Energy Efficient Multiple Antenna Communication

by

Siddharth Ray

B.Tech., EE, Indian Institute of Technology Madras, 2002
S.M., EECS, Massachusetts Institute of Technology, 2003

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2006

Author .....
Department of Electrical Engineering and Computer Science
August 21, 2006

Certified by .....
Muriel Médard
Associate Professor
Thesis Supervisor

Certified by .....
Lizhong Zheng
Associate Professor
Thesis Supervisor

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

Energy Efficient Multiple Antenna Communication

by

Siddharth Ray

Submitted to the Department of Electrical Engineering and Computer Science
on August 21, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

We consider a multiple-input, multiple-output (MIMO) wideband Rayleigh block fading channel where the channel state is unknown at the transmitter and receiver and there is only an average input power constraint. We compute the capacity and analyze its dependence on coherence length, number of antennas and receive signal-to-noise ratio (SNR) per degree of freedom. We establish conditions on the coherence length and number of antennas for the non-coherent channel to have a "near coherent" performance in the wideband regime. We also propose a signaling scheme that is near-capacity achieving in this regime. We compute the decoding error probability and study its dependence on SNR, number of antennas and coherence length. We show that error probability decays inversely with coherence length and exponentially with the product of the number of transmit and receive antennas. Moreover, in the wideband regime, channel outage dominates error probability and the critical and cut-off rates are much smaller than channel capacity.

In the second part of this thesis, we introduce the concept of a fiber aided wireless network architecture (FAWNA), which allows high-speed mobile connectivity by leveraging the speed of optical networks. Specifically, we consider a single-input, multiple-output (SIMO) FAWNA, which consists of a SIMO wireless channel interfaced with an optical fiber channel through wireless-optical interfaces. We propose a design where the received wireless signal at each interface is sampled and quantized before being sent over the fiber. The capacity of our scheme approaches the capacity of the architecture, exponentially with fiber capacity. We also show that for a given fiber capacity, there is an optimal operating wireless bandwidth and number of interfaces. We show that the optimal way to divide the fiber capacity among the interfaces is to ensure that each interface gets enough rate so that its noise is dominated by front end noise rather than by quantizer distortion. We also show that rather than dynamically change rate allocation based on channel state, a less complex, fixed rate allocation scheme can be adopted with very small loss in performance.

Thesis Supervisor: Muriel Médard
Title: Associate Professor

Thesis Supervisor: Lizhong Zheng
Title: Associate Professor

To my parents,

Chitra and Shankar Ray.

Acknowledgments

I thank my advisor, Muriel Médard, for her guidance and support throughout my graduate life at MIT. Working with such a talented person has surely been my privilege. Her excellent foresight in identifying promising research directions, has greatly benefited me through my student years. Over the past few years, she has been a great source of motivation for me, both professionally as well as personally. Muriel, you truly are a role model!

I also thank my thesis co-supervisor, Lizhong Zheng. His excellent insights, intuition and problem solving techniques have greatly helped me. It's been a pleasure working with him, and I greatly appreciate his always being available to help, especially when I hit dead ends in the course of my research.

I thank my other thesis committee members, Pierre Moulin and Pramod Viswanath, for their valuable suggestions in improving my thesis. It has been great knowing both of you.

I take this opportunity to acknowledge the staff of LIDS for all their help during the last four years. Here, I would like to especially thank Doris Inslee and Michael Lewy. Life at MIT wouldn't have been so enriching, hadn't it been for some of my friendships that developed here. Here, I would like to mention the names of Saikat Guha, Debajyoti Bera, Jay Kumar Sundararajan, Smeet Deshmukh, Shashibhushan Borade and Mithila Azad Jha. I have learned a lot from each one of you, and the fond memories of our lives at MIT will indelibly remain in my mind. I also take this opportunity to recognize my aunt, Bandana Mishra, who over the years, has been a great mentor and friend. All of you occupy a very special place in my heart.

Finally, I thank my parents, Chitra and Shankar Ray. They have always been with me and supported me. Whatever I am today, it is because of them. Words are not enough to thank them. I dedicate this thesis to my parents.

I acknowledge the following grants that supported this thesis: NSF ANI-0335256, NSF CNS-0434974 and Stanford University PY-1362.

Contents

List of Figures

Chapter 1

Introduction

This thesis addresses two broad topics in multiple antenna systems. The first is the fundamental information theoretic limits of multiple antenna systems in the wideband regime and the second is design of high-speed mobile communication networks using multiple antennas in a distributed fashion.

1.1 Multiple antenna systems in the wideband regime

Recent years have seen the emergence of high data rate, third generation wideband wireless communication standards like wideband code division multiple access (W-CDMA) and Ultra-wideband (UWB) radio. Motivated by the ever increasing demand for higher wideband wireless data rates, we consider multiple antenna communication over the wideband wireless channel.

At the cost of additional signal processing (which is getting cheaper with rapid advances in VLSI technology), multiple-input, multiple-output (MIMO) systems have been known to improve considerably performance of wireless systems in terms of reliability as well as throughput, without requiring additional resources such as bandwidth and power. However, multiple antenna research has focused primarily in the regime where the received signal-to-noise ratio (SNR) per degree of freedom is high. Such a

regime operates in essence as a narrowband regime. We now study the performance of MIMO at the other extreme, i.e., when the available bandwidth is large, which takes us to the regime where the SNR per degree of freedom is low.

In wideband channels, the available power is spread over a large number of degrees of freedom. This makes the SNR per degree of freedom low. Hence, while studying these channels, we need to focus on the low SNR regime. We will therefore use the terms "wideband" and "low SNR" interchangeably, with the understanding that the latter refers to the SNR per degree of freedom.

The study of single antenna wideband channels dates back to 1969 and early work has considered the Rayleigh fading channel model. Kennedy [4] shows that the capacity of an infinite bandwidth Rayleigh fading channel is the same as that of an infinite bandwidth additive white Gaussian noise (AWGN) channel with the same average received power. Using the results of Gallager [9], Telatar [10] obtains the capacity per unit energy for the Rayleigh fading channel as a function of bandwidth and signal energy, concluding that given an average power constraint, the Rayleigh fading and AWGN channels have the same capacity in the limit of infinite bandwidth. Telatar and Tse [24] show that this property of the channel capacity is also found in channels with general fading distributions.

Médard and Gallager [18, 28] establish that very large bandwidths yield poor performance for systems that spread the available power uniformly over time and frequency (for example DS-CDMA). They express the input process as an orthonormal expansion of basis functions localized in time and frequency. The energy and fourth moment of the coefficients scale inversely with the bandwidth and square of the bandwidth, respectively. By constraining the fourth moment (as is the case when using spread spectrum signals), they show that mutual information decays to 0 inversely with increasing bandwidth. Telatar and Tse [24] consider a wideband fading channel to be composed of a number of time-varying paths and show that the input signals needed to achieve capacity must be "peaky" in time or frequency. They also show that if white-like signals are used (as for example in spread spectrum communication), the mutual information is inversely proportional to the number of resolvable paths with

energy spread out and approaches 0 as the number of paths get large. This does not depend on whether the paths are tracked perfectly at the receiver or not. A strong coding theorem is obtained for this channel in [36]. Subramanian and Hajek [29] derive similar results as [18, 28] using the theory of capacity per unit cost, for a certain fourth-order cost function, called fourthegy.

We now consider the use of multiple antennas over these channels. MIMO channels were first studied from a capacity point of view in [13, 21]. In a Rayleigh flat-fading environment with perfect channel state information (CSI) at the receiver (coherent channel) but no CSI at the transmitter, and statistically independent propagation coefficients between all pairs of transmit and receive antennas, the multiple antenna capacity increases linearly with the smaller of the number of transmit and receive antennas, provided the signal-to-noise ratio is high [13].

When CSI is unavailable at the transmitter as well as the receiver, the channel is referred to as a non-coherent channel. In [20], Marzetta and Hochwald derive the structure of the optimal input matrix for this channel as a product of two statistically independent matrices; one of them being an isotropically distributed unitary matrix and the other being a diagonal, real and non-negative matrix. They also show that there is no gain, from the point of view of capacity, in having the number of transmit antennas be more than the coherence interval (in symbols) of the channel. Zheng and Tse [27] obtain the non-coherent MIMO capacity in the high SNR regime and show that in this regime, the number of transmit antennas required need not be more than half the coherence interval (in symbols).

In this thesis, we assume that the transmitter and receiver have no CSI. Hence, we study the non-coherent channel. We also assume Rayleigh block fading. In the limit of infinite bandwidth, Zheng and Tse [27] show that the capacities per degree of freedom for the coherent and non-coherent MIMO channels are the same, i.e.,

$$\lim_{\mathsf{SNR}\to 0} \frac{C_{\mathsf{coherent}}(\mathsf{SNR})}{\mathsf{SNR}} = \lim_{\mathsf{SNR}\to 0} \frac{C_{\mathsf{non-coherent}}(\mathsf{SNR})}{\mathsf{SNR}} = r,$$

where, r is the number of receive antennas and SNR is the average signal-to-noise ratio per degree of freedom at each receive antenna.

In this thesis, we use the following notation:

$$f(x) = O(g(x)) \quad \Leftrightarrow \quad \lim_{x \to 0} \frac{f(x)}{g(x)} = \text{non-zero constant},$$
$$f(x) = o(g(x)) \quad \Leftrightarrow \quad \lim_{x \to 0} \frac{f(x)}{g(x)} = 0.$$

The capacity can thus be expressed as:

$$C(\mathsf{SNR}) = r\mathsf{SNR} + o(\mathsf{SNR}) \text{ nats/channel use}$$

and is thus a linear function only in the limit of low SNR. As SNR increases from 0, capacity increases in a sublinear fashion, showing that low SNR communication is power efficient.

Using a Taylor series expansion, Verdú [30] shows that the second derivative of the capacity at $\mathsf{SNR} = 0$ is finite for the coherent channel. The impact on the coherent capacity of antenna correlation, Ricean factors, polarization diversity and out-of-cell interference is considered in [34]. For the non-coherent channel, Verdú [30] shows that though "flash" signaling is first order optimal, it renders the second derivative $-\infty$. Hence, the coherent and non-coherent channels have the same linear term and differ in their sublinear term. Therefore, as bandwidth is increased, the non-coherent channel capacity approaches the wideband limit slower than the coherent channel capacity.

Let us define the sublinear term for the MIMO channel with t transmit and r receive antennas as

$$\Delta^{(t,r)}(\mathsf{SNR}) \triangleq r\mathsf{SNR} - C(\mathsf{SNR}) \text{ nats/channel use.}$$

Computing the sublinear term tells us the capacity and also quantifies the convergence of the capacity function to the low SNR limit. Smaller the sublinear term, faster the convergence. Using the results of Verdú [30], the sublinear term for the Rayleigh fading coherent MIMO channel, $\Delta^{(t,r)}_{\mathsf{coherent}}(\mathsf{SNR})$, is

$$\Delta^{(t,r)}_{\mathsf{coherent}}(\mathsf{SNR}) = \frac{r(r+t)}{2t}\mathsf{SNR}^2 + o(\mathsf{SNR}^2).$$

On the other extreme, for the i.i.d Rayleigh fading non-coherent MIMO channel, the sublinear term, $\Delta^{(t,r)}_{\mathsf{i.i.d}}(\mathsf{SNR}) \gg O(\mathsf{SNR}^2)$ [30]. In this paper, we compute $\Delta^{(t,r)}_{\mathsf{i.i.d}}(\mathsf{SNR})$

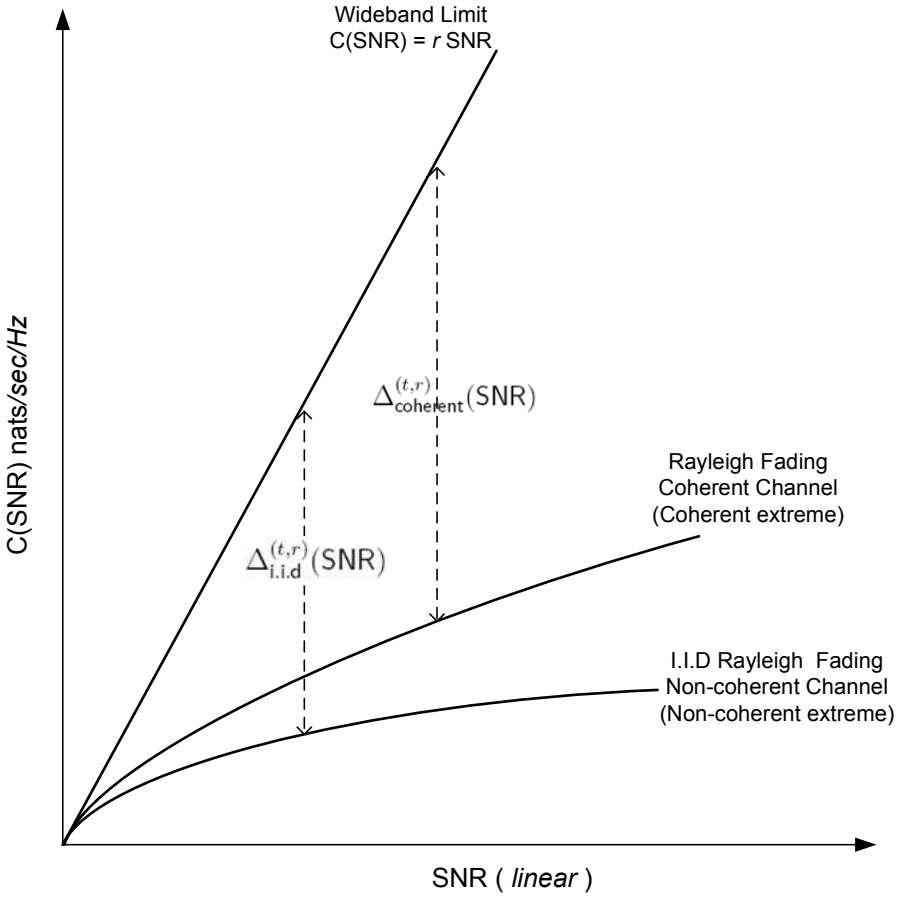

Figure 1-1: Sublinear capacity term.

and show that on-off signaling achieves capacity for the i.i.d Rayleigh fading non-coherent MIMO channel.

Figure 1-1 shows the sublinear terms for the Rayleigh fading coherent channel and the i.i.d Rayleigh fading non-coherent channel with the same number of transmit and receive antennas. A property of the non-coherent capacity is that it tends towards the coherent capacity as the coherence length increases. Hence, the sublinear term for the i.i.d Rayleigh fading non-coherent channel is the largest (non-coherent extreme), whereas, for the coherent channel, it is the smallest (coherent extreme). In this thesis, we focus on how the non-coherent MIMO channel capacity is influenced by the coherence length, number of antennas and SNR. We do so, by computing the sublinear term, which in turn tells us the capacity of the low SNR non-coherent MIMO channel of arbitrary coherence length. Thereby, we sweep the region, shown in Figure 1-1, between the coherent and non-coherent extremes.

In the low SNR regime, the sublinear term also represents the energy efficiency of communication. Let E_n and N_0 represent the energy per information nat and the noise spectral level, respectively. We have:

$$\begin{aligned}\frac{E_n}{N_0} &= \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \\ &= \frac{\mathsf{SNR}}{r\mathsf{SNR} - \Delta^{(t,r)}(\mathsf{SNR})} \\ &= \frac{1}{r}\frac{1}{1 - \frac{\Delta^{(t,r)}(\mathsf{SNR})}{r\mathsf{SNR}}}.\end{aligned}$$

Taking logarithms on both sides,

$$\log\left(\frac{E_n}{N_0}\right) \approx \frac{\Delta^{(t,r)}(\mathsf{SNR})}{r\mathsf{SNR}} - \log(r). \tag{1.1}$$

Equation (1.1) shows how energy efficiency is related to the sublinear term. The smaller the sublinear term for a channel, the more energy efficient will it be. As the non-coherent capacity is always less than the coherent capacity for the same number of transmit and receive antennas, lack of receiver CSI results in energy inefficiency. Also, note that the minimum energy (in dB) required to reliably transmit one information nat decreases logarithmically with the number of receive antennas.

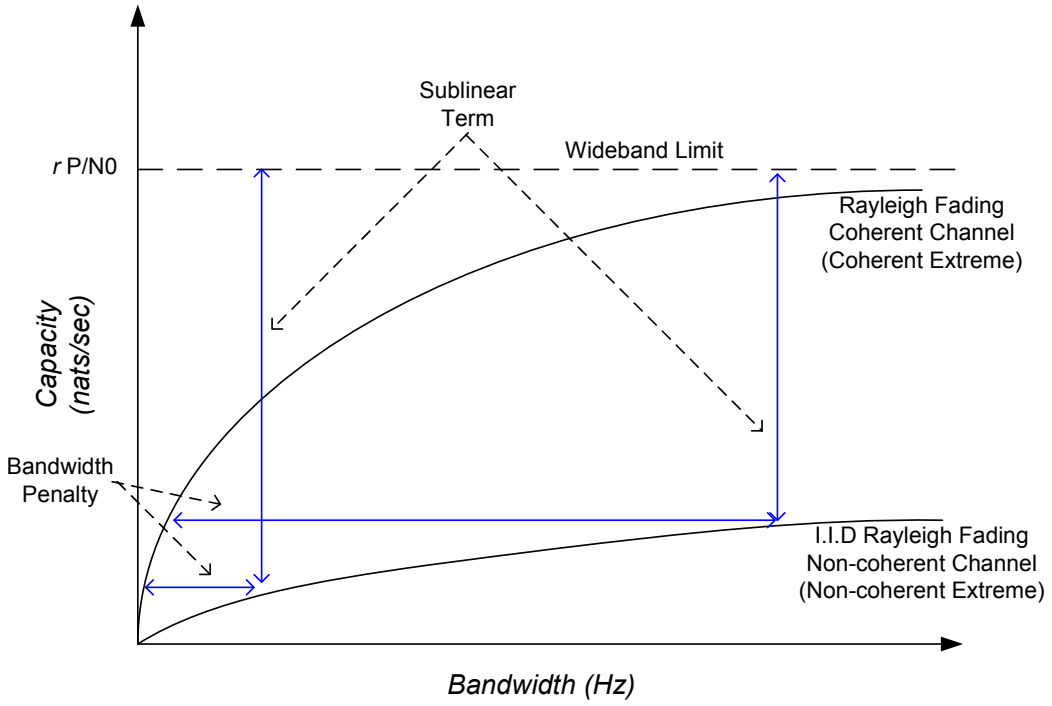

Figure 1-2: Wideband capacity.

Let us now turn to Figure 1-2, which shows how wideband capacity changes with bandwidth. We denote P as the average receive power and N_0 as the noise spectral density, which makes the wideband limit $\frac{rP}{N_0}$ nats/sec. We obtain this figure by scaling the y-axis of Figure 1-1 by the bandwidth. For the same number of transmit and receive antennas, the non-coherent capacity is less than the coherent capacity. Thus, the non-coherent channel requires a larger bandwidth in order to reliably support the same throughput as the coherent channel. We define bandwidth penalty as the extra bandwidth required by a non-coherent channel to have the same capacity as that of its coherent counterpart. Channels whose capacities converge slowly to the wideband limit have to incur large bandwidth penalties. Bandwidth penalty grows with bandwidth. Hence, for the non-coherent channel, as we approach the wideband limit, we gain in terms of energy efficiency (as the sublinear term decreases), but the bandwidth penalty becomes larger. We quantify this effect by computing the low

SNR non-coherent MIMO capacity. Studying how capacity changes with coherence length also tells us the amount of bandwidth required to achieve a "near coherent" performance. Note that since bandwidth penalty increases with decreasing coherence length, the channel at the non-coherent extreme (i.i.d Rayleigh fading non-coherent channel) has to incur the largest bandwidth penalty.

Prelov and Verdú [38] consider non-coherent communication with the input distribution constrained to be exponentially decaying. It is shown that the capacity per degree of freedom in the low SNR regime is $O(\mathsf{SNR}^2)$. Reference [39] considers the same capacity under the constraint that only the fourth and sixth order moments of the input are finite. Once again, the non-coherent capacity per degree of freedom is shown to be $O(\mathsf{SNR}^2)$. Hence, [38, 39] show that when there is a higher order (fourth and above) constraint on the input, capacity scales inversely with bandwidth. Thus, the non-coherent capacity does not approach the wideband limit and diverges from the coherent capacity as bandwidth increases. These results are akin to the single antenna channel results [18, 24, 28, 29]. Hassibi and Hochwald [32] propose a training scheme that is near-optimal in the high SNR regime. However, at low SNR, their scheme results in the rate per degree of freedom to go as $O(\mathsf{SNR}^2)$. Since the overall rate decays to 0 inversely with bandwidth, their training scheme is not desirable at low SNR.

In chapters 2 and 3 of this thesis, we consider multiple antenna communication over a wideband, non-coherent Rayleigh block fading channel. In chapter 2, we compute the capacity with only an average power constraint and consider its interaction with the coherence length of the channel, number of transmit and receive antennas and SNR. We establish how large the coherence length has to be in order for a non-coherent channel to have a "near coherent" performance at low SNR. More specifically, we show that, if the channel coherence length is above a certain antenna and SNR dependent threshold, the non-coherent and coherent capacities are the same in the low SNR regime. We show that the transmit antennas affect the sublinear capacity term and hence, the approach of capacity to the wideband limit with increasing bandwidth. We also propose a signaling scheme that is near-optimal in the wideband

regime. The capacity problem that we consider in this paper has been considered for single antenna channels by Zheng, Tse and Médard [42]. They consider the interaction between coherence length and capacity at low SNR and compute the order of the sublinear capacity term. The work in this paper builds on their work, where, we analyze the more general MIMO channel and *exactly* compute the sublinear capacity term. We use a finer scale of analysis than [42], which allows us to understand how the transmit and receive antennas affect the sublinear capacity term and hence, the approach of the non-coherent capacity to the wideband capacity limit.

We analyze the error probability for the non-coherent low SNR MIMO channel in chapter 3. The behavior of error probability for the coherent [19, 33] as well as non-coherent [22, 31] MIMO channels has been well studied in the high SNR regime. For the coherent MIMO channel with coherence length 1 symbol, the error exponent is computed by Telatar [21] for any SNR. The behavior of the error exponent for the non-coherent MIMO channel in the low SNR regime has recently been considered by Wu and Srikant in [41]. Their analysis considers the linear capacity term, $r\mathsf{SNR}$, and the error exponent is computed by fixing the coherence length and letting SNR tend to 0.

Our consideration of the effect of the interaction among SNR, number of transmit and receive antennas, and coherence length, on the error probability, yields a detailed characterization of the error probability behavior. While we consider a less general family of input distribution functions than [41], we establish results for a much wider set of operating regimes in terms of the relative values of antennas, coherence and SNR. Our analysis shows that in the low SNR regime, the critical rate as well as the cut-off rate are much smaller than the channel capacity. Moreover, the error probability decays inversely with coherence length. We introduce the notion of "diversity" in the low SNR regime and use it to show that error probability decays exponentially with the product of the number of transmit and receive antennas. Hence, in terms of reliability in the wideband regime, transmit antennas have the same importance as receive antennas. In the high SNR regime, it is well known that outage dominates the error probability. Our analysis shows that this is true even at low SNR, i.e., channel

outage dominates the error probability at low SNR.

1.2 High-speed mobile communication networks

There is a considerable demand for increasingly high-speed communication networks with mobile connectivity. Traditionally, high-speed communication has been efficiently provided through wireline infrastructure, particularly based on optical fiber, where bandwidth is plentiful and inexpensive. However, such infrastructure does not support mobility. Instead, mobile communication is provided by wireless infrastructure, most typically over the radio spectrum. However, limited available spectrum and interference effects limit mobile communication to lower data rates.

We introduce the concept of a fiber aided wireless network architecture (FAWNA), which allows high-speed mobile connectivity by leveraging the speed of optical networks. Optical networks have speeds typically in hundreds of Megabit/sec or several Gigabit/sec (Gigabit Ethernet, OC-48, OC-192, etc.). In the proposed architecture, the network coverage area is divided into zones such that an optical fiber "bus" passes through each zone. Connected to the end of the fiber is a bus controller/processor, which coordinates use of the fiber as well as connectivity to the outside world. Along the fiber are radio-optical converters (wireless-optical interfaces), which are access points consisting of simple antennas directly connected to the fiber. Each of these antennas harvest the energy from the wireless domain to acquire the full radio bandwidth in their local environment and place the associated waveform onto a subchannel of the fiber. Within the fiber, the harvested signals can be manipulated by the bus controller/processor and made available to all other antennas. In each zone, there may be one or more active wireless nodes. Wireless nodes communicate between one another, or to the outside world, by communicating to a nearby antenna. Thus any node in the network is at most two hops away from any other node, regardless of the size of the network. In general, each zone is generally covered by several antennas, and there may also be wired nodes connected directly to the fiber.

This architecture has the potential to reduce dramatically the interference effects that

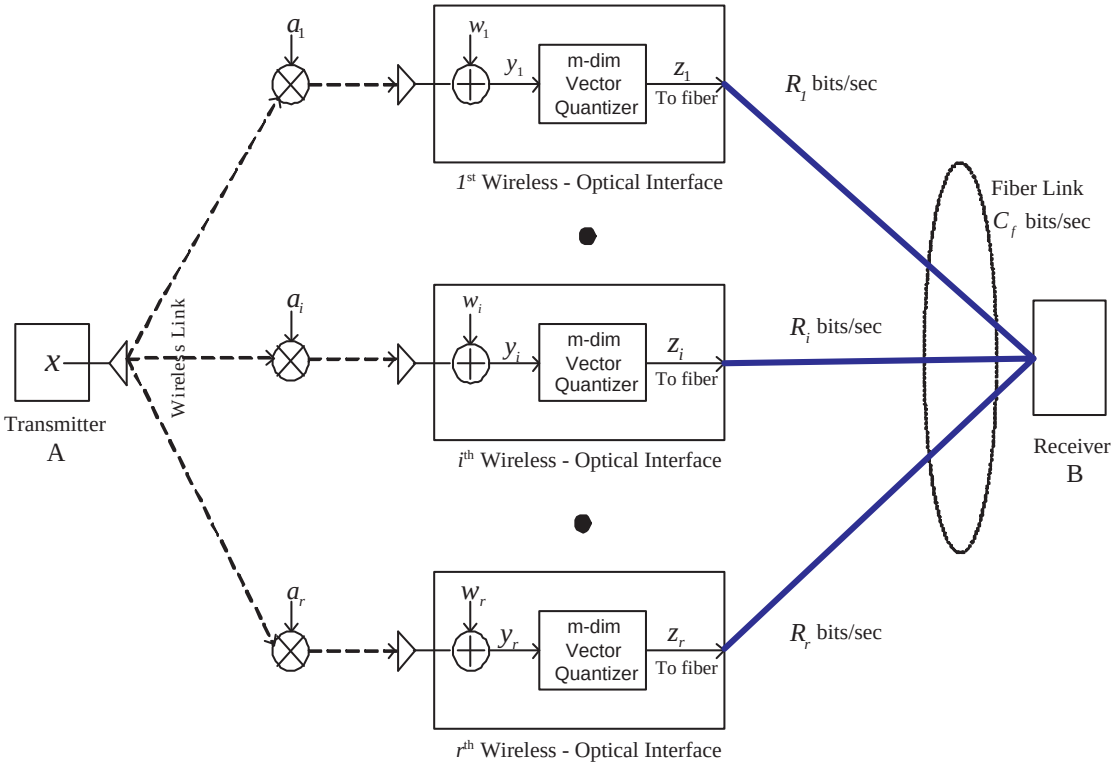

Figure 1-3: A SIMO fiber aided wireless network architecture.

limit scalability and the energy-consumption characteristics that limit battery life, in pure wireless infrastructures. A FAWNA uses the wireline infrastructure to provide a distributed means of aggressively harvesting energy from the wireless medium in areas where there is a rich, highly vascularized wireline infrastructure and distributing in an effective manner energy to the wireless domain by making use of the proximity of transmitters to reduce interference.

Tonguz and Jung [14, 16] consider the design of a similar network architecture where a given area is divided into microcells. Each microcell has a passive base station consisting of a single antenna and a laser diode. The wireless signals directly modulate this laser diode. The passive base stations are connected to a central base station using a single-mode fiber, and the signal from each microcell is decoded separately. References [26, 35, 37] consider designs which use heating, ventilation, and air conditioning (HVAC) ducts for communicating the radio frequency signals to and from the central base station.

In chapter 4, we consider a single-input, multiple-output (SIMO) fiber aided wireless network architecture. We will also refer to this as SIMO-FAWNA. Figure 1-3 shows such a link between two points A and B. The various quantities in the figure will be described in detail in chapter 4. In the two hop link, the first hop is over a wireless channel and the second, over a fiber optic channel. The links we consider are ones where the fiber optic channel capacity is larger than the wireless channel capacity.

The transmitter at A transmits information to intermediate wireless-optical interfaces over a wireless SIMO channel. The wireless-optical interfaces then relay this information to the destination, B, over a fiber optic channel. The end-to-end design is done to maximize the transmission rate from A to B. Since a FAWNA has a large number of wireless-optical interfaces, an important design objective is to keep the wireless-optical interface as simple as possible without sacrificing too much in performance.

Our problem has a similar setup, but a different objective than the CEO problem [15]. In the CEO problem, the rate-distortion tradeoff is analyzed for a given source that needs to be conveyed to the CEO through an asymptotically large number of agents. Rate-distortion theory, which uses infinite dimensional vector quantization, is used to analyze the problem. We instead compute the maximum end-to-end rate at which reliable communication is possible. In general, duality between the two problems doesn't exist. Unlike the CEO problem, the number of wireless-optical interfaces is finite and the rate (from interface to receiver B) per interface is high, owing to the fiber capacity's being large. Finite-dimensional, high resolution quantizers are used at the interfaces.

A FAWNA is an example of a channel model where quantization is performed between the source/channel encoding and decoding operations. Another example is a communication system where the receiver quantizes the incoming signal prior to decoding (Receiver implementation using a digital signal processor). Our analysis extends to all such channel models.

Let us denote the capacities of the wireless and optical channels as $C_w(P, W, r)$ and C_f bits/sec, respectively, where, P is the average transmit power at A, W is the wireless transmission bandwidth and r is the number of wireless-optical interfaces.

Since, as stated earlier, we consider links where $C_w(P, W, r) \leq C_f$, the capacity of a SIMO-FAWNA, $C_{\mathsf{SIMO}}(P, W, r, C_f)$, can be upper bounded as

$$C_{\mathsf{SIMO}}(P, W, r, C_f) < \min\left\{C_w(P, W, r), C_f\right\} = C_w(P, W, r) \text{ bits/sec.} \tag{1.2}$$

One way of communicating over a SIMO-FAWNA is to decode and re-encode at the wireless-optical interface. A major drawback of the decode/re-encode scheme is significant loss in optimality because "soft" information in the wireless signal is completely lost by decoding at the wireless-optical interface. Hence, multiple antenna gain is lost. Moreover, decoding results in the wireless-optical interface's having high complexity and requires the interface to have knowledge of the transmitter code book.

We propose a design where the wireless signal at each wireless-optical interface is sampled and quantized using a fixed rate memoryless vector quantizer before being sent over the fiber. Hence, the interfaces use a forwarding scheme. Since transmission of continuous values over the fiber is practically not possible using the type of modulation commonly associated with commercial lasers, quantization is necessary for the implementation of a forwarding scheme in a FAWNA. The proposed design has quantization between end-to-end coding and decoding. Knowledge of the transmitter code book is not required at the wireless-optical interface. The loss in "soft" information due to quantization of the wireless signal, goes to 0 asymptotically with increase in fiber capacity. The interface has low complexity, is practically implementable, is extendable to FAWNAs with large number of transmitters and interfaces and, offers adaptability to variable rates, changing channel conditions and node positions.

We show that the capacity of our scheme approaches the upper bound (1.2), exponentially with fiber capacity. The proposed scheme is thus near-optimal since, the fiber capacity is larger than the wireless capacity. Low dimensional (or even scalar) quantization can be done at the interfaces without significant loss in performance. Not only does this result in low complexity, but also smaller (or no) buffers are required, thereby further simplifying the interface. Hui and Neuhoff [12] show that asymptotically optimal quantization can be implemented with complexity increasing at most polynomially with the rate. We establish the optimal way in which fiber

capacity should be divided among the interfaces (interface rate allocation) and investigate robustness of FAWNA capacity with respect to it. We also analyze the loss from keeping rate allocation fixed (based on wireless channel statistics) rather than dynamically adjusting it according to channel state. For a SIMO-FAWNA with fixed fiber capacity, quantizer distortion as well as power efficiency increases with wireless bandwidth and number of interfaces. The two competing effects result in the existence of an optimal operating wireless bandwidth and an optimal number of wireless-optical interfaces.

Notation

Let us establish notation that will be used in this thesis. The bold type will be used to denote random quantities whereas normal type will be used to denote deterministic ones. Matrices will be denoted by capital letters and the scalar or vector components of matrices will be denoted using appropriate subscripts. Vectors will be represented by small letters with an arrow over them. All vectors are column vectors unless they have a ^{T} superscript. Scalars will be represented by small letters only. The superscript † will be used to denote the complex conjugate transpose.

Chapter 2

Wideband MIMO Capacity

In this chapter, we compute the capacity of a non-coherent MIMO channel at low SNR. The analysis shows the interaction between the number of receive and transmit antennas, coherence length of the channel and SNR, in the wideband regime. We also propose a near-capacity achieving signaling scheme for this channel.

2.1 Model

We model the wideband channel as a set of N parallel narrowband channels. In general, the narrowband channels will be correlated. We restrict our analysis to narrowband channels having independent and identical statistics, and also assume each narrowband channel as being flat faded. From [42], we see that the behavior of channels with low SNR per degree of freedom is robust to reasonable modeling assumptions and necessary simplifications. Hence, the results for a more precise MIMO channel model may not differ significantly from that of the simple model we consider in this paper.

Using the sampling theorem, the m^{th} narrowband channel at symbol time k can be represented as:

$$\vec{\mathbf{y}}[k,m] = \mathbf{H}[k,m]\vec{\mathbf{x}}[k,m] + \vec{\mathbf{w}}[k,m],$$

where $\mathbf{H}[k,m]$, $\vec{\mathbf{x}}[k,m]$, $\vec{\mathbf{w}}[k,m]$ and $\vec{\mathbf{y}}[k,m]$ are the channel matrix, input vector, noise vector and output vector, respectively, for the m^{th} narrowband channel at symbol time k. The pair (k,m) may be considered as an index for the time-frequency slot, or degree of freedom, to communicate. We denote the number of transmit and receive antennas by t and r, respectively. Hence, $\vec{\mathbf{x}}[k,m] \in \mathcal{C}^t$ and $\vec{\mathbf{y}}[k,m], \vec{\mathbf{w}}[k,m] \in \mathcal{C}^r$. The channel matrix, $\mathbf{H}[k,m]$, is a $r \times t$ complex matrix, independent of the channel input and additive noise. The entries of the channel matrix are i.i.d zero-mean complex Gaussian, with independent real and imaginary components. Equivalently, each entry of $\mathbf{H}[k,m]$ has uniformly distributed phase and Rayleigh distributed magnitude. We thus model a Rayleigh fading channel with enough separation within the transmitting and receiving antennas to achieve independence in the entries of $\mathbf{H}[k,m]$. The channel matrix is unknown at the transmitter and the receiver. However, its statistics are known to both. The noise vector, $\vec{\mathbf{w}}[k,m]$, is a zero-mean Gaussian vector with the identity as its covariance matrix, i.e., $\vec{\mathbf{w}}[k,m] \backsim \mathcal{CN}(\vec{0}, I_r)$. Since the narrowband channels are assumed to be independent, we shall omit the narrowband channel index, m, to simplify notation. The capacity of the wideband channel with power constraint P is thus N times the capacity of each narrowband channel with power constraint P/N. Hence, we can focus on the narrowband channel alone.

We further assume a block fading channel model, i.e., the channel matrix is random but fixed for the duration of the coherence time of the channel, and is i.i.d across blocks. Hence, we may omit the time index, k, and express the narrowband channel within a coherence block of length l symbols as:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{W},$$

where, $\mathbf{X} \in \mathcal{C}^{t \times l}$ has entries $\mathbf{x}_{ij}, i = 1, ..., t, j = 1, ..., l$, being the signals transmitted from the transmit antenna i at time j; $\mathbf{Y} \in \mathcal{C}^{r \times l}$ has entries $\mathbf{y}_{ij}, i = 1, ..., r, j = 1, ..., l$, being the signals received at the receive antenna i at time j; the additive noise $\mathbf{W}$ has i.i.d entries $\mathbf{w}_{ij}$, which are distributed as $\mathcal{CN}(0,1)$. The input $\mathbf{X}$ satisfies the average power constraint

$$E\Big[\mathsf{trace}\Big\{\mathbf{XX}^\dagger\Big\}\Big] = l \ \mathsf{SNR},$$

where, SNR is the average signal to noise ratio at each receive antenna per narrowband channel. As N tends to ∞, SNR tends to 0, and the narrowband channel is in the low SNR regime.

2.2 Dependence of capacity on coherence length

We first analyze the dependence of non-coherent capacity on the coherence length of the channel. In [20], the structure of the capacity achieving input matrix for our non-coherent MIMO channel model is described as

$$\mathbf{X} = \mathbf{A}\mathbf{\Phi},$$

where

$$\mathbf{A} = \begin{bmatrix} \|\vec{\mathbf{x}}_1^T\| & & & & \\ & \cdots & & & \\ & & \|\vec{\mathbf{x}}_i^T\| & & \\ & & & \cdots & \\ & & & & \|\vec{\mathbf{x}}_t^T\| \end{bmatrix},$$

is a $t \times l$ random matrix that is diagonal, real and nonnegative with identically (though possibly not independently) distributed entries and $\|\vec{\mathbf{x}}_i^T\|$ is the norm of the signal vector transmitted by the i^{th} antenna. Since these entries are identically distributed, we have $\forall i \in \{1, \ldots, t\}$

$$E[\|\vec{\mathbf{x}}_i^T\|^2] = \frac{l\ \mathsf{SNR}}{t}.$$

$\mathbf{\Phi}$ is an $l \times l$ isotropically distributed unitary matrix. The row vectors of $\mathbf{\Phi}$ are isotropic random vectors which represent the direction of the signal transmitted from the antennas. $\mathbf{A}$ and $\mathbf{\Phi}$ are statistically independent. Since this structure of the input matrix is optimal, we shall restrict our attention to inputs having such structure.

We first prove Lemma 1, which establishes two necessary conditions the input distribution must satisfy for the mutual information of the channel to be above a certain value. This lemma will be used in Theorem 1 to establish the dependence of the non-coherent capacity on the channel coherence length.

Lemma 1 *For any $\alpha \in (0,1]$ and $\gamma \in (0,\alpha)$, if there exists an input distribution on* $\mathbf{X}$ *such that*

$$\frac{1}{l} I(\mathbf{X};\mathbf{Y}) \geq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}),$$

then the following two conditions are satisfied by this distribution:

$$\frac{t}{l} E\Big[\log(1+\|\vec{\mathbf{x}}_i^T\|^2)\Big] \leq \frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}), \tag{2.1}$$

$$tE\left[\log\left(1+\frac{\|\vec{\mathbf{x}}_i^T\|^2}{l}\right)\right] \geq \mathsf{SNR} - \frac{(r+t)}{t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}), \tag{2.2}$$

for all $i \in \{1,\ldots\ldots,t\}$.

Proof: See Appendix A. □

This leads to the following theorem that describes the dependence of the non-coherent capacity on the coherence length:

Theorem 1 *Consider a non-coherent Rayleigh block fading MIMO channel with average signal to noise ratio,* SNR. *Let the block length be l and the capacity, $C(\mathsf{SNR})$. For any $\alpha \in (0,1]$ and $\gamma \in (0,\alpha)$, if*

$$C(\mathsf{SNR}) \geq C^*(\mathsf{SNR}) \triangleq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}),$$

then

$$l > l_{\mathsf{min}} \triangleq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha}.$$

Proof: See Appendix B. □

This theorem states that the coherence length must be strictly larger that l_{min} for the channel capacity to be above $C^*(\mathsf{SNR})$. Since the inequality for the coherence length is strict, this implies that a channel with capacity $C^*(\mathsf{SNR})$ will have its coherence length, l^*, strictly greater than l_{min}, i.e.,

$$l_{\mathsf{min}} < l^*.$$

2.3 Communicating using Gaussian-like signals

In this section, we propose a signaling scheme using which a rate of $C^*(\mathsf{SNR})$ is achievable if the coherence length is greater than or equal to a threshold, l^G.
We prove the following lemma, which shows that, using a Gaussian input distribution, we can achieve "near coherent" performance if the coherence length of the channel is large enough.

Lemma 2 *Consider a non-coherent Rayleigh block fading MIMO channel with average signal-to-noise ratio,* SNR*. Let the block length be* l *and the capacity,* $C(\mathsf{SNR})$*. If we use Gaussian signals over this channel, then for any* $\epsilon \in (0, 1]$*, if*

$$l \geq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(1+\epsilon)},$$

then

$$C(\mathsf{SNR}) \geq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^2 + O(\mathsf{SNR}^{2+\epsilon}).$$

Proof: We first lower bound the mutual information of the non-coherent channel MIMO channel as

$$\begin{aligned} & I(\mathbf{X};\mathbf{Y}) \\ & \quad = I(\mathbf{X};\mathbf{Y}|\mathbf{H}) + I(\mathbf{H};\mathbf{Y}) - I(\mathbf{H};\mathbf{Y}|\mathbf{X}) \\ & \quad \geq I(\mathbf{X};\mathbf{Y}|\mathbf{H}) - I(\mathbf{H};\mathbf{Y}|\mathbf{X}). \end{aligned} \tag{2.3}$$

Let us choose the distribution of $\mathbf{X}$ to be one where all the entries of $\mathbf{X}$ are i.i.d and $\mathcal{CN}(0, \frac{\mathsf{SNR}}{t})$. Note that it is exactly this distribution that achieves capacity for the coherent MIMO channel. Therefore,

$$\frac{1}{l}I(\mathbf{X};\mathbf{Y}|\mathbf{H}) = r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^2 + O(\mathsf{SNR}^3). \tag{2.4}$$

$I(\mathbf{H};\mathbf{Y}|\mathbf{X})$ is the information that can be obtained about $\mathbf{H}$ from observing $\mathbf{Y}$, conditioned on $\mathbf{X}$ being known. Therefore

$$I(\mathbf{H};\mathbf{Y}|\mathbf{X})$$

$$
\begin{aligned}
&= h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Y}|\mathbf{X},\mathbf{H}) \\
&= rtE\log(1+\|\vec{\mathbf{x}}_i^T\|^2) \\
&\leq rt\log\Big(1+\frac{l}{t}\mathsf{SNR}\Big), \qquad (2.5)
\end{aligned}
$$

where we have used Jensen's inequality to get the upper bound in (2.5). Combining (2.3), (2.4) and (2.5), and noting that

$$C(\mathsf{SNR}) \geq \frac{1}{l} I(\mathbf{X};\mathbf{Y}),$$

we obtain:

$$C(\mathsf{SNR}) \geq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^2 - r\frac{t}{l}\log\Big(1+\frac{l}{t}\mathsf{SNR}\Big) + O(\mathsf{SNR}^3). \qquad (2.6)$$

For any $\epsilon \in (0,1]$, let us choose

$$l = \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(1+\epsilon)}.$$

Therefore,

$$
\begin{aligned}
&r\frac{t}{l}\log\Big(1+\frac{l}{t}\mathsf{SNR}\Big) \\
&\quad= r\frac{(r+t)^2}{t}\mathsf{SNR}^{2(1+\epsilon)}\log\Big(1+\frac{t}{(r+t)^2\mathsf{SNR}^{1+2\epsilon}}\Big) \\
&\quad= r\frac{(r+t)^2}{t}\mathsf{SNR}^{2(1+\epsilon)}\log\Big(\frac{t}{(r+t)^2\mathsf{SNR}^{1+2\epsilon}}\Big) + o(\mathsf{SNR}^{2(1+\epsilon)}) \\
&\quad= r\frac{(r+t)^2}{t}\mathsf{SNR}^{2(1+\epsilon)}\log\Big(\frac{t}{(r+t)^2}\Big) \\
&\qquad\qquad + r\frac{(r+t)^2}{t}(1+2\epsilon)\mathsf{SNR}^{2+\epsilon}\Big[\mathsf{SNR}^{\epsilon}\log\Big(\frac{1}{\mathsf{SNR}}\Big)\Big] + o(\mathsf{SNR}^{2(1+\epsilon)}) \\
&\quad\leq r\frac{(r+t)^2}{t}\mathsf{SNR}^{2(1+\epsilon)}\log\Big(\frac{t}{(r+t)^2}\Big) + r\frac{(r+t)^2}{t}(1+2\epsilon)\mathsf{SNR}^{2+\epsilon} + o(\mathsf{SNR}^{2(1+\epsilon)}) \qquad (2.7) \\
&\quad= O(\mathsf{SNR}^{2+\epsilon}).
\end{aligned}
$$

In (2.7), we use the fact that since $\epsilon > 0$ and $\mathsf{SNR} \to 0$, $\mathsf{SNR}^{\epsilon}\log(\frac{1}{\mathsf{SNR}}) \ll 1$. Since $r\frac{t}{l}\log\Big(1+\frac{l}{t}\mathsf{SNR}\Big)$ decreases monotonically with l, we have that

$$
\begin{aligned}
l &\geq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(1+\epsilon)} \\
\Rightarrow r\frac{t}{l}\log\Big(1+\frac{l}{t}\mathsf{SNR}\Big) &\leq O(\mathsf{SNR}^{2+\epsilon}).
\end{aligned}
$$

Combining this with (2.6) completes the proof. □

We now introduce an input distribution that has a flashy as well as a continuous nature. A similar input distribution was first introduced in [42] for achieving the *order* of the sublinear capacity term for a single-input, single-output non-coherent Rayleigh block fading channel.

For a given $\alpha \in (0, 1]$, transmit in only $\delta(\mathsf{SNR}) = \mathsf{SNR}^{1-\alpha}$ fraction of the blocks. As we are in the low signal to noise ratio regime, $\delta(\mathsf{SNR}) \in (\mathsf{SNR}, 1]$. Since we concentrate the power only over a fraction of the blocks, the signal to noise ratio for the blocks in which we transmit increases to SNR_b, where,

$$\mathsf{SNR}_b \triangleq \frac{\mathsf{SNR}}{\delta(\mathsf{SNR})} = \mathsf{SNR}^{\alpha}.$$

In the blocks that we choose to transmit, let the entries of the input matrix $\mathbf{X}$ be i.i.d $\mathcal{CN}(0, \frac{\mathsf{SNR}_b}{t})$. Note that as we increase α from 0 to 1, the fraction of blocks that we transmit increases from SNR to 1. Therefore, as α increases, the distribution changes from a peaky to a continuous one. We call this type of signaling *Peaky Gaussian* signaling. We establish the following theorem:

Theorem 2 *Consider a non-coherent Rayleigh block fading MIMO channel with average signal to noise ratio,* SNR*. Let the block length be* l *and the capacity,* $C(\mathsf{SNR})$*. If we use Peaky Gaussian signals over this channel, then for any* $\alpha \in (0, 1]$ *and* $\epsilon \in (0, \alpha]$*, if*

$$l \geq l^G \triangleq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(\alpha+\epsilon)},$$

then

$$C(\mathsf{SNR}) \geq C^*(\mathsf{SNR}) = r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\epsilon}).$$

Proof: Let us use Peaky Gaussian signaling for communicating over the non-coherent MIMO channel. We can now apply Lemma 2 to the blocks over which we choose to transmit. Note that these blocks have a signal to noise ratio of SNR_b. Thus, for any

$\epsilon^{'} \in (0,1]$, if

$$\begin{aligned} l &\geq \frac{t^2}{(r+t)^2}\mathsf{SNR}_b{}^{-2(1+\epsilon^{'})} \\ &= \tfrac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(\alpha+\alpha\epsilon^{'})}, \end{aligned}$$

then

$$C(\mathsf{SNR}_b) \geq r\mathsf{SNR}_b - \frac{r(r+t)}{2t}\mathsf{SNR}_b{}^2 + O(\mathsf{SNR}_b{}^{2+\epsilon^{'}}).$$

Since we are transmitting in $\delta(\mathsf{SNR})$ fraction of the blocks,

$$\begin{aligned} &C(\mathsf{SNR}) \\ &\quad = \delta(\mathsf{SNR}) \cdot C(\mathsf{SNR}_b) \\ &\quad \geq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\alpha\epsilon^{'}}). \end{aligned}$$

Note that for $\epsilon^{'} \in (0,1]$, $\alpha\epsilon^{'} \triangleq \epsilon \in (0,\alpha]$. This completes the proof. □

Thus, we see that using Peaky Gaussian signals a rate of $C^*(\mathsf{SNR})$ is achievable if the coherence length is greater than or equal to l^G.

To achieve reliably any rate, the required coherence length using Peaky Gaussian signaling is strictly greater than the required length (Theorem 1) using the optimal input distribution. Thus, if l^* is the coherence length needed to have a capacity of $C^*(\mathsf{SNR})$,

$$l_{\mathsf{min}} < l^* \leq l^G.$$

However, for $\alpha \in (0,1]$, as $\epsilon \to 0$, $l^G \to l_{\mathsf{min}}$ and Peaky Gaussian signaling is near-optimal for the non-coherent MIMO channel.

From Theorems 1 and 2, we see that for any $\alpha \in (0,1]$ and $\epsilon \in (0,\alpha]$, if

$$\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha} < l \leq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(\alpha+\epsilon)}, \tag{2.8}$$

the sublinear capacity term is

$$\Delta^{(t,r)}(\mathsf{SNR}) = \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\epsilon}).$$

We summarize this result in the following theorem:

Theorem 3 *Consider a non-coherent Rayleigh block fading MIMO channel with average signal to noise ratio,* SNR*. For any* $\alpha \in (0,1]$ *and* $\epsilon \in (0,\alpha]$*, the capacity of the channel is*

$$C(\mathsf{SNR}) = r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\epsilon})$$

if and only if there exists a $\sigma \in (0,\epsilon]$ *such that*

$$l = \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(\alpha+\sigma)}.$$

This theorem tells us the capacity of a non-coherent MIMO channel in the low SNR regime and shows its dependence on the coherence length of the channel, number of receive and transmit antennas and SNR. Note that the transmit antennas affect the sublinear capacity term. Peaky Gaussian signals are near-optimal when communicating over this channel. σ is used in the theorem to parameterize (2.8). This theorem leads to the following corollary:

Corollary 1 *Consider a non-coherent Rayleigh block fading MIMO channel with average signal to noise ratio* SNR*. For any* $\alpha \in (0,1]$ *and* $\epsilon \in (0,\alpha]$*,*

$$\Delta^{(t,r)}(\mathsf{SNR}) = \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\epsilon})$$

if and only if there exists a $\sigma \in (0,\epsilon]$ *such that*

$$l = \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(\alpha+\sigma)}.$$

In Theorem 3 and Corollary 1, α is an indicator of how close the channel capacity is to the coherent extreme. The coherent channel corresponds to the case when $\alpha = 1$ and the i.i.d non-coherent channel corresponds to the case when $\alpha \to 0$. We have also seen that Peaky Gaussian signals are optimal for the non-coherent MIMO channel. Thus, with a channel coherence length of $l \sim \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha}$, one should transmit Gaussian signals in $\delta = \mathsf{SNR}^{1-\alpha}$ fraction of the blocks. At the coherent extreme, $\delta = \mathsf{SNR}^0$ and one should transmit in all the blocks in order to achieve capacity. On the other hand, for the i.i.d Rayleigh fading channel (non-coherent extreme), one should only transmit in $\delta = \mathsf{SNR}^1$ fraction of the blocks. We shall study the non-coherent extreme

with a finer scaling later on in the thesis.

Let us eliminate the parameter α from Corollary 1. The sublinear capacity term then becomes

$$\Delta^{(t,r)}(\mathsf{SNR}) = \frac{r}{2\sqrt{l}}\mathsf{SNR} + o(\frac{\mathsf{SNR}}{\sqrt{l}}).$$

From (1.1), we have

$$\log\left(\frac{E_n}{N_0}\right) \propto \sqrt{\frac{1}{l}}.$$

Hence, the minimum energy required to reliably transmit an information bit decreases inversely with the square root of the coherence length of the channel. These results apply for $\alpha \in (0, 1]$. For channels whose coherence length is larger that $\frac{t^2}{(t+r)^2}\mathsf{SNR}^{-2}$ symbols, the sublinear capacity term remains $O(\mathsf{SNR}^2)$. Let us now focus on the coherent and non-coherent extremes.

Coherent Extreme

In this case $\alpha = 1$ and from Theorem 3, we know that for $\epsilon \in (0, 1]$,

$$C(\mathsf{SNR}) = r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^2 + O(\mathsf{SNR}^{2+\epsilon}),$$

iff there exists a $\sigma \in (0, \epsilon]$ such that

$$l = \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2(1+\sigma)}. \tag{2.9}$$

We see that, provided the coherence length is large enough, the non-coherent capacity is the same as the coherent capacity in the low SNR regime. Moreover, Peaky Gaussian signaling is now completely continuous. Hence, when $l \geq \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2}$, the coherent and non-coherent capacities are the same in the low SNR regime and continuous Gaussian signals are optimal for both.

Non-coherent Extreme

From Theorem 3, we see that as $\alpha \to 0$, $l \to 1$ and we have an i.i.d Rayleigh fading channel. In order to get the exact value of the sublinear capacity term for

this channel, we need to know the precise value of α, which is not possible by this asymptotic analysis. We do the precise analysis in Appendix C and show that the capacity is[1]

$$C(\mathsf{SNR}) = r\mathsf{SNR} - \Delta^{(t,r)}_{\text{i.i.d}}(\mathsf{SNR}),$$

where

$$\Delta^{(t,r)}_{\text{i.i.d}}(\mathsf{SNR}) \doteq \frac{r\mathsf{SNR}}{\log(\frac{r}{\mathsf{SNR}})}.$$

Capacity is achieved using a single transmit antenna with on-off signaling, that becomes increasingly "flashy" at low SNR. This is consistent with our asymptotic analysis, which shows that only SNR fraction of the blocks should be used for transmission in the non-coherent extreme. Hence, we see that on-off signaling is optimal for the single-input, single-output [25] as well as the MIMO i.i.d Rayleigh fading channels.

[1]Definition of ($\doteq$): Let $f(\mathsf{SNR})$ and $g(\mathsf{SNR})$ be functions of SNR. We denote $f(\mathsf{SNR}) \doteq g(\mathsf{SNR})$ if

$$\lim_{\mathsf{SNR}\to 0} \frac{\log f(\mathsf{SNR})}{\log g(\mathsf{SNR})} = 1.$$

Chapter 3

Wideband MIMO Error Probability

In this chapter, we study the block error probability for the non-coherent MIMO channel, $P_{\mathsf{error}}^{\mathsf{block}}$, when Peaky Gaussian signaling is used at the transmitter and maximum-likelihood decoding is used at the receiver.

3.1 Error probability expression

The error probability is the average over the ensemble of codes when Peaky Gaussian signaling is used. The coherence blocks used for transmission are independently coded. Hence, the error probability can be expressed as:

$$\begin{aligned} & P_{\mathsf{error}}^{\mathsf{block}} \\ & \quad = \Pr(\text{error}|\text{Block used for transmission}) \cdot \Pr(\text{Block used for transmission}) \\ & \qquad + \Pr(\text{error}|\text{Block not used for transmission}) \cdot \Pr(\text{Block not used for transmission}). \end{aligned}$$

Since we use Peaky Gaussian signaling and the receiver is assumed to have perfect knowledge of the blocks that are being used for transmission, we have

$$\begin{aligned} \Pr(\text{Block used for transmission}) &= \delta(\mathsf{SNR}), \\ \Pr(\text{error}|\text{Block not used for transmission}) &= 0. \end{aligned}$$

Hence,

$$P_{\mathsf{error}}^{\mathsf{block}} = \delta(\mathsf{SNR}) \cdot \Pr(\text{error}|\text{Block used for transmission}).$$

If we consider the input matrix transmitted in a block, $\mathbf{X}$, as a super symbol of dimension $t \times l$, the channel is memoryless, since, for each use of the channel an independent realization of $\mathbf{H}$ is drawn. Hence, using the results in [2], the error probability can be upper bounded as

$$\Pr(\text{error}|\text{Block used for transmission}) \leq \exp[-E_r(R)],$$

where, $E_r(R)$ is the random coding error exponent for the super symbol channel:

$$E_r(R) = \max_{\rho \in [0,1]} \Big\{ E_0(\rho) - \rho R \Big\},$$

where,

$$E_0(\rho) = -\log \int \left[\int q(X) p(Y|X)^{\frac{1}{1+\rho}} dX \right]^{1+\rho} dY,$$

$q(X)$ is the distribution of $\mathbf{X}$, R is the transmission rate in nats per block used for transmission and $\mathbf{Y}$ is the channel's output matrix.

Since the signaling is Gaussian in a block used for transmission,

$$q(X) = \frac{1}{\pi^{lt}} \exp\Big[- \mathsf{trace}(X^\dagger X) \Big].$$

The range of R for which $E_r(R)$ is positive is:

$$0 \leq R \leq \frac{l}{\delta(\mathsf{SNR})} \cdot C(\mathsf{SNR}) \triangleq C^{\mathsf{block}}(\mathsf{SNR}), \tag{3.1}$$

where, $C^{\mathsf{block}}(\mathsf{SNR})$ is the non-coherent capacity per block.

If we express l as

$$l = \frac{t^2}{(r+t)^2} \mathsf{SNR}^{-2\nu}, \quad \nu > 0,$$

then, from the results in the capacity section,

$$\begin{aligned}
\delta(\mathsf{SNR}) &= \mathsf{SNR}^{1-\min\{1,\nu\}}, \\
C(\mathsf{SNR}) &= r\mathsf{SNR} - \frac{r(r+t)}{2t} \mathsf{SNR}^{1+\min\{1,\nu\}} + o(\mathsf{SNR}^{1+\min\{1,\nu\}}), \\
C^{\mathsf{block}}(\mathsf{SNR}) &= \frac{t^2}{(r+t)^2} \mathsf{SNR}^{-2\nu} \left[r\mathsf{SNR}^{\min\{1,\nu\}} - \frac{r(r+t)}{2t} \mathsf{SNR}^{2\min\{1,\nu\}} \right] + o(\mathsf{SNR}^{-2[\nu - \min\{1,\nu\}]}).
\end{aligned}$$

The signal-to-noise ratio in a block used for transmission, SNR_b, is:

$$\mathsf{SNR}_b \triangleq \frac{\mathsf{SNR}}{\delta(\mathsf{SNR})} = \mathsf{SNR}^{\min\{1,\nu\}}.$$

Note that for $\nu \leq 1$, $\nu = \alpha$. By letting ν exceed 1, we analyze the error probability behavior for coherence lengths larger than that required for coherent performance. The error probability result is summarized in the following theorem:

Theorem 4 *The block error probability for a non-coherent Rayleigh block fading MIMO channel, $P_{\mathsf{error}}^{\mathsf{block}}$, can be upper bounded as:*

$$P_{\mathsf{error}}^{\mathsf{block}} \leq [\mathsf{SNR}^{1-\min\{1,\nu\}}] \cdot \exp[-E_r(R)],$$

where,

$$\begin{aligned}
E_r(R) &= rt \log\left(1 + \frac{t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{2(t+r)^2}\right) - R - o(1) && R \in [0, R_{\mathsf{critical}}] \\
&= rt \log\left(1 + \frac{\rho^* t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho^*)}\right) - \rho^* R - o(1) && R \in [R_{\mathsf{critical}}, C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR})] \\
&= o(1) && R \in [C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}), C^{\mathsf{block}}(\mathsf{SNR})] \\
&= 0 && R \in [C^{\mathsf{block}}(\mathsf{SNR}), \infty),
\end{aligned}$$

and

$$\begin{aligned}
\nu &= \frac{\log\left(\frac{(r+t)^2}{t^2} l\right)}{2\log\left(\frac{1}{\mathsf{SNR}}\right)}, \\
\rho^* &= \frac{1}{2}\left[\sqrt{1 + 4\left(\frac{rt}{R} - \frac{(t+r)^2\mathsf{SNR}^{2\nu-\min\{1,\nu\}}}{t}\right)} - 1\right], \\
R_{\mathsf{critical}} &= rt/2 + o(1), \\
C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}) &= \frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\nu}\Big[r\mathsf{SNR}^{\min\{1,\nu\}} - 2\frac{r(r+t)}{\sqrt{t}}\mathsf{SNR}^{\nu+\frac{\min\{1,\nu\}}{2}} \\
&\qquad - \frac{r(r+t)}{2t}\mathsf{SNR}^{2\min\{1,\nu\}} + o\Big(\mathsf{SNR}^{\min\{\nu+\frac{\min\{1,\nu\}}{2}, 2\min\{1,\nu\}\}}\Big)\Big].
\end{aligned}$$

Proof: See Section 3.3. □

3.2 Dependence of error probability on various parameters

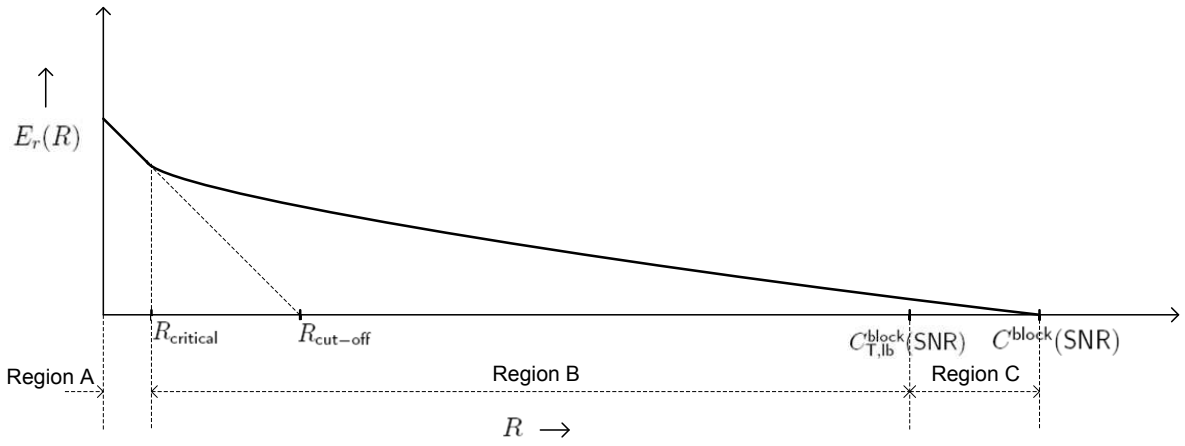

Figure 3-1: Random coding error exponent for the non-coherent MIMO channel at low SNR.

Theorem 4 divides the range of rates for which $E_r(R)$ is positive into three regions: A, B and C, which is illustrated in Figure 3-1. Let us consider region A: $R \in [0, R_{\mathsf{critical}}]$. Since $R_{\mathsf{critical}} = O(1)$ and $C^{\mathsf{block}}(\mathsf{SNR}) = O(\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]})$, the critical rate is much smaller than the channel capacity:

$$R_{\mathsf{critical}} \ll C^{\mathsf{block}}(\mathsf{SNR}).$$

Region A is an $O(\mathsf{SNR}^{[2\nu-\min\{1,\nu\}]})$ fraction of the capacity and is very small in the wideband regime. The cut-off rate, $R_{\mathsf{cut-off}}$, is given by

$$R_{\mathsf{cut-off}} = E_r(0) \doteq rt \cdot [2\nu - \min\{1,\nu\}] \cdot \log\left(\frac{1}{\mathsf{SNR}}\right).$$

Since the cut-off rate is an $O\Big(\mathsf{SNR}^{[2\nu-\min\{1,\nu\}]} \cdot \log(\frac{1}{\mathsf{SNR}})\Big)$ fraction of the capacity, it is much smaller than the capacity in the wideband regime:

$$R_{\mathsf{cut-off}} \ll C^{\mathsf{block}}(\mathsf{SNR}).$$

Let us consider the third region over which $E_r(R)$ is positive, region C:
$R \in [C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}), C^{\mathsf{block}}(\mathsf{SNR})]$. This interval is a $[C^{\mathsf{block}}(\mathsf{SNR}) - C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR})]/C^{\mathsf{block}}(\mathsf{SNR})$ fraction of the capacity, where,

$$\frac{C^{\mathsf{block}}(\mathsf{SNR}) - C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR})}{C^{\mathsf{block}}(\mathsf{SNR})} = \begin{cases} O(\mathsf{SNR}^{\nu - \frac{\min\{1,\nu\}}{2}}) & \nu \leq \frac{3}{2} \\ \\ o(\mathsf{SNR}) & \nu > \frac{3}{2} \end{cases}.$$

Hence, region C is also a very small fraction of the capacity in the wideband regime. Therefore, we can conclude that it is region B: $R \in [R_{\mathsf{critical}}, C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR})]$, that dominates the range of rates in the wideband regime.

From Theorem 4, the error probability in Region B can be expressed as:

$$P_{\mathsf{error}}^{\mathsf{block}} \sim \mathsf{SNR}^{1-\min\{1,\nu\}} \cdot \left[\frac{R}{l \cdot \mathsf{SNR}^{\min\{1,\nu\}}}\right]^{rt}. \tag{3.2}$$

To observe this, let us consider the error exponent for

$$R = l \cdot r\mathsf{SNR}^{\kappa}, \quad \min\{1,\nu\} < \kappa < 2\nu. \tag{3.3}$$

This rate lies in Region B and the optimum ρ is

$$\rho^* = O\Big(\frac{1}{R}\Big).$$

Substituting in Theorem 4, we observe

$$\begin{aligned} &E_r(R) \\ &\quad = rt \log\left(1 + \frac{\rho^* t \mathsf{SNR}^{-[2\nu - \min\{1,\nu\}]}}{(t+r)^2(1+\rho^*)}\right) - \rho^* R - o(1)\Big|_{(\rho^* = O(\frac{1}{R}))} \\ &\quad \doteq rt \log\left[\frac{l \cdot \mathsf{SNR}^{\min\{1,\nu\}}}{R}\right]. \end{aligned}$$

For $\nu \leq 1$, $\mathsf{SNR}^{\min\{1,\nu\}} \propto 1/\sqrt{l}$. Hence, for a fixed rate R, the error probability decays *inversely* with the coherence length in the following way:

$$P_{\mathsf{error}}^{\mathsf{block}} \propto \begin{cases} \left(\frac{1}{l}\right)^{\frac{rt-1}{2}} & l \leq \mathsf{SNR}^{-2} \\ \\ \left(\frac{1}{l}\right)^{rt} & l > \mathsf{SNR}^{-2} \end{cases}.$$

Let us now examine the effect of antennas on the error probability. To analyze this, we propose a definition of "diversity" in the low SNR / wideband regime.

Let $\mathcal{P}$ and W be the total received power and system bandwidth, respectively. High SNR diversity [19], $d_H(W)$, is commonly defined as:

$$d_H(W) \triangleq -\lim_{\mathcal{P}\to\infty} \frac{\log(P_{\mathsf{error}}^{\mathsf{block}}(\mathcal{P},\mathcal{W}))}{\log(\mathcal{P})}.$$

This definition describes the asymptotic behavior of error probability with received power, for fixed bandwidth.

In the low SNR/wideband regime, we define diversity, $d_L(\mathcal{P})$, as:

$$d_L(\mathcal{P}) \triangleq -\lim_{W\to\infty} \frac{\log(P_{\mathsf{error}}^{\mathsf{block}}(\mathcal{P},\mathcal{W}))}{\log(W)}.$$

This definition describes the asymptotic behavior of error probability with bandwidth, for fixed received power. Since $\mathsf{SNR} \propto 1/W$, an equivalent definition of low SNR diversity is [1]:

$$d_L \triangleq \lim_{\mathsf{SNR}\to 0} \frac{\log(P_{\mathsf{error}}^{\mathsf{block}}(\mathcal{P},\mathsf{SNR}))}{\log(\mathsf{SNR})}. \tag{3.4}$$

From (3.2, 3.3), we have

$$d_L = r\cdot t\cdot\Big[\kappa - \min\{1,\nu\}\Big] + 1 - \min\{1,\nu\}.$$

Hence, we conclude that the decay in error probability is *exponential* with the product of the number of transmit and receive antennas, rt. Similarly to the high SNR regime, the product of the number of transmit and receive antennas comes about as a diversity factor in the low SNR regime. Hence, we conjecture that rt is a diversity factor for a MIMO channel at any SNR.

In the capacity section of this paper, we have seen that receive antennas have greater significance than transmit antennas since, the former affects the linear as well as the sublinear capacity term whereas, the latter affects only the sublinear term. However, since the error probability decays exponentially with rt, the transmit antennas have the same importance as receive antennas in terms of reliability. This emphasizes the

[1] We omit the argument of $d_L(.)$ for simplicity.

importance of multiple transmit antennas in the wideband regime.

Let us now consider channel outage in the low SNR regime. For a block fading channel, outage occurs in a coherence block when the channel matrix is so ill-conditioned that the block mutual information cannot support the target block data rate. We denote the outage probability as P_{outage}, and present a heuristic computation to show that

$$P_{\mathsf{error}}^{\mathsf{block}} \sim \mathsf{SNR}^{1-\min\{1,\nu\}} \cdot P_{\mathsf{outage}}.$$

Thus we see that in the low SNR/wideband regime, for rates away from capacity, the error probability is dominated by the outage probability. Hence, like at high SNR, channel outage is the major source for errors even at low SNR.

Heuristic Proof: The outage probability can be upper bounded using a training based scheme (This scheme is described in detail in the proof of Theorem 4). We directly state the channel model when this scheme is used for data transmission (the first t symbols are used for training):

$$\vec{\mathbf{y}}_i = \mathbf{H}^{'}\vec{\mathbf{x}}_i + \vec{\mathbf{v}}_i^{'}, \quad i = t+1, \ldots, l,$$

where, $\mathbf{H}^{'}$ has i.i.d $\mathcal{CN}(0,1)$ entries and is perfectly known at the receiver (this is the MMSE channel estimate) , $\vec{\mathbf{v}}_i^{'}$ is a zero-mean noise vector having the covariance matrix

$$E[\vec{\mathbf{v}}_i^{'}\vec{\mathbf{v}}_i^{'\dagger}] = I_r,$$

and $\{\vec{\mathbf{x}}_i\}$ are i.i.d complex Gaussian vectors:

$$\vec{\mathbf{x}}_i \sim \mathcal{CN}(0, \frac{f^*(\mathsf{SNR})}{t}\ I_t),$$

where,

$$f^*(\mathsf{SNR}) = \mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}}).$$

Now,

$$P_{\mathsf{outage}}$$

$$
\begin{aligned}
&= \Pr\left(I(\vec{\mathbf{x}}_{t+1},\ldots,\vec{\mathbf{x}}_l;\vec{\mathbf{y}}_{t+1},\ldots,\vec{\mathbf{y}}_l|\mathbf{H}')<R\right)\\
&\leq \Pr\left(\log\det\left(I_t+\frac{f^*(\mathsf{SNR})}{t}\mathbf{H}'^{\dagger}\mathbf{H}'\right)<\frac{R}{l-t}\right) \qquad (3.5)\\
&\leq \Pr\left(\log\left(1+\frac{f^*(\mathsf{SNR})}{t}\mathsf{trace}(\mathbf{H}'^{\dagger}\mathbf{H}')\right)<\frac{R}{l-t}\right) \qquad (3.6)\\
&\sim \Pr\left(\chi^2_{rt}<\frac{R}{lf^*(\mathsf{SNR})}\right). \qquad (3.7)
\end{aligned}
$$

Equation (3.5) follows since the mutual information is minimized if $\{\vec{\mathbf{v}}_i'\}$ are i.i.d complex Gaussian [23, 32]. In (3.6), we use the inequality:

$$\det\left(I_t+\frac{f^*(\mathsf{SNR})}{t}\mathbf{H}'^{\dagger}\mathbf{H}'\right)\geq 1+\frac{f^*(\mathsf{SNR})}{t}\mathsf{trace}(\mathbf{H}'^{\dagger}\mathbf{H}').$$

In (3.7), χ^2_{rt} represents $\mathsf{trace}(\mathbf{H}'^{\dagger}\mathbf{H}')$ and is a chi-squared random variable with rt degrees of freedom. Hence, if we choose the rate in Region B as in (3.3), we have for low SNR,

$$\frac{R}{lf^*(\mathsf{SNR})}\ll 1.$$

Hence,

$$
\begin{aligned}
P_{\mathsf{outage}} &\sim \left[\frac{R}{l\cdot\mathsf{SNR}^{\min\{1,\nu\}}}\right]^{rt}\\
\Rightarrow P_{\mathsf{error}}^{\mathsf{block}} &\sim \mathsf{SNR}^{1-\min\{1,\nu\}}\cdot P_{\mathsf{outage}}.
\end{aligned}
$$

□

3.3 Error probability derivation

In this section, we prove Theorem 4.

Upper Bound to $E_r(R)$

We first establish an upper bound to $E_r(R)$ by providing the receiver perfect knowledge of $\mathbf{H}$. Let us denote the random coding error exponent for this coherent channel by $E_r^U(R)$. Since the error probability for the coherent channel cannot be greater that the channel without knowledge of $\mathbf{H}$, we have

$$E_r(R)\leq E_r^U(R), \qquad (3.8)$$

where,

$$E_r^U(R) = \max_{\rho \in [0,1]} \Big\{ E_0^U(\rho) - \rho R \Big\},$$

and

$$E_0^U(\rho) = -\log \int \int \left[\int q(X) p(Y, H|X)^{\frac{1}{1+\rho}} dX \right]^{1+\rho} dY\, dH.$$

The computation of $E_0^U(\rho)$, when $l = 1$, is done in [21]. Here, we do the computation for arbitrary l. The following lemma specifies an upper bound to $E_0^U(\rho)$:

Lemma 3

$$E_0^U(\rho) \leq rt \log \left(1 + \frac{\rho t \mathsf{SNR}^{-[2\nu - \min\{1,\nu\}]}}{(t+r)^2(1+\rho)} \right).$$

Proof: Since $\mathbf{H}$ is independent of $\mathbf{X}$,

$$p(Y, H|X) = p(H) p(Y|X, H).$$

Hence, $E_0^U(\rho)$ can be expressed as

$$E_0^U(\rho) = -\log \left(E_{\mathbf{H}} \left[\int \left[\int q(X) p(Y|X, \mathbf{H})^{\frac{1}{1+\rho}} dX \right]^{1+\rho} dY \right] \right).$$

The conditional probability $p(Y|X, H)$ is given by

$$p(Y|X, H) = \Big(\frac{\mathsf{SNR}_b}{\pi t} \Big)^{rl} \exp \left[-\frac{\mathsf{SNR}_b}{t} \mathsf{trace} \Big\{ \Big(HX - Y \Big)^{\dagger} \Big(HX - Y \Big) \Big\} \right].$$

Defining B as

$$\begin{aligned} B &\triangleq \frac{\mathsf{SNR}_b}{t(1+\rho)} H^{\dagger} H. \\ \Rightarrow B^{-1} &= \frac{t(1+\rho)}{\mathsf{SNR}_b} (H^{\dagger})^{-1} (H)^{-1}. \end{aligned}$$

In the proof of this lemma, for any matrix M, we use M^{-1} to denote its pseudoinverse. Now,

$$\begin{aligned} &\int q(X) p(Y|X, H)^{\frac{1}{1+\rho}} dX \\ &\quad = \int \frac{1}{\pi^{lt}} \exp \Big[-\mathsf{trace}(X^{\dagger} X) \Big] \Big(\frac{\mathsf{SNR}_b}{\pi t} \Big)^{\frac{rl}{1+\rho}} \exp \Big[-\frac{\mathsf{SNR}_b}{t(1+\rho)} \mathsf{trace} \Big\{ (HX - Y)^{\dagger} (HX - Y) \Big\} \Big] dX \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\pi^{lt}}\Big(\frac{\mathsf{SNR}_b}{\pi t}\Big)^{\frac{rl}{1+\rho}} \int \exp\Big[-\mathsf{trace}\Big\{X^\dagger(I_t+B)X - \frac{\mathsf{SNR}_b}{t(1+\rho)}(X^\dagger H^\dagger Y + Y^\dagger H X - Y^\dagger Y)\Big\}\Big]dX \\
&= \frac{1}{\pi^{lt}}\Big(\frac{\mathsf{SNR}_b}{\pi t}\Big)^{\frac{rl}{1+\rho}} \exp\Big[-\frac{\mathsf{SNR}_b}{t(1+\rho)}\mathsf{trace}\Big\{Y^\dagger(I_r-(I_r+B^{-1})^{-1})Y\Big\}\Big] \int \exp\Big[-\frac{\mathsf{SNR}_b}{t(1+\rho)} \\
&\qquad\qquad \cdot\mathsf{trace}\Big\{X^\dagger H^\dagger(B^{-1}+I_r)HX - X^\dagger H^\dagger Y - Y^\dagger H X + Y^\dagger(I_r+B^{-1})^{-1}Y\Big\}\Big]dX \\
&= \Big(\frac{\mathsf{SNR}_b}{\pi t}\Big)^{\frac{rl}{1+\rho}} \exp\Big[-\frac{\mathsf{SNR}_b}{t(1+\rho)}\mathsf{trace}\Big\{Y^\dagger(I_r-(I_r+B^{-1})^{-1})Y\Big\}\Big] \det(I_t+B)^{-l}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\int \left[\int q(X)p(Y|X,H)^{\frac{1}{1+\rho}}dX\right]^{1+\rho} dY \\
&\quad= \Big(\frac{\mathsf{SNR}_b}{\pi t}\Big)^{rl} \det(I_t+B)^{-l(1+\rho)} \int \exp\Big[-\frac{\mathsf{SNR}_b}{t}\mathsf{trace}\Big\{Y^\dagger(I_r-(I_r+B^{-1})^{-1})Y\Big\}\Big]dY \\
&\quad= \det(I_t+B)^{-l(1+\rho)} \det\Big(I_r-(I_r+B^{-1})^{-1}\Big)^{-l} \\
&\quad= \det(I_t+B)^{-\rho l} \\
&\quad= \det\left(I_t+\frac{\mathsf{SNR}_b}{t(1+\rho)}H^\dagger H\right)^{-\rho l}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&E_0^U(\rho) \\
&\quad= -\log E_{\mathbf{H}}\left[\det\left(I_t+\frac{\mathsf{SNR}_b}{t(1+\rho)}\mathbf{H}^\dagger\mathbf{H}\right)^{-\rho l}\right] \qquad (3.9)\\
&\quad= -\log E_{\mathbf{H}}\left[\exp\left(-\rho l \log\det\left(I_t+\frac{\mathsf{SNR}_b}{t(1+\rho)}\mathbf{H}^\dagger\mathbf{H}\right)\right)\right] \\
&\quad\le -\log E_{\mathbf{H}}\left[\exp\left(-\frac{\rho l \mathsf{SNR}_b}{t(1+\rho)}\mathsf{trace}(\mathbf{H}^\dagger\mathbf{H})\right)\right] \qquad (3.10)\\
&\quad= -\log\int_0^\infty \frac{x^{rt-1}}{(rt-1)!}\exp\left(-\Big(1+\frac{\rho l\mathsf{SNR}_b}{t(1+\rho)}\Big)x\right)dx \\
&\quad= rt\log\left(1+\frac{\rho l\mathsf{SNR}_b}{t(1+\rho)}\right).
\end{aligned}
$$

To obtain (3.10), we use the following inequality:

$$
\begin{aligned}
&\log\det\left(I_t+\frac{\mathsf{SNR}_b}{t(1+\rho)}\mathbf{H}^\dagger\mathbf{H}\right) \\
&\qquad= \sum_{i=1}^{\min(t,r)} \log\left(1+\frac{\mathsf{SNR}_b}{t(1+\rho)}\lambda_i(\mathbf{H}^\dagger\mathbf{H})\right)
\end{aligned}
$$

$$\leq \frac{\mathsf{SNR}_b}{t(1+\rho)} \sum_{i=1}^{\min(t,r)} \lambda_i(\mathbf{H}^\dagger\mathbf{H})$$
$$= \frac{\mathsf{SNR}_b}{t(1+\rho)} \mathsf{trace}(\mathbf{H}^\dagger\mathbf{H}).$$

$\lambda_i(\mathbf{H}^\dagger\mathbf{H})$ is the i^{th} eigenvalue of the random matrix $\mathbf{H}^\dagger\mathbf{H}$. Hence, $E_0^U(\rho)$ can be upper bounded as:

$$E_0^U(\rho)$$
$$\leq rt \log\left(1 + \frac{\rho l \mathsf{SNR}_b}{t(1+\rho)}\right)$$
$$= rt \log\left(1 + \frac{\rho t \mathsf{SNR}^{-[2\nu - \min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right).$$

This completes the proof of the lemma. □

Combining (3.8) with Lemma 3, we obtain an upper bound for $E_r(R)$:

$$E_r(R) \leq \max_{\rho\in[0,1]} \left\{ rt \log\left(1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right) - \rho R \right\}. \tag{3.11}$$

Since $E_r(R)$ is positive over the rate range (3.1), any upper bound to it will also be positive over (3.1). In fact, since perfect knowledge of $\mathbf{H}$ at the receiver increases capacity, the upper bound is positive over a rate range larger than (3.1).

Lower Bound to $E_r(R)$

We now use a training based scheme to obtain a lower bound on $E_r(R)$. Since this is one of the possible schemes that can be used for the non-coherent channel, the random coding error exponent for this scheme, $E_r^L(R)$, can be upper bounded as

$$E_r^L(R) \leq E_r(R). \tag{3.12}$$

We rewrite the channel model within one coherence block as

$$\vec{\mathbf{y}}_i = \mathbf{H}\vec{\mathbf{x}}_i + \vec{\mathbf{w}}_i, \quad i = 1, \ldots, l. \tag{3.13}$$

The channel matrix, $\mathbf{H}$, is constant within the block. The total energy available in the block is:

$$\mathsf{E}_{\mathsf{total}} = l \cdot \mathsf{SNR}_b.$$

We use the first t symbols of the block for training[2] using $\gamma \in (0,1)$ fraction of the total energy. The remaining fraction is used for communicating data. The training and data communication phases use independent input signals. Hence, the energy used for training is:

$$\mathsf{E}_{\mathsf{training}} = \gamma \mathsf{E}_{\mathsf{total}} = \gamma l \mathsf{SNR}_b.$$

The following training sequence is used:

$$[\vec{\mathbf{x}}_1 \dots \vec{\mathbf{x}}_t] = \sqrt{\frac{\mathsf{E}_{\mathsf{training}}}{t}}\ I_t.$$

This training scheme makes $\mathbf{y}_{i,j}$ a sufficient statistic for estimating $\mathbf{h}_{i,j}$. The receiver computes the minimum mean-squares error (MMSE) estimate of $\mathbf{H}$ from $[\vec{\mathbf{y}}_1 \dots \vec{\mathbf{y}}_t]$. Using $\hat{\mathbf{h}}_{i,j}$ and $\tilde{\mathbf{h}}_{i,j}$ to denote the estimate and estimation error of $\mathbf{h}_{i,j}$, respectively, we have for $i \in \{1,\dots,r\}, j \in \{1,\dots,t\}$:

$$\hat{\mathbf{h}}_{i,j} \sim \mathcal{CN}\Big(0, \frac{\frac{\mathsf{E}_{\mathsf{training}}}{t}}{1+\frac{\mathsf{E}_{\mathsf{training}}}{t}}\Big),$$
$$\tilde{\mathbf{h}}_{i,j} \sim \mathcal{CN}\Big(0, \frac{1}{1+\frac{\mathsf{E}_{\mathsf{training}}}{t}}\Big),$$

and, $\hat{\mathbf{h}}_{i,j}, \tilde{\mathbf{h}}_{i,j}$ are independent because the estimation is MMSE. Moreover, the sets $\{\hat{\mathbf{h}}_{i,j}\}$ and $\{\tilde{\mathbf{h}}_{i,j}\}$ have independent elements. Thus, representing the estimate and estimation error of the channel matrix as $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$, respectively, we have

$$\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}},$$

where, $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are independent matrices, each with i.i.d Gaussian entries.
For the remaining $l-t$ symbols within the same block, $\mathsf{E}_{\mathsf{total}} - \mathsf{E}_{\mathsf{training}} = (1-\gamma) l \mathsf{SNR}_b$ energy is used to send data using an i.i.d Gaussian code. The channel in this phase can be represented as

$$\vec{\mathbf{y}}_i = \hat{\mathbf{H}}\vec{\mathbf{x}}_i + \underbrace{\tilde{\mathbf{H}}\vec{\mathbf{x}}_i + \vec{\mathbf{w}}_i}_{\vec{\mathbf{v}}_i}, \qquad i = t+1,\dots,l. \tag{3.14}$$

[2]We assume $l > t$.

$\{\vec{\mathbf{x}}_i\}$ are i.i.d complex Gaussian vectors:

$$\vec{\mathbf{x}}_i \sim \mathcal{CN}\left(0, \frac{(1-\gamma)l\mathsf{SNR}_b}{(l-t)t}\ I_t\right).$$

$\tilde{\mathbf{H}}\vec{\mathbf{x}}_i$ is the noise due to the estimation error from the training phase coupled with the input signal. Combining the additive white noise with the noise due to estimation error, we have

$$\vec{\mathbf{v}}_i \triangleq \tilde{\mathbf{H}}\vec{\mathbf{x}}_i + \vec{\mathbf{w}}_i.$$

Note that $\vec{\mathbf{v}}_i$ is uncorrelated with $\hat{\mathbf{H}}\vec{\mathbf{x}}_i$. However, since $\hat{\mathbf{H}}\vec{\mathbf{x}}_i$ contains $\vec{\mathbf{x}}_i$, $\vec{\mathbf{v}}_i$ is not independent of $\hat{\mathbf{H}}\vec{\mathbf{x}}_i$. Its covariance matrix is

$$\begin{aligned}
&E\left[\vec{\mathbf{v}}_i\vec{\mathbf{v}}_i^{\dagger}\right] \\
&\quad = E\left[\tilde{\mathbf{H}}\vec{\mathbf{x}}_i\vec{\mathbf{x}}_i^{\dagger}\tilde{\mathbf{H}}^{\dagger}\right] + I_r \\
&\quad = \left[\frac{(1-\gamma)l\mathsf{SNR}_b}{t(l-t)}.\frac{t}{1+\frac{\mathsf{E}_{\mathsf{training}}}{t}} + 1\right]\cdot I_r \\
&\quad = \left[\frac{t(1-\gamma)l\mathsf{SNR}_b}{(l-t)(t+\gamma l\mathsf{SNR}_b)} + 1\right]\cdot I_r.
\end{aligned}$$

The channel in (3.14) can be normalized to:

$$\vec{\mathbf{y}}_i = \mathbf{H}'\vec{\mathbf{x}}_i + \vec{\mathbf{v}}_i', \quad i = t+1, \ldots, l, \tag{3.15}$$

where, $\mathbf{H}'$ has i.i.d $\mathcal{C}N(0,1)$ entries and is perfectly known at the receiver (this is the MMSE estimate), $\vec{\mathbf{v}}_i'$ is a zero-mean noise vector having the covariance matrix

$$E\left[\vec{\mathbf{v}}_i'\vec{\mathbf{v}}_i'^{\dagger}\right] = I_r,$$

and $\{\vec{\mathbf{x}}_i\}$ are i.i.d complex Gaussian vectors:

$$\vec{\mathbf{x}}_i \sim \mathcal{CN}(0, \frac{f(\gamma, \mathsf{SNR})}{t}\ I_t),$$

where

$$f(\gamma, \mathsf{SNR}) = \frac{\frac{\gamma l\mathsf{SNR}_b}{t+\gamma l\mathsf{SNR}_b}\cdot\frac{(1-\gamma)l\mathsf{SNR}_b}{(l-t)}}{\frac{t(1-\gamma)l\mathsf{SNR}_b}{(l-t)(t+\gamma l\mathsf{SNR}_b)} + 1}.$$

Now,

$$\begin{aligned} & f(\gamma, \mathsf{SNR}) \\ & \quad = l\mathsf{SNR}_b^2 \cdot \frac{\gamma(1-\gamma)l}{t(1-\gamma)l\mathsf{SNR}_b + (l-t)(t+\gamma l\mathsf{SNR}_b)} \\ & \quad \geq l\mathsf{SNR}_b^2 \cdot \frac{\gamma(1-\gamma)}{\gamma l\mathsf{SNR}_b + t(1+\mathsf{SNR}_b)} \\ & \quad = \frac{l\mathsf{SNR}_b^2}{t(1+\mathsf{SNR}_b)} \cdot \frac{\gamma(1-\gamma)}{1+\gamma\frac{l\mathsf{SNR}_b}{t(1+\mathsf{SNR}_b)}}. \end{aligned} \tag{3.16}$$

Define

$$f^*(\mathsf{SNR}) \triangleq \max_{\gamma\in(0,1)} f(\gamma, \mathsf{SNR}).$$

Using (3.16), we get a lower bound to $f^*(\mathsf{SNR})$:

$$\begin{aligned} & f^*(\mathsf{SNR}) \\ & \quad \geq \frac{l\mathsf{SNR}_b^2}{t(1+\mathsf{SNR}_b)} \cdot \max_{\gamma\in(0,1)} \left\{ \frac{\gamma(1-\gamma)}{1+\gamma\frac{l\mathsf{SNR}_b}{t(1+\mathsf{SNR}_b)}} \right\} \\ & \quad = \mathsf{SNR}_b \left[1 - \frac{2\sqrt{1+\frac{l\mathsf{SNR}_b}{t(1+\mathsf{SNR}_b)}}}{\frac{l\mathsf{SNR}_b}{t(1+\mathsf{SNR}_b)}} + \frac{2}{\frac{l\mathsf{SNR}_b}{t(1+\mathsf{SNR}_b)}} \right] \\ & \quad = \mathsf{SNR}^{\min\{1,\nu\}} - 2\frac{(r+t)}{\sqrt{t}}\mathsf{SNR}^{\nu+\frac{\min\{1,\nu\}}{2}} + o\Big(\mathsf{SNR}^{\nu+\frac{\min\{1,\nu\}}{2}}\Big) \triangleq f_{LB}^*(\mathsf{SNR}). \end{aligned}$$

Note that

$$f_{LB}^*(\mathsf{SNR}) = \mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}}). \tag{3.17}$$

The random coding error exponent for this scheme is

$$E_r^L(R) = \max_{\rho\in[0,1]} \Big\{ E_0^L(\rho) - \rho R \Big\}, \tag{3.18}$$

where,

$$E_0^L(\rho) = \max_{\gamma\in(0,1)} \left\{ -\log \int\int \left[\int q(X)p(Y,H^{'}|X)^{\frac{1}{1+\rho}} dX \right]^{1+\rho} dY\, dH^{'} \right\}.$$

Since the training and data communication phases use independent input signals, $\mathbf{H}^{'}$ is independent of $\mathbf{X}$. Thus

$$p(Y,H^{'}|X) = p(H^{'})p(Y|X,H^{'}),$$

and

$$E_0^L(\rho) = -\log\left(E_{\mathbf{H}'}\left[\int\left[\int q(X)p(Y|X,\mathbf{H}')^{\frac{1}{1+\rho}}dX\right]^{1+\rho}dY\right]\right).$$

The following lemma specifies a lower bound to $E_0^L(\rho)$:

Lemma 4

$$E_0^L(\rho) \geq rt\log\left(1+\frac{\rho t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right) - o(1).$$

Proof: Reference [23, 32] shows that capacity of the channel in (3.15) is minimized if $\{\vec{\mathbf{v}}_i\}$ are i.i.d Gaussian:

$$\vec{\mathbf{v}}_i \sim \mathcal{CN}(0, I_r), \quad i = t+1, \ldots, l.$$

We conjecture that this noise distribution also minimizes error exponent. With this assumption, the error exponent for this channel with i.i.d Gaussian noise is similar to that of the coherent channel ($E_0^U(\rho)$ in (3.9), with SNR_b replaced by $f(\gamma,\mathsf{SNR})$ and l replaced by $l-t$). Hence, we obtain the following lower bound:

$$\begin{aligned}
&E_0^L(\rho)\\
&\quad\geq \max_{\gamma\in(0,1)}\left\{-\log E_{\mathbf{H}'}\left[\det\left(I_t+\frac{f(\gamma,\mathsf{SNR})}{t(1+\rho)}\mathbf{H}'^{\dagger}\mathbf{H}'\right)^{-\rho(l-t)}\right]\right\}\\
&\quad= -\log E_{\mathbf{H}'}\left[\det\left(I_t+\frac{\max_{\gamma\in(0,1)} f(\gamma,\mathsf{SNR})}{t(1+\rho)}\mathbf{H}'^{\dagger}\mathbf{H}'\right)^{-\rho(l-t)}\right]\\
&\quad= -\log E_{\mathbf{H}'}\left[\det\left(I_t+\frac{f^*(\mathsf{SNR})}{t(1+\rho)}\mathbf{H}'^{\dagger}\mathbf{H}'\right)^{-\rho(l-t)}\right]\\
&\quad\geq -\log E_{\mathbf{H}'}\left[\det\left(I_t+\frac{f^*_{LB}(\mathsf{SNR})}{t(1+\rho)}\mathbf{H}'^{\dagger}\mathbf{H}'\right)^{-\rho(l-t)}\right]\\
&\quad\geq -\log E_{\mathbf{H}'}\left[\left(1+\frac{f^*_{LB}(\mathsf{SNR})}{t(1+\rho)}\mathsf{trace}(\mathbf{H}'^{\dagger}\mathbf{H}')\right)^{-\rho(l-t)}\right] \qquad (3.19)\\
&\quad= -\log\int_0^\infty \frac{x^{rt-1}\exp(-x)}{(rt-1)!}\left(1+\frac{f^*_{LB}(\mathsf{SNR})}{t(1+\rho)}x\right)^{-\rho(l-t)}dx\\
&\quad= -\log\int_0^\infty \frac{x^{rt-1}}{(rt-1)!}\exp\left[-x-\rho(l-t)\log\left(1+\frac{f^*_{LB}(\mathsf{SNR})}{t(1+\rho)}x\right)\right]dx\\
&\quad= -\log C, \qquad (3.20)
\end{aligned}$$

where,

$$C \triangleq \int_0^\infty \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \rho(l-t)\log\left(1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}x\right)\right] dx.$$

Equation (3.19) holds owing to the following inequality:

$$\begin{aligned}
&\det\left(I_t + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}\mathbf{H}'^{\dagger}\mathbf{H}'\right) \\
&= \prod_{i=1}^{\min(t,r)} \left(1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}\lambda_i(\mathbf{H}'^{\dagger}\mathbf{H}')\right) \\
&\geq 1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)} \sum_{i=1}^{\min(t,r)} \lambda_i(\mathbf{H}'^{\dagger}\mathbf{H}') \\
&= 1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}\mathsf{trace}(\mathbf{H}'^{\dagger}\mathbf{H}'). \qquad (3.21)
\end{aligned}$$

$\lambda_i(\mathbf{H}'^{\dagger}\mathbf{H}')$ is the i^{th} eigenvalue of the random matrix $\mathbf{H}'^{\dagger}\mathbf{H}'$.

We now compute an upper bound to C. Splitting the range of integration, we have

$$C = C_1 + C_2, \qquad (3.22)$$

where,

$$\begin{aligned}
C_1 &= \int_0^{2t(1+\rho)} \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \rho(l-t)\log\left(1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}x\right)\right] dx, \\
C_2 &= \int_{2t(1+\rho)}^\infty \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \rho(l-t)\log\left(1 + \frac{f_{LB}^*(\mathsf{SNR})}{t(1+\rho)}x\right)\right] dx.
\end{aligned}$$

Expanding the logarithmic function, C_1 can be upper bounded as:

$$\begin{aligned}
&C_1 \\
&\leq \int_0^{2t(1+\rho)} \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \frac{\rho(l-t)f_{LB}^*(\mathsf{SNR})x}{t(1+\rho)}\left(1 - \frac{f_{LB}^*(\mathsf{SNR})}{2t(1+\rho)}x\right)\right] dx \\
&\leq \int_0^{2t(1+\rho)} \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \frac{\rho(l-t)f_{LB}^*(\mathsf{SNR})(1-f_{LB}^*(\mathsf{SNR}))}{t(1+\rho)}x\right] dx \\
&\leq \int_0^{\infty} \frac{x^{rt-1}}{(rt-1)!} \exp\left[-x - \frac{\rho(l-t)f_{LB}^*(\mathsf{SNR})(1-f_{LB}^*(\mathsf{SNR}))}{t(1+\rho)}x\right] dx \\
&= \left[1 + \frac{\rho(l-t)f_{LB}^*(\mathsf{SNR})(1-f_{LB}^*(\mathsf{SNR}))}{t(1+\rho)}\right]^{-rt} \\
&= \left[1 + \frac{\rho}{t(1+\rho)}\left[\left(l-t\right)\left(\mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}})\right)\right.\right.
\end{aligned}$$

$$\cdot \Big(1 - \mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}})\Big)\Big]\Big]^{-rt} \quad (3.23)$$

$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)} \left[(1 - \frac{(t+r)^2}{t}\mathsf{SNR}^{2\nu})(1+o(1)) \right.\right.$$
$$\left.\left. \cdot (1 - \mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}}))\right]\right]^{-rt}$$
$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\Big(1+o(1)\Big)\right]^{-rt}$$
$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right]^{-rt}\Big[1+o(1)\Big]. \quad (3.24)$$

Note that equation (3.23) follows from (3.17). Now, C_2 can be upper bounded as:

$$\begin{aligned} C_2 & \\ &\leq \int_{2t(1+\rho)}^{\infty} \frac{x^{rt-1}}{(rt-1)!} \exp\Big[-x - \rho(l-t)\log\Big(1 + 2f^*_{LB}(\mathsf{SNR})\Big)\Big]\, dx \\ &\leq \int_{0}^{\infty} \frac{x^{rt-1}}{(rt-1)!} \exp\Big[-x - \rho(l-t)\log\Big(1 + 2f^*_{LB}(\mathsf{SNR})\Big)\Big]\, dx \\ &= \exp\Big[-\rho(l-t)\log\Big(1 + 2f^*_{LB}(\mathsf{SNR})\Big)\Big]. \end{aligned} \quad (3.25)$$

Combining (3.22, 3.24, 3.25), we get the upper bound for C as:

$$C \quad (3.26)$$
$$\leq \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right]^{-rt}\Big[1+o(1)\Big] + \exp\Big[-\rho(l-t)\log\Big(1 + 2f^*_{LB}(\mathsf{SNR})\Big)\Big]$$
$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right]^{-rt}\Big[1+o(1)\Big]\Big[1 + \Big[1+o(1)\Big]\Big[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\Big]^{rt}$$
$$\cdot \exp\Big[-\rho\Big(\frac{t^2}{(t+r)^2}\mathsf{SNR}^{-2\nu} - t\Big)\log(1 + 2(\mathsf{SNR}^{\min\{1,\nu\}} + o(\mathsf{SNR}^{\min\{1,\nu\}})))\Big]\Big] \quad (3.27)$$
$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right]^{-rt}\Big[1+o(1)\Big]\Big[1 + \Big[1+o(1)\Big]\cdot o(1)\Big]$$
$$= \left[1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right]^{-rt}\Big[1+o(1)\Big]. \quad (3.28)$$

Equation (3.27) follows from (3.17). From (3.20, 3.28), we get a lower bound to $E_0^L(\rho)$ as:

$$E_0^L(\rho) \geq rt\log\left(1 + \frac{\rho t \mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right) - o(1). \quad (3.29)$$

This completes the proof of the lemma. □

Combining (3.12, 3.18) with Lemma 4, we obtain a lower bound for $E_r(R)$:

$$E_r(R) \geq \max_{\rho \in [0,1]} \left\{ rt \log \left(1 + \frac{\rho t \mathsf{SNR}^{-[2\nu - \min\{1,\nu\}]}}{(t+r)^2(1+\rho)} \right) - \rho R \right\} - o(1). \tag{3.30}$$

Since the training based scheme has a lower capacity than the non-coherent capacity, the range of rates for which the error exponent for the training based scheme is positive, is reduced from (3.1). We compute a lower bound to the capacity for this scheme.

Letting $\{\vec{\mathbf{v}}_i^{'}\}$ be i.i.d white Gaussian vectors, i.e., $\vec{\mathbf{v}}_i^{'} \sim \mathcal{CN}(0, I_r)$ in (3.15), we can lower bound [23, 32] the capacity per block used for transmission for this training based scheme, $C_{\mathsf{T}}^{\mathsf{block}}(\mathsf{SNR})$, as

$$\begin{aligned}
&C_{\mathsf{T}}^{\mathsf{block}}(\mathsf{SNR}) \\
&\quad \geq (l-t) \max_{\gamma \in (0,1)} \left\{ E_{\mathbf{H}'} \left[\log \det \left(I_t + \frac{f(\gamma, \mathsf{SNR})}{t} \mathbf{H}'^{\dagger} \mathbf{H}' \right) \right] \right\} \\
&\quad = (l-t) E_{\mathbf{H}'} \left[\log \det \left(I_t + \frac{\max_{\gamma \in (0,1)} f(\gamma, \mathsf{SNR})}{t} \mathbf{H}'^{\dagger} \mathbf{H}' \right) \right] \\
&\quad = (l-t) E_{\mathbf{H}'} \left[\log \det \left(I_t + \frac{f^*(\mathsf{SNR})}{t} \mathbf{H}'^{\dagger} \mathbf{H}' \right) \right] \\
&\quad \geq (l-t) E_{\mathbf{H}'} \left[\log \det \left(I_t + \frac{f_{LB}^*(\mathsf{SNR})}{t} \mathbf{H}'^{\dagger} \mathbf{H}' \right) \right] \\
&\quad \geq (l-t) \left[r f_{LB}^*(\mathsf{SNR}) - \frac{r(r+t)}{2t} f_{LB}^{*2}(\mathsf{SNR}) \right] \\
&\quad \geq \frac{t^2}{(t+r)^2} \mathsf{SNR}^{-2\nu} \Big[r \mathsf{SNR}^{\min\{1,\nu\}} - 2 \frac{r(r+t)}{\sqrt{t}} \mathsf{SNR}^{\nu + \frac{\min\{1,\nu\}}{2}} \\
&\qquad\qquad - \frac{r(r+t)}{2t} \mathsf{SNR}^{2\min\{1,\nu\}} + o\Big(\mathsf{SNR}^{\min\{\nu + \frac{\min\{1,\nu\}}{2}, 2\min\{1,\nu\}\}} \Big) \Big] \triangleq C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}).
\end{aligned} \tag{3.31}$$

Equation (3.31) follows from (3.17). Hence, the lower bound to $E_r(R)$ in (3.30) is positive in the range

$$0 \leq R \leq C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}).$$

Combining the upper and lower bounds

Combining (3.11, 3.30), we have

$$\begin{aligned} E_r(R) &= \max_{\rho\in[0,1]}\left\{rt\log\left(1+\frac{\rho t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right)-\rho R\right\}-o(1), \\ 0 &\le R\le C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}). \end{aligned} \tag{3.32}$$

Let

$$\rho^*=\arg\max_{\rho\in[0,1]}\left\{rt\log\left(1+\frac{\rho t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2(1+\rho)}\right)-\rho R\right\}.$$

We compute ρ^* in the regime of low SNR as:

$$\rho^*=\begin{cases} 1 & 0\le R\le R_{\mathsf{critical}} \\ \frac{1}{2}\left[\sqrt{1+4\left(\frac{rt}{R}-\frac{(t+r)^2\mathsf{SNR}^{2\nu-\min\{1,\nu\}}}{t}\right)}-1\right] & R_{\mathsf{critical}}\le R\le C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}) \end{cases},$$

where, R_{critical}, the critical rate [2], is

$$R_{\mathsf{critical}}=rt/2+o(1).$$

Substituting ρ^* in (3.32), we have for $0\le R\le R_{\mathsf{critical}}$,

$$E_r(R)=rt\log\left(1+\frac{t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{2(t+r)^2}\right)-R-o(1), \tag{3.33}$$

and, for $R_{\mathsf{critical}}\le R\le C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR})$,

$$\begin{aligned} &E_r(R) \\ &\quad= rt\log\left(1+\left[\frac{\sqrt{1+4\left(\frac{rt}{R}-\frac{(t+r)^2\mathsf{SNR}^{2\nu-\min\{1,\nu\}}}{t}\right)}-1}{\sqrt{1+4\left(\frac{rt}{R}-\frac{(t+r)^2\mathsf{SNR}^{2\nu-\min\{1,\nu\}}}{t}\right)}+1}\right]\left[\frac{t\mathsf{SNR}^{-[2\nu-\min\{1,\nu\}]}}{(t+r)^2}\right]\right) \\ &\qquad -\frac{R}{2}\left[\sqrt{1+4\left(\frac{rt}{R}-\frac{(t+r)^2\mathsf{SNR}^{2\nu-\min\{1,\nu\}}}{t}\right)}-1\right]-o(1). \end{aligned} \tag{3.34}$$

For $R \in \left(C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}), C^{\mathsf{block}}(\mathsf{SNR})\right]$, the lower bound (3.30) to $E_r(R)$, is 0. However, the upper bound (3.11) is $o(1)$ in this range. Hence, we can say that

$$E_r(R) = o(1) \;\text{ for }\; C_{\mathsf{T,lb}}^{\mathsf{block}}(\mathsf{SNR}) \leq R \leq C^{\mathsf{block}}(\mathsf{SNR}). \tag{3.35}$$

Equations (3.33 - 3.35) characterize the random coding error exponent for the non-coherent channel. This completes the proof of Theorem 4.

Chapter 4

Fiber Aided Wireless Network Architecture

In this chapter, we introduce the concept of a Fiber Aided Wireless Network Architecture (FAWNA), which allows high-speed mobile connectivity by leveraging the speed of optical networks. Specifically, we consider a single-input, multiple-output FAWNA (SIMO-FAWNA). Figure 4-1 (same as Figure 1-3) shows such a link between two points A and B. The various quantities in the figure will be described in detail in this chapter. Unless specified otherwise, all logarithms in this chapter are to the base 2.

4.1 Model and communication scheme

There are r wireless-optical interfaces and each of them is equipped with a single antenna. The interfaces relay the wireless signals they receive from the transmitter, to the receiver at B, over an optical fiber. Communication over the fiber is interference free, which may be achieved, for example, using Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA).

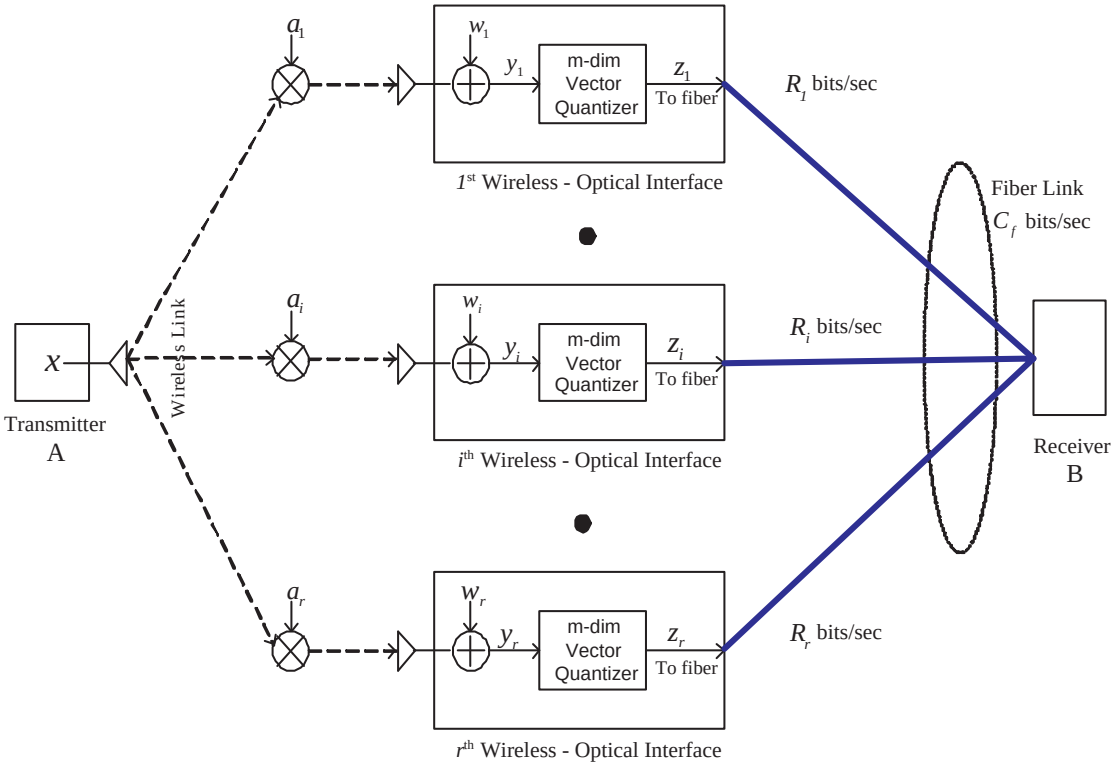

Figure 4-1: A SIMO fiber aided wireless network architecture.

4.1.1 Wireless Channel

We use a linear model for the wireless channel between A and the wireless-optical interfaces:

$$\vec{\mathbf{y}} = \vec{\mathbf{a}}\mathbf{x} + \vec{\mathbf{w}}, \tag{4.1}$$

where, $\mathbf{x} \in \mathcal{C}, \vec{\mathbf{w}}, \vec{\mathbf{y}} \in \mathcal{C}^r$ are the channel input, additive noise and output, respectively. We assume ergodic block fading where, $\vec{\mathbf{a}} \in \mathcal{C}^r$ is the channel state that is random but fixed for the coherence time of the channel and changes independently from block to block. The channel state is independent of the channel input and the additive noise, and is perfectly known at the receiver at B but not at the transmitter. $\mathbf{a}_i$ denotes the channel gain from the transmitter to the i^{th} wireless-optical interface. The additive noise, $\vec{\mathbf{w}} \sim \mathcal{CN}(0, N_0 I_r)$, is independent of the channel input and $N_0/2$ is the double-sided white noise spectral density. The channel input, $\mathbf{x}$, satisfies the average power

constraint

$$E[|\mathbf{x}|^2] = P/W,$$

where, P and W are the average transmit power at A and wireless bandwidth, respectively. Hence, the ergodic wireless channel capacity is

$$C_w(P, W, r) = WE\left[\log\left(1 + \frac{\|\vec{\mathbf{a}}\|^2 P}{N_0 W}\right)\right], \tag{4.2}$$

and W symbols are transmitted over the wireless channel every second.

4.1.2 Fiber Optic Channel

The fiber optic channel between the wireless-optical interfaces and the receiver can reliably support a rate of C_f bits/sec. Communication over the fiber is interference free and the i^{th} interface communicates at a rate of R_i bits/sec with the receiver at B. Let us define the set of all rate vectors satisfying

$$0 < R_i \leq C_f \qquad \text{for } i \in \{1, \ldots, r\}, \tag{4.3}$$

$$\sum_{i=1}^{r} R_i = C_f, \tag{4.4}$$

as $\mathcal{S}$. Fiber channel coding is performed at the wireless-optical interfaces to reliably achieve the rate vectors in $\mathcal{S}$. Note that the code required for the fiber is a very low complexity one. An example of a code that may be used is the 8B10B code, which is commonly used in Ethernet. Hence, fiber channel coding does not significant increase the complexity at the wireless-optical interface. We assume error free communication over the fiber for all sum rates below fiber capacity. To keep the interfaces simple, source coding is not done at the interfaces. We show later that since fiber capacity is large compared to the wireless capacity, the loss from not performing source coding is negligible.

4.1.3 Communication Scheme

The input to the wireless channel, $\mathbf{x}$, is a zero mean circularly symmetric complex Gaussian random variable, $\mathbf{x} \sim \mathcal{CN}(0, P/W)$. Note that it is this input distribution

that achieves the capacity of our wireless channel model. At each wireless-optical interface, the output from the antenna is first converted from passband to baseband and then sampled at the Nyquist rate of W complex samples/sec. The random variable, $\mathbf{y}_i$, represents the output from the sampler at the i^{th} interface. Fixed rate memoryless m-dimensional vector quantization is performed on these samples at a rate of R_i/W bits/complex sample. The quantized complex samples are subsequently sent over the fiber after fiber channel coding and modulation. Thus, the fiber is required to support reliably a rate of R_i bits/sec from the i^{th} wireless-optical interface to the receiver at B.

The quantizer noise at the i^{th} interface, $\mathbf{q}_i$, is modeled as being additive. Hence, the two-hop channel between A and B can be modeled as:

$$\vec{\mathbf{z}} = \vec{\mathbf{a}}\mathbf{x} + \vec{\mathbf{w}} + \vec{\mathbf{q}}, \tag{4.5}$$

where, $\vec{\mathbf{q}} = [\mathbf{q}_1, \ldots, \mathbf{q}_r]^T$. The interfaces have noise from two sources, receiver front end (front end noise $\vec{\mathbf{w}}$) and distortion introduced by their quantizers ($\vec{\mathbf{q}}$). The quantizer at each interface is an optimal fixed rate memoryless m-dimensional high resolution vector quantizer. Hence, its distortion-rate function is given by the Zador-Gersho function [1, 6, 8]:

$$E[|\mathbf{q}_i|^2] = E[|\mathbf{y}_i|^2] M_m \beta_m 2^{-\frac{R_i}{W}} = \left(N_0 + \frac{E\left[|\mathbf{a}_i|^2\right] P}{W} \right) M_m \beta_m 2^{-\frac{R_i}{W}}. \tag{4.6}$$

M_m is the Gersho's constant, which is independent of the distribution of $\mathbf{y}_i$, and β_m is the Zador's factor, which depends on the distribution of $\mathbf{y}_i$. Since fiber channel capacity is large, the assumption that the quantizer is a high resolution one is valid. Hence, for all i, $R_i/W \gg 1$. Also, as this quantizer is an optimal fixed rate memoryless vector quantizer, references [3, 5, 6, 7, 11] show that the following hold:

$$\begin{aligned} E[\mathbf{q}_i] &= 0, \\ E[\mathbf{z}_i \mathbf{q}_i^*] &= 0, \\ E[\mathbf{y}_i \mathbf{q}_i^*] &= -E[|\mathbf{q}_i|^2]. \end{aligned}$$

Therefore,

$$E[|\mathbf{z}_i|^2] = E[|\mathbf{y}_i|^2] - E[|\mathbf{q}_i|^2].$$

We denote the SIMO-FAWNA ergodic capacity using our scheme as $C_q(P, W, r, m, C_f)$. This can be expressed as

$$\begin{aligned} & C_q(P, W, r, m, C_f) \\ & \quad = WI(\mathbf{x}; \vec{\mathbf{z}}|\vec{\mathbf{a}}) \\ & \quad = WE\left[I(\mathbf{x}; \vec{\mathbf{z}}|\vec{\mathbf{a}} = \vec{a})\right] \\ & \quad = E\left[C_q^b(P, W, \vec{\mathbf{a}}, r, m, C_f)\right]. \end{aligned} \tag{4.7}$$

where, $C_q^b(P, W, \vec{a}, r, m, C_f) \triangleq WI(\mathbf{x}; \vec{\mathbf{z}}|\vec{a})$.

Since we consider links where $C_w(P, W, r) \leq C_f$, we obtain the following upper bound:

$$C_q(P, W, r, m, C_f) < C_w(P, W, r). \tag{4.8}$$

We show later in this chapter that $C_q(P, W, r, m, C_f)$ approaches this upper bound, exponentially with fiber capacity, and hence, is near optimal. Observe that the wireless-optical interfaces have low complexity and do not require knowledge of the transmitter code book. They are extendable to FAWNAs with large number of transmitters and interfaces and offer adaptability to variable rates, changing channel conditions and node positions.

4.2 Interface Rate Allocation

In this section, we address two questions: First, how should rates be allocated to the interfaces in a coherence block and second, since channel state varies independently from block to block, is there significant loss in not computing the optimal rate allocation every block?

To answer the first question, consider the channel within a block interval. The channel state in this block takes the realization $\vec{a}$. We establish the following theorem:

Theorem 5 *For any interface rate allocation, $\vec{R}$, we have*

$$\begin{aligned} & C_q^b(P, W, \vec{a}, r, m, C_f) \\ & \quad \geq W \log\left(\frac{1}{1 - \frac{P}{N_0 W}\vec{v}^{\dagger} M^{-1} \vec{v}}\right) \triangleq C_{q,LB}^b(P, W, \vec{a}, r, m, \vec{R}), \end{aligned} \tag{4.9}$$

where, $\vec{v}$ is specified for $i \in \{1, \ldots, r\}$ as

$$v_i = a_i(1 - M_m\beta_m 2^{-\frac{R_i}{W}}),$$

and M is specified for $i \in \{1, \ldots, r\}, j \in \{1, \ldots, r\}$ as

$$M_{ij} = \begin{cases} \frac{a_i a_j^* P}{N_0 W}\left(1 - M_m\beta_m 2^{-\frac{R_i}{W}}\right)\left(1 - M_m\beta_m 2^{-\frac{R_j}{W}}\right) & \text{for } i \neq j, \\ \left(1 + \frac{|a_i|^2 P}{N_0 W}\right)\left(1 - M_m\beta_m 2^{-\frac{R_i}{W}}\right) & \text{for } i = j. \end{cases}$$

Proof: See Appendix D. □

In the next section, we show that the supremum of the lower bound (4.9) over all rate vectors in $\mathcal{S}$, approaches $C_q^b(P, W, \vec{a}, r, m, C_f)$, exponentially with fiber capacity. Hence, we consider this lower bound alone for finding the optimal interface rate allocation.

The optimal rate allocation for this block is given by

$$\vec{R}^*(\vec{a}) = \arg\max_{\vec{R}\in\mathcal{S}} \left[C_{q,LB}^b(P, W, \vec{a}, r, m, \vec{R})\right]. \tag{4.10}$$

To understand optimal rate allocation, let us consider a SIMO-FAWNA with two interfaces[1], fiber capacity 200 Mbps, channel state $\vec{a} = [1\ \frac{1}{2}]^T$, $\frac{P}{N_0} = 100 \times 10^6$, $W = 5$ MHz and $M_m\beta_m = 1$. Since $R_2 = C_f - R_1$, it suffices to consider the capacity with respect to R_1 alone. The plot of $C_{q,LB}^b(P, W, \vec{a}, r, m, \vec{R})$ with respect to R_1 is shown in figure 4-2.

We can divide the plot into three regions. The first region is from 0 Mbps to 50 Mbps, where the first interface has low rate[2] and the second has high rate. Thus, noise at the first interface is quantizer distortion dominated whereas at the second interface is front end noise dominated. Hence, as we increase the rate for the first interface, the distortion at the first interface decreases and overall capacity increases. The reduction in rate at the second interface due to increase in R_1 has negligible effect on capacity since front end noise still dominates at the second interface.

The second region is from 50 Mbps to 170 Mbps. In this region, the rates for both

[1]Even though we consider a two interface SIMO-FAWNA, results generalize to SIMO-FAWNAs with any number of interfaces.

[2]Whenever we mention "low rate", the rate considered is always high enough for the high resolution quantizer model to be valid.

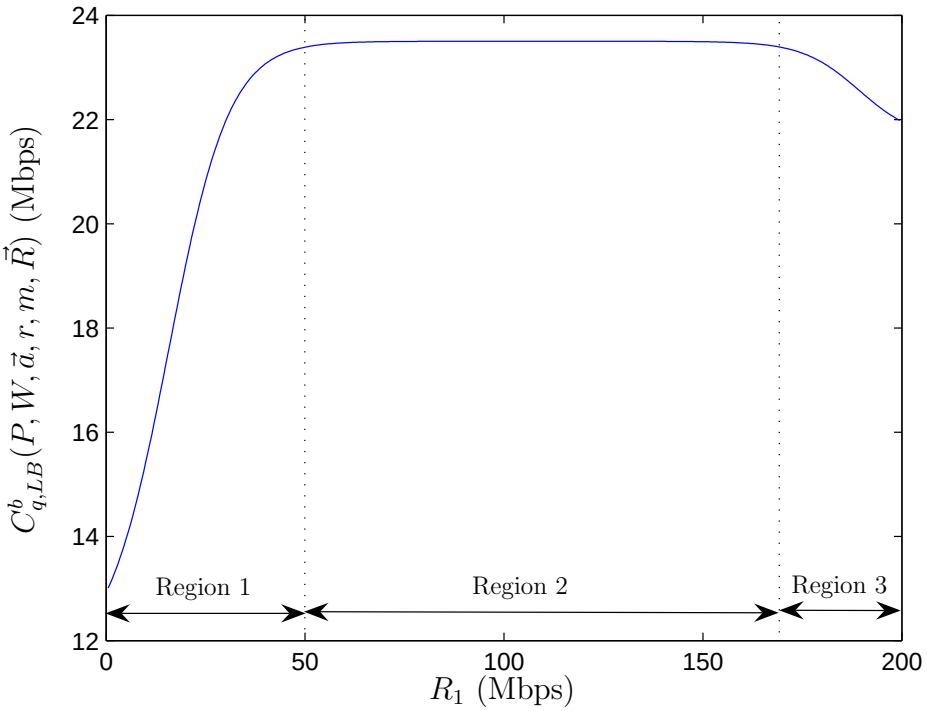

Figure 4-2: Interface rate allocation for a two interface SIMO-FAWNA.

interfaces are high enough for front end noise to dominate. Since quantizer distortion is low with respect to the front end noise at both interfaces, capacity is almost invariant to rate allocation. Observe that capacity is maximum in this region and the size of this region is much larger than that of the first and third.

The third region is from 170 Mbps to 200 Mbps and here, the first interface has high rate and the second has low rate. Therefore, noise at the first interface is front end noise dominated whereas at the second interface is quantizer distortion dominated. An increase in rate for the first interface results in decrease in rate for the second interface. This decrease in rate results in an increase in quantizer distortion at the second interface, which results in overall capacity decrease.

The channel gain at the first interface is higher than that at the second interface. Hence, compared to the second interface, the first interface requires more rate to bring its quantizer's distortion below the front end noise power. Also, reduction in quantizer distortion at the first interface results in higher capacity gains than reduc-

tion in quantizer distortion at the second interface. This can been seen from the asymmetric nature of the plot in figure 4-2 around $R_1 = 100$ Mbps.

We see that the optimum interface rate allocation for a FAWNA is to ensure that each interface gets enough rate for it to lower its quantizer distortion to the point where its noise is front end noise dominated. Wireless-optical interfaces seeing higher channel gains require higher rates to bring down their quantizer distortion. After this requirement is met, FAWNA capacity is almost invariant to allocation of left over fiber capacity. This can be seen from the near flat capacity curve in the second region of the plot in figure 4-2. Thus, any interface rate allocation that ensures that noise at none of the wireless-optical interfaces is quantizer distortion dominated, is near optimal.

Since fiber capacity is large compared to the wireless capacity, the fraction of fiber capacity needed to bring down the distortion for the interfaces so that none of them is quantizer distortion limited, is small. Therefore, the set of interface rate vectors for which $C^b_{q,LB}(P, W, \vec{a}, r, m, \vec{R})$ is near maximum, is large and there is considerable flexibility in allocating rates across the interfaces. Therefore, we see that large fiber capacity brings robustness to interface rate allocation in a FAWNA. For example, from figure 4-2, we see that even an equal rate allocation for the two interface SIMO-FAWNA is near-optimal.

We now address the second question posed at the beginning of this section: Since channel state changes independently from block to block, is there significant loss in not computing the optimal rate allocation every block? First, consider the case where interface rate allocation is dynamic, i.e., done in every block. The optimal rate allocation vector for the block is given by (4.10) and it depends on the channel realization (state). The ergodic capacity lower bound of a SIMO-FAWNA with dynamic rate allocation is given by

$$C^{\mathsf{D}}_{q,LB}(P, W, r, m, C_f) = E\left[C^b_{q,LB}\left(P, W, \vec{\mathbf{a}}, r, m, \vec{R}^*(\vec{\mathbf{a}})\right)\right].$$

Consider the same two interface SIMO-FAWNA as in the previous question but with channel state $\vec{\mathbf{a}} = [\mathbf{h}_1 \ \frac{1}{2}\mathbf{h}_2]^T$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are i.i.d $\mathcal{CN}(0, 1)$. For this wireless-

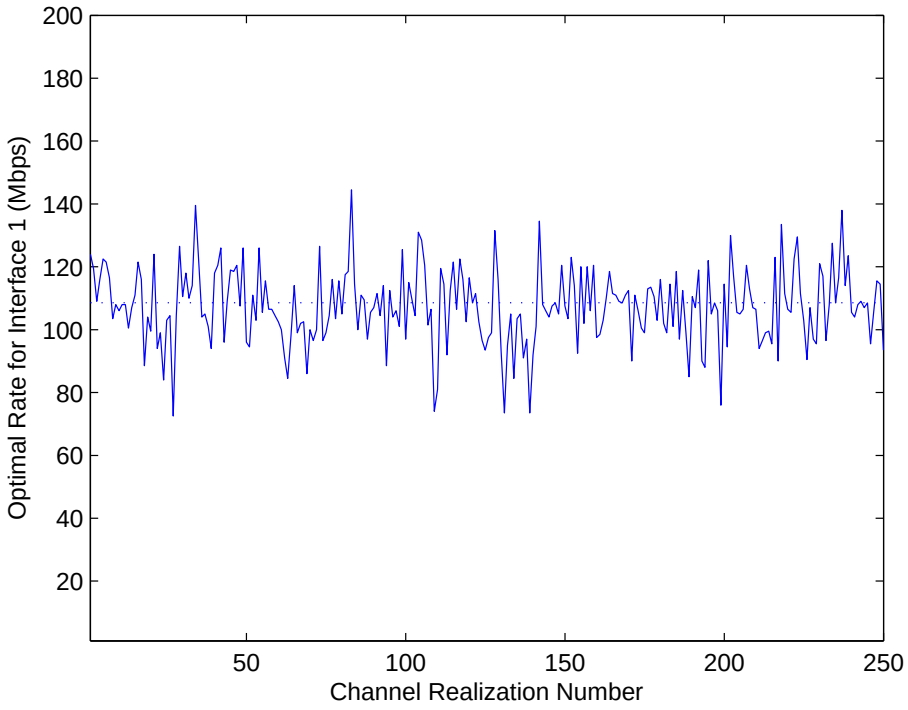

Figure 4-3: Dynamic rate allocation.

optical channel, we compute $C^{\mathsf{D}}_{q,LB}(P, W, r, m, C_f) \sim 21.4$ Mbps . Figure 4-3 shows how the optimal rate for the first interface, R_1^*, changes with channel realization. Since the average channel gain at the first interface is larger than that at the second, the mean of the observations in the figure is above half the fiber capacity.

Dynamic rate allocation involves computation of the optimal rate allocation vector at the receiver at B and updating the interfaces with optimal values of rates, every coherence block. This considerably increases the complexity in a FAWNA. In order to simplify, we consider static rate allocation, i.e., interface rate allocation is computed based on wireless channel statistics and fixed forever. The interface rate allocation vector is chosen as one that maximizes the ergodic capacity lower bound:

$$\vec{R}^*_{\mathsf{S}} = \arg\max_{\vec{R}\in\mathcal{S}} E\left[C^b_{q,LB}\left(P, W, \vec{\mathbf{a}}, r, m, \vec{R}\right)\right].$$

Hence, the ergodic capacity lower bound of a SIMO-FAWNA with static rate allocation is

$$C^{\mathsf{S}}_{q,LB}(P, W, r, m, C_f) = E\left[C^b_{q,LB}\left(P, W, \vec{\mathbf{a}}, r, m, \vec{R}^*_{\mathsf{S}}\right)\right].$$

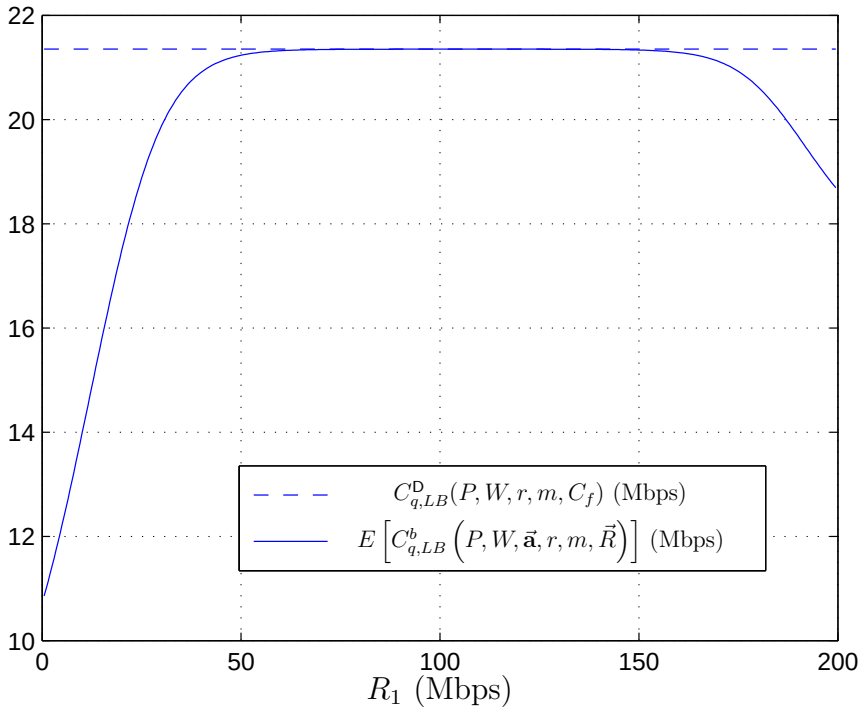

Figure 4-4: Near-optimality of static rate allocation.

Note that this is sub-optimal to dynamic rate allocation:

$$C^{\mathsf{S}}_{q,LB}(P,W,r,m,C_f) \leq C^{\mathsf{D}}_{q,LB}(P,W,r,m,C_f) \leq C_q(P,W,r,m,C_f).$$

For the two interface SIMO-FAWNA, figure 4-4 shows how ergodic capacity changes with R_1. Since the ergodic capacity is the capacity averaged over channel realizations, this plot is similar to that in figure 4-2. From figure 4-4, we observe that $C^{\mathsf{S}}_{q,LB}(P,W,r,m,C_f) = 21.35$ Mbps and the near-optimal rates for interface 1 are $R^*_{\mathsf{S},1} \sim [72, 142]$ Mbps.

Note that the loss from static rate allocation is very small. Moreover, the set of static rate allocation vectors for which this loss is very small, is large. For this example, the loss is only 50 Kbps or 0.23% of capacity, and all rates from 72 Mbps to 142 Mbps are close to optimal for interface 1. Though the SIMO-FAWNA capacity is sensitive to quantizer distortion, large fiber capacity ensures that the interfaces always have enough rate so that they are never distortion limited over the typical set of channel realizations. This robustness of FAWNA capacity to interface rate allocation makes

static rate allocation near-optimal. Observe from figure 4-4 that even equal rate allocation is near-optimal. This near-optimality of static rate allocation translates to considerable reduction in FAWNA complexity.

4.3 Effect of various parameters on performance

In this section, we analyze the effect of quantizer dimension, fiber capacity, transmit power, number of interfaces and wireless bandwidth on the performance of our scheme. To simplify analysis, we set the wireless channel gain $\vec{\mathbf{a}} = \mathbf{g} \cdot \vec{1}$, where, $\vec{1}$ is a r dimensional column vector with all ones and $\mathbf{g}$ is a complex random variable. For this channel, all interfaces have the same instantaneous received power. Hence, an equal interface rate allocation is optimal:

$$\vec{R}^*(\mathbf{g} \cdot \vec{1}) = \vec{R}^*_{\mathsf{S}} = \frac{C_f}{r} \cdot \vec{1},$$

and there is no loss from static interface rate allocation. Hence,

$$C^{\mathsf{S}}_{q,LB}(P, W, r, m, C_f) = C^{\mathsf{D}}_{q,LB}(P, W, r, m, C_f).$$

Since the ergodic capacity using dynamic rate allocation is the same as that using static rate allocation, we will remove the superscript to simplify notation and denote the ergodic capacity lower bound as $C_{q,LB}(P, W, r, m, C_f)$. From Theorem 5, we can express this lower bound as

$$C_{q,LB}(P, W, r, m, C_f) = WE\left[\log\left(1 + \frac{r|\mathbf{g}|^2(1 - M_m\beta_m 2^{-\frac{C_f}{rW}})\frac{P}{N_0W}}{1 + \frac{|\mathbf{g}|^2 P M_m \beta_m 2^{-\frac{C_f}{rW}}}{N_0W}}\right)\right]. \quad (4.11)$$

We show in this section that the lower bound (4.11) approaches the upper bound $C_w(P, W, r)$ in (4.8), exponentially with fiber capacity. Hence, since the fiber capacity is large, the lower bound almost completely characterizes $C_q(P, W, r, m, C_f)$ and we consider this alone for analysis.

4.3.1 Effect of quantizer dimension

We now study the effect of quantizer dimension, m, on the performance of the proposed scheme. Since Gaussian signaling is used for the wireless channel, the input to the quantizer at the interface is a correlated Gaussian random vector. Zador's factor and Gersho's constant obey the following property:

$$M_\infty \beta_\infty \leq M_m \beta_m \leq M_1 \beta_1 \leq M_1 \beta_1^G,$$

where, β_1^G is the Zador's factor for an i.i.d Gaussian source and $\beta_1 \leq \beta_1^G$. $M_m \beta_m$ decreases with increase in m. Since $M_1 = \frac{1}{12}$, $M_\infty = \frac{1}{2\pi e}$, $\beta_1^G = 6\sqrt{3}\pi$ and $\beta_\infty = 2\pi e$,

$$1 \leq M_m \beta_m \leq \frac{\pi\sqrt{3}}{2}.$$

The lower bound corresponds to fixed rate infinite dimensional vector quantization whereas, the upper bound corresponds to fixed rate scalar quantization.
In (4.11), $\frac{r|\mathbf{g}|^2(1-M_m\beta_m 2^{-\frac{C_f}{rW}})\frac{P}{N_0 W}}{1+\frac{|\mathbf{g}|^2 P M_m \beta_m 2^{-\frac{C_f}{rW}}}{N_0 W}}$ decreases monotonically with increase in $M_m\beta_m$. Hence, $C_{q,LB}(P, W, r, m, C_f)$ increases with m and can be lower and upper bounded as

$$C_{q,LB}(P, W, r, 1, C_f) \leq C_{q,LB}(P, W, r, m, C_f) \leq C_{q,LB}(P, W, r, \infty, C_f),$$

where, $C_{q,LB}(P, W, r, 1, C_f)$ and $C_{q,LB}(P, W, r, \infty, C_f)$ correspond to ergodic capacity lower bounds for fixed rate scalar and infinite dimensional vector quantization at the interfaces, respectively. Reduction in quantizer dimension reduces complexity at the interface but results in a capacity penalty. The maximum loss in capacity occurs when fixed rate scalar quantizers are used at the wireless-optical interfaces.

4.3.2 Effect of fiber capacity

We now analyze the effect of fiber capacity on the performance of a SIMO-FAWNA. Define

$$\Phi(C_f) \triangleq C_w(P, W, r) - C_{q,LB}(P, W, r, m, C_f).$$

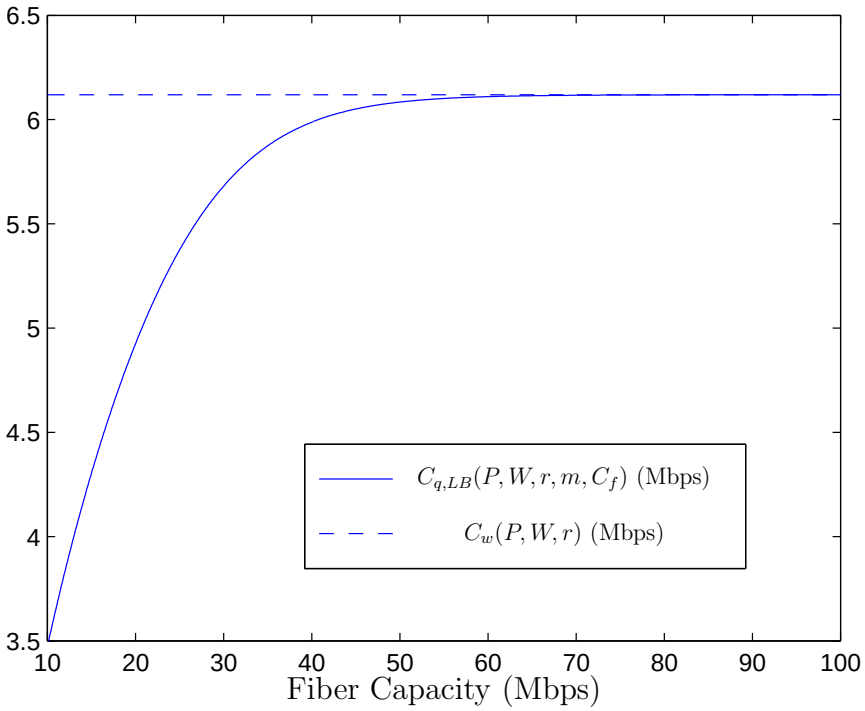

Figure 4-5: Dependence of SIMO-FAWNA capacity on fiber capacity.

From (4.2, 4.11), we have

$$\Phi(C_f) = -E\left[\log\left(1 - \frac{r|\mathbf{g}|^2\frac{P}{N_0W}(1+\frac{|\mathbf{g}|^2P}{N_0W})M_m\beta_m 2^{-\frac{C_f}{rW}}}{1+\frac{|\mathbf{g}|^2PM_m\beta_m 2^{-\frac{C_f}{rW}}}{N_0W}}\right)\right].$$

Now,

$$\begin{aligned}
&\Phi(C_f)\\
&\quad\le -E\left[\log\left(1-\frac{r|\mathbf{g}|^2P}{N_0W}\left(1+\frac{|\mathbf{g}|^2P}{N_0W}\right)M_m\beta_m 2^{-\frac{C_f}{rW}}\right)\right]\\
&\quad= O(2^{-C_f}),
\end{aligned}$$

and

$$\begin{aligned}
&\Phi(C_f)\\
&\quad\ge -E\left[\log\left(1-\frac{r|\mathbf{g}|^2\frac{P}{N_0W}(1+\frac{|\mathbf{g}|^2P}{N_0W})M_m\beta_m 2^{-\frac{C_f}{rW}}}{1+\frac{|\mathbf{g}|^2PM_m\beta_m}{N_0W}}\right)\right]\\
&\quad= O(2^{-C_f}).
\end{aligned}$$

Hence,

$$\Phi(C_f) = O(2^{-C_f}).$$

Therefore,

$$C_{q,LB}(P, W, r, m, C_f) = C_w(P, W, r) - O(2^{-C_f}).$$

This implies that the ergodic capacity lower bound using the proposed scheme approaches the capacity upper bound (4.8), *exponentially* with fiber capacity. Also observe that $\Phi(\infty) = 0$. Note that though our scheme simply quantizes and forwards the wireless signals without source coding, we see that it is near-optimal since the fiber capacity is much larger than the wireless capacity. This behavior is illustrated in figure 4-5, which is a plot of $C_{q,LB}(P, W, r, m, C_f)$ and the upper bound (4.8), versus fiber capacity. In the plot, we set $\mathbf{g} \sim \mathcal{CN}(0, 1)$, $W = 1$ Mhz, $M_m\beta_m = 1$, $r = 5$ and $\frac{P}{N_0} = 25 \times 10^6$ sec^{-1}. Note that the fiber capacity required to achieve good performance is not large for an optical fiber, which have speeds in the order of Gigabit/sec.

4.3.3 Effect of transmit power

An increase in transmit power, P, leads to two competing effects. The first is increase in receive power at the interfaces, which increases capacity. The second is increase in quantizer distortion, which reduces capacity. The ergodic capacity lower bound of our scheme, $C_{q,LB}(P, W, r, m, C_f)$, increases monotonically with $\frac{r|\mathbf{g}|^2(1-M_m\beta_m 2^{-\frac{C_f}{rW}})\frac{P}{N_0W}}{1+\frac{|\mathbf{g}|^2 P M_m\beta_m 2^{-\frac{C_f}{rW}}}{N_0W}}$, which in turn increases monotonically with P. Hence, the first effect always dominates and the ergodic capacity lower bound of our scheme *increases* with transmit power.

4.3.4 Effect of number of wireless-optical interfaces

Let us focus on the effect of the number of interfaces, r, on $C_{q,LB}(P, W, r, m, C_f)$. Since the quantization rate at the interface is never allowed to go below 1, the maximum

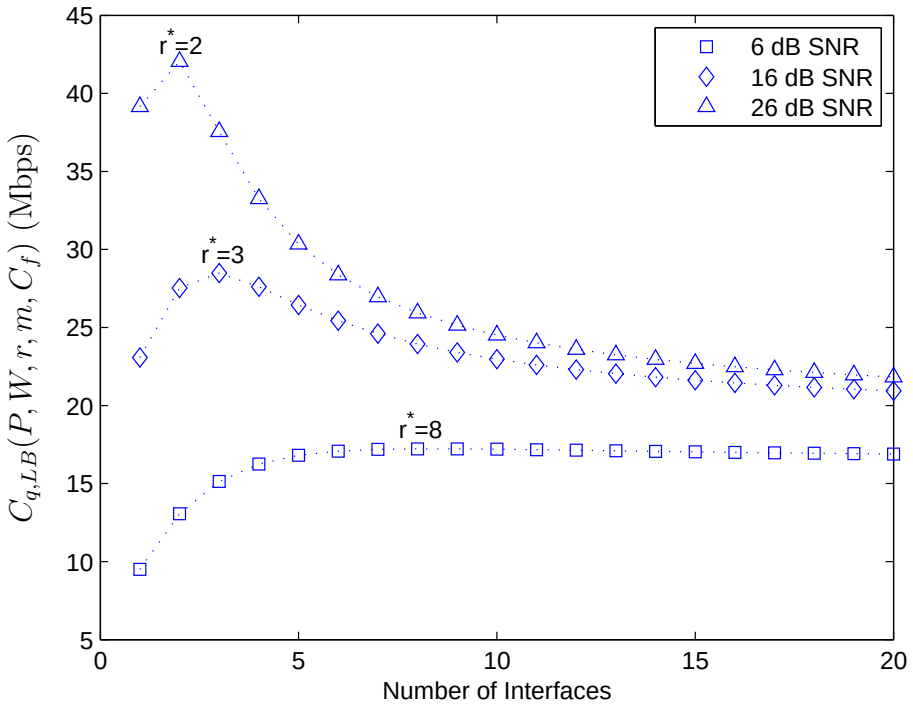

Figure 4-6: Effect of the number of interfaces on $C_{q,LB}(P, W, r, m, C_f)$.

number of interfaces possible is $r_{\max} = \lfloor \frac{C_f}{W} \rfloor$. Keeping all other variables fixed, the optimal number of interfaces, r^*, is given by

$$r^* = \arg \max_{r \in \{1,2,\ldots,r_{\max}\}} C_{q,LB}(P, W, r, m, C_f).$$

For fixed wireless bandwidth and fiber capacity, an increase in the number of interfaces leads to two competing effects. First, capacity increases owing to receive power gain from the additional interfaces. Second, quantizer distortion increases owing to additional interfaces sharing the same fiber, which results in capacity reduction. Hence, capacity doesn't increase monotonically with the number of antennas. Obtaining an analytical expression for r^* is difficult. However, r^* can easily be found by numerical techniques. Figure 4-6 is a plot of $C_{q,LB}(P, W, r, m, C_f)$ versus r for $\mathbf{g} \sim \mathcal{CN}(0, 1)$, $W = 5$ Mhz, $M_m \beta_m = 1$, $C_f = 100$ Mbps. Note that, for this example, $r_{\max} = 20$. Plots are obtained for $\frac{P}{N_0} = 20 \times 10^6$ sec^{-1}, 200×10^6 sec^{-1} and 2000×10^6 sec^{-1}, which correspond to average interface signal-to-noise ratio (SNR) of 6 dB, 16 dB and 20 dB, respectively. The corresponding values of r^* are 8, 3 and 2, respectively. Ob-

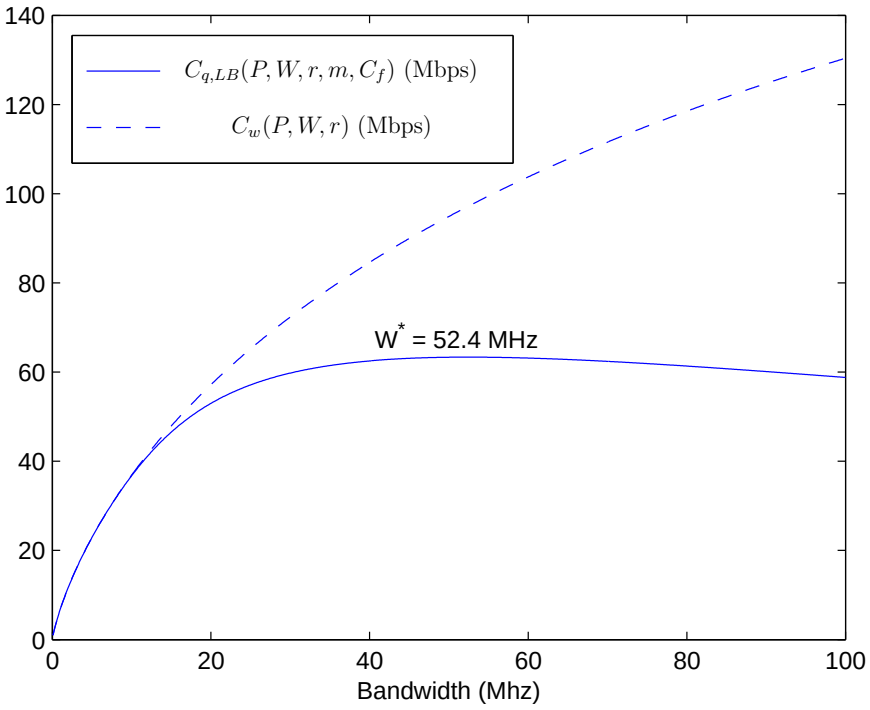

Figure 4-7: Dependence of SIMO-FAWNA capacity on wireless bandwidth.

serve that r^* decreases with increase in average interface SNR. This happens because, when average interface SNR is low, it becomes more important to gain power rather than to have fine quantization. On the other hand, when average interface SNR is high, the latter is more important. Hence, as average interface SNR decreases, r^* tends towards r_{max}.

4.3.5 Effect of wireless bandwidth

We now analyze the effect of wireless bandwidth, W, on $C_{q,LB}(P,W,r,m,C_f)$. Since the quantization rate is never allowed to go below 1, the maximum possible bandwidth is C_f/r. For fixed fiber capacity and number of interfaces, the optimal bandwidth of operation, W^*, is given by

$$W^* = \arg \max_{W \in \left[0, \frac{C_f}{r}\right]} C_{q,LB}(P,W,r,m,C_f).$$

Since quantizer distortion as well as power efficiency increases with W, the behavior of the capacity lower bound with bandwidth is similar to that with the number of interfaces. Note that the quantization rate at each interface decays inversely with bandwidth. When the operating bandwidth is lowered from W^*, the capacity lower bound is lowered because the reduction in power efficiency is more than the reduction in quantizer distortion. On the other hand, when the operating bandwidth is increased from W^*, the loss in capacity from increased quantizer distortion is more than the capacity gain from increased power efficiency.

The optimal bandwidth, W^*, can be found by numerical techniques. Figure 4-7 shows the plot of the capacity lower bound and the upper bound (4.8) for $\mathbf{g} \sim \mathcal{CN}(0,1)$, $C_f = 200$ Mbps, $M_m\beta_m = 1$, $r = 2$ and $\frac{P}{N_0} = 100 \times 10^6$ sec^{-1}. The optimal bandwidth for this case is $W^* = 52.4$ Mhz.

Chapter 5

Conclusion

In chapters 2 and 3, we compute the capacity and error probability for the non-coherent wideband MIMO channel. The effect on capacity and reliability of coherence length and number of transmit and receive antennas is examined. The analysis shows that, though the number of transmit antennas does not affect the linear capacity term, it does affect the sublinear capacity term, i.e., the approach of capacity to the wideband limit with increasing bandwidth. We also establish conditions on the channel coherence length and number of antennas for the non-coherent capacity to be the same as the coherent capacity in the wideband regime. We show that the error probability decays inversely with coherence length and exponentially with product of the number of transmit and receive antennas. This highlights the importance of multiple transmit antennas, besides multiple receive antennas, in the low SNR regime. An interesting observation is that outage probability dominates the error probability even at low SNR.

Recent research [43, 44] shows that for single antenna channels in the wideband regime, an energy limited jammer has no effect on the wideband capacity limit nor the error exponent. Future work may consider extending these results to the multiple antennas case.

In chapter 4, we study a SIMO-FAWNA from a capacity view point and propose a near-optimal design. We show that an optimal interface rate allocation is one which ensures that each interface gets enough rate so that its noise is dominated by front

end noise rather than quantizer distortion. Capacity is almost invariant to the way in which left over fiber capacity is allocated. Hence, large fiber capacity ensures robustness of SIMO-FAWNA capacity to interface rate allocation. This robustness has an important implication on design, rather than dynamically change interface rate allocation based on channel state, a fixed rate allocation scheme can be adopted with very small loss in capacity. This results in considerable reduction in system complexity. We also show that for a given fiber capacity, there is an optimal operating wireless bandwidth and an optimal number of wireless-optical interfaces. The wireless-optical interfaces have low complexity and do not require knowledge of the transmitter code book. The design also has extendability to FAWNAs with large number of transmitters and interfaces and, offers adaptability to variable rates, changing channel conditions and node positions.

Future research may consider FAWNAs with multiple transmitters and examine the performance of various multiple access schemes. For the multiple transmitters scenario, interference reduction and tradeoff between the various system parameters are interesting topics for study.

Appendix A

Proof of Lemma 1

Proof of (2.1): For any $\alpha \in (0, 1]$ and $\gamma \in (0, \alpha)$, let there exist an input distribution on $\mathbf{X}$ that satisfies the following:

$$\frac{1}{l} I(\mathbf{X}; \mathbf{Y}) \geq r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}). \tag{A.1}$$

Let $\mathbf{Y}^G$ be a matrix with i.i.d complex Gaussian entries and satisfying

$$E[\mathsf{trace}(\mathbf{Y}^G \mathbf{Y}^{G\dagger})] = E[\mathsf{trace}(\mathbf{Y}\mathbf{Y}^{\dagger})].$$

Hence, $h(\mathbf{Y}) \leq h(\mathbf{Y}^G)$ and the entries of $\mathbf{Y}^G$ i.i.d $\mathcal{CN}(0, 1 + \mathsf{SNR})$. Moreover, conditioned on $\mathbf{X}$, the row vectors of $\mathbf{Y}$ are i.i.d $\mathcal{CN}(0, \mathbf{X}\mathbf{X}^{\dagger} + I_l)$. We can thus upper bound the mutual information as

$$\begin{aligned} & I(\mathbf{X}; \mathbf{Y}) \\ & \quad = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) \\ & \quad \leq h(\mathbf{Y}^G) - h(\mathbf{Y}|\mathbf{X}) \\ & \quad = rl \log(1 + \mathsf{SNR}) - rtE[\log(1 + \|\vec{\mathbf{x}}_i^T\|^2)] \\ & \quad \leq rl\mathsf{SNR} - rtE[\log(1 + \|\vec{\mathbf{x}}_i^T\|^2)]. \end{aligned} \tag{A.2}$$

Combining (A.1) and (A.2) and noting that the norms of the input vectors $\|\vec{\mathbf{x}}_i^T\|$ are identically distributed, we see that if the input distribution satisfies (A.1), then it necessarily satisfies the first condition (2.1).

Proof of (2.2): Observing the structure of the optimal input [20] for the non-coherent MIMO channel, we can upper bound the mutual information as

$$I(\mathbf{X};\mathbf{Y}) \leq I(\mathbf{A};\mathbf{Y}|\mathbf{\Phi}) + I(\mathbf{\Phi};\mathbf{Y}|\mathbf{A}), \tag{A.3}$$

where, $I(\mathbf{A};\mathbf{Y}|\mathbf{\Phi})$ is the information conveyed by the norm of the transmitted signal vectors given that the receiver has side information about their directions, and $I(\mathbf{\Phi};\mathbf{Y}|\mathbf{A})$ is the information conveyed by the direction of these vectors when the receiver has side information about their norm. We establish upper bounds on these two terms.

Upper bound for $I(\mathbf{A};\mathbf{Y}|\mathbf{\Phi})$:
When the receiver has side information about $\mathbf{\Phi}$, it can filter out noise orthogonal to the subspace spanned by the row vectors of $\mathbf{\Phi}$ to obtain an equivalent channel

$$\begin{aligned}
&\mathbf{Y}\mathbf{\Phi}^{\dagger} \\
&\quad = \mathbf{H}\mathbf{X}\mathbf{\Phi}^{\dagger} + \mathbf{W}\mathbf{\Phi}^{\dagger} \\
&\quad = \mathbf{H}\mathbf{A} + \mathbf{W}^{'},
\end{aligned}$$

where $\mathbf{W}^{'}$ has the same distribution as $\mathbf{W}$ and there is no loss in information since $\mathbf{Y}\mathbf{\Phi}^{\dagger}$ is a sufficient statistic for estimating $\mathbf{X}$ from $\mathbf{Y}$. Therefore

$$\begin{aligned}
&I(\mathbf{A};\mathbf{Y}|\mathbf{\Phi}) \\
&\quad = I(\mathbf{A};\mathbf{Y}\mathbf{\Phi}^{\dagger}|\mathbf{\Phi}) \\
&\quad = I(\mathbf{A};\mathbf{H}\mathbf{A} + \mathbf{W}^{'}|\mathbf{\Phi}) \\
&\quad \leq \sum_{i=1}^{t} I(\|\vec{\mathbf{x}}_i^T\|;\vec{\mathbf{h}}_i\|\vec{\mathbf{x}}_i^T\| + \vec{\mathbf{w}}^{'T}_i|\vec{\boldsymbol{\phi}}_i^T) \\
&\quad \leq \sum_{i=1}^{t}\sum_{j=1}^{r} I(\|\vec{\mathbf{x}}_i^T\|;\mathbf{h}_{ij}\|\vec{\mathbf{x}}_i^T\| + \mathbf{w}^{'}_{ij}|\vec{\boldsymbol{\phi}}_i^T),
\end{aligned} \tag{A.4}$$

where the last two inequalities follow from the chain rule of mutual information and the fact that conditioning reduces entropy. In order to get an upper bound on $I(\|\vec{\mathbf{x}}_i^T\|;\mathbf{h}_{ij}\|\vec{\mathbf{x}}_i^T\| + \mathbf{w}^{'}_{ij}|\vec{\boldsymbol{\phi}}_i^T)$, we need to maximize this mutual information with the

average power constraint, $l\mathsf{SNR}/t$, and the constraint specified by (2.1). If we relax the latter constraint (2.1), then the mutual information is that of a single-input, single-output i.i.d Rayleigh fading channel with average power constraint $l\mathsf{SNR}/t$. From [25], we know that this mutual information is maximized by an on-off distribution of the form

$$\|\vec{\mathbf{x}}_i^T\|^2 = \begin{cases} \frac{l}{\zeta}\frac{\mathsf{SNR}}{t} & w.p. \quad \zeta \\ 0 & w.p. \quad 1-\zeta \end{cases}$$

for $\forall i \in \{1, \ldots, t\}$ and some $\zeta > 0$. This signaling scheme becomes increasingly "flashy" as the SNR gets low, i.e., $\zeta \to 0$ as $\mathsf{SNR} \to 0$. Hence, (A.4) becomes

$$\begin{aligned} & I(\mathbf{A}; \mathbf{Y}|\mathbf{\Phi}) \\ & \quad \leq \sum_{i=1}^{t}\sum_{j=1}^{r} I(\|\vec{\mathbf{x}}_i^T\|; \mathbf{h}_{ij}\|\vec{\mathbf{x}}_i^T\| + \mathbf{w}_{ij}^{'}|\vec{\boldsymbol{\phi}}_i^T) \\ & \quad \leq \sum_{i=1}^{t}\sum_{j=1}^{r} H(\|\vec{\mathbf{x}}_i^T\|) \\ & \quad \approx rt\zeta \log(\frac{1}{\zeta}), \end{aligned}$$

where the approximation is valid since we are in the low signal to noise ratio regime and $\zeta \to 0$ as $\mathsf{SNR} \to 0$. Therefore, we have

$$\frac{1}{l} I(\mathbf{A}; \mathbf{Y}|\mathbf{\Phi}) \leq \frac{rt\zeta}{l} \log(\frac{1}{\zeta}). \tag{A.5}$$

However, this on-off distribution minimizes (2.1) also and hence the extra constraint does not change the optimal input. Therefore, it suffices to consider on-off signals. Hence, (2.1) becomes

$$\begin{aligned} & \frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}) \\ & \quad \geq \frac{t\zeta}{l} \log(1 + \frac{l}{\zeta t}\mathsf{SNR}) \\ & \quad \geq \frac{t\zeta}{l}\Big[\log(\frac{1}{\zeta}) - \log(\frac{t}{l\mathsf{SNR}})\Big] \\ & \quad \approx \frac{t\zeta}{l} \log(\frac{1}{\zeta}), \end{aligned} \tag{A.6}$$

where the approximation is valid since $\frac{l}{\zeta}\frac{\mathsf{SNR}}{t} \gg 1$ as $\mathsf{SNR} \to 0$, i.e. the peak amplitude becomes very large as the signal to noise ratio tends to 0. Combining (A.5) and (A.6), we have

$$\frac{1}{l} I(\mathbf{A}; \mathbf{Y}|\mathbf{\Phi}) \leq \frac{r(r+t)}{2t} \mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}). \tag{A.7}$$

Upper bound for $I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A})$:

We can upper bound $I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A})$ in terms of the mutual information of a single-input, single-output channel, i.e.

$$I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A}) \leq \sum_{i=1}^{t} \sum_{j=1}^{r} I(\vec{\boldsymbol{\phi}}_i^T; \vec{\mathbf{y}}_j^T | \mathbf{A}, \vec{\boldsymbol{\phi}}_1^T, \ldots, \vec{\boldsymbol{\phi}}_{i-1}^T, \vec{\boldsymbol{\phi}}_{i+1}^T, \ldots, \vec{\boldsymbol{\phi}}_t^T, \vec{\mathbf{y}}_1^T, \ldots, \vec{\mathbf{y}}_{j-1}^T, \vec{\mathbf{y}}_{j+1}^T, \ldots, \vec{\mathbf{y}}_r^T).$$

The term inside the double summation represents the mutual information of the channel between the i^{th} transmit antenna and j^{th} receive antenna when no other antenna is present and the norm of $\vec{\mathbf{x}}_i^T$ is known at the receiver. Since the input vectors are identically distributed and the channel matrix has i.i.d entries, the mutual information between any pair of transmit and receive antennas given that the other antennas are absent will be the same. Hence, for all $i \in \{1, \ldots, t\}$ and $j \in \{1, \ldots, r\}$,

$$I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A}) \leq rtI(\vec{\boldsymbol{\phi}}_i^T; \vec{\mathbf{y}}_j^T | \mathbf{A}, \vec{\boldsymbol{\phi}}_1^T, \ldots, \vec{\boldsymbol{\phi}}_{i-1}^T, \vec{\boldsymbol{\phi}}_{i+1}^T, \ldots, \vec{\boldsymbol{\phi}}_t^T, \vec{\mathbf{y}}_1^T, \ldots, \vec{\mathbf{y}}_{j-1}^T, \vec{\mathbf{y}}_{j+1}^T, \ldots, \vec{\mathbf{y}}_r^T).$$

We may thus consider the single-input, single-output channel between the i^{th} transmit antenna and j^{th} receive antenna:

$$\vec{\mathbf{y}}_j^T = \mathbf{h}_{ij} \|\vec{\mathbf{x}}_i^T\| + \vec{\mathbf{w}}_j^T.$$

Hence,

$$\begin{aligned} & I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A}) \\ & \quad \leq rtI(\vec{\mathbf{x}}_i^T; \vec{\mathbf{y}}_j^T | \|\vec{\mathbf{x}}_i^T\|) \\ & \quad = rtE\left[I(\vec{\mathbf{x}}_i^T; \vec{\mathbf{y}}_j^T | \|\vec{x}_i^T\|)\right]. \end{aligned} \tag{A.8}$$

Since $I(\vec{\mathbf{x}}_i^T; \mathbf{y}_j^T | \|\vec{x}_i^T\|)$ is the mutual information of a single-input, single-output channel over l channel uses, it has a power constraint of $\frac{\|\vec{x}_i^T\|}{l}$. This mutual information can

be upper bounded by the capacity of AWGN channel with the same power constraint, i.e.

$$I(\vec{\mathbf{x}}_i^T; \vec{\mathbf{y}}_j^T | \|\vec{x}_i^T\|) \leq l \log\left(1 + \frac{\|\vec{x}_i^T\|}{l}\right). \tag{A.9}$$

Combining (A.8) with (A.9), we obtain

$$I(\mathbf{\Phi}; \mathbf{Y}|\mathbf{A}) \leq rtlE\left[\log\left(1 + \frac{\|\vec{\mathbf{x}}_i^T\|}{l}\right)\right]. \tag{A.10}$$

From (A.3), (A.7) and (A.10), we obtain our upper bound to $I(\mathbf{X}; \mathbf{Y})$ as

$$\frac{1}{l} I(\mathbf{X}; \mathbf{Y}) \leq rtE\left[\log\left(1 + \frac{\|\vec{\mathbf{x}}_i^T\|}{l}\right)\right] + \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}). \tag{A.11}$$

Combining (A.11) with (A.1) and noting that all the input vectors have identically distributed norms, we see that the input distribution satisfying (A.1) satisfies the second constraint (2.2) also. This completes the proof of the lemma. □

Appendix B

Proof of Theorem 1

For any $\alpha \in (0,1]$ and $\gamma \in (0,\alpha)$, let

$$C(\mathsf{SNR}) \geq C^*(\mathsf{SNR}) = r\mathsf{SNR} - \frac{r(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}).$$

This implies that there exists a probability distribution on $\mathbf{X}$ such that

$$I(\mathbf{X};\mathbf{Y}) \geq C^*(\mathsf{SNR}).$$

From Lemma 1, we know that this distribution must satisfy the following constraints for all $i \in \{1,\ldots,t\}$:

$$\frac{t}{l}E\Big[\log(1+\|\vec{\mathbf{x}}_i^T\|^2)\Big] \leq \frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}), \tag{B.1}$$

$$tE\Big[\log\Big(1+\frac{\|\vec{\mathbf{x}}_i^T\|^2}{l}\Big)\Big] \geq \mathsf{SNR} - \frac{(r+t)}{t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}). \tag{B.2}$$

Using these constraints, we establish a necessary condition on the coherence length. As the norms of the transmitted signals are identically distributed, it suffices to consider only one of them. Therefore, we shall omit the subscript, i, and define random variable $\mathbf{b}$ as

$$\mathbf{b} \triangleq \frac{t\|\vec{\mathbf{x}}^T\|^2}{l\mathsf{SNR}}.$$

The two constraints become

$$\frac{t}{l}E\Big[\log\Big(1+\frac{\mathbf{b}l\mathsf{SNR}}{t}\Big)\Big] \leq \frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}), \tag{B.3}$$

$$tE\Big[\log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big] \geq \mathsf{SNR} - \frac{(r+t)}{t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}). \tag{B.4}$$

Moreover, as

$$E[\|\vec{\mathbf{x}}^T\|^2] = \frac{l\mathsf{SNR}}{t}, \tag{B.5}$$

$$\Rightarrow E[\mathbf{b}] = 1. \tag{B.6}$$

Note that (B.4, B.6) do not depend on the coherence length, l, whereas (B.3) does. Also, the left hand side of (B.3) is a monotonically decreasing function of l. Thus, to find how large the coherence length must be, we need to find the distribution on $\mathbf{b}$ that minimizes the left hand side of (B.3) subject to the constraints (B.4, B.6). Using this distribution for $\mathbf{b}$, we can obtain the necessary condition on the coherence length from (B.3).

For any $\beta > 0$, we can express (B.4) as

$$\begin{aligned}
&\frac{\mathsf{SNR}}{t} - \frac{(r+t)}{t^2}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}) \\
&\quad \le E\Big[\log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big] \\
&\quad = \Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta})E\Big[\log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} \ge t\mathsf{SNR}^{-\beta}\Big] \\
&\qquad\qquad + \Pr(\mathbf{b} < t\mathsf{SNR}^{-\beta})E\Big[\log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} < t\mathsf{SNR}^{-\beta}\Big] \\
&\quad \le \Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta})E\Big[\log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} \ge t\mathsf{SNR}^{-\beta}\Big] \\
&\qquad\qquad + \Pr(\mathbf{b} < t\mathsf{SNR}^{-\beta})E\Big[\frac{\mathbf{b}\mathsf{SNR}}{t}\Big|\mathbf{b} < t\mathsf{SNR}^{-\beta}\Big] \\
&\quad = \frac{\mathsf{SNR}}{t} - \Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta})E\Big[\frac{\mathbf{b}\mathsf{SNR}}{t} - \log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} \ge t\mathsf{SNR}^{-\beta}\Big].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta})E\Big[\frac{\mathbf{b}\mathsf{SNR}}{t} - \log\Big(1+\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} \ge t\mathsf{SNR}^{-\beta}\Big] \\
\le \frac{(r+t)}{t^2}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}).
\end{aligned} \tag{B.7}$$

When $\beta \ge 1$, $\mathbf{b} \ge t\mathsf{SNR}^{-\beta}$ implies $\frac{\mathbf{b}\mathsf{SNR}}{t} \gg 1$, which makes $\frac{\mathbf{b}\mathsf{SNR}}{t} \gg \log(1+\frac{\mathbf{b}\mathsf{SNR}}{t})$ (since $\mathsf{SNR} \ll 1$). Hence, $\forall \beta \ge 1$

$$\Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta})E[\mathbf{b}|\mathbf{b} \ge t\mathsf{SNR}^{-\beta}] \le \frac{(r+t)}{t}\mathsf{SNR}^{\alpha} + O(\mathsf{SNR}^{\alpha+\gamma}) = o(1). \tag{B.8}$$

From Markov's inequality, $\forall \beta \ge 1$

$$\Pr(\mathbf{b} \ge t\mathsf{SNR}^{-\beta}) \le \frac{\mathsf{SNR}^{\beta}}{t} = o(1). \tag{B.9}$$

When $\beta < 1$, $\mathbf{b} < t\mathsf{SNR}^{-\beta}$ implies $\frac{\mathbf{b}\mathsf{SNR}}{t} \ll 1$. Hence, (B.7) can be expressed as

$$\begin{aligned}
&\frac{(r+t)}{t^2}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}) \\
&\quad \geq \Pr(\mathbf{b} \geq t\mathsf{SNR}^{-\beta}) E\Big[\frac{\mathbf{b}\mathsf{SNR}}{t} - \log\Big(1 + \frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|\mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big] \\
&\quad \geq \Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big) E\Big[\frac{\mathbf{b}\mathsf{SNR}}{t} - \log\Big(1 + \frac{\mathbf{b}\mathsf{SNR}}{t}\Big)\Big|t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big] \\
&\quad \geq \Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big) E\Big[\frac{1}{2}\Big(\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)^2 - \frac{1}{3}\Big(\frac{\mathbf{b}\mathsf{SNR}}{t}\Big)^3\Big|t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big] \\
&\quad \geq \Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big)\Big[\frac{1}{2}\mathsf{SNR}^{2(1-\beta)} - \frac{1}{3}\mathsf{SNR}^{3(1-\beta)}\Big].
\end{aligned}$$

Thus, $\forall \beta \in (0,1)$

$$\Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big) \leq \frac{2(r+t)}{t^2}\frac{\mathsf{SNR}^{2\beta-(1-\alpha)}}{1-\frac{2}{3}\mathsf{SNR}^{1-\beta}} + o\Big(\frac{\mathsf{SNR}^{2\beta-(1-\alpha)}}{1-\frac{2}{3}\mathsf{SNR}^{1-\beta}}\Big). \quad \text{(B.10)}$$

Let us divide the interval $[t\mathsf{SNR}^{-\beta}, t\mathsf{SNR}^{-1}]$, $\beta \in (0,1)$, into $K > 1$ finite intervals so that each interval is of length

$$\varepsilon = \frac{t(\mathsf{SNR}^{-1} - \mathsf{SNR}^{-\beta})}{K}.$$

Now, for any $\varepsilon > 0$

$$\begin{aligned}
&E\Big[\mathbf{b}|t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big]\Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big) \\
&\quad = \sum_{i=1}^{K} E\Big[\mathbf{b}|t\mathsf{SNR}^{-(\beta+i\varepsilon)} \geq \mathbf{b} \geq t\mathsf{SNR}^{-[\beta+(i-1)\varepsilon]}\Big]\Pr\Big(t\mathsf{SNR}^{-(\beta+i\varepsilon)} \geq \mathbf{b} \geq t\mathsf{SNR}^{-[\beta+(i-1)\varepsilon]}\Big) \\
&\quad \leq t\sum_{i=1}^{K} \mathsf{SNR}^{-(\beta+i\varepsilon)}\Pr\Big(t\mathsf{SNR}^{-(\beta+i\varepsilon)} \geq \mathbf{b} \geq t\mathsf{SNR}^{-[\beta+(i-1)\varepsilon]}\Big) \\
&\quad \leq t\sum_{i=1}^{K} \mathsf{SNR}^{-(\beta+i\varepsilon)}\Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-[\beta+(i-1)\varepsilon]}\Big) \\
&\quad \leq \frac{2(r+t)}{t}\sum_{i=1}^{K} \mathsf{SNR}^{-(\beta+i\varepsilon)}\Big[\frac{\mathsf{SNR}^{[2(\beta+(i-1)\varepsilon)-(1-\alpha)]}}{1-\frac{2}{3}\mathsf{SNR}^{[1-(\beta+(i-1)\varepsilon)]}} + o\Big(\frac{\mathsf{SNR}^{[2(\beta+(i-1)\varepsilon)-(1-\alpha)]}}{1-\frac{2}{3}\mathsf{SNR}^{[1-(\beta+(i-1)\varepsilon)]}}\Big)\Big] \quad \text{(B.11)} \\
&\quad = \frac{2(r+t)}{t}\sum_{i=1}^{K}\Big[\frac{\mathsf{SNR}^{[\beta-(1-\alpha)+(i-2)\varepsilon]}}{1-\frac{2}{3}\mathsf{SNR}^{[1-(\beta+(i-1)\varepsilon)]}} + o\Big(\frac{\mathsf{SNR}^{[\beta-(1-\alpha)+(i-2)\varepsilon]}}{1-\frac{2}{3}\mathsf{SNR}^{[1-(\beta+(i-1)\varepsilon)]}}\Big)\Big].
\end{aligned}$$

Equation (B.10) is used to obtain (B.11). Let $\beta = 1-\alpha+2\varepsilon$, where $\varepsilon \in (0, \frac{\alpha}{2})$. Then,

$$\begin{aligned}
&E[\mathbf{b}|t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}]\Pr(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}) \\
&\qquad \leq \frac{2(r+t)}{t}\sum_{i=1}^{K}\Big[\frac{\mathsf{SNR}^{i\varepsilon}}{1-\frac{2}{3}\mathsf{SNR}^{\alpha-(i+1)\varepsilon}} + o\Big(\frac{\mathsf{SNR}^{i\varepsilon}}{1-\frac{2}{3}\mathsf{SNR}^{\alpha-(i+1)\varepsilon}}\Big)\Big].
\end{aligned}$$

Since,

$$\frac{\mathsf{SNR}^{i\varepsilon}}{1-\frac{2}{3}\mathsf{SNR}^{\alpha-(i+1)\varepsilon}} = o(1) \quad \forall i \in \{1,\ldots,K\},$$

we have

$$E\Big[\mathbf{b}|t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-(1-\alpha+2\varepsilon)}\Big] \Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-(1-\alpha+2\varepsilon)}\Big) \leq o(1). \tag{B.12}$$

From (B.10), we have for $\beta = 1-\alpha+2\varepsilon$, where $\varepsilon \in (0, \frac{\alpha}{2})$,

$$\Pr\Big(t\mathsf{SNR}^{-1} \geq \mathbf{b} \geq t\mathsf{SNR}^{-\beta}\Big) \leq \frac{2(r+t)}{t^2}\frac{\mathsf{SNR}^{1-\alpha+4\varepsilon}}{1-\frac{2}{3}\mathsf{SNR}^{\alpha-2\varepsilon}} + o\Big(\frac{\mathsf{SNR}^{1-\alpha+4\varepsilon}}{1-\frac{2}{3}\mathsf{SNR}^{\alpha-2\varepsilon}}\Big) = o(1). \tag{B.13}$$

From (B.8, B.9, B.12, B.13), we know that for $0 < \epsilon < \alpha$

$$\Pr(\mathbf{b} \geq t\mathsf{SNR}^{-(1-\alpha+\epsilon)}) = o(1),$$
$$\Pr(\mathbf{b} \geq t\mathsf{SNR}^{-(1-\alpha+\epsilon)})E[\mathbf{b}|\mathbf{b} \geq t\mathsf{SNR}^{-(1-\alpha+\epsilon)}] = o(1),$$

which implies

$$\Pr(\mathbf{b} \leq t\mathsf{SNR}^{-(1-\alpha+\epsilon)}) = O(1),$$
$$\Pr(\mathbf{b} \leq t\mathsf{SNR}^{-(1-\alpha+\epsilon)})E[\mathbf{b}|\mathbf{b} \leq t\mathsf{SNR}^{-(1-\alpha+\epsilon)}] = O(1).$$

Hence, the distribution on $\mathbf{b}$ that minimizes $E\Big[\log\Big(1+\frac{\mathbf{b}l\mathsf{SNR}}{t}\Big)\Big]$, subject to the constraints (B.4, B.6), is the on-off distribution:

$$\mathbf{b} = \begin{cases} t\mathsf{SNR}^{-(1-\alpha+\epsilon)} & w.p.\ \ \eta \\ 0 & w.p.\ \ 1-\eta \end{cases}$$

where

$$\eta = \frac{\mathsf{SNR}^{1-\alpha+\epsilon}}{t}.$$

Hence, with this on-off distribution on $\mathbf{b}$, (B.3) becomes

$$\frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha} + O(\mathsf{SNR}^{1+\alpha+\gamma}) \geq \frac{\mathsf{SNR}^{1-\alpha+\epsilon}}{l}\log\Big(1+l\mathsf{SNR}^{\alpha-\epsilon}\Big). \tag{B.14}$$

Now,

$$\begin{aligned}
&\frac{\mathsf{SNR}^{1-\alpha+\epsilon}}{l}\log\left(1+l\mathsf{SNR}^{\alpha-\epsilon}\right)\Big|_{l=\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha+\epsilon}} \\
&\quad=\frac{(r+t)^2}{t^2}\mathsf{SNR}^{1+\alpha}\log\left(1+\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-\alpha}\right) \\
&\quad\gg\frac{(r+t)}{2t}\mathsf{SNR}^{1+\alpha}+O(\mathsf{SNR}^{1+\alpha+\gamma}).
\end{aligned}$$

Thus, with $l=\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha+\epsilon}$, (B.14) is not satisfied. However, the right hand side of (B.14) is a monotonically decreasing function of l. Hence, for the constraint in (B.14) to be met

$$l>\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha+\epsilon}\quad\forall\epsilon\in(0,\alpha).$$

Thus, we see that if an input distribution satisfies (B.1, B.2, B.5), then the coherence length must necessarily obey

$$l>\frac{t^2}{(r+t)^2}\mathsf{SNR}^{-2\alpha}\triangleq l_{\mathsf{min}}.$$

This completes the proof of the theorem. □

Appendix C

Capacity of an i.i.d Rayleigh fading MIMO channel

We compute the capacity of an i.i.d Rayleigh fading MIMO channel when CSI is unavailable at both the transmitter as well as the receiver. It is shown in [20] that increasing the number of transmit antennas beyond the coherence length does not increase capacity. Hence, from a capacity point of view, it suffices to use only one transmit antenna ($t = 1$). We will therefore consider the capacity of a single-input, multiple-output (SIMO) channel.

We will use on-off signaling to communicate over the channel. This signaling scheme is later proved to be optimal for the i.i.d Rayleigh fading MIMO channel. We specify the signaling as

$$\mathbf{x} = \begin{cases} \sqrt{A} & w.p.\ \ \omega \\ 0 & w.p.\ \ 1-\omega \end{cases} \tag{C.1}$$

where, $A \in \Re^+$ and $\omega = \frac{\mathsf{SNR}}{A}$. With this signaling, we have the following probability distributions

$$\begin{aligned} p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) &= \frac{1}{\pi^r} \exp(-\|\vec{y}\|^2), \\ p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y}) &= \frac{1}{\pi(1+A)^r} \exp\Big(-\frac{\|\vec{y}\|^2}{1+A}\Big). \end{aligned}$$

The mutual information $I(\mathbf{x};\vec{\mathbf{y}})$ can be written as $I(\mathbf{x};\vec{\mathbf{y}}) = h(\vec{\mathbf{y}}) - h(\vec{\mathbf{y}}|\mathbf{x})$. Now,

$$
\begin{aligned}
&h(\vec{\mathbf{y}}) \\
&\quad = -\int p_{\vec{\mathbf{y}}}(\vec{y}) \log(p_{\vec{\mathbf{y}}}(\vec{y}))d\vec{y} \\
&\quad = -\int \left[(1-\omega)p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) + \omega p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})\right] \log \left[(1-\omega)p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) + \omega p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})\right] d\vec{y} \\
&\quad = -(1-\omega)\int p_{\vec{\mathbf{y}}|\mathbf{x}=0} \log \left[(1-\omega)p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) \left(1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right)\right] d\vec{y} \\
&\quad -\omega \int p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}} \log \left[(1-\omega)p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) \left(1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right)\right] d\vec{y} \\
&\quad = -\log(1-\omega) - (1-\omega)\int p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) \log \left[1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right] d\vec{y} \\
&\quad -\omega \int p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y}) \log \left[1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right] d\vec{y} + h(\vec{\mathbf{y}}|\mathbf{x}) + \omega D(p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}||p_{\vec{\mathbf{y}}|\mathbf{x}=0}).
\end{aligned}
$$

The divergence $D(p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}||p_{\vec{\mathbf{y}}|\mathbf{x}=0})$ is the divergence between two Gaussian random vectors and is therefore

$$D(p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}||p_{\vec{\mathbf{y}}|\mathbf{x}=0}) = r(A - \log(1+A)).$$

The expression for the mutual information becomes

$$I(\mathbf{x};\vec{\mathbf{y}}) = r\mathsf{SNR} - r\mathsf{SNR}\frac{\log(1+A)}{A} - \log\left(1 - \frac{\mathsf{SNR}}{A}\right) - I(\mathsf{SNR}, A),$$

where,

$$I(\mathsf{SNR}, A) \;=\; I_1(\mathsf{SNR}, A) + I_2(\mathsf{SNR}, A), \tag{C.2}$$

$$
\begin{aligned}
I_1(\mathsf{SNR}, A) &= (1-\omega)\int p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y}) \log \left[1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right] d\vec{y}, \\
I_2(\mathsf{SNR}, A) &= \omega \int p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y}) \log \left[1 + \frac{\omega}{1-\omega}\frac{p_{\vec{\mathbf{y}}|\mathbf{x}=\sqrt{A}}(\vec{y})}{p_{\vec{\mathbf{y}}|\mathbf{x}=0}(\vec{y})}\right] d\vec{y}.
\end{aligned}
$$

At low SNR, A takes very high values. Therefore, the mutual information can be written as

$$I(\mathbf{x};\vec{\mathbf{y}}) = r\mathsf{SNR} - r\mathsf{SNR}\frac{\log(1+A)}{A} - I(\mathsf{SNR}, A) \tag{C.3}$$

We will now compute $I_1(\mathsf{SNR}, A)$ and $I_2(\mathsf{SNR}, A)$. Let us define ς^* to be such that

$$\frac{\mathsf{SNR}}{A(1+A)^r} \exp\left(\frac{A\varsigma^*}{1+A}\right) = 1.$$

Note that

$$\frac{\varsigma^*}{1+A} = \frac{\log(A)}{A} + r\frac{\log(1+A)}{A} + \frac{\log(\frac{1}{\mathsf{SNR}})}{A}.$$

Thus,

$$\lim_{A\to\infty} \frac{\varsigma^*}{1+A} = 0.$$

We shall use this in future derivations.

Computing $I_1(\mathsf{SNR}, A)$:

$$I_1(\mathsf{SNR}, A) = \frac{A - \mathsf{SNR}}{A\ \pi^r} \int \exp(-\|\vec{y}\|^2) \log\left[1 + \frac{\mathsf{SNR}}{(A-\mathsf{SNR})(1+A)^r} \exp\left(\frac{A}{1+A}\|\vec{y}\|^2\right)\right] d\vec{y}.$$

Converting to spherical coordinates in $2r$ dimensions, we have for large A,

$$\begin{aligned}
&I_1(\mathsf{SNR}, A) \\
&\quad= \frac{1}{(r-1)!} \int_0^\infty \varsigma^{r-1} \exp(-\varsigma) \log\left[1 + \frac{\mathsf{SNR}}{A(1+A)^r} \exp\left(\frac{A\varsigma}{1+A}\right)\right] d\varsigma + o(\mathsf{SNR}^2) \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!} \int_0^\infty \varsigma^{r-1} \exp(-(\varsigma-\varsigma^*)) \log\left[1 + \exp\left(\frac{A(\varsigma-\varsigma^*)}{1+A}\right)\right] d\varsigma + o(\mathsf{SNR}^2) \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!} \int_{-\varsigma^*}^\infty (\varsigma+\varsigma^*)^{r-1} \exp(-\varsigma) \log\left[1 + \exp\left(\frac{A\varsigma}{1+A}\right)\right] d\varsigma + o(\mathsf{SNR}^2) \\
&\quad= I_{11}(\mathsf{SNR}, A) + I_{12}(\mathsf{SNR}, A) + o(\mathsf{SNR}^2), \qquad \text{(C.4)}
\end{aligned}$$

where,

$$\begin{aligned}
I_{11}(\mathsf{SNR}, A) &= \frac{\exp(-\varsigma^*)}{(r-1)!} \int_{-\varsigma^*}^0 (\varsigma+\varsigma^*)^{r-1} \exp(-\varsigma) \log\left[1 + \exp\left(\frac{A\varsigma}{1+A}\right)\right] d\varsigma, \\
I_{12}(\mathsf{SNR}, A) &= \frac{\exp(-\varsigma^*)}{(r-1)!} \int_0^\infty (\varsigma+\varsigma^*)^{r-1} \exp(-\varsigma) \log\left[1 + \exp\left(\frac{A\varsigma}{1+A}\right)\right] d\varsigma.
\end{aligned}$$

Computing $I_{11}(\mathsf{SNR}, A)$:

$$
\begin{aligned}
&I_{11}(\mathsf{SNR}, A) \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\log\left[1+\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma \\
&\quad\leq \frac{\exp(-\varsigma^*)}{(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\exp\left(\frac{A\varsigma}{1+A}\right)d\varsigma \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp\left(-\frac{\varsigma}{1+A}\right)d\varsigma \\
&\quad= \frac{(1+A)^r}{(r-1)!}\exp\left(-\frac{A\varsigma^*}{1+A}\right)\int_{0}^{\frac{\varsigma^*}{1+A}}\varsigma^{r-1}\exp(-\varsigma)d\varsigma \\
&\quad= \frac{(1+A)^r}{(r-1)!}\exp\left(-\frac{A\varsigma^*}{1+A}\right)\left[\Gamma(r)-\Gamma\left(r,\frac{\varsigma^*}{1+A}\right)\right] \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A),
\end{aligned}
$$

where,

$$
I_{11}^{U}(\mathsf{SNR}, A) = \frac{\mathsf{SNR}}{A} - (1+A)^r\left[\frac{\mathsf{SNR}}{A(1+A)^r}\right]^{1+\frac{1}{A}}\left[\sum_{j=0}^{r-1}\frac{[\frac{\varsigma^*}{1+A}]^j}{j!}\right].
$$

Also,

$$
\begin{aligned}
&I_{11}(\mathsf{SNR}, A) \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\log\left[1+\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma \\
&\quad\geq I_{11}^{U}(\mathsf{SNR}, A) - \frac{\exp(-\varsigma^*)}{2(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\exp\left(\frac{2A\varsigma}{1+A}\right)d\varsigma \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{\exp(-\varsigma^*)}{2(r-1)!}\int_{-\varsigma^*}^{0}(\varsigma+\varsigma^*)^{r-1}\exp\left(\frac{A-1}{A+1}\,\varsigma\right)d\varsigma \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{\exp(-\varsigma^*)}{2(r-1)!}\int_{0}^{\varsigma^*}\varsigma^{r-1}\exp\left(\frac{A-1}{A+1}\,(\varsigma-\varsigma^*)\right)d\varsigma \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{1}{2(r-1)!}\exp\Big(-\frac{2A\varsigma^*}{A+1}\Big)\Big[\frac{A+1}{A-1}\Big]^r\int_{0}^{(\frac{A-1}{A+1})\varsigma^*}\varsigma^{r-1}\exp(\varsigma)d\varsigma \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{(-1)^{r-1}}{2(r-1)!}\exp\Big(-\frac{2A\varsigma^*}{A+1}\Big)\Big[\frac{A+1}{A-1}\Big]^r\left[\Gamma(r,-\frac{A-1}{A+1}\varsigma^*)-\Gamma(r)\right] \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{(-1)^{r-1}}{2}\Big[\frac{A+1}{A-1}\Big]^r\left[\exp(-\varsigma^*)\sum_{j=0}^{r-1}\frac{[-\frac{A-1}{A+1}\varsigma^*]^j}{j!} - \exp\Big(-\frac{2A\varsigma^*}{A+1}\Big)\right] \\
&\quad= I_{11}^{U}(\mathsf{SNR}, A) - \frac{(-1)^{r-1}}{2}\Big[\frac{A+1}{A-1}\Big]^r\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{1+A}}\left[\sum_{j=0}^{r-1}\frac{[-\frac{A-1}{A+1}\varsigma^*]^j}{j!}\right] + o(\mathsf{SNR}^2).
\end{aligned}
$$

Let

$$I_{11}^{L}(\mathsf{SNR},A)=\frac{(-1)^{r-1}}{2}\Big[\frac{A+1}{A-1}\Big]^{r}\Big[\frac{\mathsf{SNR}}{A(1+A)^{r}}\Big]^{1+\frac{1}{1+A}}\left[\sum_{j=0}^{r-1}\frac{\big[-\frac{A-1}{A+1}\varsigma^{*}\big]^{j}}{j!}\right].$$

Thus, we have

$$I_{11}^{U}(\mathsf{SNR},A)-I_{11}^{L}(\mathsf{SNR},A)+o(\mathsf{SNR}^{2})\leq I_{11}(\mathsf{SNR},A)\leq I_{11}^{U}(\mathsf{SNR},A).$$

Since $\frac{\varsigma^{*}}{1+A}\to 0$ as $A\to\infty$, we have

$$\begin{aligned}\lim_{A\to\infty}I_{11}^{U}(\mathsf{SNR},A)&=0,\\ \lim_{A\to\infty}I_{11}^{L}(\mathsf{SNR},A)&=0,\\ \Rightarrow\lim_{A\to\infty}I_{11}(\mathsf{SNR},A)&=0.\end{aligned}\tag{C.5}$$

Computing $I_{12}(\mathsf{SNR},A)$:

$$\begin{aligned}&I_{12}(\mathsf{SNR},A)\\ &\quad=\frac{\exp(-\varsigma^{*})}{(r-1)!}\int_{0}^{\infty}(\varsigma+\varsigma^{*})^{r-1}\exp(-\varsigma)\log\left[1+\exp\Big(\frac{A\varsigma}{1+A}\Big)\right]d\varsigma\\ &\quad=\frac{\exp(-\varsigma^{*})}{(r-1)!}\int_{0}^{\infty}(\varsigma+\varsigma^{*})^{r-1}\exp(-\varsigma)\left[\frac{A\varsigma}{1+A}+\log\left[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\right]\right]d\varsigma\\ &\quad=I_{12}^{1}(\mathsf{SNR},A)+I_{12}^{2}(\mathsf{SNR},A),\end{aligned}\tag{C.6}$$

where,

$$\begin{aligned}I_{12}^{1}(\mathsf{SNR},A)&=\frac{\exp(-\varsigma^{*})}{(r-1)!}\Big[\frac{A}{1+A}\Big]\int_{0}^{\infty}\varsigma(\varsigma+\varsigma^{*})^{r-1}\exp(-\varsigma)d\varsigma,\\ I_{12}^{2}(\mathsf{SNR},A)&=\frac{\exp(-\varsigma^{*})}{(r-1)!}\int_{0}^{\infty}(\varsigma+\varsigma^{*})^{r-1}\exp(-\varsigma)\log\left[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\right]d\varsigma.\end{aligned}$$

Now,

$$\begin{aligned}&I_{12}^{1}(\mathsf{SNR},A)\\ &\quad=\frac{\exp(-\varsigma^{*})}{(r-1)!}\Big[\frac{A}{1+A}\Big]\int_{0}^{\infty}\varsigma(\varsigma+\varsigma^{*})^{r-1}\exp(-\varsigma)d\varsigma,\\ &\quad=\frac{1}{(r-1)!}\Big[\frac{A}{1+A}\Big]\int_{\varsigma^{*}}^{\infty}(\varsigma-\varsigma^{*})\varsigma^{r-1}\exp(-\varsigma)d\varsigma,\\ &\quad=\frac{1}{(r-1)!}\Big[\frac{A}{1+A}\Big]\Big[\Gamma(r+1,\varsigma^{*})-\varsigma^{*}\Gamma(r,\varsigma^{*})\Big]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(r-1)!}\Big[\frac{A}{1+A}\Big]\Big[\Gamma(r,\varsigma^*)(r-\varsigma^*) + (\varsigma^*)^r \exp(-\varsigma^*)\Big] \\
&= \exp(-\varsigma^*)\Big[\frac{A}{1+A}\Big]\Big[\frac{r-\varsigma^*}{(r-1)!}\Big[\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\Big] + \frac{(\varsigma^*)^r}{(r-1)!}\Big] \\
&= \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\Big[\frac{A}{1+A}\Big]\Big[\frac{r-\varsigma^*}{(r-1)!}\Big[\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\Big] + \frac{(\varsigma^*)^r}{(r-1)!}\Big] \\
&= I_{12}^{1A}(\mathsf{SNR}, A)\ \mathsf{SNR}^{1+\frac{1}{A}},
\end{aligned}
$$

where,

$$
I_{12}^{1A}(\mathsf{SNR}, A) = \Big[\frac{1}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\Big[\frac{A}{1+A}\Big]\Big[\frac{r-\varsigma^*}{(r-1)!}\Big[\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\Big] + \frac{(\varsigma^*)^r}{(r-1)!}\Big].
$$

Since $\frac{\varsigma^*}{1+A} \to 0$ as $A \to \infty$, we have

$$
\lim_{A\to\infty} I_{12}^{1A}(\mathsf{SNR}, A) = 0.
$$

Thus,

$$
\lim_{A\to\infty} I_{12}^{1}(\mathsf{SNR}, A) = 0. \tag{C.7}
$$

We now compute $I_{12}^{2}(\mathsf{SNR}, A)$.

$$
\begin{aligned}
&I_{12}^{2}(\mathsf{SNR}, A) \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\log\Big[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\Big]\,d\varsigma \\
&\quad\le \frac{\exp(-\varsigma^*)}{(r-1)!}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\exp\Big(-\frac{A\varsigma}{1+A}\Big)d\varsigma \\
&\quad= \frac{\exp(-\varsigma^*)}{(r-1)!}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\left(-\Big(\frac{1+2A}{1+A}\Big)\varsigma\right)d\varsigma \\
&\quad= \frac{1}{(r-1)!}\Big[\frac{1+A}{1+2A}\Big]^r \exp\Big(\frac{A\varsigma}{1+A}\Big)\int_{(\frac{1+2A}{1+A})\varsigma^*}^\infty \varsigma^{r-1}\exp(-\varsigma)d\varsigma \\
&\quad= \frac{1}{(r-1)!}\Big[\frac{1+A}{1+2A}\Big]^r \exp\Big(\frac{A\varsigma}{1+A}\Big)\Gamma\left(r, (\frac{1+2A}{1+A})\varsigma^*\right) \\
&\quad= \Big[\frac{1+A}{1+2A}\Big]^r \exp(-\varsigma^*)\sum_{j=0}^{r-1}\frac{[(\frac{1+2A}{1+A})\varsigma^*]^j}{j!} \\
&\quad= I_{12}^{2U}(\mathsf{SNR}, A),
\end{aligned}
$$

where,

$$I_{12}^{2U}(\mathsf{SNR},A)=\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\Big[\frac{1+A}{1+2A}\Big]^r\left[\sum_{j=0}^{r-1}\frac{[(\frac{1+2A}{1+A})\varsigma^*]^j}{j!}\right].$$

Also,

$$\begin{aligned}
&I_{12}^{2}(\mathsf{SNR},A)\\
&\quad\geq I_{12}^{2U}(\mathsf{SNR},A)-\frac{\exp(-\varsigma^*)}{2(r-1)!}\int_0^\infty(\varsigma+\varsigma^*)^{r-1}\exp(-\varsigma)\exp\Big(-\frac{2A\varsigma}{1+A}\Big)d\varsigma\\
&\quad= I_{12}^{2U}(\mathsf{SNR},A)-\frac{1}{2(r-1)!}\exp\Big(\frac{2A\varsigma}{1+A}\Big)\int_{\varsigma^*}^\infty\varsigma^{r-1}\exp\left(-\Big(\frac{1+3A}{1+A}\Big)\varsigma\right)d\varsigma\\
&\quad= I_{12}^{2U}(\mathsf{SNR},A)-\frac{1}{2(r-1)!}\Big[\frac{1+A}{1+3A}\Big]^r\exp\Big(\frac{2A\varsigma}{1+A}\Big)\Gamma\left(r,(\frac{1+3A}{1+A})\varsigma^*\right)\\
&\quad= I_{12}^{2U}(\mathsf{SNR},A)-\frac{\exp(-\varsigma^*)}{2}\Big[\frac{1+A}{1+3A}\Big]^r\left[\sum_{j=0}^{r-1}\frac{[(\frac{1+3A}{1+A})\varsigma^*]^j}{j!}\right]\\
&\quad= I_{12}^{2U}(\mathsf{SNR},A)-\frac{1}{2}I_{12}^{2L}(\mathsf{SNR},A),
\end{aligned}$$

where,

$$I_{12}^{2L}(\mathsf{SNR},A)=\Big[\frac{\mathsf{SNR}}{A(1+A)^r)}\Big]^{1+\frac{1}{A}}\Big[\frac{1+A}{1+3A}\Big]^r\left[\sum_{j=0}^{r-1}\frac{[(\frac{1+3A}{1+A})\varsigma^*]^j}{j!}\right].$$

Since $\frac{\varsigma*}{1+A}\to 0$ as $A\to\infty$,

$$\begin{aligned}
\lim_{A\to\infty} I_{12}^{2L}(\mathsf{SNR},A)&=0,\\
\lim_{A\to\infty} I_{12}^{2U}(\mathsf{SNR},A)&=0,\\
\Rightarrow\lim_{A\to\infty} I_{12}^{2}(\mathsf{SNR},A)&=0.
\end{aligned}\tag{C.8}$$

Substituting (C.7) and (C.8) in (C.6), we have

$$\lim_{A\to\infty} I_{12}(\mathsf{SNR},A)=0.\tag{C.9}$$

Substituting (C.5) and (C.9) in (C.4), we obtain

$$I_1(\mathsf{SNR},A)=o(\mathsf{SNR}^2).\tag{C.10}$$

Computing $I_2(\mathsf{SNR},A)$:

$$I_2(\mathsf{SNR},A)=\frac{\mathsf{SNR}}{\pi^r A(1+A)^r}\int\exp\left(-\frac{\|\vec{y}\|^2}{1+A}\right)\log\left[1+\frac{\mathsf{SNR}}{(A-\mathsf{SNR})(1+A)^r}\exp\left(\frac{A}{1+A}\|\vec{y}\|^2\right)\right]d\vec{y}.$$

Converting to spherical coordinates in $2r$ dimensions, we have for large A,

$$
\begin{aligned}
&I_2(\mathsf{SNR}, A) \\
&\quad = \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\infty} \varsigma^{r-1}\exp\left(-\frac{\varsigma}{1+A}\right)\log\left[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma \\
&\quad = I_{21}(\mathsf{SNR}, A) + I_{22}(\mathsf{SNR}, A) + o(\mathsf{SNR}^2), \qquad \text{(C.11)}
\end{aligned}
$$

where,

$$
\begin{aligned}
I_{21}(\mathsf{SNR}, A) &= \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\left(-\frac{\varsigma}{1+A}\right)\log\left[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma, \\
I_{22}(\mathsf{SNR}, A) &= \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_{\varsigma^*}^{\infty} \varsigma^{r-1}\exp\left(-\frac{\varsigma}{1+A}\right)\log\left[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma.
\end{aligned}
$$

Computing $I_{21}(\mathsf{SNR}, A)$:

$$
\begin{aligned}
&I_{21}(\mathsf{SNR}, A) \\
&\quad = \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\left(-\frac{\varsigma}{1+A}\right)\log\left[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\left(\frac{A\varsigma}{1+A}\right)\right]d\varsigma \\
&\quad \le \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\frac{\mathsf{SNR}}{A(1+A)^r}\exp\Big(\frac{A\varsigma}{1+A}\Big)d\varsigma \\
&\quad = \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^2\frac{1}{(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\Big(\frac{A-1}{A+1}\,\varsigma\Big)d\varsigma \\
&\quad = \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^2\Big[\frac{A+1}{A-1}\Big]^r\frac{1}{(r-1)!}\int_0^{(\frac{A-1}{A+1})\varsigma^*} \varsigma^{r-1}\exp(\varsigma)d\varsigma \\
&\quad = \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^2\Big[\frac{A+1}{A-1}\Big]^r\frac{(-1)^{r-1}}{(r-1)!}\left[\Gamma\left(r, -\Big(\frac{A-1}{A+1}\Big)\,\varsigma^*\right)-\Gamma(r)\right] \\
&\quad = \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^2\Big[\frac{A+1}{A-1}\Big]^r\frac{(-1)^{r-1}}{(r-1)!}\left[(r-1)!\exp\left(\frac{A-1}{A+1}\,\varsigma^*\right)\sum_{j=0}^{r-1}\frac{-(\frac{A-1}{A+1})\varsigma^*}{j!}-(r-1)!\right] \\
&\quad = \Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^2\Big[\frac{A+1}{A-1}\Big]^r\frac{(-1)^{r-1}}{(r-1)!}\left[(r-1)!\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{\frac{1}{A}-1}\sum_{j=0}^{r-1}\frac{-(\frac{A-1}{A+1})\varsigma^*}{j!}-(r-1)!\right] \\
&\quad = I_{21}^U(\mathsf{SNR}, A)\ \mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2),
\end{aligned}
$$

where,

$$
I_{21}^U(\mathsf{SNR}, A) = \frac{(-1)^r}{[A(1+A)^r]^{1+\frac{1}{A}}}\Big[\frac{A+1}{A-1}\Big]^r\sum_{j=0}^{r-1}\frac{[-(\frac{A-1}{A+1})\varsigma^*]^j}{j!}. \qquad \text{(C.12)}
$$

Now,

$$
\begin{aligned}
&I_{21}(\mathsf{SNR}, A) \\
&\quad = \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\log\left[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\Big(\frac{A\varsigma}{1+A}\Big)\right]d\varsigma \\
&\quad \geq \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\frac{\mathsf{SNR}}{A(1+A)^r}\exp\Big(\frac{A\varsigma}{1+A}\Big)d\varsigma \\
&\quad -\frac{\mathsf{SNR}}{2A(A+1)^r(r-1)!}\int_0^{\varsigma^*} \varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\exp\Big(\frac{A\varsigma}{1+A}\Big)\Big]^2 d\varsigma \\
&\quad = I_{21}^U(\mathsf{SNR}, A)\ \mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2) \\
&\qquad\qquad -\frac{1}{2}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^3\frac{1}{(r-1)!}\int_0^{\varsigma^*}\varsigma^{r-1}\exp\Big(\frac{2A-1}{1+A}\varsigma\Big)d\varsigma \\
&\quad = I_{21}^U(\mathsf{SNR}, A)\ \mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2) \\
&\qquad\qquad -\frac{1}{2}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^3\frac{(-1)^r}{(r-1)!}\Big[\frac{A+1}{2A-1}\Big]^r\left[\Gamma(r,-(\frac{2A-1}{1+A})\varsigma^*)-\Gamma(r)\right] \\
&\quad = I_{21}^U(\mathsf{SNR}, A)\ \mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2) \\
&\quad -\frac{1}{2}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^3\frac{(-1)^r}{(r-1)!}\Big[\frac{A+1}{2A-1}\Big]^r\left[(r-1)!\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{\frac{1}{A}-2}\sum_{j=0}^{r-1}\frac{[-(\frac{2A-1}{A+1})\varsigma^*]^j}{j!}-(r-1)!\right] \\
&\quad = \Big[I_{21}^U(\mathsf{SNR}, A)-\frac{1}{2}I_{21}^L(\mathsf{SNR}, A)\Big]\mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2),
\end{aligned}
$$

where,

$$
I_{21}^L(\mathsf{SNR}, A) = \frac{(-1)^r}{[A(1+A)^r]^{1+\frac{1}{A}}}\Big[\frac{A+1}{2A-1}\Big]^r\sum_{j=0}^{r-1}\frac{[-(\frac{2A-1}{A+1})\varsigma^*]^j}{j!}. \tag{C.13}
$$

Combining (C.12, C.13), we have

$$
\begin{aligned}
&\Big[I_{21}^U(\mathsf{SNR}, A)-\frac{1}{2}I_{21}^L(\mathsf{SNR}, A)\Big]\mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2) \\
&\qquad\qquad \leq I_{21}(\mathsf{SNR}, A) \ \leq\ \Big[I_{21}^U(\mathsf{SNR}, A)\Big]\mathsf{SNR}^{1+\frac{1}{A}} + o(\mathsf{SNR}^2).
\end{aligned}
$$

At low SNR, $\frac{\varsigma^*}{1+A} \to 0$ as $A \to \infty$. Thus, we have

$$
\begin{aligned}
\lim_{A\to\infty} I_{21}^U(\mathsf{SNR}, A) = 0, \\
\lim_{A\to\infty} I_{21}^L(\mathsf{SNR}, A) = 0.
\end{aligned}
$$

Therefore,

$$
I_{21}(\mathsf{SNR}, A) = o(\mathsf{SNR}^2). \tag{C.14}
$$

Computing $I_{22}(\mathsf{SNR}, A)$:

$$
\begin{aligned}
&I_{22}(\mathsf{SNR}, A) \\
&\quad = \frac{\mathsf{SNR}}{A(A+1)^r(r-1)!}\int_{\varsigma^*}^{\infty}\varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\log\Big[1+\frac{\mathsf{SNR}}{A(1+A)^r}\exp\Big(\frac{A\varsigma}{1+A}\Big)\Big]\,d\varsigma \\
&\quad = \frac{1}{(r-1)!}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\int_{\varsigma^*}^{\infty}\varsigma^{r-1}\exp\Big(-\frac{(\varsigma-\varsigma^*)}{1+A}\Big)\log\Big[1+\exp\Big(\frac{A(\varsigma-\varsigma^*)}{1+A}\Big)\Big]\,d\varsigma \\
&\quad = \frac{1}{(r-1)!}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\int_{0}^{\infty}(\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\log\Big[1+\exp\Big(\frac{A\varsigma}{1+A}\Big)\Big]\,d\varsigma \\
&\quad = \frac{1}{(r-1)!}\Big[\frac{\mathsf{SNR}}{A(1+A)^r}\Big]^{1+\frac{1}{A}}\int_{0}^{\infty}(\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big) \\
&\qquad\qquad\qquad\qquad\cdot\Big[\frac{A\varsigma}{1+A}+\log\Big[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\Big]\Big]\,d\varsigma \\
&\quad = \frac{1}{(r-1)!}\Big[\frac{1}{A(1+A)^r}\Big]^{\frac{1}{A}}\mathsf{SNR}^{1+\frac{1}{A}}\Big[I_{22}^{1}(\mathsf{SNR}, A)+I_{22}^{2}(\mathsf{SNR}, A)\Big], \qquad \text{(C.15)}
\end{aligned}
$$

where,

$$
\begin{aligned}
I_{22}^{1}(\mathsf{SNR}, A) &= \frac{1}{(1+A)^{r+1}}\int_{0}^{\infty}\varsigma(\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)d\varsigma, \\
I_{22}^{2}(\mathsf{SNR}, A) &= \frac{1}{A(1+A)^r}\int_{0}^{\infty}(\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\log\Big[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\Big]\,d\varsigma.
\end{aligned}
$$

Now,

$$
\begin{aligned}
&I_{22}^{1}(\mathsf{SNR}, A) \\
&\quad = \frac{1}{(1+A)^{r+1}}\int_{0}^{\infty}\varsigma(\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)d\varsigma \\
&\quad = \frac{1}{(1+A)^{r+1}}\exp\Big(\frac{\varsigma^*}{1+A}\Big)\Big[\int_{\varsigma^*}^{\infty}\varsigma^{r}\exp\Big(-\frac{\varsigma}{1+A}\Big)d\varsigma-\varsigma^*\int_{\varsigma^*}^{\infty}\varsigma^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)d\varsigma\Big] \\
&\quad = \frac{1}{(1+A)^{r+1}}\exp\Big(\frac{\varsigma^*}{1+A}\Big)\Big[(1+A)^{r+1}\Gamma(r+1,\frac{\varsigma^*}{1+A})-\varsigma^*(1+A)^{r}\Gamma(r,\frac{\varsigma^*}{1+A})\Big] \\
&\quad = \frac{1}{(1+A)^{r+1}}\exp\Big(\frac{\varsigma^*}{1+A}\Big)\Big[\Big[r(1+A)^{r+1} \\
&\qquad\qquad -\varsigma^*(1+A)^{r}\Big]\Gamma(r,\frac{\varsigma^*}{1+A})+(1+A)(\varsigma^*)^{r}\exp\Big(-\frac{\varsigma^*}{1+A}\Big)\Big] \\
&\quad = \frac{A}{1+A}I_{22}^{1A}(r, \mathsf{SNR}, A),
\end{aligned}
$$

where,

$$
I_{22}^{1A}(r, \mathsf{SNR}, A) = r!\left[\sum_{j=0}^{r-1}\frac{[\frac{\varsigma^*}{1+A}]^j}{j!}\right]\Big[1-\frac{\varsigma^*}{r(1+A)}\Big]+\Big[\frac{\varsigma^*}{1+A}\Big]^{r}.
$$

Since $\frac{\varsigma^*}{1+A} \to 0$ as $A \to \infty$,

$$\lim_{A\to\infty} I_{22}^{1A}(r, \mathsf{SNR}, A) = r! \ . \tag{C.16}$$

Thus,

$$I_{22}^{1}(r, \mathsf{SNR}, A) = \frac{Ar!}{1+A}. \tag{C.17}$$

We now compute $I_{22}^{2}(\mathsf{SNR}, A)$.

$$\begin{aligned}
&I_{22}^{2}(\mathsf{SNR}, A) \\
&\quad= \frac{1}{A(1+A)^r}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\log\Big[1+\exp\Big(-\frac{A\varsigma}{1+A}\Big)\Big]d\varsigma \\
&\quad\le \frac{1}{A(1+A)^r}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\exp\Big(-\frac{A\varsigma}{1+A}\Big)d\varsigma \\
&\quad= \frac{1}{A(1+A)^r}\exp(\varsigma^*)\int_{\varsigma^*}^\infty \varsigma^{r-1}\exp(-\varsigma)d\varsigma \\
&\quad= \frac{1}{A(1+A)^r}\exp(\varsigma^*)\Gamma(r,\varsigma^*) \\
&\quad= \frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}.
\end{aligned} \tag{C.18}$$

Also,

$$\begin{aligned}
&I_{22}^{2}(\mathsf{SNR}, A) \\
&\quad\ge \frac{1}{A(1+A)^r}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\exp\Big(-\frac{A\varsigma}{1+A}\Big)d\varsigma \\
&\quad-\frac{1}{2A(1+A)^r}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{\varsigma}{1+A}\Big)\exp\Big(-\frac{2A\varsigma}{1+A}\Big)d\varsigma \\
&\quad= \left[\frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\right] - \frac{1}{2A(1+A)^r}\int_0^\infty (\varsigma+\varsigma^*)^{r-1}\exp\Big(-\frac{(1+2A)\varsigma}{1+A}\Big)d\varsigma \\
&\quad= \left[\frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\right] - \frac{1}{2A(1+A)^r}\exp\Big(\frac{(1+2A)\varsigma^*}{1+A}\Big)\Big[\frac{1+A}{1+2A}\Big]^r\Gamma\left(r, \frac{(1+2A)\varsigma^*}{1+A}\right) \\
&\quad= \left[\frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\right] - \frac{(r-1)!}{2A(1+A)^r}\Big[\frac{1+A}{1+2A}\Big]^r\sum_{j=0}^{r-1}\frac{[(\frac{1+2A}{1+A})\varsigma^*]^j}{j!}.
\end{aligned} \tag{C.19}$$

Now, as $A \to \infty$, $\frac{\varsigma^*}{1+A} \to 0$. Therefore,

$$\lim_{A\to\infty}\frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!} = 0,$$

$$\lim_{A\to\infty}\left[\frac{(r-1)!}{A(1+A)^r}\sum_{j=0}^{r-1}\frac{(\varsigma^*)^j}{j!}\right]-\frac{(r-1)!}{2A(1+A)^r}\Big[\frac{1+A}{1+2A}\Big]^r\sum_{j=0}^{r-1}\frac{[(\frac{1+2A}{1+A})\varsigma^*]^j}{j!}=0.$$

As both the upper bound (C.18) and lower bound (C.19) go to 0, we have

$$\lim_{A\to\infty} I_{22}^2(\mathsf{SNR},A)=0. \tag{C.20}$$

Substituting (C.17) and (C.20) in (C.15), we have

$$\begin{aligned}
&I_{22}(\mathsf{SNR},A)\\
&\quad=\frac{1}{(r-1)!}\Big[\frac{1}{A(1+A)^r}\Big]^{\frac{1}{A}}\mathsf{SNR}^{1+\frac{1}{A}}\Big[\frac{Ar!}{1+A}\Big]+o(\mathsf{SNR}^2),\\
&\quad=rA^{-\frac{r+1}{A}}\mathsf{SNR}^{1+\frac{1}{A}}+o(\mathsf{SNR}^2).
\end{aligned}$$

Therefore,

$$I_2(\mathsf{SNR},A)=rA^{-\frac{r+1}{A}}\mathsf{SNR}^{1+\frac{1}{A}}+o(\mathsf{SNR}^2). \tag{C.21}$$

Substituting (C.10) and (C.21) in (C.3) and (C.2), we obtain

$$I(\mathbf{x};\vec{\mathbf{y}})=r\mathsf{SNR}-r\mathsf{SNR}\frac{\log(1+A)}{A}-rA^{-\frac{r+1}{A}}\mathsf{SNR}^{1+\frac{1}{A}}+o(\mathsf{SNR}^2).$$

Let the capacity of the channel be $C(\mathsf{SNR})$. Since on-off signaling may not be optimal for the channel, we will denote the highest achievable rate using on-off signaling as $C_{\mathsf{on-off}}(\mathsf{SNR})$. $C_{\mathsf{on-off}}(\mathsf{SNR})$ is given by

$$\begin{aligned}
&C_{\mathsf{on-off}}(\mathsf{SNR})\\
&\quad=\max_A I(\mathbf{x};\vec{\mathbf{y}})\\
&\quad=r\mathsf{SNR}[1-M^*(\mathsf{SNR})]+o(\mathsf{SNR}^2),
\end{aligned}\tag{C.22}$$

where,

$$\begin{aligned}
&M^*(\mathsf{SNR})\\
&\quad=\min_A\Big[\frac{\log(1+A)}{A}+A^{-\frac{r+1}{A}}\mathsf{SNR}^{\frac{1}{A}}\Big]\\
&\quad=\min_A\Big[\frac{\log(A)}{A}+A^{-\frac{r+1}{A}}\mathsf{SNR}^{\frac{1}{A}}\Big].
\end{aligned}$$

The last equality holds since A is large. Let us denote

$$M(A, \mathsf{SNR}) = \min_{A} \Big[\frac{\log(A)}{A} + A^{-\frac{r+1}{A}} \mathsf{SNR}^{\frac{1}{A}} \Big].$$

We will prove the following lemma to get a lower bound on $M^*(\mathsf{SNR})$.

Lemma 5

$$M^*(\mathsf{SNR}) \geq M_L(\mathsf{SNR}) \triangleq \frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})}. \tag{C.23}$$

Proof: We shall prove this by contradiction. Let there be an A_1 such that the lemma does not hold. Since $\frac{\log(A_1)}{A_1} \geq 0$ and $A_1^{-\frac{r+1}{A_1}} \mathsf{SNR}^{\frac{1}{A_1}} \geq 0$, we have,

$$\frac{\log(A_1)}{A_1} \leq M_L(\mathsf{SNR}), \tag{C.24}$$

$$A_1^{-\frac{r+1}{A_1}} \mathsf{SNR}^{\frac{1}{A_1}} \leq M_L(\mathsf{SNR}). \tag{C.25}$$

If (C.24) holds, we have

$$A_1 \geq \log(\frac{r}{\mathsf{SNR}}).$$

Moreover,

$$\begin{aligned} & A_1^{-\frac{r+1}{A_1}} \mathsf{SNR}^{\frac{1}{A_1}} \\ & \quad \geq A_1^{-\frac{r+1}{A_1}} \Big[\frac{\mathsf{SNR}}{r}\Big]^{\frac{1}{A_1}} \\ & \quad = \exp(-\frac{(r+1)\log(A_1)}{A_1}) \Big[\frac{\mathsf{SNR}}{r}\Big]^{-\frac{1}{\log(\frac{\mathsf{SNR}}{r})}} \\ & \quad \geq \exp[-(r+1)M_L(\mathsf{SNR})]e^{-1}. \end{aligned}$$

As $\mathsf{SNR} \to 0$, we have

$$\begin{aligned} \exp[-(r+1)M_L(\mathsf{SNR})]e^{-1} &\gg M_L(\mathsf{SNR}), \\ \Rightarrow A_1^{-\frac{r+1}{A_1}} \mathsf{SNR}^{\frac{1}{A_1}} &\gg M_L(\mathsf{SNR}). \end{aligned}$$

This contradicts (C.25), which completes the proof. □

To get an upper bound for $M^*(\mathsf{SNR})$, we pick a value of A. Let

$$A_2 = \frac{\log(\frac{r}{\mathsf{SNR}})}{\log\log(\frac{r}{\mathsf{SNR}})}.$$

Now,

$$M^*(\mathsf{SNR}) \leq \frac{\log(A_2)}{A_2} + A_2^{-\frac{r+1}{A_2}} \mathsf{SNR}^{\frac{1}{A_2}}. \tag{C.26}$$

We have

$$\begin{aligned}
&\frac{\log(A_2)}{A_2} \\
&\quad = \frac{\left[\log\log(\frac{r}{\mathsf{SNR}}) - \log\log\log(\frac{r}{\mathsf{SNR}})\right]\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})} \\
&\quad \leq \frac{\left[\log\log(\frac{r}{\mathsf{SNR}})\right]^2}{\log(\frac{r}{\mathsf{SNR}})}, \qquad \text{(C.27)}
\end{aligned}$$

and

$$\begin{aligned}
&A_2^{-\frac{r+1}{A_2}} \mathsf{SNR}^{\frac{1}{A_2}} \qquad \text{(C.28)} \\
&\quad = \left[\frac{r}{A_2^{r+1}}\right]^{\frac{1}{A_2}} \left[\frac{\mathsf{SNR}}{r}\right]^{\frac{1}{A_2}}, \\
&\quad \leq \left[\max_r \frac{r}{A_2^{r+1}}\right]^{\frac{1}{A_2}} \left[\frac{\mathsf{SNR}}{r}\right]^{\frac{1}{A_2}}, \\
&\quad \leq \left[\frac{1}{eA_2\log(A_2)}\right]^{\frac{1}{A_2}} \left[\frac{\mathsf{SNR}}{r}\right]^{\frac{1}{A_2}}, \\
&\quad \leq \left[\frac{\mathsf{SNR}}{r}\right]^{\frac{1}{A_2}} \qquad \text{(C.29)} \\
&\quad = \left[\frac{\mathsf{SNR}}{r}\right]^{-\frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{\mathsf{SNR}}{r})}} \\
&\quad = \frac{1}{\log(\frac{r}{\mathsf{SNR}})}. \qquad \text{(C.30)}
\end{aligned}$$

Equation (C.29) holds since $A_2 \gg 1$ for $\mathsf{SNR} \to 0$, which makes

$$\left[\frac{1}{eA_2\log(A_2)}\right]^{\frac{1}{A_2}} \leq [1]^{\frac{1}{A_2}} = 1.$$

Combining (C.26, C.27, C.30), we have

$$M^*(\mathsf{SNR}) \leq \frac{[\log\log(\frac{r}{\mathsf{SNR}})]^2 + 1}{\log(\frac{r}{\mathsf{SNR}})}. \tag{C.31}$$

Finally, using (C.22), Lemma 5 and (C.31), we have

$$\begin{aligned}
&r\mathsf{SNR} - r\mathsf{SNR}\frac{[\log\log(\frac{r}{\mathsf{SNR}})]^2 + 1}{\log(\frac{r}{\mathsf{SNR}})} + o(\mathsf{SNR}^2) \\
&\qquad \leq C_{\mathsf{on-off}}(\mathsf{SNR}) \leq r\mathsf{SNR} - r\mathsf{SNR}\frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})} + o(\mathsf{SNR}^2).
\end{aligned} \tag{C.32}$$

Since on-off signalling may not be optimal

$$C_{\mathsf{on-off}}(\mathsf{SNR}) \leq C(\mathsf{SNR}). \tag{C.33}$$

As conditioning reduces entropy, we can express the input-output mutual information as

$$I(\mathbf{x};\vec{\mathbf{y}}) \leq \sum_{k=1}^{r} I(\mathbf{x};\mathbf{y}_k). \tag{C.34}$$

Each term on the right hand side of (C.34) is maximized by an on-off distribution [25] and, we know from [42] that, with this distribution, the mutual information $\forall k \in \{1 \dots r\}$ is

$$I(\mathbf{x};\mathbf{y}_k) \leq \mathsf{SNR} - \mathsf{SNR}\frac{\log\log(\frac{1}{\mathsf{SNR}})}{\log(\frac{1}{\mathsf{SNR}})} + o(\mathsf{SNR}^2).$$

Hence, we can upper bound the capacity as

$$C(\mathsf{SNR}) \leq r\mathsf{SNR} - r\mathsf{SNR}\frac{\log\log(\frac{1}{\mathsf{SNR}})}{\log(\frac{1}{\mathsf{SNR}})} + o(\mathsf{SNR}^2).$$

Since,

$$\frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})} \leq \frac{\log\log(\frac{1}{\mathsf{SNR}})}{\log(\frac{1}{\mathsf{SNR}})},$$

we have

$$C(\mathsf{SNR}) \leq r\mathsf{SNR} - r\mathsf{SNR}\frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})} + o(\mathsf{SNR}^2). \tag{C.35}$$

Combining (C.32, C.33, C.35), we obtain

$$\begin{aligned} r\mathsf{SNR} - r\mathsf{SNR}\frac{\left[\log\log(\frac{r}{\mathsf{SNR}})\right]^2 + 1}{\log(\frac{r}{\mathsf{SNR}})} + o(\mathsf{SNR}^2) \\ \leq C(\mathsf{SNR}) \leq r\mathsf{SNR} - r\mathsf{SNR}\frac{\log\log(\frac{r}{\mathsf{SNR}})}{\log(\frac{r}{\mathsf{SNR}})} + o(\mathsf{SNR}^2). \end{aligned} \tag{C.36}$$

We now introduce a notation for the approximation that ignores higher order logarithm functions. Let $f(\mathsf{SNR})$ and $g(\mathsf{SNR})$ be functions of SNR. We shall denote

$$f(\mathsf{SNR}) \doteq g(\mathsf{SNR})$$

if

$$\lim_{\mathsf{SNR}\to 0} \frac{\log f(\mathsf{SNR})}{\log g(\mathsf{SNR})} = 1.$$

With this scaling, the inequalities in (C.36) become equalities and the capacity can be expressed as

$$C(\mathsf{SNR}) = r\mathsf{SNR} - \Delta_{\mathsf{i.i.d}}^{(t,r)}(\mathsf{SNR}),$$

where,

$$\Delta_{\mathsf{i.i.d}}^{(t,r)}(\mathsf{SNR}) \doteq \frac{r\mathsf{SNR}}{\log(\frac{r}{\mathsf{SNR}})}.$$

We also see that on-off signaling (C.1) is capacity achieving for the i.i.d Rayleigh fading MIMO channel in the wideband regime (keeping in mind our scaling, which ignores higher order logarithm functions.)

Appendix D

Proof of Theorem 5

The channel state realization is $\vec{\mathbf{a}} = \vec{a}$. Consider another channel which differs from our model, (4.5), in only the quantization noise distribution. For this channel

$$\begin{aligned} \vec{\mathbf{z}}^G &= \vec{a}\mathbf{x} + \vec{\mathbf{w}} + \vec{\mathbf{q}}^G, \\ \mathbf{q}_i^G &\sim \mathcal{C}N(0, E[|\mathbf{q}_i|^2]), \quad i \in \{1, \ldots, r\}, \end{aligned} \tag{D.1}$$

the quantization noise is jointly Gaussian with the input $\mathbf{x}$ and the following are satisfied:

$$E[\mathbf{z}_i^G \mathbf{q}_i^{G*}] = E[\mathbf{z}_i \mathbf{q}_i^*] = 0, \tag{D.2}$$

$$E[\mathbf{y}_i \mathbf{q}_i^{G*}] = E[\mathbf{y}_i \mathbf{q}_i^*] = -E[|\mathbf{q}_i|^2]. \tag{D.3}$$

Given the channel realization $\vec{a}$, since $\vec{\mathbf{q}}^G$ is jointly Gaussian with $\mathbf{x}$, $\vec{\mathbf{z}}^G$ is also jointly Gaussian with $\mathbf{x}$. References [17, 40] show that noise that is jointly Gaussian with the input, minimizes mutual information. Hence, we obtain the following bound:

$$I(\mathbf{x}; \vec{\mathbf{z}}^G | \vec{a}) \leq I(\mathbf{x}; \vec{\mathbf{z}} | \vec{a}). \tag{D.4}$$

Let $\hat{\mathbf{x}}_{llse}(\vec{\mathbf{z}}^G)$ be the linear least-squares error (LLSE) estimate of $\mathbf{x}$ from $\vec{\mathbf{z}}^G$ and $\mathbf{e}_{llse}$, the corresponding estimation error. Hence, $\mathbf{x}$ can be expressed as:

$$\mathbf{x} = \hat{\mathbf{x}}_{llse}(\vec{\mathbf{z}}^G) + \mathbf{e}_{llse}.$$

Since $\mathbf{x}$ and $\vec{\mathbf{z}}^G$ are jointly Gaussian, minimum mean-squares estimation (MMSE) is the same as LLSE estimation, and the estimation error is Gaussian and independent

of the estimate. The variance of the estimation error is denoted as $\lambda_{llse} = E[|\mathbf{e}_{llse}|^2]$. We now compute the lower bound to the SIMO-FAWNA block capacity using our proposed scheme:

$$\begin{aligned}
&\frac{1}{W}C_q^b(P, W, \vec{a}, r, m, C_f) \\
&\quad = I(\mathbf{x}; \vec{\mathbf{z}}|\vec{a}) \\
&\quad \geq I(\mathbf{x}; \vec{\mathbf{z}}^G|\vec{a}) \qquad \text{(D.5)} \\
&\quad = h(\mathbf{x}) - h(\mathbf{e}_{llse}) \\
&\quad = \log\Big(\frac{P}{\lambda_{llse}W}\Big). \qquad \text{(D.6)}
\end{aligned}$$

We use (D.4) to obtain the inequality in (D.5). In order to compute the lower bound (D.6), λ_{llse} needs to be computed. This can be expressed as:

$$\lambda_{llse} = E[|\mathbf{x}|^2] - E[\mathbf{x}\vec{\mathbf{z}}^{G\dagger}]\Lambda_{\vec{\mathbf{z}}^G}^{-1}E[\mathbf{x}^*\vec{\mathbf{z}}^G], \qquad \text{(D.7)}$$

where, $\Lambda_{\vec{\mathbf{z}}^G}$ is the autocorrelation matrix of $\vec{\mathbf{z}}^G$.

From our channel and quantizer models, we have the following Markov chains for $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, r\}$, $i \neq j$:

$$\mathbf{x} \leftrightarrow \mathbf{y}_i \leftrightarrow \mathbf{q}_i^G, \qquad \text{(D.8)}$$

$$\mathbf{y}_i \leftrightarrow \mathbf{x} \leftrightarrow \mathbf{q}_j^G, \qquad \text{(D.9)}$$

$$\mathbf{q}_i^G \leftrightarrow \mathbf{x} \leftrightarrow \mathbf{q}_j^G. \qquad \text{(D.10)}$$

Using the first Markov chain, (D.8), we obtain:

$$\begin{aligned}
&E[\mathbf{x}\mathbf{q}_i^{G*}] \\
&\quad = E_{\mathbf{y}_i}\Big[E[\mathbf{x}\mathbf{q}_i^{G*}|\mathbf{y}_i]\Big] \\
&\quad = E_{\mathbf{y}_i}\Big[E[\mathbf{x}|\mathbf{y}_i]E[\mathbf{q}_i^{G*}|\mathbf{y}_i]\Big] \\
&\quad = E_{\mathbf{y}_i}\Big[\frac{E[\mathbf{x}\mathbf{y}_i^*]}{E[|\mathbf{y}_i|^2]}\mathbf{y}_i E[\mathbf{q}_i^{G*}|\mathbf{y}_i]\Big] \qquad \text{(D.11)} \\
&\quad = \frac{E[\mathbf{x}\mathbf{y}_i^*]E[\mathbf{y}_i\mathbf{q}_i^{G*}]}{E[|\mathbf{y}_i|^2]} \\
&\quad = -\frac{a_i^* P M_m \beta_m 2^{-\frac{R_i}{W}}}{W}. \qquad \text{(D.12)}
\end{aligned}$$

Since $\mathbf{x}$ and $\mathbf{y}_i$ are jointly Gaussian random variables, we obtain (D.11). Equation (D.12) follows from the quantizer properties (4.6, D.3), and our wireless channel model (4.1).

Using the second Markov chain, (D.9), we obtain for $i \neq j$:

$$\begin{aligned}
&E[\mathbf{y}_i \mathbf{q}_j^{G*}] \\
&\quad = E_{\mathbf{x}}\Big[E[\mathbf{y}_i \mathbf{q}_j^{G*}|\mathbf{x}]\Big] \\
&\quad = E_{\mathbf{x}}\Big[E[\mathbf{y}_i|\mathbf{x}]E[\mathbf{q}_j^{G*}|\mathbf{x}]\Big] \\
&\quad = E_{\mathbf{x}}\Big[\frac{E[\mathbf{x}^*\mathbf{y}_i]}{E[|\mathbf{x}|^2]}\mathbf{x}E[\mathbf{q}_j^{G*}|\mathbf{x}]\Big] \qquad \text{(D.13)} \\
&\quad = \frac{E[\mathbf{x}^*\mathbf{y}_i]E[\mathbf{x}\mathbf{q}_j^{G*}]}{E[|\mathbf{x}|^2]} \\
&\quad = -\frac{a_i a_j^* P M_m \beta_m 2^{-\frac{R_j}{W}}}{W}. \qquad \text{(D.14)}
\end{aligned}$$

Since $\mathbf{x}$ and $\mathbf{y}_i$ are jointly Gaussian random variables, we obtain (D.13). Equation (D.14) follows from (D.12), and our wireless channel model (4.1).

The third Markov chain, (D.10), gives us for $i \neq j$:

$$\begin{aligned}
&E[\mathbf{q}_i^G \mathbf{q}_j^{G*}] \\
&\quad = E_{\mathbf{x}}\Big[E[\mathbf{q}_i^G \mathbf{q}_j^{G*}|\mathbf{x}]\Big] \\
&\quad = E_{\mathbf{x}}\Big[E[\mathbf{q}_i^G|\mathbf{x}]E[\mathbf{q}_j^{G*}|\mathbf{x}]\Big] \\
&\quad = E_{\mathbf{x}}\Big[\frac{E[\mathbf{x}^*\mathbf{q}_i^G]}{E[|\mathbf{x}|^2]}\mathbf{x}E[\mathbf{q}_j^{G*}|\mathbf{x}]\Big] \qquad \text{(D.15)} \\
&\quad = \frac{E[\mathbf{x}^*\mathbf{q}_i^G]E[\mathbf{x}\mathbf{q}_j^{G*}]}{E[|\mathbf{x}|^2]} \\
&\quad = \frac{a_i a_j^* P M_m^2 \beta_m^2 2^{-\frac{R_i+R_j}{W}}}{W}. \qquad \text{(D.16)}
\end{aligned}$$

Since $\mathbf{x}$ and $\mathbf{q}_i^G$ are jointly Gaussian random variables, we obtain (D.15). We obtain (D.16) from (D.12) and our wireless channel model (4.1).

From (4.1) and (D.12), we obtain:

$$E[\mathbf{x}\mathbf{z}_i^{G*}] = E[\mathbf{x}\mathbf{y}_i^*] + E[\mathbf{x}\mathbf{q}_i^{G*}] = \frac{P a_i^*(1 - M_m\beta_m 2^{-\frac{R_i}{W}})}{W} \triangleq \frac{P}{W}v_i^*. \qquad \text{(D.17)}$$

We now compute $\Lambda_{\vec{\mathbf{z}}^G}$. From (4.6, D.3), for $i \in \{1, \ldots, r\}$,

$$\begin{aligned}
&E[|\mathbf{z}_i^G|^2] \\
&\quad = E[|\mathbf{y}_i|^2] + E[\mathbf{y}_i \mathbf{q}_i^{G*}] + E[\mathbf{y}_i^* \mathbf{q}_i^G] + E[|\mathbf{q}_i|^2] \\
&\quad = E[|\mathbf{y}_i|^2](1 - M_m \beta_m 2^{-\frac{R_i}{W}}) \\
&\quad = N_0 \Big(1 + \frac{|a_i|^2 P}{N_0 W}\Big) \Big(1 - M_m \beta_m 2^{-\frac{R_i}{W}}\Big) \triangleq N_0 M_{ii}, \qquad \text{(D.18)}
\end{aligned}$$

and for $i \in \{1, \ldots, r\}, j \in \{1, \ldots, r\}, i \neq j$

$$\begin{aligned}
&E[\mathbf{z}_i^G \mathbf{z}_j^{G*}] \\
&\quad = E[\mathbf{y}_i \mathbf{y}_j^*] + E[\mathbf{y}_i \mathbf{q}_j^{G*}] + E[\mathbf{y}_j^* \mathbf{q}_i^G] + E[\mathbf{q}_i^G \mathbf{q}_j^{G*}] \\
&\quad = a_i a_j^* \frac{P}{W} - \frac{a_i a_j^* P M_m \beta_m 2^{-\frac{R_j}{W}}}{W} - \frac{a_i a_j^* P M_m \beta_m 2^{-\frac{R_i}{W}}}{W} + \frac{a_i a_j^* P M_m^2 \beta_m^2 2^{-\frac{R_i + R_j}{W}}}{W} \\
&\quad = \frac{a_i a_j^* P}{W} \Big(1 - M_m \beta_m 2^{-\frac{R_i}{W}}\Big) \Big(1 - M_m \beta_m 2^{-\frac{R_j}{W}}\Big) \triangleq N_0 M_{ij}. \qquad \text{(D.19)}
\end{aligned}$$

Combining equations (D.6, D.7, D.17, D.18, D.19) completes the proof. □

Bibliography

[1] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions", *Ph.D. Dissertation, Stanford University*, 1963.

[2] R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications", *IEEE Transactions on Information Theory*, Vol. 11, No. 1, pp. 3-18, January 1965.

[3] J. G. Dunn, "The performance of a class of n dimensional quantizers for a Gaussian source", *Proceedings of Columbia Symposium on Signal Transmission Processing*, Columbia University, NY 1965.

[4] R. S. Kennedy, "Fading Dispersive Communication Channels", New York: *Wiley Interscience*, 1969.

[5] J. A. Bucklew and N. C. Gallagher, Jr., "A note on optimum quantization", *IEEE Transactions on Information Theory*, vol. IT-25, pp. 365-366, May 1979.

[6] A. Gersho, "Asymptotically optimal block quantization", *IEEE Transactions on Information Theory*, vol. IT-25, pp. 373-380, July 1979.

[7] N. C. Gallagher and J. A. Bucklew, "Properties of minimum mean squared error block quantizers", *IEEE Transactions on Information Theory*, vol. IT-28, pp. 105-107, January 1982.

[8] P. L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension", *IEEE Transactions on Information Theory*, vol. IT-28, pp. 139-148, March 1982.

[9] R. G. Gallager, "Energy limited channels: Coding, multiaccess and spread spectrum", LIDS Technical Report, LIDS-P-1714, MIT, Cambridge, MA, November 1987.

[10] I. E. Telatar, "Coding and multiaccess for the energy limited Rayleigh fading channel", *Master's Thesis*, MIT, 1988.

[11] A. Gersho and R. M. Gray, "Vector Quantization and Signal Compression", *Kluwer*, Boston, MA, 1992.

[12] D. Hui and D. L. Neuhoff, "On the complexity of scalar quantization", *ISIT 1995*, p. 372, 1995.

[13] G. J. 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.

[14] H. Jung and O. K. Tonguz, "Comparison of Fiber and Coaxial Link for Access Network in Microcellular PCS", *IEE Electronics Letters*, vol. 32, no. 5, pp. 425-426, February 1996.

[15] T. Berger, Z. Zhang and H. Viswanathan, "The CEO Problem", *IEEE Transactions on Information Theory*, vol. 42, pp. 887-902, May 1996.

[16] O. K. Tonguz and H. Jung, "Personal Communications Access Networks using Subcarrier Multiplexed Optical Links", *Joint Special Issue of the IEEE Journal of Lightwave Technology and IEEE Journal on Selected Areas of Communications on Multiwavelength Optical Technology and Networks*, vol. 14, no. 6, pp. 1400-1409, June 1996.

[17] M. Médard, "Capacity of Correlated Jamming Channels", *Thirty-fifth Annual Allerton Conference on Communication, Control and Computing*, Allerton, Illinois, September 1997.

[18] R. G. Gallager and M. Médard, "Bandwidth scaling for fading channels", *International Symposium on Information Theory*, p. 471, Ulm, Germany, 1997.

[19] V. Tarokh, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Transactions on Information Theory*, Vol. 44, pp. 744-765, March 1998.

[20] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in flat fading", *IEEE Transactions on Information Theory*, Vol. 45, pp. 139-157, January 1999.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels", *European Transactions on Telecommunications*, 10(6):585-95, November/December 1999.

[22] B. M. Hochwald and T. L. Marzetta, "Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading", *IEEE Transactions on Information Theory*, Vol. 46, pp. 543-564, March 2000.

[23] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel", *IEEE Transactions on Information Theory*, Vol. 46, May 2000.

[24] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels", *IEEE Transactions on Information Theory*, Vol. 46, July 2000.

[25] I. C. Abou-Fayçal, M. D. Trott and S. Shamai, "The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels", *IEEE Transactions on Information Theory*, Vol. 47, pp. 1290-1301, May 2001.

[26] D. Stancil, O. K. Tonguz, A. Xhafa, P. Nikitin, A. Cepni, and D. Brodtkorb, "High-speed Internet Access via HVAC Ducts: A New Approach", *IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 6, pp. 3604-3607, San Antonio, Texas, November 2001.

[27] L. Zheng and D. N. C. Tse, "Communication on the Grassman Manifold: A geometric approach to the noncoherent multiple-antenna channel", *IEEE Transactions on Information Theory*, Vol. 48, pp. 359-83, February 2002.

[28] M. Médard and R.G. Gallager, "Bandwidth scaling for fading multipath channels", *IEEE Transactions on Information Theory*, Vol. 48, April 2002.

[29] V. G. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and Capacity", *IEEE Transactions on Information Theory*, Vol. 48, pp. 908-27, April 2002.

[30] S. Verdú, "Spectral efficiency in the wideband regime", *IEEE Transactions on Information Theory*, Vol. 48, pp. 1319-1343, June 2002.

[31] L. Zheng and D. N. C. Tse, "The Diversity-Multiplexing Tradeoff for Noncoherent Multiple Antenna Channels", *Allerton Conference on Communication, Control and Computing*, Allerton, Illinois, 2002.

[32] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?", *IEEE Transactions on Information Theory*, Vol. 49, pp. 951-963, April 2003.

[33] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels", *IEEE Transactions on Information Theory*, Vol. 49, pp. 1073-1096, May 2003.

[34] A. Lozano, A. M. Tulino and S. Verdú, "Multiple-Antenna capacity in the Low-Power Regime", *IEEE Transactions on Information Theory*, Vol. 49, No. 10, October 2003.

[35] P. Nikitin, D. D. Stancil, O. K. Tonguz, A. Cepni, A. Xhafa, and D. Brodtkorb, "Impulse Response Characteristics of the HVAC Duct Communication Channel", *IEEE Transactions on Communications*, vol. 51, no. 10, pp. 1736-1742, October 2003.

[36] D. S. Lun, M. Médard and I. C. Abou-Fayçal, "On the Performance of Peaky Capacity-Achieving Signaling on Multipath Fading Channels", *IEEE Transaction on Communications*, Vol. 52, pp. 931-938, June 2004.

[37] O. K. Tonguz, D. D. Stancil, A. Xhafa, A. G. Cepni, P. Nikitin, and D. Brodtkorb, "A Simple Path Loss Prediction Model for HVAC Systems", *IEEE Transactions on Vehicular Technology*, July 2004.

[38] V. V. Prelov and S. Verdú, "Second-Order Asymptotics of Mutual Information", *IEEE Transactions on Information Theory*, Vol. 50, pp. 1567-1580, August 2004.

[39] C. Rao and B. Hassibi, "Analysis of Multiple-Antenna Wireless Links at Low SNR", *IEEE Transactions on Information Theory*, Vol. 50(9), pp. 2123-2130, September 2004.

[40] A. Kashyap, T. Başar and R. Srikant, "Correlated Jamming on MIMO Gaussian Fading Channels", *IEEE Transactions on Information Theory*, vol. 50, pp. 2119-2123, September 2004.

[41] X. Wu and R. Srikant, "MIMO Channels in the Low SNR Regime: Communication Rate, Error Exponent and Signal Peakiness", *Information Theory Workshop 2004*, San Antonio, Texas, 2004.

[42] L. Zheng, D. N. C. Tse and M. Médard, "Channel Coherence in the low SNR Regime", *IEEE Transactions on Information Theory*, to appear.

[43] S. Ray, P. Moulin and M. Médard, "On Jamming in the Wideband Regime", *2006 International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, July 2006.

[44] S. Ray, P. Moulin and M. Médard, "On Optimal Signaling and Jamming Strategies in Wideband Fading Channels", *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 2006*, Cannes, France, July 2006.