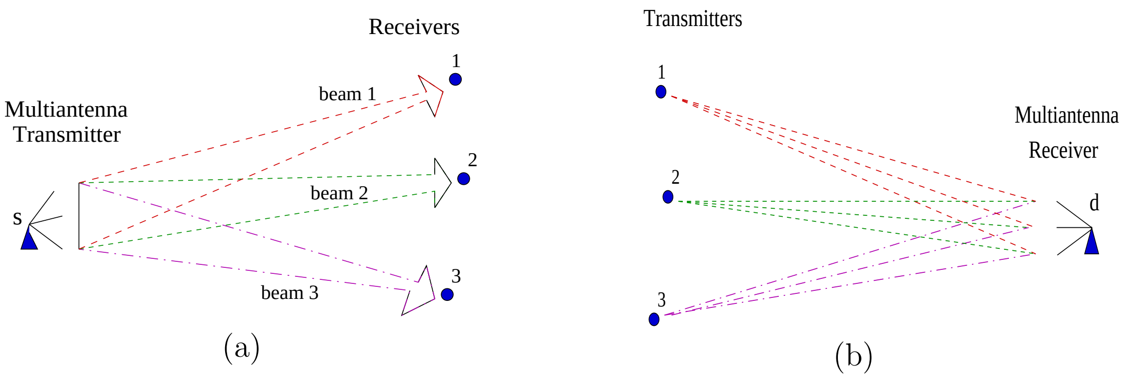

Figure 3-2: Value of coordination: (a)Broadcast (b)Multi-access

3.2.1 Multiple Access to a Multiantenna Receiver

There are n different transmitter nodes and each has a single antenna with power P/n. Each node wants to send separate information to a multiantenna destination node with n antennas (Fig. 3-2(b)). The channel equation here is the same as the point-to-point MIMO channel,

$$\mathbf{y} = \mathsf{H}\mathbf{x} + \mathbf{w}.$$

The only difference here is that the transmit antennas are now connected to separate nodes. Hence each antenna i chooses its codeword $\mathbf{x_i}(m_i)$ based on the separate message m_i it wants to transmit. These messages are uniformly chosen from $1 \leq m_i \leq M$, where M is the number of possible messages. The messages $m_1 \cdots m_n$ chosen and hence the codewords transmitted by different transmitters are independent of each other. Note the subtle difference from the point-to-point MIMO case, where the space-time code was constructed by choosing Gaussian symbols independently for each antenna. In that case, the codewords for all the transmit antennas are simultaneously decided by the overall message to be transmitted.

VBLAST [10] is a code originally developed for point-to-point MIMO. It decomposes the original data into n separate parts and allocates each part to one of the n antennas. Thus each antenna has a separate message to send. Hence, VBLAST can also be implemented in multiple access of a multiantenna receiver. It is known that all the degrees of freedom can be achieved by VBLAST. To see this, consider a suboptimal version of VBLAST, where the codeword from each antenna is decoded by projecting the received vector into a subspace free of interference from other transmit antennas. Thus the given vector channel of dimension n is converted to n parallel scalar channels, where the input of the i'th parallel channel is the same as the input of the i'th antenna

$$\mathbf{y}'_i = \mathsf{h}'_\mathsf{i}\mathbf{x}_i + \mathbf{w}'_i \qquad 1 \leq i \leq n$$

If the minimum singular value[3] of the vector channel H is $\sigma_{\min}$, none of the n parallel scalar channels can be weaker than $\sigma_{\min}$ i.e. $|\mathsf{h}'_\mathsf{i}| \geq \sigma_{\min}$ for $1 \leq i \leq n$. Hence the sum rate achieved by VBLAST is

$$\begin{aligned} R &= \sum_{i=1}^{n} \mathcal{E}\left[\log(1 + |\mathsf{h}'_\mathsf{i}|^2\mathsf{SNR})\right] \\ &\geq \sum_{i=1}^{n} \mathcal{E}\left[\log(1 + \sigma^2_{\min}\mathsf{SNR})\right] \end{aligned}$$

We know that $\sigma^2_{\min}$ is an exponential random variable with unit mean [22] as the entries of H are i.i.d. complex Gaussian with variance 1. Hence at high SNR, the RHS above is approximately $n \log \mathsf{SNR}$ in the sense explained before (after Eq. 2.9). Hence the LHS is lower bounded by $n \log \mathsf{SNR}$ and all the degrees of freedom can be achieved in this multiple access channel by VBLAST.

3.2.2 Broadcast from a Multiantenna Transmitter

Consider n different users, each having a single receive antenna (Fig. 3-2(a)). Each user wants to receive distinct information from a multiantenna source node with n antennas and total power P. The channel equation is the same as the point-to-point MIMO channel. The difference is that, the different n receive antennas are now connected to separate users. Thus each user has to decode based only on its own received signals.

[3]**Singular Value Decomposition:** Any complex matrix H can be decomposed as $\mathsf{H} = U\Sigma V$, where $\Sigma = \mathrm{diag}(\sigma_n, \cdots, \sigma_1)$ is a diagonal matrices with all non-negative entries. Each σ_i is called a singular value of H. U and V are complex unitary matrices. Note that squares of the singular values of H are the eigenvalues of $\mathsf{H}\mathsf{H}^\dagger$. A matrix is called singular when (at least) one of its singular values is zero. Similarly, a matrix is called near-singular when (at least) one of its singular values is near-zero.

Assume that the transmitter knows the channel state H. Let signal u_i be intended for user $1 \leq i \leq n$ i.e. $\mathbf{u} = [u_1 \cdots u_n]^T$ is to be sent to the users. The transmitter transmits $\mathbf{x} = c\,\mathsf{H}^{-1}\mathbf{u}$, where c is a fixed number chosen to satisfy the transmit power constraint. Then the received vector is $c\mathbf{u} + \mathbf{w}$, so the signal received by each user i (i.e. $cu_i + w_i$) is free of interference. We have converted this vector broadcast channel into n parallel (complex) AWGN channels, each providing 1 degree of freedom.

Care needs to be taken though: the transmitter should be shut off when H is near-singular. Otherwise the transmitter will need infinite power for implementing this channel inversion strategy. To be precise, we will shut off the transmitter when the smallest eigenvalue of $\mathsf{H}\mathsf{H}^\dagger$ is smaller than some sufficiently small $\epsilon > 0$. However, this only happens with probability ϵ for small values of ϵ. It is because the smallest eigenvalue of $\mathsf{H}\mathsf{H}^\dagger$ has an exponential distribution with mean 1 [20]. As all the n degrees of freedom are achieved when the transmitter is not shut off, $n(1-\epsilon)$ degrees of freedom are achieved on average at SNR going to infinity. Thus almost all the n degrees of freedom can be achieved by choosing ϵ small enough. This strategy of pre-inverting the channel is essentially sending information to the receivers in non-interfering beams, hence we call it a "beamforming" strategy[4].

3.2.3 Combining Beamforming and VBLAST

We saw that removing coordination at any one side of a MIMO system does not reduce the total achievable degrees of freedom. For multiple access, VBLAST can be used. Similarly for broadcast, beamforming can be performed if the transmitter knows the channel state. Are these the only two network structures where lack of coordination does not prevent achieving all the degrees of freedom or we can claim the same for a larger class of networks?

Now we simply combine the two strategies for broadcast and multiple access (Fig. 3-3). The source splits its data into n equal rate sub-streams. These sub-streams are beamformed to individual relay nodes simultaneously. Then each relay node decodes the sub-stream beamformed towards it, and retransmits the message to the

[4] Refer to [7, 8, 9] for a detailed analysis.

destination through the multiple access channel.

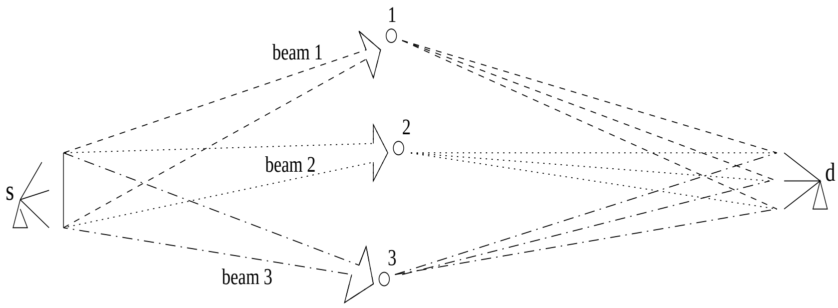

Figure 3-3: Combining beamforming with VBLAST

As seen before, beamforming ensures that there is no interference between the n data-streams, so it enables the source to transmit information reliably at a rate of approximately $\log \mathsf{SNR}$ to each of the relays. Similarly, as the destination is assumed to know the channel state, it can reliably decode each of the n data-streams of data rate $\log \mathsf{SNR}$ from each relay. Hence we get the following lemma.

Lemma 1 *All the n degrees of freedom can be achieved when only one relay layer exists between the multiantenna source and destination.*

Note that multiple antennas at the source and destination made it possible to have multiple spatial channels between them. Only one such spatial channel is available if the source and destination have only one antenna on them (as in the Gupta-Kumar and Xie-Kumar models [3, 6]). This is another significant difference between our network model and other network models.

Lemma 1 can be easily extended to the network in Fig. 3-4. It has nodes with

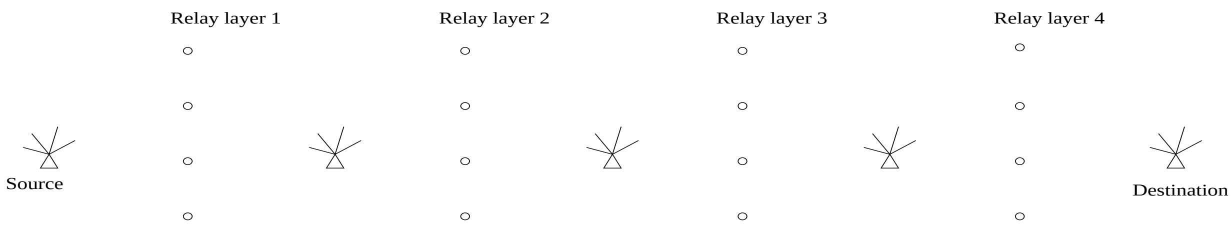

Figure 3-4: An extension of the single relay layer network

multiple antennas placed between any two layers of relay nodes. In this network, each multiple antenna node decodes the full message and again separates it into n data-streams which are beamformed to the nodes in the next relay layer. Here in every stage of the network operation, we have full coordination either at the transmitter side or at the receiver side due to the multiantenna nodes placed in between any two relay layers. Thus all the degrees of freedom can be achieved in this network as well.

In this strategy, every multiantenna node needs to know the state of its adjacent MIMO channels. This is needed for beamforming and VBLAST at each stage. Moreover, wireless networks rarely have these intermediate multiantenna nodes. Hence we go back to the general case in Fig. 1-3 or 3-1, where these intermediate multiantenna nodes are absent. The transmission from one relay layer to another does not have coordination at either end. The transmit and receive antennas of the relay nodes cannot coordinate to act like multiantenna. Hence the previous scheme of alternate beamforming and VBLAST cannot work. This is because each relay node in the first layer does not know the messages received by other nodes in that layer, so the first relay layer cannot perform beamforming to the next layer. Similarly, each node in the second relay layer does not know the received signals at other nodes in that layer, so it cannot decode the full message of rate $n \log \mathsf{SNR}$.

If we insist on using a *separation-based* strategy, where every relay node decodes the full message before transmitting [3, 6], only one degree of freedom can be obtained. This is because the single-antenna relay nodes can at most decode reliably at a rate of $\log \mathsf{SNR}$ [1]. It becomes interesting to investigate if all the degrees of freedom can still be achieved using some *non-separation-based* strategy when more than one relay layers are present. This would be a significant performance gain over the separation based strategy. This motivates the next chapter.

Chapter 4

General Case: Any Number of Layers

Previous chapter discussed the difficulty in achieving all the degrees of freedom in the multiple layer relay network under consideration—the lack of coordination between the relay nodes within a layer. The optimal network operation is unknown for this network. However, it is clear that no more than n degrees of freedom can be achieved since the source-destination nodes have only n antennas each. In this section, we propose a particular network operation (which need not be optimal). We show that this strategy achieves all possible n degrees of freedom for any fixed number of layers.

4.1 Network Operation

We fix the functions of all the relay nodes and by doing so convert the network to a point-to-point MIMO channel. Each relay node just re-transmits a scaling of the received symbol (with the added noise). Although each relay node could perform a different scaling, for simplicity we assume a fixed scaling $\sqrt{n}$ at every relay node, that is the transmitted signal by node i in layer m is given by $\mathbf{x}_{mi} = \mathbf{y}_{mi}/\sqrt{n}$. This scaling of $\sqrt{n}$ is to ensure that all relay nodes obey the average power constraint. Thus a relay does not do any kind of decoding (even partial) and all the decoding is done at the destination, which employs an optimal decoding rule. This is a highly desirable feature since it simplifies the network operation immensely and only the source and destination perform the computationally intensive tasks.

A code of rate R and length N has 2^{NR} codewords denoted by[1] $\tilde{\mathbf{x}}_0^N(i)$ for $1 \leq i \leq 2^{NR}$. Note that each codeword $\tilde{\mathbf{x}}_0^N(i)$ is a $n \times N$ matrix. A randomly generated Gaussian code of length N and rate R is prepared beforehand as explained later and conveyed to the source and the destination. Once chosen, the same code is used throughout the communication process. The source selects the codeword $\tilde{\mathbf{x}}_0^N(i)$, corresponding to the message i to be transmitted; the 2^{NR} possible messages are assumed to be equiprobable. The destination selects one of the 2^{NR} possible codewords by performing ML decoding on the received signal.

In this randomly generated Gaussian space-time code, all nN scalar symbols (in space and time) of any given codeword $\tilde{\mathbf{x}}_0^N(i)$ are generated i.i.d. according to a complex Gaussian distribution[2] of variance ρP for some $\rho < 1$. The choice of symbols for any one codeword is independent of the other codewords. Such a randomly generated code using the capacity achieving distribution has arbitrarily small error probability for any R less than the capacity, provided the codelength N is large enough [12].

Note that, each source antenna uses only a fraction ρ of its available power P. Since SNR was defined as P/σ^2, the effective SNR is $\rho\,\mathsf{SNR}$. In spite of the reduction in effective SNR, the number of degrees of freedom achieved remain unaltered as seen before in Section 2.2.1. This is because

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log(\rho \cdot \mathsf{SNR})}{\log \mathsf{SNR}} = 1 \tag{4.1}$$

for any $\rho > 0$. Thus the number of degrees of freedom measured in $\log \mathsf{SNR}$ and $\log(\rho\mathsf{SNR})$ are the same in the limit of high SNR. The next section shows that the choice of scaling $\sqrt{n}$ ensures that the signal component of the power transmitted by each relay node is the same as the source antenna power ρP. The remaining power $(1-\rho)P$ at each relay node is available for the noise component, because both signal

[1] Recall that we denoted the vector transmitted by layer i by $\mathbf{x}_i$. Hence the codewords transmitted by the source are denoted by $\tilde{\mathbf{x}}_0^N(i)$, where N denotes the length of the codeword.

[2] Recall from chapter 2 that this was the capacity achieving distribution for MIMO channel.

and noise is forwarded to the next layer in our network operation.

4.2 All Relay Nodes Obey Power Constraints

Let the random state $\mathbf{H}$ denote the entire network's channel state $(\mathbf{H}_0, \mathbf{H}_1 \cdots, \mathbf{H}_k)$, and let its realization i.e. sample value be denoted by $H = (H_0, \cdots, H_k)$. Each layer retransmits (after scaling down by $\sqrt{n}$) the received vector, so the transmitted vector by the jth layer is

$$\mathbf{x}_j = \frac{(\mathbf{H}_{j-1}\mathbf{H}_{j-2}...\mathbf{H}_0)}{\sqrt{n}^j}\mathbf{x}_0 + \left(\frac{\mathbf{w}_j}{\sqrt{n}} + \sum_{i=1}^{j-1} \frac{\mathbf{H}_{j-1}\mathbf{H}_{j-2}\cdots\mathbf{H}_i}{\sqrt{n}^{j+1-i}}\mathbf{w}_i \right) \quad (4.2)$$

Recall that $\mathcal{E}\left[\mathbf{x}_0\mathbf{x}_0^\dagger\right] = \rho P I$. Moreover, the noise at different relay nodes is i.i.d. and also independent of the signal, so the covariance matrix of the transmitted vector $\mathbf{x}_j$ conditioned on the channel realization is given by

$$\mathcal{E}_{\mathbf{x}_j}\left[\mathbf{x}_j\mathbf{x}_j^\dagger|\mathbf{H}=H\right] = \frac{\rho P}{\sqrt{n}^{2j}}(H_{j-1}H_{j-2}\cdots H_0)(H_{j-1}H_{j-2}\cdots H_0)^\dagger \quad (4.3)$$

$$+ \frac{\sigma^2}{\sqrt{n}^2}I + \sum_{i=1}^{j-1}\frac{\sigma^2}{\sqrt{n}^{2(j-i+1)}}(H_{j-1}H_{j-2}\cdots H_i)(H_{j-1}H_{j-2}\cdots H_i)^\dagger.$$

The average power transmitted by layer j is the average of $||\mathbf{x}_j||^2$.

$$\begin{aligned}\mathcal{E}\left[||\mathbf{x}_j||^2\right] &= \mathcal{E}_{\mathbf{H}}\left[\text{trace}\left(\mathcal{E}_{\mathrm{x_j}}\left[\mathrm{x_j x_j^\dagger}|\mathbf{H}=\mathrm{H}\right]\right)\right] \\ &= \rho P n + \sigma^2 + \sigma^2\sum_{1}^{j-1} 1 = \rho P n + \sigma^2 j\end{aligned}$$

The last step follows by using Eq. 4.3 with the following fact (proved in appendix B)

$$\mathcal{E}_{\mathbf{H}}\left[\text{trace}\left((\mathrm{H_{j-1}H_{j-2}\cdots H_0})(\mathrm{H_{j-1}H_{j-2}\cdots H_0})^\dagger\right)\right] = n^{j+1}$$

By symmetry, the transmitted power by any layer has equal contributions from each node in that layer. Thus the power transmitted by each node in the jth layer is equal to $\rho P + j\sigma^2/n$. The noise component in the transmitted power can be ignored when SNR goes to infinity in the sense that

$$\rho P + j\sigma^2 \leq \rho P + k\sigma^2 < (\rho+\delta)P \qquad \text{for any } \delta > 0 \ \text{ and large enough } P/\sigma^2.$$

Thus the power constraint at each relay node is satisfied under the proposed network operation. Note that ρ can be chosen arbitrarily close to 1, so for simplicity we henceforth assume $\rho = 1$ i.e. each source antenna uses the full available power P.

4.3 Achievable Performance

As every layer simply retransmits the received vector after scaling it down by $\sqrt{n}$, the received vector at the destination is

$$\begin{aligned} \mathbf{y}_{k+1} &= \frac{(\mathbf{H}_k\mathbf{H}_{k-1}...\mathbf{H}_0)}{\sqrt{n}^k}\mathbf{x}_0 + \left(\mathbf{w}_{k+1} + \sum_{i=1}^{k}\frac{\mathbf{H}_k\mathbf{H}_{k-1}\cdots\mathbf{H}_i}{\sqrt{n}^{k+1-i}}\mathbf{w}_i\right) && (4.4)\\ &= \frac{(\mathbf{H}_k\mathbf{H}_{k-1}...\mathbf{H}_0)}{\sqrt{n}^k}\mathbf{x}_0 + \mathbf{w}'_{k+1} && (4.5)\end{aligned}$$

Thus the proposed network operation makes the network looks like a point-to-point MIMO system, where the effective channel matrix is the product (or concatenation) of $k+1$ Gaussian random matrices. The above equations also show that the effective noise $\mathbf{w}'_{k+1}$ contains the inherent noise at the layer $k+1$ (i.e. receive antennas) as well as the noise accumulated from all the relay layers. Evidently, there are two potential roadblocks to achieving all the degrees of freedom. First, all the degrees of freedom will not be achieved if the accumulated noise over all the relay layers is too large. Second, all the degrees of freedom will not be achieved if the effective channel matrix is near-singular.

Conditioned on a channel realization, the effective noise $\mathbf{w}'_{k+1}$ is a jointly Gaussian vector with correlated components. It is independent of the transmitted signal. Distribution of this effective Gaussian noise is completely described by its covariance matrix. As noise at different relay nodes is i.i.d., the covariance matrix is given by

$$\begin{aligned}\mathcal{E}_{\mathbf{w}'_{k+1}}\left[\mathbf{w}'_{k+1}\mathbf{w}'^{\dagger}_{k+1}|\mathbf{H}=H\right] &= \sigma^2 I + \sum_{i=1}^{k}\frac{\sigma^2}{\sqrt{n}^{2(k+1-i)}}(H_k H_{k-1}\cdots H_i)(H_k H_{k-1}\cdots H_i)^{\dagger} \\ &\triangleq \sigma^2 G_H G_H^{\dagger}\end{aligned} \tag{4.6}$$

Now the capacity of this effective channel in Eq. 4.5 can be written as follows [1],

$$\mathcal{E}_{\mathbf{H}}\left[\log\det\left(I+\frac{\mathsf{SNR}}{n^k}(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)^{\dagger}\right)\right] \tag{4.7}$$

$$= \mathcal{E}_{\mathbf{H}}\left[\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\mu_i\right)\right] \tag{4.8}$$

where μ_i represents the ith eigenvalue of $(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)^{\dagger}$. It is the resultant channel matrix from source to destination which ensures i.i.d. white noise at the receive antennas. This noise whitening at the receiver is done by the matrix $\mathbf{G_H}^{-1}$ (see Eq. 4.6). As SNR grows large, this capacity expression can be approximated as $n\log\mathsf{SNR}$ (shown in appendix B).

Theorem 2 *The proposed network operation achieves all the n degrees of freedom. In other words, it achieves a rate R such that*

$$\lim_{\mathsf{SNR}\to\infty}\frac{R}{\log\mathsf{SNR}} = n \quad \textit{(that is } R\approx n\log\mathsf{SNR} \textit{ in our notation.)}$$

Thus at high SNR, our *non-separation-based* strategy achieves a rate that is n times that of any *separation-based* strategy. We now explain intuitively why the two roadblocks to achieving all the degrees of freedom do not matter at high SNR. The effective noise level at the receiver is essentially k times larger compared to the original noise level, due to noise accumulation through k relay layers. This reduces the effective SNR by a factor of k, but it does not reduce the degrees of freedom because k is fixed and thus $\lim_{\mathsf{SNR}\to\infty}\frac{\log(\mathsf{SNR}/k)}{\log\mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty}\frac{\log\mathsf{SNR}}{\log\mathsf{SNR}}$.

Second, the concatenation of channels $\prod_i \mathbf{H}_i$ becomes near-singular with essentially the same probability as that of an individual channel $\mathbf{H}_i$. This is because the smallest singular value of $\prod_i \mathbf{H}_i$ is at least as large as the product of the smallest singular values of each $\mathbf{H}_i$ ([19]: Thm. H.1).

Simulation: We saw how the proposed network operation achieves the maximum achievable degrees of freedom in a point-to-point system[3]. A simple simulation was done to check this theoretical result. We compared the probability of error between a *separation-based* strategy and our *non-separation-based* strategy for $n = 2$, $k = 4$, and SNR of $40dB$. Both strategies had a rate of $2bps/Hz$. The *separation-based* strategy used only one node in each layer, which decoded the received message and then retransmitted it to the node in next layer. The proposed *non-separation-based* strategy used the Alamouti space-time code [17] (instead of a randomly generated Gaussian code). The *non-separation-based* strategy achieved a probability of error 8.28×10^{-5} as compared to the 6.57×10^{-4} achieved by the *separation-based* strategy.

[3]In fact, all the degrees of freedom can also be achieved when not only the overall average power is constrained to P, but also the power within each fading block is constrained. This is stops relays from transmitting very large power (compared to P) in any fading block. The power constraint in this case is

$$\mathcal{E}\left[\, ||\mathbf{x}_{mi}||^2 |\ \mathbf{H}\right] \leq P \qquad 1 \leq i \leq n, 0 \leq m \leq k \tag{4.9}$$

In this case, all the degrees of freedom are achieved essentially because, when the signal is being forwarded across the layers, its very large amplification is extremely rare. See appendix D for more details.

Chapter 5

Tradeoffs: Rate, Network size and SNR

In the previous chapter, we saw that all the degrees of freedom can be achieved for any fixed number of layers. Nonetheless, extremely high SNR might be needed to achieve that depending on the number of relay layers. Note that for any fixed value of $\mathsf{SNR} > 0$, more and more noise gets accumulated with increasing k. The end-to-end mutual information will be driven to zero eventually when k goes to infinity. Therefore the SNR required for achieving all the degrees of freedom keeps getting higher with increasing network size, i.e. increasing number of layers. Thus the measure of degrees of freedom is hiding the detrimental effect of increasing network size by assuming very high SNR. In fact, in the simulation in the previous chapter, the *separation-based* strategy outperformed our *non-separation-based* strategy for lower values of SNR.

On one hand, no communication is possible when the number of layers i.e. the network size goes to infinity and the SNR is fixed. On the other hand, all the degrees of freedom are achievable when the SNR goes to infinity and the network size is fixed. This raises an interesting question: what happens between these two extremes? That is the case when both SNR and network size are *comparable* to each other in some sense. One also wonders how to compare these two different parameters of SNR and network size[1]. What scaling should be used for the comparison? It is clear that whether all the degrees of freedom are achieved or not does not depend on the network size or the SNR by itself. It depends on how these two quantities are relative to each other. If the SNR is not high enough for the network size, fewer degrees of freedom will be achieved. In this chapter, we study this interaction between SNR, network size and achievable rate i.e. degrees of freedom.

5.1 Asymptotic formulation and results

It is natural but difficult to explore the loss in performance due to increasing the number of layers at any fixed finite SNR. This is because the exact eigenvalue distribution of the effective channel matrix is unknown. Therefore, we explore the ways in which the number of layers k can grow to infinity when the SNR is going to infinity so that all the degrees of freedom can be achieved. In other words, we investigate the set of functions $k(\mathsf{SNR})$ which could denote the number of layers so that all the degrees of freedom are achievable. This function may go to infinity with SNR going to infinity. Now taking SNR to infinity enables us to formally find out how fast the SNR should grow with increasing number of layers to achieve all the degrees of freedom. The following theorem is proved in the appendix.

Theorem 3 *In this network, all the n degrees of freedom can be achieved for any function $k(\mathsf{SNR})$ for which*

$$\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR})}{\log \mathsf{SNR}} = 0$$

This confirms Theorem 2, where the number of layers is a fixed number not growing with SNR.

If the condition in Theorem 3 is not satisfied, the achievable degrees of freedom are reduced. In other words, if one is not aiming for all the degrees of freedom then a larger network can be reached. How large the network can be, if one is aiming for some $r < n$ degrees of freedom, is given by this sufficient condition.

[1] This is similar to asking "How to compare oranges with mangoes?"

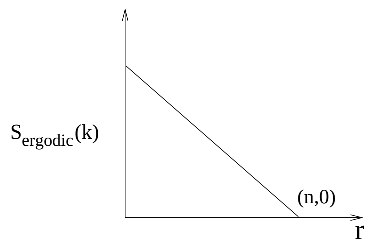

Figure 5-1: Tradeoff between achievable degrees of freedom and network size penalty

Theorem 4 *In this network, r degrees of freedom can be achieved for any function $k(\mathsf{SNR})$ for which*

$$\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR})}{\log \mathsf{SNR}} \leq \varphi(n)(n-r) \tag{5.1}$$

where $\varphi(n)$ is a fixed positive parameter of the point-to-point MIMO channel defined as $(n\log(n) - E_{\mathbf{H}_m} \log\det(\mathbf{H}_m\mathbf{H}_m^H))^{-1}$.

In our notation, this says that the achievable rate $R = r\log\mathsf{SNR}$ decreases linearly with increasing number of layers as

$$R \approx n\log\mathsf{SNR} - k/\varphi(n) \tag{5.2}$$

We define the LHS of Eq. (5.1) as the network's *size penalty* $S_{\text{ergodic}}(k)$ with respect to ergodic capacity. Figure 5-1 plots this linear tradeoff between achievable rate in terms of degrees of freedom and the network size in terms of the size penalty on ergodic capacity.

Note that theorems 3, 4 are only sufficient conditions. First, there might be other network strategies (though it seems unlikely), which perform better with increasing network size than our strategy. Second reason is that we have analyzed a lower bound to the achievable rate of our strategy. However, we conjecture that these theorems should also be necessary conditions as the SNR goes to infinity.

5.2 Illustrative example: n=1 case

Theorems 3 and 4 are proved in the appendix B, but here we consider the toy case of $n = 1$ to illustrate the essential reasoning with minimum mathematics (Fig. 5.2). Obviously using the proposed *non-separation-based* strategy is non-sensical here. Each relay can simply decode the full message and forward it to the next relay. In an arbitrarily large sized network, this *separation-based* strategy achieves the exact capacity of this network given by $\mathcal{E}_{|h_0|}[\log(1 + |h_0|^2\mathsf{SNR}]$, where $|h_0|$ is a Rayleigh distributed random variable. Consequently, it also achieves all of the single degree of freedom available. In fact, this network is often used to highlight the advantage of digital systems over analog systems-because there is no noise accumulation in digital systems as opposed to the analog systems.

On the contrary, in the $n > 1$ case, no *separation-based* strategy can achieve all the degrees of freedom due to lack of coordination between the relay nodes. Note that there is no issue of coordination in the $n = 1$ case. We can only rely on *non-separation-based* strategies for achieving all the degrees of freedom when $n > 1$. Thus coordination (or lack of it) is an key issue in the choice between *separation-based* strategies and *non-separation-based* strategies.

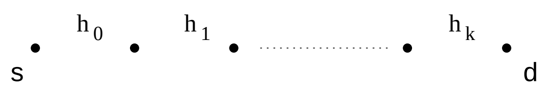

Figure 5-2: Case of $n = 1$: h_i denotes the channel gain from relay i to relay $i + 1$.

We first ignore one of the two possible roadblocks for achieving all the degrees of freedom by our *non-separation-based* strategy—assume that the concatenation of channels is never near-singular. Then the only cause of performance loss is the noise accumulation. This is the case for a cascade of AWGN channels, where each h_i is not a random variable but identically equal to 1. Hence the concatenated channel gain is 1- never near-singular, so the only roadblock for rate is the noise accumulation. Due to the noise accumulated through the relay layers, the effective noise variance increases to $(k+1)\sigma^2$. Hence the number of achievable degrees of freedom is given by

$$
\begin{aligned}
r &= \lim_{\mathsf{SNR}\to\infty} \frac{\log\left(1+\frac{\mathsf{SNR}}{k+1}\right)}{\log \mathsf{SNR}} \\
&= 1 - \lim_{\mathsf{SNR}\to\infty} \frac{\log k}{\log \mathsf{SNR}}
\end{aligned}
$$

Hence the size penalty function in the AWGN case is defined as $\lim_{\mathsf{SNR}\to\infty} \frac{\log k}{\log \mathsf{SNR}}$.

However, there was one more reason for rate loss: concatenation of channels. Hence consider the fading case, where each h_i is a complex Gaussian random variable with variance 1. Now assume that the effect of noise accumulation can be ignored and the only cause of performance loss is the concatenation of channels being near-singular. Then the maximum achievable rate is given by

$$
\begin{aligned}
R &= \mathcal{E}\left[\log\left(1+\prod_{i=0}^{k} |h_i|^2\, \mathsf{SNR}\right)\right] \\
&\approx \mathcal{E}\left[\log\left(\prod_{i=0}^{k} |h_i|^2\, \mathsf{SNR}\right)\right] \qquad \text{(high SNR)} \\
&= \log \mathsf{SNR} + (k+1)\mathcal{E}\left[\log |h_0|^2\right]
\end{aligned}
$$

Thus the achievable number of degrees of freedom is given by[2]

$$
\begin{aligned}
r = \lim_{\mathsf{SNR}\to\infty} \frac{R}{\log \mathsf{SNR}} &= 1 + \lim_{\mathsf{SNR}\to\infty} \frac{k}{\log \mathsf{SNR}} \cdot \mathcal{E}\left[\log |h_0|^2\right] \\
&= 1 - S_{\text{ergodic}}(k)/\varphi(1)
\end{aligned}
$$

This explains the effect of concatenation of channels. Note that the size penalty function in the AWGN case ($\lim_{\mathsf{SNR}\to\infty} \frac{\log k}{\log \mathsf{SNR}}$) is much smaller than that in the fading case ($\lim_{\mathsf{SNR}\to\infty} \frac{k}{\log \mathsf{SNR}}$). In fact, the former is 0 for any finite value of the later. This indicates that we can ignore the effect of noise accumulation in analyzing the fading case. More importantly, this shows that the dominant reason for rate loss is the concatenation of channels, not noise accumulation.

[2] Note that $\mathcal{E}\left[\log |h_0|^2\right]$ is a negative number because $|h_0|^2$ is an exponential random variable with mean 1.

5.3 Some Remarks

Going back to Eq. 5.2, we can say that the SNR requirement for maintaining a certain data rate R increases geometrically with linearly increasing network size. To be precise, if the number of layers increases from k_1 to k_2 then the following rule of thumb gives the new SNR_2 required to maintain the rate:

$$\log(\mathsf{SNR}_2/\mathsf{SNR}_1) = \frac{k_2 - k_1}{n\varphi(n)}$$

Thus the required SNR gets extremely large with increasing network size. Theorem 4 also tells that our *non-separation-based* strategy outperforms any *separation-based* strategy when the SNR is sufficiently large or network size is sufficiently small. Recall that a *separation-based* strategy achieves at most one degree of freedom.

A similar mathematical problem was studied in [13] in the context of a point-to-point channel. The channel equation was

$$\mathbf{y} = (\mathbf{H}_k \mathbf{H}_{k-1} \cdots \mathbf{H}_0)\mathbf{x} + \mathbf{w}$$

where each $\mathbf{H}_i$ is an independent random Gaussian matrix and $\mathbf{w}$ is Gaussian noise with covariance matrix $\sigma^2 I$. In that paper however, the size n of those random matrices goes to infinity instead of being a fixed finite number. It showed that negligible degrees of freedom are obtained if k grows without any bound. Now Theorem 4 suggests that even for the finite n case, no degrees of freedom can be achieved when k grows too fast. That is because the size penalty on ergodic capacity is infinity when k grows *too fast* compared to the SNR.

Chapter 6

Tradeoff between Rate, Diversity and Network Size

Earlier chapters focused on the ergodic capacity, which averages mutual information conditioned on the channel state, over all channel states. Achieving the ergodic capacity is difficult because it requires that the codelength should span a large number of fading blocks. In practice, we are also interested in the decoding error probability when the codelength only spans some finite number of fading blocks. As in section 2.2.2, we analyze the particular case when the codelength spans only one fading block ("short code"). We have seen in section 2.2.2 that this analysis easily extends to the analysis of codes which span L number of fading blocks. We have also seen that the error probability of a short code is approximated well by the outage probability. In other words, the error probability and outage probability have the same SNR exponent for a short code[15]. Recall the notation $f(\mathsf{SNR}) \doteq g(\mathsf{SNR})$ as a shorthand for

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log(f(\mathsf{SNR}))}{\log \mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty} \frac{\log(g(\mathsf{SNR}))}{\log \mathsf{SNR}}$$

In this notation, $P_e(\mathsf{SNR}) \doteq P_{\text{out}}(\mathsf{SNR}) \doteq \mathsf{SNR}^{-d}$ for a short code of diversity d. Nevertheless, analyzing the outage probability is easier than the error probability. Hence we will study the outage probability for evaluating the diversity d.

If r degrees of freedom are achieved in our network, recalling Eq. (4.5-4.8) gives

$$P_{\text{out}}(\mathsf{SNR}) \;=\; P\left(\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\mu_i\right) < r\log\mathsf{SNR}\right) \tag{6.1}$$

where μ_i represents the ith eigenvalue of $(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)(\mathbf{G_H}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)^{\dagger}$.

This chapter first considers the case when the network size is fixed. Then for reasons explained later, it studies the case of increasing network size where the number of layers could be denoted by $k(\mathsf{SNR})$ which satisfies

$$\lim_{\mathsf{SNR}\to\infty}\frac{k(\mathsf{SNR})\log\log\mathsf{SNR}}{\log\mathsf{SNR}} < \infty$$

In both the cases, we show (in appendix D) that the SNR exponent of the outage probability is unchanged even if all the relays are noise-free. In the case where the relays are noise-free, the noise at the destination is already i.i.d., so the noise whitening filter $\mathbf{G_H}^{-1} \equiv I$. In other words, the SNR exponent of Eq. 6.1 is unchanged if each μ_i is replaced by the ith eigenvalue of $(\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_0)(\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_0)^{\dagger}$, denoted by ν_i. Hence the diversity achieved at rate $r\log\mathsf{SNR}$ is also equal to

$$d(r) = \lim_{\mathsf{SNR}\to\infty}\frac{-\log P\left(\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\nu_i\right) < r\log\mathsf{SNR}\right)}{\log\mathsf{SNR}} \tag{6.2}$$

Thus the effect of noise accumulation across the relay layers can be ignored for diversity calculations. This is because the accumulated noise essentially reduces the SNR by a factor of k. However, this does not reduce the achievable degrees of freedom as the SNR goes to infinity. This implies that the rate-diversity tradeoff is mainly influenced by concatenation of channels, not noise accumulation. This fact was also observed regarding ergodic capacity in the previous chapter.

Theorem 5 *At high SNR, the rate-diversity tradeoff in our network is unchanged when all the relay nodes are noise-free.*

Therefore, this rate-diversity tradeoff is also congruent to that in the following point-to-point concatenated MIMO channel.

$$\mathbf{y} = (\mathbf{H}_k \mathbf{H}_{k-1} \cdots \mathbf{H}_0)\mathbf{x} + \mathbf{w}$$

where each $\mathbf{H}_i$ is independent of others and has i.i.d. complex Gaussian entries of variance $1/n$. The noise $\mathbf{w}$ is a Gaussian random vector with covariance matrix $\sigma^2 I$. This is also the point-to-point channel studied in [13]. Such similarity between these two problems was also observed for the tradeoff between ergodic capacity and network size (Theorem 4). Thus whatever is true regarding the tradeoffs in this concatenated point-to-point channel is also true for our network.

6.1 Fixed Network Size

As the effects of noise accumulation can be ignored, we only need to study the rate diversity tradeoff of a channel which is a concatenation of $k+1$ i.i.d. Gaussian matrices. First, we state the following property[1].

Lemma 6 *Let σ_i^A and σ_i^B denote the singular values of matrices A and B respectively. They are arranged in increasing order, for example: $\sigma_1^A \leq \sigma_2^A \cdots \leq \sigma_n^A$. Then the singular values of AB are lower bounded as follows*

$$\sigma_i^{AB} \geq \sigma_1^B \sigma_i^A \qquad 1 \leq i \leq n \tag{6.3}$$

More generally[2], for any $1 \leq m \leq n$

$$\sigma_i^{AB} \geq \sigma_m^B \sigma_{i-m+1}^A \qquad m \leq i \leq n \tag{6.4}$$

Now we explain the intuition behind this property. Let $U_A \Sigma_A V_A$ and $U_B \Sigma_B V_B$ be the singular value decomposition of A and B respectively. Singular values of AB

[1] A proof is given in the appendix as we could not find this in existing literature.

[2] Note that Eq. 6.3 is already known [19]. Lemma 6 in this thesis can thus be viewed as its generalization.

are same as those of $\Sigma_A V_A U_B \Sigma_B$, since multiplication by an unitary matrix does not change the singular values.

One can think of the matrix Σ_A as an ellipsoid in a complex n dimensional space, with its axes collinear to the coordinate axes. Lets denote this ellipsoid by ξ_{Σ_A}. It corresponds to a pdf contour surface of an n dimensional Gaussian with independent components. More specifically, consider a dummy complex Gaussian row vector $\mathbf{z}$ with covariance matrix I. Ellipsoid ξ_{Σ_A} represents a pdf contour of $\mathbf{z}\Sigma_A$ such that the length of each axis of ξ_{Σ_A} is equal to a singular value σ_i^A. Multiplying Σ_A by the unitary matrix $V_A U_B$ corresponds to a rotation of ξ_{Σ_A} in its n dimensional space. This rotated ellipsoid $\xi_{\Sigma_A V_A U_B}$ represents the pdf contour of $\mathbf{z}\Sigma_A V_A U_B$. Then multiplying $\Sigma_A V_A U_B$ with a diagonal matrix Σ_B corresponds to scaling (stretching or shrinking) the rotated ellipsoid $\xi_{\Sigma_A V_A U_B}$ along each jth coordinate axis by a factor σ_j^B. The resultant ellipsoid $\xi_{\Sigma_A V_A U_B \Sigma_B}$ corresponds to the pdf contour of $\mathbf{z}\Sigma_A V_A U_B \Sigma_B$. The axis-lengths of this resultant ellipsoid $\xi_{\Sigma_A V_A U_B \Sigma_B}$ are equal to the singular values of $\Sigma_A V_A U_B \Sigma_B$, which are same as those of AB.

Illustration: We demonstrate this idea with the example of a real two-dimensional case. Let $\mathbf{z}$ be a real two dimensional vector with i.i.d. Gaussian entries of unit variance. Also let $\Sigma_A = \text{diag}(2,1)$, $\Sigma_B = \text{diag}(1,4)$ and let $V_A U_B$ be a rotation by $\pi/4$.

$$V_A U_B = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Figure 6-1 shows the elliptical pdf contours of $\mathbf{z}$, $\mathbf{z}\Sigma_A$, $\mathbf{z}\Sigma_A V_A U_B$ and $\mathbf{z}\Sigma_A V_A U_B \Sigma_B$. They are denoted by ξ_I, ξ_{Σ_A}, $\xi_{\Sigma_A V_A U_B}$ and $\xi_{\Sigma_A V_A U_B \Sigma_B}$, respectively.

Lemma 6 says that the axis-lengths of $\xi_{\Sigma_A V_A U_B \Sigma_B}$ are minimized when the axis of the rotated ellipsoid $\xi_{\Sigma_A V_A U_B}$ are collinear to the coordinate axes. Since the axes of the original ellipsoid ξ_{Σ_A} are along the set of coordinate axes, the optimal rotation $V_A U_B$ is simply permuting the axis-lengths of the original ellipsoid ξ_{Σ_A}, while maintaining their orientations. In algebraic terminology, the singular values of AB are minimized when $V_A U_B$ equals a permutation matrix.

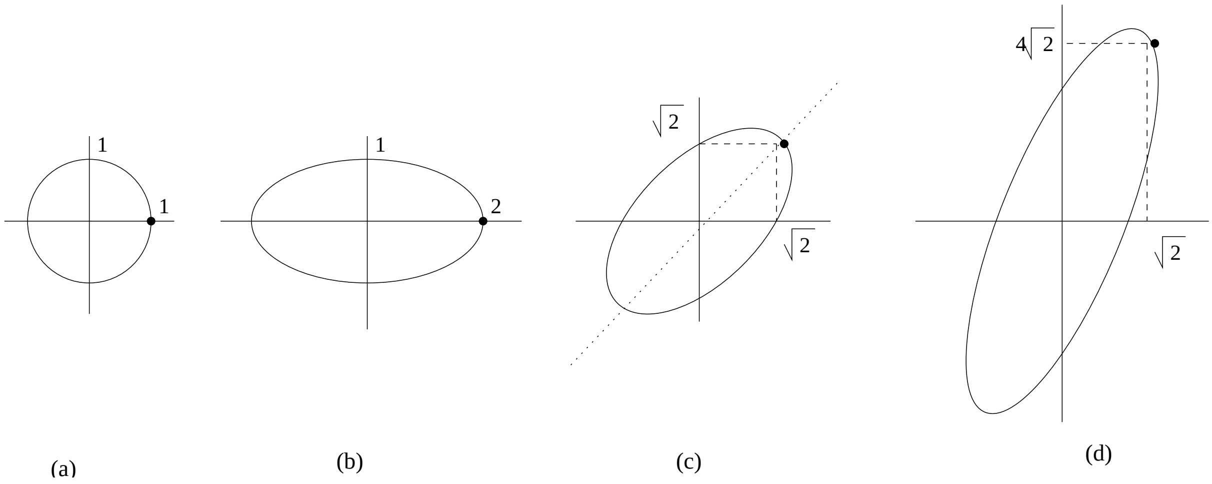

Figure 6-1: Elliptical pdf contours of two dimensional real Gaussian random vectors: (a) ξ_I (b) ξ_{Σ_A} (Stretching) (c) $\xi_{\Sigma_A V_A U_B}$ (Rotation) (d) $\xi_{\Sigma_A V_A U_B \Sigma_B}$ (Stretching a rotation)

For example, consider the case of the smallest singular value σ_1^{AB} of AB. The length of the smallest axis of the original ellipsoid ξ_{Σ_A} is σ_1^A and the same is true for the rotated ellipsoid $\xi_{\Sigma_A V_A U_B}$, because rotation does not change any axis length. The smallest axis of the resultant ellipsoid $\xi_{\Sigma_A V_A U_B \Sigma_B}$ is at least $\sigma_1^A \sigma_1^B$. It is because scaling the rotated ellipsoid $\xi_{\Sigma_A V_A U_B}$ along each coordinate axis i by a factor of σ_i^B means that any direction is at least scaled by $\sigma_{\min}^B \triangleq \sigma_1^B$. Hence $\xi_{\Sigma_A V_A U_B}$ is scaled by at least a factor of σ_1^B in each direction. Evidently, the minimum of σ_1^{AB} is attained i.e. the smallest axis-length of resultant ellipsoid $\xi_{\Sigma_A V_A U_B \Sigma_B}$ is minimized when the smallest axis of the rotated ellipsoid is aligned to the coordinate axis which gets scaled by σ_1^B. The above lemma follows from generalizing this visualization. Thus this lemma is essentially claiming that the axis-lengths of the resultant ellipsoid are minimized when the scalings on the rotated ellipsoid act directly along its axes.

We can rewrite Lemma 6 as

$$\min \sigma_i^{AB} = \min_{\Pi_A(i)+\Pi_B(i)\geq i+1} \sigma_{\Pi_A(i)}^A \sigma_{\Pi_B(i)}^B \tag{6.5}$$

where Π_A, Π_B are permutation functions on the set $\{1, 2 \cdots n\}$. Any particular choice of Π_A and Π_B fixes the right hand of the above equation for all i. Hence we have to search for Π_A and Π_B which jointly minimize all the singular values of AB.

Now we define α_i in terms of the substitution $\nu_i \triangleq (\frac{\mathsf{SNR}}{n^k})^{-\alpha_i}$ and let snr be defined as $\frac{\mathsf{SNR}}{n^k}$. This snr is equal to the SNR level for the transmission from the last relay layer to the destination. The scaling by $\sqrt{n}$ at each of the k relay layers has reduced the initial SNR by a factor of n^k. Note however that $\mathsf{SNR} \doteq \mathsf{snr}$. This implies that the number of degrees of freedom measured in $\log \mathsf{SNR}$ and $\log \mathsf{snr}$ are the same. Thus although the SNR is reduced across the layers, it does not reduce the degrees of freedom achieved. We use λ_i^j to denote the ith smallest eigenvalue of $\mathbf{H}_j \mathbf{H}_j^\dagger$. We also define β_i^j in terms of the substitution $\lambda_i^j \triangleq \mathsf{snr}^{-\beta_i^j}$.

By repeated application of Eq. 6.5 and application of Laplace's principle we get

$$\begin{aligned} P(\nu_i < \mathsf{snr}^{-\theta_i} :\ 1 \leq i \leq n) &= P(\alpha_i > \theta_i :\ 1 \leq i \leq n) && (6.6) \\ &\doteq \max_{\Pi_0, \Pi_1 \cdots \Pi_k} P\left(\sum_{j=0}^{k} \beta^j_{\Pi_j(i)} > \theta_i :\ 1 \leq i \leq n \right) && (6.7) \end{aligned}$$

where each Π_j denotes a permutation function. Laplace's principle [21] is essentially saying that the predominant way in which $(\alpha_i > \theta_i :\ 1 \leq i \leq n)$ happens is when the ellipsoids of all the channel matrices are aligned to each other according to the permutations $(\Pi_0, \Pi_1 \cdots \Pi_k)$. That is where the eigenvalues of the concatenated channel are minimized. Only the alignment according to the optimal set of permutations dominates this probability. All other reasons for this event are much less probable than this one dominating cause.

From [15], we know the distribution of eigenvalues of matrix $\mathbf{H}_j \mathbf{H}_j^\dagger$ is given by

$$p(\beta_1^j, \beta_2^j \cdots \beta_n^j) \doteq \prod_{i=1}^{n} \mathsf{SNR}^{-(2i-1)\beta_i^j} \doteq \prod_{i=1}^{n} \mathsf{snr}^{-(2i-1)\beta_i^j} \qquad (6.8)$$

Recall that every channel $\mathbf{H}_j$ is independent of all other channels, so its ith eigenvalue λ_i^j is independent of all $\lambda_{i'}^{j'}$ for $j \neq j'$. Hence $\beta_i^j = -\frac{\log \lambda_i^j}{\log \mathsf{snr}}$ is independent of all $\beta_{i'}^{j'}$ when $j \neq j'$.

Hence the joint distribution of the set of all β_i^j is given by

$$p(\beta_i^j : 1 \le i \le n, 0 \le j \le k) \doteq \prod_{j=0}^{k}\prod_{i=1}^{n} \mathsf{snr}^{-(2i-1)\beta_i^j} \tag{6.9}$$

Now lets go back to the outage probability. We can write $(1+\mathsf{snr}\nu_i) \doteq \mathsf{snr}^{(1-\alpha_i)^+}$, where $(x)^{(+)}$ denotes $\max\{0,x\}$. Now the outage probability from Eq. 6.2 is given by

$$\begin{aligned} P_{\text{out}}(\mathsf{SNR}) = P\left(\prod_i (1+\frac{\mathsf{SNR}}{n^k}\nu_i) \le \mathsf{SNR}^r\right) &\doteq P\left(\prod_i \mathsf{snr}^{(1-\alpha_i)^+} \le \mathsf{SNR}^r\right) && (6.10)\\ &\doteq P\left(\prod_i \mathsf{SNR}^{(1-\alpha_i)^+} \le \mathsf{SNR}^r\right) && (6.11)\\ &= P\left(\sum_i (1-\alpha_i)^+ \le r\right) && (6.12) \end{aligned}$$

Applying Laplace's principle to the above equation along similar lines as in [15], together with Eq. 6.7 and Eq. 6.8 we get

$$P_{\text{out}}(\mathsf{SNR}) \doteq \min_{\Pi_0,\Pi_1\cdots\Pi_k} \min_{\beta} \left(\prod_{j=0}^{k}\prod_{i=1}^{n} \mathsf{snr}^{-(2i-1)\beta_i^j}\right) \tag{6.13}$$

where β is the following constraint for outage to occur

$$\sum_{i=1}^{n}\left(1-\sum_{j=0}^{k}\beta_{\Pi_j(i)}^j\right)^+ \le r$$

Thus we have reduced the diversity calculation the following linear optimization.

Theorem 7 *At r degrees of freedom, this network achieves a diversity equal to*

$$d(r) = \min_{\Pi_0,\Pi_1\cdots\Pi_k} \min_{\beta} \left(\sum_{j=0}^{k}\sum_{i=1}^{n} (2i-1)\beta_i^j\right) \qquad \textit{where}$$

β denotes all $(\beta_i^j : 1 \le i \le n, 0 \le j \le k)$ satisfying $\sum_{i=1}^{n}\left(1-\sum_{j=1}^{k+1}\beta_{\Pi_j(i)}^j\right)^+ \le r$.

Example: Consider the case of $k = 1$ and $n = 3$, for illustration. For $r = 2$, the above minimization is attained at[3] $\beta_1^0 = \beta_1^1 = 0.5$ and every other β_i^j being 0. The permutations are having $\Pi_0(1) = \Pi_1(1) = 1$. That is the smallest axis of (ellipsoids of) each matrix should be collinear with each other. This gives $d(2) = 1$. For r between 2 and 3, the diversity is the line segment joining $d(2)$ and $d(3) = 0$. In general, one only needs to find $d(r)$ for $r = 0, 1 \cdots n$. Diversity at non-integer r is obtained by linear interpolation of the diversity at neighboring integers to r.

For $r = 1$ in our example, the minimum is attained when $\beta_1^0 = \beta_1^1 = 1$ and all other β_i^j are 0. The permutation should have $\Pi_0(1) = 2 = \Pi_1(2)$ and $\Pi_0(2) = 1 = \Pi_1(1)$. That is the smallest axis of (ellipsoids of) each matrix is aligned to the second smallest axis of the other matrix. This gives $d(1) = 2$. For $r = 0$, the minimum is attained when $\beta_1^0 = \beta_1^1 = 1$ and $\beta_2^0 = \beta_2^1 = 0.5$. The permutation functions are $\Pi_0(1) = 3 = \Pi_1(3)$, $\Pi_0(2) = 2 = \Pi_1(2)$ and $\Pi_0(3) = 1 = \Pi_1(1)$. This aligns the smallest axis of first matrix to largest axis of the other matrix and vice versa. The second-smallest axes of both matrices are collinear. The diversity attained is $d(0) = 5$.

In general for $k = 1$ and any n, at integer r degrees of freedom, the minimum is attained when the smallest $n - r$ singular values of each of two matrices are aligned in the reverse order to each other. Meaning, that ith smallest axis of first matrix is aligned to $n - r + i + 1$th smallest axis of second matrix. The smallest $\lfloor \frac{n-r}{2} \rfloor$ singular values of both matrices are of the order snr^{-1} i.e. for $j = 0, 1$

$$\beta_i^j = 1 \quad \text{for} \quad 1 \leq i \leq \lfloor \frac{n-r}{2} \rfloor$$

where $\lfloor x \rfloor$ is the largest integer not greater than x. For odd values of $n - r$, both matrices have $\beta_i^j = 0.5$ for $i = \lfloor \frac{n-r}{2} \rfloor + 1$. All other β_i^j are zero. Nevertheless, keep in mind that the described optimal solution is not unique. Figure 6-3 plots the tradeoffs for different values of n when $k = 1$.

[3] The optimal Π_0, Π_1 and β_i^js stated here are not unique as we will see later.

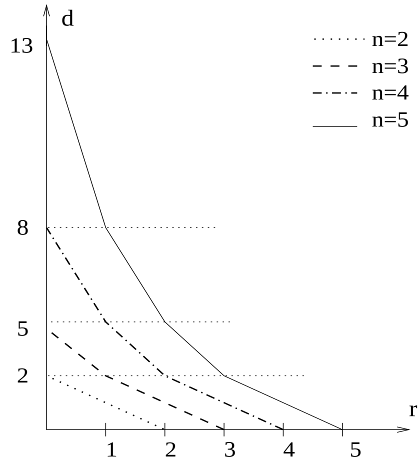

Figure 6-2: Rate-diversity tradeoff for various n when $k = 1$

6.1.1 Diversity Calculations and Typical Outage Events

From Theorem 7, how to find the diversity for the general case when $k \geq 1$ is not immediately clear. The following story explains an easy method for diversity calculation at integer r degrees of freedom. Moreover, it also explains the typical (or dominant) cause of outage events. For non-integer r, linear interpolation of the diversity at neighboring integers of r gives the diversity.

> King A has ranks 1 to n for horses with $k+1$ horses of each rank. The jth horse of rank i is called β_i^j. King A has to send $n-r$ horses to another king. A horse of rank i costs $(2i-1)$. To minimize the total cost, king A first sends all his lowest rank (i.e. rank 1) horses. If more horses are needed, he sends the rank 2 horses and so on. The cost incurred for sending $n-r$ horses in this way equals the diversity achieved.

Evidently, if $k+1 \geq n$ only the lowest rank horses need to be sent. Thus we get the following corollary.

Corollary 8 *This network achieves a diversity of $n-r$ at r degrees of freedom for any fixed number k of relay layers satisfying $k \geq n-1$.*

This is a much worse rate-diversity tradeoff compared to the usual point-to-point MIMO channel (i.e. the $k = 0$ case), as seen in Figure 6.1.1. It is expected because the king A in the point-to-point channel case has only 1 soldier of each rank. Hence sending $n-r$ horses is costlier for him compared to a king having $k+1$ horses of each rank. Thus a product of multiple i.i.d. random Gaussian matrices makes it much easier to have an outage, compared to a single random Gaussian matrix. For the same reason, for any r, n and k, the point-to-point MIMO channel will always achieve more (or equal) diversity compared to the concatenated channel.

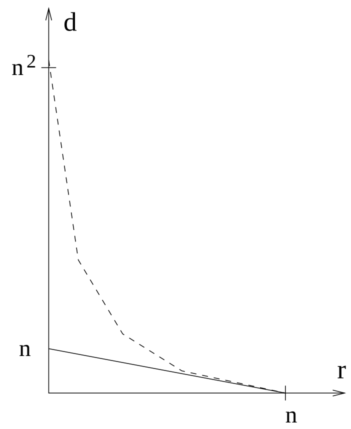

Figure 6-3: Rate-diversity tradeoffs when $k+1 \geq n$. Continuous line shows the tradeoff for our network and dotted line shows the same for point-to-point MIMO channel.

The above story also gives us some insight into the most dominant i.e. the typical cause of outage. Giving $n-r$ horses denotes destroying $n-r$ degrees of freedom. Choosing horse j of rank i denotes λ_i^j being of the order of SNR^{-1} (in the $\doteq$ sense). Thus each horse has the ability to destroy one separate degree of freedom completely. The cost of horse j of rank i signifies the likelihood of having $\lambda_i^j \doteq \mathsf{SNR}^{-1}$. More the cost $(2i-1)$ of a horse, it is more unlikely to have $\lambda_i^j \doteq \mathsf{SNR}^{-1}$. More specifically, this has a probability of the order of $\mathsf{SNR}^{-(2i-1)}$; that is, less likely for higher i. If the total cost of sending some $n-r$ horses is d_1, then the likelihood of loosing $n-r$ degrees of freedom is of the order of SNR^{-d_1}. The optimal choice of the $n-r$ horses (as explained in the above story) minimizes d_1 and hence corresponds to the dominant cause of loosing $n-r$ degrees of freedom. Other choices which cost more than the optimal cost $d(r)$ are much less likely to be the cause of loosing $n-r$ degrees of freedom.

In the dominant cause of loosing $n-r$ degrees of freedom, note that two near-zero eigenvalues λ_i^j (of the order of SNR^{-1}) cannot be aligned in the same direction. The product matrix will have an eigenvalue $\nu_i \doteq \mathsf{SNR}^{-2}$ if they are aligned in the same direction. This is an atypical (rare) way for destroying a degree of freedom which only requires $\nu_i \doteq \mathsf{SNR}^{-1}$. Thus typically, each near-zero eigenvalue of the order of SNR^{-1} destroys exactly one separate degree of freedom. Equivalently, each near-zero eigenvalue ν_i^j has acted on a different axis of the resultant ellipsoid of $\mathbf{H}_k \cdots \mathbf{H}_0$. Hence typically, there are $n-r$ near-zero eigenvalues λ_i^j of the order of SNR^{-1} and each of them destroys a separate degree of freedom. The optimal choice of choosing the horses of the lowest possible ranks corresponds to the dominant way in which $n-r$ eigenvalues out of the set $\{\lambda_i^j\}$ of all the $(k+1)n$ eigenvalues are of the order of SNR^{-1}.

6.2 Increasing Network Size

In the last section, we saw that the diversity is given by $d(r) = n - r$, when k is large ($k \geq n-1$). How is this affected when k is not a fixed number but grows to infinity with SNR? Along similar lines as in the previous chapter we ask the following question. What is the set of functions $k(\mathsf{SNR})$, which could denote the number of layers so that a diversity d can be achieved with r degrees of freedom? For a reason which will be clear later, we restrict ourselves to $k(\mathsf{SNR})$ which satisfy,

$$\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR}) \log\log \mathsf{SNR}}{\log \mathsf{SNR}} < \infty \tag{6.14}$$

In Eq. 6.8, we had ignored a multiplicative factor of[4] $(\log \mathsf{SNR})^n \tau$, because it did not matter then in the $\doteq$ sense of equality. It cannot be ignored now when k goes to infinity with SNR. Considering this factor, the joint distribution of all β_i^j in Eq. 6.9

[4] τ is a constant depending on n.

is modified to

$$
\begin{aligned}
p(\beta_i^j : 1 \le i \le n, 0 \le j \le k) &\doteq \tau^{k+1} (\log \mathsf{SNR})^{n(k+1)} \prod_{j=0}^{k} \prod_{i=1}^{n} \mathsf{SNR}^{-(2i-1)\beta_i^j} \\
&\doteq \tau^{k+1} (\log \mathsf{SNR})^{n(k+1)} \prod_{j=0}^{k} \prod_{i=1}^{n} \mathsf{snr}^{-(2i-1)\beta_i^j} \\
&\doteq (\log \mathsf{SNR})^{n(k+1)} \left(\prod_{j=0}^{k} \prod_{i=1}^{n} \mathsf{snr}^{-(2i-1)\beta_i^j} \right)
\end{aligned}
$$

Second step follows because we still have $\mathsf{SNR} \doteq \mathsf{snr} = \mathsf{SNR}/n^k$ as before, since $k/\log \mathsf{SNR} \to 0$ if Eq. 6.14 is satisfied. Last step follows since $\tau^{k+1} \doteq 1$, again because $k/\log \mathsf{SNR} \to 0$.

Now the outage probability in Eq. 6.13 will be multiplied by an extra factor of $(\log \mathsf{SNR})^{n(k+1)}$. Hence the diversity achieved is given by

$$
d(r) = -\left(\lim_{\mathsf{SNR}\to\infty} \frac{\log (\log \mathsf{SNR})^{n(k+1)}}{\log \mathsf{SNR}} \right) + \min_{\Pi_0, \Pi_1 \cdots \Pi_k} \min_{\beta} \left(\sum_{j=0}^{k} \sum_{i=1}^{n} (2i-1)\beta_i^j \right)
$$

where β means the same as in Theorem 7. Lets denote the term in the first bracket above by $S_{\text{outage}}(k)$. Thus the diversity achieved is $S_{\text{outage}}(k)$ less compared to the calculation in Theorem 7. Since for large k, Theorem 7 yields to $d = n - r$, we get the following theorem.

Theorem 9 *The following diversity can be achieved in this network when r degrees of freedom are to be achieved:*

$$
d(r) = (n - r - S_{outage}(k))^{+} \tag{6.15}
$$

where $S_{outage}(k)$ is equal to $\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR}) n \log\log \mathsf{SNR}}{\log \mathsf{SNR}}$, which represents the size penalty function in the outage formulation.

In other words, r degrees of freedom and diversity d can be achieved for those functions $k(\mathsf{SNR})$, which satisfy $S_{\text{outage}}(k) \le n - r - d$. Note that as expected, $S_{\text{outage}}(k) = 0$ when k is fixed as in Corollary 8.

6.2.1 Illustrative example: n=1 case

Once again consider the simple case of $n = 1$. Let $h_0, h_1 \cdots h_k$ denote the channel gains as the previous chapter and substitute $|h_j|^2 \triangleq \mathsf{SNR}^{-\gamma_j}$. As the effect of noise accumulation can be ignored, the outage probability is given by

$$\begin{aligned}
P_{\text{out}}(\mathsf{SNR}) &\doteq P\left(\log\left(1+\prod_j |h_j|^2\ \mathsf{SNR}\right) < r\log\mathsf{SNR}\right) \\
&\doteq P\left(\left(1-\sum_j \gamma_j\right)^+ < r\right) \\
&= \int_{(1-\sum_j \gamma_j)^+ < r} p(\gamma_0, \gamma_1, \cdots \gamma_k)\ \ d\gamma_0 d\gamma_1 \cdots d\gamma_k
\end{aligned}$$

However, every $|h_j|^2$ is an exponential random variable with mean 1 because every h_j is a circular symmetric complex Gaussian. Hence the distribution of $\gamma_j = -\frac{\log|h_j|^2}{\log\mathsf{SNR}}$ is given by

$$p(\gamma_j) = (\log\mathsf{SNR})\ \mathsf{SNR}^{-\gamma_j} \exp(-\mathsf{SNR}^{-\gamma_j})$$

For negative γ_j, the last factor above decays exponentially fast to zero with increasing SNR. Hence we may neglect those cases in the integral for outage probability. For $\gamma_j > 0$, it converges to 1 and converges to $1/e$ for $\gamma_j = 0$. Either way, the $\exp(-\mathsf{SNR}^{-\gamma_j})$ is equal to 1 in the $\doteq$ sense.

Now note that the $p(\gamma_0, \gamma_1, \cdots \gamma_k) = \prod_j p(\gamma_j)$, since all channels are independent of each other. The integral for outage probability now reduces to the following integral over all nonnegative γ_j

$$\begin{aligned}
P_{\text{out}}(\mathsf{SNR}) &\doteq \int_{(1-\sum_j \gamma_j)^+ < r} (\log\mathsf{SNR})^{k+1}\mathsf{SNR}^{-\sum_j \gamma_j}\ d\gamma_0 \cdots d\gamma_k \\
&\doteq (\log\mathsf{SNR})^{k+1}\ \mathsf{SNR}^{-(1-r)}
\end{aligned}$$

Hence we get

$$d(r) = -\left(\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR})\log\log\mathsf{SNR}}{\log\mathsf{SNR}}\right) + 1 - r$$

This explains the nature of the size penalty function in the outage formulation.

6.2.2 Some Remarks

Comparing size penalty on ergodic capacity, $S_{\text{ergodic}}(k)$, in Eq. (5.1) to $S_{\text{outage}}(k)$ in Eq. (6.15), we observe that the size penalty in the outage formulation can be significantly higher than that on the ergodic capacity. For example, if

$$k(\mathsf{SNR}) = \bar{k} \triangleq \frac{\log \mathsf{SNR}}{\log\log \mathsf{SNR}} \tag{6.16}$$

$S_{\text{outage}}(\bar{k}) = n$, which implies that at any positive diversity gain, the supported outage capacity yields 0 degrees of freedom. We can thus view $\bar{k}$ as an upper limit on the supportable network size, when coding over a single fading block. This explains why we started with the assumption in Eq. 6.14.

Note that the ergodic size penalty for $\bar{k}$ number of layers is $S_{\text{ergodic}}(\bar{k}) = 0$, which means that the ergodic capacity yields all the n degrees of freedom. However, one has to code over a large number of blocks to get close to the ergodic capacity. This behavior is different from the point-to-point case. In a point-to-point MIMO channel, the outage capacity approaches the ergodic capacity ($\approx n \log \mathsf{SNR}$) as the required diversity gain approaches 0. In contrast, in a large network system (k comparable to $\bar{k}$), the outage capacity, at any positive diversity requirement, is in general much smaller than the ergodic capacity. Therefore, the network throughput is mostly restricted by outage events.

Another important observation is that the maximum diversity for a network is n instead n^2 as in the point-to-point case discussed in chapter 2. Reason being, at $r = 0$, a typical outage event for the network is that n out of the $k+1$ channel matrices lose 1 degree of freedom each, rather than only one channel matrix losing all n degrees of freedom[5]. Losing one degree of freedom by each of those n matrices is much more common than loosing all n degrees of freedom by a single matrix. In short, the dependence on many random matrices makes it much easier to loose degrees of freedom in this network compared to the point-to-point case.

It is also worth noting that the network size in terms of $S_{\text{outage}}(k)$ and diversity

[5] In terms of the horse story, it means that n horses of the lowest rank are sent out.

d can be traded off. That means when r degrees of freedom are achieved, either $S_{\text{outage}}(k)$ equal to $n - r$ can be achieved if no diversity is needed or vice versa. Of course any diversity and $S_{\text{outage}}(k)$ which add up to $n - r$ can also be achieved.

As an endnote, we can extend all the diversity calculations (for short codes) in this chapter to the case where the codelength spans L fading blocks. This is along similar lines as in chapter 2. The error probability in that case is given by $P_e \doteq \mathsf{SNR}^{-Ld(r)}$, where $d(r)$ is the achievable diversity for this network when r degrees of freedom are achieved. It is because now an error happens essentially when all the L fading blocks are in outage.

Chapter 7

Discussion

All the results in this thesis were derived for SNR going to infinity. Nonetheless, we can also deduce some rules of thumb to be used in practice by substituting actual values of parameters like SNR and k in the size penalty functions for ergodic or outage formulation. For example, consider an engineer who wants to achieve the same diversity and degrees of freedom after the number of layers are increased from k_1 to k_2. The new SNR_2 to ensure this is obtained by equating the "empirical" value of the size penalty function, i.e. $\frac{nk_1 \log\log \mathsf{SNR}_1}{\log \mathsf{SNR}_1} = \frac{nk_2 \log\log \mathsf{SNR}_2}{\log \mathsf{SNR}_2}$. However, the reader should keep in mind that all the results we obtained were based on a particular *non-separation-based* strategy for our network. There might be other *non-separation-based* strategies (though unlikely) that achieve better tradeoffs.

In chapters 5 and 6, the number of layers ($k(\mathsf{SNR})$) grew to infinity. The total power of all the nk relays also grew with the network size as $nk(\mathsf{SNR})P$. In some situations, this is not possible and the total power has to remain constant irrespective of the network size. If this total power available to this growing network is constrained to P, the results in chapters 5 and 6 remain unchanged. Because $P \doteq \frac{P}{nk(\mathsf{SNR})}$ for the cases in those chapters. This is when total power is equally divided amongst all the relays. Essentially, the SNR is so large that dividing it by $nk(\mathsf{SNR})$ does not reduce the rate significantly.

The main result (Theorem 2) can be trivially extended to a network of j source-destination pairs, which has all the source-destination pairs separated by the same *relays-in-layers* network. In that case, each source-destination pair can have a rate of $\frac{N}{j}\log\mathsf{SNR}$, which can be achieved by time-sharing.

Our network model can be also extended to a situation where the source and destination do not have multiple antennas. In that case, a multiantenna is mimicked by using multiple time-slots. Each of these time-slots acts like a virtual antenna, and every technique for multi-antenna systems can be employed on this virtual multi-antenna system. As long as the time-slots are separated enough to experience different fading realizations, they replace the role of independently faded spatial paths in multiantenna systems.

7.1 Increasing n case

We have assumed the number n of relays per layer to be a fixed number throughout the thesis. Thus the SNR is much larger than n. However, to get a better idea of this requirement, we can think of n growing with SNR as $n(\mathsf{SNR})$, using a similar line as $k(\mathsf{SNR})$ in chapters 5 and 6.

Now consider the case of point-to-point MIMO channel discussed in chapter 2. We apply a result from [22] about the smallest eigenvalue λ_1^{H} of $\mathsf{H}\mathsf{H}^{\dagger}$, where H is a random matrix with i.i.d. complex Gaussian entries. It says that $\mathcal{E}_{\mathsf{H}}\left[\log(\lambda_1^{\mathsf{H}}n)\right]$ tends to a constant when n grows to infinity. Applying this to the capacity of this channel (from Eq. 2.7 and 2.8)

$$\begin{aligned} C_{\mathrm{MIMO}} = \sum_{i=1}^{n}\mathcal{E}_{\mathsf{H}}\left[\log(1+\mathsf{SNR}\lambda_i^{\mathsf{H}})\right] &\geq n\,\mathcal{E}\left[\log(1+\mathsf{SNR}\lambda_1^{\mathsf{H}})\right] \\ &\geq n\,\mathcal{E}\left[\log(\mathsf{SNR}\lambda_1^{\mathsf{H}})\right] \\ &= n\,\mathcal{E}\left[\log(\frac{\mathsf{SNR}}{n}) + \log n\lambda_1^{\mathsf{H}}\right] \\ &\approx n\left(\log\mathsf{SNR} - \log n + \text{ (a constant)}\right) \end{aligned}$$

This expression is approximately $n\log\mathsf{SNR}$ when the following condition is satisfied

$$\lim_{\mathsf{SNR}\to\infty}\frac{\log n(\mathsf{SNR})}{\log\mathsf{SNR}} = 0 \tag{7.1}$$

Note that this is only a sufficient condition for achieving all the degrees of freedom. In fact, it is not even clear if achieving all the degrees of freedom becomes more difficult with growing matrix size. It might be the case that all the degrees of freedom are achieved for all functions $n(\mathsf{SNR})$. Thus we obtained a sufficient condition on the number of antennas with respect to SNR, for achieving all the degrees of freedom in a point-to-point MIMO system.

This condition is also true for achieving all the degrees of freedom in our network when number of layers k is fixed. More generally, when k is also growing with SNR, the following sufficient condition (for achieving all the degrees of freedom) is obtained along similar lines as in Theorem 3

$$\lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR}) \log n(\mathsf{SNR})}{\log \mathsf{SNR}} = 0$$

7.2 Implications to Ad-Hoc Networks

The main result (Theorem 2) can be extended to a multi-access network where n separate users with 1 antenna each are communicating with a destination having n antennas. The n users and the destination are separated by k layers of relays. This is essentially the same as performing VBLAST in our original single source-destination network. Hence each user can achieve one degree of freedom.

This can be applied to an static ad-hoc network as in [23]. There the entire network of area A has M single antenna nodes spread randomly with an uniform distribution over the area. The network was divided into cells of area $a(M)A$. In other words, each cell occupied a fraction $a(M)$ of the total network area which depends on the total number of nodes.

If there exists a constant c_1 and an integer M_1 such that $a(M) \geq c_1 \log M/M$ for all $M \geq M_1$, then each cell almost surely has $Ma(M) \pm \sqrt{2Ma(M)\log M}$ nodes [23]. In particular, we consider the case when $a(M) \gg \log M/M$ and hence the number of nodes in each cell is given by[1] $n \approx Ma(M)$.

[1] Recall our notation of $\approx$ and $\gg$, which means the ratio of the their two sides goes to one and infinity respectively as M tends to infinity.

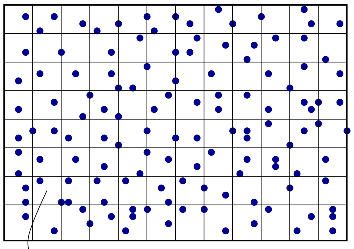

Figure 7-1: Ad-hoc network divided into multiple cells

In that model, transmissions from a cell can only reach the neighboring cells. In their communication scheme, if any node in a particular cell i is transmitting a message to cell i', then all neighboring cells of i' (other than cell i) remain quiet to ensure interference-free reception of the message. Each cell transmits at regularly spaced time-slots leaving other slots to the neighboring cells. However, only one node in a cell transmits at a time. The message from each source reaches its destination through the sequence of cells on the line joining the source-destination pair (Fig. 7-2). This is a *separation-based* strategy where the message is decoded before transmitting in the next hop.

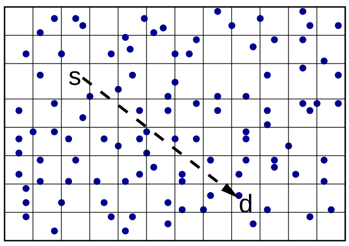

Figure 7-2: Path of a message from its source to destination: The message is hopped through the cells on the dotted line which joins the source-destination pair.

This framework fits well into our network model by considering that each relay layer is formed by the nodes in a cell on the line joining the source destination pair. Now instead of only one node in a cell transmitting at a time, we let all the nodes in a cell transmit simultaneously. They simply retransmit the received signals from the previous cell/layer without decoding. Somewhat optimistically, let us assume that each cell also has a multiantenna receiver with n antennas. Moreover, we assume that this multiantenna receiver knows the MIMO channels in the path of the message for each destination in the cell. After receiving the messages of all the destinations in its cell, it beamforms those messages to their respective destinations. Incorporating our *non-separation-based* strategy will essentially give an n fold increase in the achievable rate at high SNR. The delay analysis will remain the same as in [23].

The throughput $T(M)$ and delay $D(M)$ for the strategy in [23] is given by

$$T(M) \approx \frac{c_2}{M\sqrt{a(M)}} \quad \text{and} \quad D(M) \approx c_3/\sqrt{a(M)}$$

where c_2 and c_3 are some constants. At high SNR, incorporating our *non-separation-based* strategy gives the following n times improvement for the throughput in the throughput-delay tradeoff for $a(M) \gg \log M/M$.

$$T(M) \approx c_2\sqrt{a(M)} \quad \text{and} \quad D(M) \approx c_3/\sqrt{a(M)} \tag{7.2}$$

The delay to throughput ratio in the *separation-based* strategy of [23] is of the order of M. Equivalently, the delay is essentially M times the throughput. With our strategy the delay to throughput ratio is order of magnitude better because for our strategy

$$D(M)/T(M) \approx \frac{c_3}{c_2 a(M)} \ll M/\log M \ll M$$

In addition to this improved delay-throughput tradeoff, the maximum achievable throughput here is far greater than the maximum throughput for *separation-based* strategies in [23, 3]. The maximum throughput possible there was of the order of $1/\sqrt{M \log M}$. By choosing a fixed value of $a(M)$, our strategy can give an essentially constant throughput which does not decrease with increasing number of nodes M.

The previous discussion may seem to imply that ad-hoc networks can be made scalable i.e. the throughput per node need not decrease with increasing number of nodes. However, one has to keep in mind that it had many optimistic assumptions such as: having (extremely) high SNR, having no interference between non-neighboring cells and having a multiantenna receiver in each cell with all the required channel knowledge. Nevertheless, this analysis suggests that in some cases, significant gains might be obtained in practice by using *non-separation-based* strategies—particularly at high SNR.

7.3 Summary

We showed that in the limit of high SNR, all the degrees of freedom can be achieved whether or not there is of coordination amongst relay nodes; that is, the lack of coordination amongst relay nodes costs nothing asymptotically. We found the tradeoffs between rate, network size and SNR. We also studied the rate diversity tradeoff for this network, and how it is affected by increasing network size. Penalty functions for increasing network size were derived for the ergodic and outage formulations.

This thesis is another example (like [16]), where a very simple network operation gives asymptotically optimal performance. It also suggests that in practice, exploring *non-separation-based* strategies has the potential to provide significant improvements—specially at high SNR.

Bibliography

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecommunications*, vol. 10, pp. 585–595, Nov/Dec 1999.

[2] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[3] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. on Info. Theory*, Vol. 46, no. 2, pp. 388–404, March 2002.

[4] L. Xie and P. Kumar, "A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation," submitted to *IEEE Trans. on Info. Theory.*

[5] O. Lèvêque and I. Telatar, " "Information theoretic upper bounds on the capacity of large extended ad-hoc wireless networks", submitted to *IEEE Trans. on Info. Theory.*

[6] P. Gupta, P. Kumar, "Towards an information theory of large networks: An achievable rate region," submitted to *IEEE Trans. on Info. Theory.*

[7] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," submitted to *IEEE Trans. on Info. Theory.*

[8] P. Viswanath and D. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Downlink-Uplink Duality," to appear in *IEEE Trans. on Info. Theory*, Aug. 2003.

[9] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," submitted to *IEEE Trans. on Info. Theory.*

[10] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela,"V-BLAST: An Architecture for realizing very high data rates over the rich-scattering wireless channel," *URSI International Symposium on Signals, Systems and Electronics: Conference Proceedings*, pp. 295-300, 1998.

[11] M. Gastpar, G. Kramer, and P. Gupta, "The multiple relay channel: coding and antenna-clustering capacity," *Proc. IEEE Intl. Symp. Info. Theory*, (Lausanne, Switzerland), p. 136, July 2002.

[12] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[13] R. Muller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels ," *IEEE Trans. on Info. Theory*, Vol. 48, no. 7, pp. 2086–2091, July 2002.

[14] V. Tarokh V, N. Seshadri, and A. R. Calderbank, "Space-time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory*, Vol. 44, No. 2, pp. 744-765, Mar. 1998.

[15] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels ," *IEEE Trans. on Info. Theory*, Vol. 49, no. 5, pp. 1073–1096, May 2003.

[16] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian Relay Networks," submitted to *IEEE Trans. on Info. Theory.*

[17] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE JSAC*, pp. 1451–1458, Oct 1998.

[18] C. Rao, *Linear Statistical Inference and its Applications.* John Wiley and Sons, 1965.

[19] A. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* Academic Press, 1979.

[20] R. Muirhead, *Aspects of Multivariate Statistical Theory.* John Wiley and Sons, 1982.

[21] P. Dupuis, R. Ellis, *A weak convergence approach to the theory of large deviations.* John Wiley and Sons, 1997.

[22] A. Edelman, "Eigenvalues and condition numbers of random matrices," *PhD Thesis*, MIT, May 1989.

[23] A. Gamal, J. Mammen, B. Prabhakar and D. Shah, "Throughput Delay Tradeoff in Wireless Networks," *Proc. IEEE INFOCOM*, Hong Kong, March 2004.

Appendix A

Proof of Lemma 6

Let $U_A\Sigma_A V_A$ and $U_B\Sigma_B V_B$ be the singular value decomposition of A and B, respectively. Assume these singular values to be arranged in increasing order, for example $\sigma_1^A \leq \sigma_1^A \cdots \leq \sigma_n^A$. Now singular values of AB are same as those of $\Sigma_A V_A U_B \Sigma_B$, since multiplication by an unitary matrix does not change the singular values. We use Q for $V_A U_B$ for simpler expressions. We now write

$$\begin{aligned}
\Sigma_B &\equiv \mathrm{diag}(\sigma_n^B, \sigma_{n-1}^B, \cdots, \sigma_1^B) \\
&= \mathrm{diag}(\sigma_1^B, \sigma_1^B, \cdots, \sigma_1^B) \\
&+ \mathrm{diag}(\sigma_n^B - \sigma_1^B, \sigma_{n-1}^B - \sigma_1^B, \cdots \sigma_2^B - \sigma_1^B, 0) \\
&\triangleq \sigma_1^B I + \overline{\Sigma}_B
\end{aligned}$$

Then we have $$\Sigma_A Q \Sigma_B = \sigma_1^B \Sigma_A Q + \Sigma_A Q \overline{\Sigma}_B \triangleq \Omega_1 + \Omega_2$$

Now eigenvalues of $(\Omega_1 + \Omega_2)(\Omega_1 + \Omega_2)^\dagger$ are at least as large as eigenvalues of $\Omega_1\Omega_1^\dagger$. First, because $\Omega_1\Omega_2^\dagger + \Omega_2\Omega_1^\dagger$ and $\Omega_2\Omega_2^\dagger$ are positive semi-definite matrices. Second, because eigenvalues of sum of two positive semi-definite matrices (say C and D) are at least as large as the eigenvalues of any one of those semi-definite matrices (C or D). Hence singular values of $\Omega_1 + \Omega_2$ are at least as large as those of Ω_1. However, singular values of $\Omega_1 = \sigma_1^B \Sigma_A Q$ are same as those of $\sigma_1^B \Sigma_A$. Hence we get the following lower bound on singular values of AB denoted by σ_i^{AB},

$$\sigma_i^{AB} \geq \sigma_1^B \sigma_i^A$$

Similarly for a more general version of this result, let $1 \leq m \leq n$

$$\begin{aligned}
\Sigma_B &= \mathrm{diag}(\sigma_m^B, \cdots, \overset{\text{m'th place}}{\sigma_m^B}, 0 \cdots 0) \\
&+ \mathrm{diag}(\sigma_n^B - \sigma_m^B, \cdots \sigma_{m+1}^B - \sigma_m^B, 0, \sigma_{m-1}^B, \cdots, \sigma_1^B) \\
&\triangleq \sigma_m^B \hat{I^m} + \hat{\Sigma_B} \\
\text{Hence} \qquad \Sigma_A Q \Sigma_B &= \sigma_m^B \Sigma_A Q \hat{I^m} + \Sigma_A Q \hat{\Sigma_B} \triangleq \hat{\Omega}_1 + \hat{\Omega}_2
\end{aligned}$$

Here $\hat{I^m}$ is a diagonal matrix whose last $m-1$ diagonal entries are zero and other diagonal entries are 1. Using exactly the same arguments as before, we see that singular values of AB are at least as large as those of $\sigma_m^B \Sigma_A Q \hat{I^m}$. By Poincare separation theorem ([18]: Sec. 1f.2), we know that jth largest eigenvalue of $\Sigma_A Q \hat{I^m}$ is at least as large as $(j+m-1)$th largest eigenvalue of Σ_A. It is because multiplying by $\hat{I^m}$ can at most nullify the largest $m-1$ singular values of $\Sigma_A Q$ (i.e. of Σ_A).

To understand this, one can think of the diagonal matrix Σ_A as an ellipsoid in a complex n dimensional space, with its axis collinear to the coordinate axis. The length of each axis is equal to a singular value σ_i^A. Multiplying Σ_A by an unitary matrix corresponds to a rotation of the ellipsoid in its n dimensional space. Then multiplying $\Sigma_A Q$ with any diagonal matrix D corresponds to stretching or shrinking this rotated ellipsoid along each jth coordinate axis by a factor d_j, which is the j the diagonal entry of D. The lengths of the axis of this resultant ellipsoid are equal to the singular values of $\Sigma_A Q \hat{I^m}$. Evidently, the worst rotation Q of the original ellipsoid is when the largest $m-1$ axis of the original ellipsoid are shrunk to zero by $\hat{I^m}$. Hence the jth largest singular value of $\Sigma_A Q \hat{I^m}$ is at least equal to the $(j+m-1)$th largest singular value of Σ_A.

Note that the jth largest singular value of a matrix is its $(n+1-j)$th smallest singular value. Replacing $n+1-j$ by i gives

$$\sigma_i^{AB} \geq \sigma_m^B \sigma_{i-m+1}^A \qquad m \leq i \leq n \qquad \square$$

Appendix B

Proof of Theorems 2, 3 and 4

We saw the achievable rate for our network operation is

$$R = \mathcal{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{I} + \frac{\mathsf{SNR}}{\sqrt{n}^{2k}}(\mathbf{G}_{\mathbf{H}}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)(\mathbf{G}_{\mathbf{H}}^{-1}\mathbf{H}_k\mathbf{H}_{k-1}..\mathbf{H}_1\mathbf{H}_0)^{\dagger}\right)\right]$$

where $\mathbf{H}$ denotes the network's channel state $(\mathbf{H}_0, \cdots, \mathbf{H}_k)$. Let $\lambda_i^{l,p,\cdots q}$ denote the ith smallest eigenvalue of $(\mathbf{H}_l\mathbf{H}_p..\mathbf{H}_q)(\mathbf{H}_l\mathbf{H}_p..\mathbf{H}_q)^{\dagger}$. Note that smallest singular value of $\mathbf{G}_{\mathbf{H}}^{-1}$ is the inverse of the largest singular value $\mathbf{G}_{\mathbf{H}}$. Hence by Theorem 6 we get,

$$\begin{aligned}
R &\geq \mathcal{E}_{\mathbf{H}}\left[\sum_{i=1:n}\log\left(1 + \frac{\mathsf{SNR}}{n^k}\frac{\lambda_i^{k,..0}}{\lambda_{\max}^{G}}\right)\right] \\
&\geq \mathcal{E}_{\mathbf{H}}\left[\sum_{i=1:n}\log\left(1 + \frac{\mathsf{SNR}}{n^k}\lambda_i^{k,..0}\left(1 + \sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)^{-1}\right)\right]
\end{aligned}$$

because the largest eigenvalue of $AA^{\dagger} + BB^{\dagger}$ is at most the sum of largest eigenvalues of $AA^{\dagger}$ and $BB^{\dagger}$. Here $\lambda_{\max}^{G}$ denoted the largest eigenvalue of $\mathbf{G}_{\mathbf{H}}\mathbf{G}_{\mathbf{H}}{}^{\dagger}$. The above lower bound on rate is itself smaller than

$$\begin{aligned}
&\geq \mathcal{E}_{\mathbf{H}}\left[\sum_{i=1:n}\left(\log\left(1 + \frac{\mathsf{SNR}}{n^k}\lambda_i^{k,..0}\right) - \log\left(1 + \sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)\right)\right] \\
&\geq \mathcal{E}_{\mathbf{H}}\left[\sum_{i=1:n}\left(\log\left(\frac{\mathsf{SNR}}{n^k}\lambda_i^{k,..0}\right) - \log\left(1 + \sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)\right)\right] \\
&\geq n\log\mathsf{SNR} - (k+1)\left(\log n - \mathcal{E}_{\mathbf{H}}\left[\log\det(\mathbf{H}_m\mathbf{H}_m^{H})\right]\right) - n\mathcal{E}_{\mathbf{H}}\left[\log\left(1 + \sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)\right] \\
&\geq n\log\mathsf{SNR} - (k+1)\left(\log n - \mathcal{E}_{\mathbf{H}}\left[\log\det(\mathbf{H}_m\mathbf{H}_m^{H})\right]\right) - n\log\left(1 + \sum_{j=1:k}\mathcal{E}_{\mathbf{H}}\left[\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right]\right)
\end{aligned}$$

The second last step is because all $\mathbf{H}_i$s are i.i.d. and determinant (i.e. product of eigenvalues) of a product of matrices is same as the product of their determinants. The last step follows by Jensen's inequality.

Noticing that sum of eigenvalues i.e. trace of a positive semi-definite matrix is always larger than its largest eigenvalue we get,

$$
\begin{aligned}
n \log\left(1+\sum_{j=1:k} \mathcal{E}_{\mathbf{H}}\left[\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right]\right) &\leq n \log\left(1+\sum_{j=1:k} \mathcal{E}_{\mathbf{H}}\left[\frac{\mathrm{trace}(\mathbf{H}_{\mathrm{k,k-1..j}}\mathbf{H}^{\mathrm{H}}_{\mathrm{k,k-1..j}})}{n^{(k-j+1)}}\right]\right) \\
&= n \log\left(1+\sum_{j=1:k} \mathcal{E}_{\mathbf{H}}\left[\mathrm{trace}\frac{(\mathbf{H}_{1,2..\mathrm{j}}\mathbf{H}^{\mathrm{H}}_{1,2..\mathrm{j}})}{\mathrm{n}^{\mathrm{j}}}\right]\right) \\
&= n \log\left(1+\sum_{j=1:k} n\right) \\
&= n \log(1+kn)
\end{aligned}
$$

An inductive argument is used for proving the second last step above i.e.

$$
\mathcal{E}_{\mathbf{H}}\left[\mathrm{trace}\left(\frac{\mathbf{H}_{1,2..\mathrm{j}}\mathbf{H}^{\mathrm{H}}_{1,2..\mathrm{j}}}{\mathrm{n}^{\mathrm{j}}}\right)\right] = n \quad \forall j \tag{3}
$$

For $j=1$, since all n^2 elements of $\mathbf{H}_1$ are i.i.d. complex Gaussian with variance 1, we have $\mathcal{E}_{\mathbf{H}}\left[\mathrm{trace}\left(\frac{\mathbf{H}_1\mathbf{H}_1^{\dagger}}{\mathrm{n}}\right)\right] = n^2.1$.

Consider two independent $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ with i.i.d. entries of variance 1. We have $\mathcal{E}_{\mathbf{AB}}\left[||(\mathbf{AB})(i,j)||^2\right] = n$. Similarly, all entries of $\mathbf{A}\frac{\mathbf{B}}{\sqrt{\mathbf{n}}}$ will be i.i.d. of variance 1, i.e. same as the variance of entries of $\mathbf{A}$ and $\mathbf{B}$. This is true for all distributions of entries of $\mathbf{A}$ and $\mathbf{B}$, as long as their entries are independent of each other and have variance 1. Since all $\mathbf{H}_i$'s are independent of each other, $\mathbf{H}_1\mathbf{H}_2..\mathbf{H}_j$ is independent of $\mathbf{H}_{j+1}$. Hence[2] $\mathcal{E}_{\mathbf{H}}\left[||(\mathbf{H}_1\frac{\mathbf{H}_2}{\sqrt{n}}\cdots\frac{\mathbf{H}_{j-1}}{\sqrt{n}})\frac{\mathbf{H}_j}{\sqrt{n}}||_F^2\right] = n^2.1$. The desired result follows.

The following lower bound on achievable degrees of freedom is obtained using

[2]Frobenius norm of a matrix $||\mathbf{A}||_F$ is defined as $||\mathbf{A}||_F^2 = \sum_{i,j}|\mathbf{A}(i,j)|^2$

Eq. 3 and previously obtained lower bound on the rate.

$$\begin{aligned} r = \lim_{\mathsf{SNR}\to\infty} \frac{R}{\log \mathsf{SNR}} &\geq n - \varphi(n) \lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR})+1}{\log \mathsf{SNR}} - \lim_{\mathsf{SNR}\to\infty} \frac{n \log(kn+1)}{\log \mathsf{SNR}} \\ &= n - S_{\text{ergodic}}(k) - n \lim_{\mathsf{SNR}\to\infty} \frac{\log k(\mathsf{SNR})}{\log \mathsf{SNR}} \end{aligned}$$

For any finite value of $S_{\text{ergodic}}(k)$, the last term in the above expression equals zero. Thus we get Theorems 3 and 4. The size penalty $S_{\text{ergodic}}(k)$ is zero for any fixed k, hence all the degrees of freedom can be achieved. Thus Theorem 2 is proved.

Appendix C

Proof of Theorem 5

Recall that each $\mu_i \leq \frac{\lambda_i^{k..0}}{\lambda_{\max}^G}$ by Theorem 6. Hence,

$$\begin{aligned}
P_{\text{out}}(\mathsf{SNR}) &\leq P\left(\sum_{i=1:n} \log\left(1+\frac{\mathsf{SNR}}{n^k}\frac{\lambda_i^{k..0}}{\lambda_{\max}^G}\right) < r\log\mathsf{SNR}\right)\\
&\leq P\left(\sum_{i=1:n} \log\left(1+\frac{\mathsf{SNR}}{n^k}\frac{\lambda_i^{k..0}}{\lambda_{\max}^G}\right) < r\log\mathsf{SNR}\right)\\
&\leq P\left(\sum_{i=1:n} \log\left(1+\frac{\mathsf{SNR}}{n^k}\lambda_i^{k..0}\left(1+\sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)^{-1}\right) < r\log\mathsf{SNR}\right)\\
&\leq P\left(\sum_{i=1:n} \log\left(1+\frac{\mathsf{SNR}}{n^k}\lambda_i^{k..0}\right) - n\log\left(1+\sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right) < r\log\mathsf{SNR}\right)
\end{aligned}$$

Denote the term $n\log\left(1+\sum_{j=1:k}\frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}}\right)$ by θ. Let δ be an arbitrarily small positive number.

$$\begin{aligned}
&= P(\theta \geq \delta\log\mathsf{SNR})P\left(\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\lambda_i^{k..0}\right) - \theta < r\log\mathsf{SNR} \mid \theta \geq \delta\log\mathsf{SNR}\right)\\
&+ P(\theta < \delta\log\mathsf{SNR})P\left(\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\lambda_i^{k..0}\right) - \theta < r\log\mathsf{SNR} \mid \theta < \delta\log\mathsf{SNR}\right)
\end{aligned}$$

$$\begin{aligned}
&\leq P(\theta \geq \delta\log\mathsf{SNR})\\
&+P\left(\sum_{i=1:n}\log\left(1+\frac{\mathsf{SNR}}{n^k}\lambda_i^{k..0}\right) < (r+\delta)\log\mathsf{SNR} \mid \theta < \delta\log\mathsf{SNR}\right) \qquad (4)
\end{aligned}$$

We will now show that the first term above is much smaller than the second term. In other words, the SNR exponent is (much) more negative for the first term than the second term in the limit of high SNR.

$$\begin{aligned}
P(\theta \geq \delta \log \mathsf{SNR}) &= P\left(1 + \sum_{j=1:k} \frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}} \geq \mathsf{SNR}^{\delta/n}\right) \\
&\leq P\left(\bigcup_{j=1:k} \frac{\lambda_{\max}^{k,k-1..j}}{n^{(k-j+1)}} \geq \frac{\mathsf{SNR}^{\delta/n}-1}{k}\right) \\
&\leq P\left(\bigcup_{j=1:k} \frac{\lambda_{\max}^{k}}{n}\frac{\lambda_{\max}^{k-1}}{n}..\frac{\lambda_{\max}^{j}}{n} \geq \frac{\mathsf{SNR}^{\delta/n}-1}{k}\right) \\
&= P\left(\bigcup_{j=1:k} \frac{\lambda_{\max}^{1}}{n}\frac{\lambda_{\max}^{2}}{n}..\frac{\lambda_{\max}^{j}}{n} \geq \frac{\mathsf{SNR}^{\delta/n}-1}{k}\right)
\end{aligned}$$

Since SNR is very large, $\mathsf{SNR}^{\mathsf{d}/n}/2 > 1$. Hence we get,

$$\begin{aligned}
P(\theta \geq \delta \log \mathsf{SNR}) &= P\left(\bigcup_{j=1:k} \frac{\lambda_{\max}^{1}}{n}\frac{\lambda_{\max}^{2}}{n}..\frac{\lambda_{\max}^{j}}{n} \geq \frac{\mathsf{SNR}^{\delta/n}}{2k}\right) \\
&\leq \sum_{j=1:k} P\left(\frac{\lambda_{\max}^{1}}{n}\frac{\lambda_{\max}^{2}}{n}..\frac{\lambda_{\max}^{j}}{n} \geq \frac{\mathsf{SNR}^{\delta/n}}{2k}\right)
\end{aligned}$$

Now we substitute $\frac{\mathsf{SNR}^{\delta/n}}{2k} = L(\mathsf{SNR})$ for simpler expressions. Now recalling the fact that geometric mean is at most equal to arithmetic mean,

$$\begin{aligned}
P(\theta \geq \delta \log \mathsf{SNR}) &\leq \sum_{j=1:k} P\left(\frac{\lambda_{\max}^{1} + \lambda_{\max}^{2} + \cdots + \lambda_{\max}^{j}}{jn} \geq L(\mathsf{SNR})^{1/j}\right) \\
&\leq \sum_{j=1:k} P\left(\frac{\mathrm{trace}(\mathbf{H}_1\mathbf{H}_1^{\dagger} + \mathbf{H}_2\mathbf{H}_2^{\dagger} + \cdots + \mathbf{H}_{\mathrm{j}}\mathbf{H}_{\mathrm{j}}^{\dagger})}{jn} \geq L(\mathsf{SNR})^{1/j}\right) \qquad (5)
\end{aligned}$$

For fixed number of layers, the sum of j traces above is the sum of squared norms of jn^2 complex Gaussian random variables. Hence the j term in the above summation is equal to $P(\chi > jnL(\mathsf{SNR})^{1/j})$, where χ is a Chi-squared random variable with $2jn^2$ degrees of freedom. We know that tail of a Chi-squared distribution are exponential. In other words,

$$\lim_{\mathsf{SNR}\to\infty} \frac{-\log P(\chi > jnL(\mathsf{SNR})^{1/j})}{\log \mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty} \frac{jnL(\mathsf{SNR})^{1/j}}{\log \mathsf{SNR}} \tag{6}$$

This limit equals infinity after substituting the value of $L(\mathsf{SNR})$. As all the terms in the summation in Eq. 5, the SNR exponent of that summation also is infinite. Thus the SNR exponent of the first term in Eq. 4 is infinite. The second term must have a finite SNR exponent, otherwise infinite diversity will be achieved. Now choosing arbitrarily small δ in Eq. 4 proves the theorem for fixed number of layers.

Now consider the case when number of layers goes to infinity such that $S_{\text{outage}}(k) = \lim_{\mathsf{SNR}\to\infty} \frac{k(\mathsf{SNR})n \log\log \mathsf{SNR}}{\log \mathsf{SNR}}$ is finite[3]. There are two possibilities in this case. Either the largest term of the summation in Eq. 5 is for some finite j or it is for[4] $j = k$. If it is for some finite j, the SNR exponent of that term is given by Eq. 6. To find the SNR exponent for the case $j = k$, we apply central limit theorem to the sum of traces. Hence the distribution of $\chi/\sqrt{k}$ becomes a Gaussian with mean $\sqrt{k}n^2$ and variance equal to the variance of trace$(\mathbf{H}_1\mathbf{H}_1^\dagger)$, denoted by τ^2. Thus the SNR exponent of the last term of the summation in Eq. 5 is

$$\begin{aligned}
&\lim_{\mathsf{SNR}\to\infty} \frac{-\log P(\chi > knL(\mathsf{SNR})^{1/k})}{\log \mathsf{SNR}} \\
= &\lim_{\mathsf{SNR}\to\infty} \frac{-\log P\left(\frac{\chi}{\sqrt{k}} - \sqrt{k}n^2 > \sqrt{k}nL(\mathsf{SNR})^{1/k} - \sqrt{k}n^2\right)}{\log \mathsf{SNR}} \\
\leq &\lim_{\mathsf{SNR}\to\infty} \frac{-\log P\left(\frac{\chi}{\sqrt{k}} - \sqrt{k}n^2 > \sqrt{k}nL(\mathsf{SNR})^{1/k}\right)}{\log \mathsf{SNR}} \\
= &\lim_{\mathsf{SNR}\to\infty} \frac{-\log Q(\frac{\sqrt{k}nL(\mathsf{SNR})^{1/k}}{\tau})}{\log \mathsf{SNR}}
\end{aligned}$$

$$= \lim_{\mathsf{SNR}\to\infty} \frac{(\sqrt{k}nL(\mathsf{SNR})^{1/k})^2}{2\tau^2 \log \mathsf{SNR}} \tag{7}$$

[3]It turns out from Theorem 9 that no diversity is achieved any way if this limit larger than 1.
[4]Henceforth we use simply k for $k(\mathsf{SNR})$.

The last equality follows due to the fact that $\lim_{x\to\infty}\frac{Q(x)}{\exp(-x^2/2)}=1$. Comparing Eq. 6 to Eq. 7, tells that the smallest exponent is when the $j=k$. Take the ratio of the two SNR exponents to see this,

$$\begin{aligned}\frac{2\tau^2}{n}\frac{jL(\mathsf{SNR})^{1/j}}{kL(\mathsf{SNR})^{2/k}} &\geq \frac{2\tau^2}{n}\frac{L(\mathsf{SNR})^{\frac{1}{j}-\frac{2}{k}}}{k}\\ &\geq \frac{2\tau^2}{n}\frac{L(\mathsf{SNR})^{\frac{1}{k}}}{k}\end{aligned}$$

This goes to infinity as SNR goes to infinity[5]. Thus in the sum in Eq. 5, the last term is the largest term i.e. the term with the smallest SNR exponent.

Another way to show this is by observing that $jL(\mathsf{SNR})^{1/j}$ is a decreasing sequence for large enough $L(\mathsf{SNR})$. This can be shown by differentiating it with respect to j. Using this fact, we get the following upper bound on the sum in Eq. 5.

$$P(\theta\geq\delta\log\mathsf{SNR})\leq kP\left(\frac{\mathrm{trace}(\mathbf{H}_1\mathbf{H}_1^\dagger+\mathbf{H}_2\mathbf{H}_2^\dagger+\cdots+\mathbf{H}_\mathrm{k}\mathbf{H}_\mathrm{k}^\dagger)}{kn}\geq L(\mathsf{SNR})^{1/k}\right)$$

Now comparing the SNR exponent of the two sides by using Eq. 7,

$$\begin{aligned}\lim_{\mathsf{SNR}\to\infty}\frac{-\log P(\theta\geq\delta\log\mathsf{SNR})}{\log\mathsf{SNR}} &\geq \lim_{\mathsf{SNR}\to\infty}\frac{-\log k}{\log\mathsf{SNR}}+\lim_{\mathsf{SNR}\to\infty}\frac{(\sqrt{k}nL(\mathsf{SNR})^{1/k})^2}{2\tau^2\log\mathsf{SNR}}\\ &= \frac{n^2}{2\tau^2}\lim_{\mathsf{SNR}\to\infty}\frac{k(\mathsf{SNR})L(\mathsf{SNR})^{2/k}}{\log\mathsf{SNR}} \qquad \text{(As } S_{\mathrm{outage}}(k) \text{ is finite)}\\ &= \infty\end{aligned}$$

Last step follows by substituting the expression for $L(\mathsf{SNR})$ and recalling that $S_{\mathrm{outage}}(k)$ is finite. Thus Theorem 5 is true even when k goes to infinity such that $S_{\mathrm{outage}}(k)$ is finite.

[5]This is because the $S_{\mathrm{outage}}(k)$ to is finite.

Appendix D

Proof of Footnote 3 in Chapter 4

Let each source antenna transmit power P/C, instead of P. This C denotes a large but fixed number. Now we use the fact that all diagonal entries of a positive semi-definite matrix are upper bounded by it largest eigenvalue. Now recalling Eq. 4.3 with P replaced by P/C. As the maximum eigenvalue of a sum of positive semi-definite matrices is at most the sum of their maximum eigenvalues, the power transmitted by any relay node in jth layer is channel state H is upper bounded by

$$\begin{aligned}
\mathcal{E}_{\mathbf{x}_j}\left[\mathbf{x}_j\mathbf{x}_j^{\dagger}|\mathbf{H}=H\right] &\leq \frac{P}{Cn^j}\lambda_{\max}^{j-1,j-2\cdots 0}+\frac{\sigma^2}{n}+\sum_{i=1}^{j-1}\frac{\sigma^2}{n^{j-i+1}}\lambda_{\max}^{j-1\cdots i} \\
&\leq \frac{P}{Cn^j}\prod_{l=0}^{j-1}\lambda_{\max}^{l}+\frac{\sigma^2}{n}+\sum_{i=1}^{j-1}\frac{\sigma^2}{n^{j-i+1}}\prod_{l=i}^{j-1}\lambda_{\max}^{l}
\end{aligned}$$

Let Φ be such that $\Pr(\lambda_{\max}^{l}>\Phi)=\epsilon$, where ϵ is arbitrarily small. As discussed before, the noise component has no contribution to the transmitted power. Hence the power transmitted by any node in jth layer is lower than $\frac{P}{C}\frac{\Phi^j}{n^j}$, with at least a probability of $(1-\epsilon)^j$. Thus if $C=\frac{\Phi^k}{n^K}$, all layers will obey the power constraint with at least a probability of $(1-\epsilon)^k$. For simplifying the proof, we make an unrealistic (and pessimistic) assumption that the network is turned *off* in a fading block where the power constraint is not satisfied. Then the network will be *on* almost always i.e. with probability $(1-\epsilon)^k$. Also note that this C is a fixed number, so the degrees of freedom achieved are unchanged when SNR goes to infinity (see Eq. 4.1). This strategy can thus achieve at least $n(1-\epsilon)^k$ degrees of freedom.

Hence by proper choice of ϵ (and hence C), any $r<n$ degrees of freedom can be achieved. In practice however, the network need not be turned off when

the power constraint is not being satisfied. For example, all relays could scale down the received signal little more before transmission than the usual $\sqrt{n}$ factor. This strategy will achieve any $r \leq n$ degrees of freedom. Thus having a power constraint on each fading block does not hamper achieving all the degrees of freedom.

However, this result is true only if the number of layers is fixed. It is not valid when it goes to infinity as in chapters 5 and 6. It is because, $(1-\epsilon)^k$ goes to zero as k goes to infinity for any fixed ϵ or (equivalently) any fixed C.