DISSERTATION

# SIGNAL DESIGN, DIVERSITY, AND CAPACITY IN MULTI-ACCESS COMMUNICATION SYSTEMS

SUBMITTED BY

ZHIFEI FAN

ELECTRICAL AND COMPUTER ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

COLORADO STATE UNIVERSITY

FORT COLLINS, COLORADO

FALL 2006

COLORADO STATE UNIVERSITY

October 19, 2006

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ZHIFEI FAN ENTITLED SIGNAL DESIGN, DIVERSITY, AND CAPACITY IN MULTI-ACCESS COMMUNICATION SYSTEMS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Donald Estep

Edwin K. P. Chong

Peter J. Brockwell

Mahmood R. Azimi-Sadjadi

Adviser Louis SCHARF

Department Head/Director Anthony A. Maciejewski

**ABSTRACT OF DISSERTATION**

SIGNAL DESIGN, DIVERSITY, AND CAPACITY IN MULTI-ACCESS COMMUNICATION SYSTEMS

In this dissertation, we exploit degrees of freedom in time and frequency to trade-off capacity and diversity in time-frequency-spreading channels. We then design signals to maximize channel performance.

We consider the trade-off between multiplexing gain and diversity gain for the block-fading sub-channel model. The channel vector may be rank deficient with arbitrary covariance structure. We derive this trade-off by considering the scaling law of the ergodic capacity, which represents the multiplexing gain, and the error probability, which determines the diversity gain at high SNR. With a fixed multiplexing gain, we give an upper bound and a lower bound on the maximum diversity gain, and give an optimization procedure to get the exact maximum diversity gain.

We also address the trade-off between multiplexing gain and diversity gain for the frequency-selective channel. Similarly, we derive this trade-off by considering the scaling law of the ergodic capacity, which determines the multiplexing gain, and the error probability, which determines the diversity gain at high SNR. It is proved that this trade-off only depends on the number of independent taps of the equivalent FIR channel filter. The error probability is bounded by the outage probability and the error probability without outage. The scaling law of the outage probability and the error probability without outage at high SNR are derived, as are approximations of the outage probability at both low and high SNR.

Besides the theoretical research on capacity and diversity, we investigate signal design, i.e. joint analog precoder and equalizer design, for multichannel data transmission over the frequency-selective channel. The design goal is to maximize mutual information rate, minimize the mean square error, or minimize the bit error rate subject to a transmit power constraint. We assume a continuous channel model with precoder transmissions for M subchannels that lie in an $n$-dimensional linear subspace of $L^2(\mathcal{R})$. We first design the subspace according to the channel characteristics, and then design the precoders as functions in this subspace. After the design of the optimal precoder and equalizer, we explore the geometry of these designs. We show that all of these precoder and equalizer designs are, in fact, decompositions of a virtual two-channel problem into a system of canonical coordinates, wherein variables in the canonical message channel are correlated only pairwise with corresponding variables in the canonical measurement channel. This finding clarifies the geometry of precoder and equalizer designs and illustrates that they decompose the two-channel communication problem into what might be called the Shannon channel.

We also investigate joint precoder and equalizer designs for a CDMA multiuser, multi-path system, and design the precoder to get a simplified receiver design for an MMSE equalizer on the receiver side, using a warp convergence property for special matrices.

Zhifei Fan

Electrical and Computer Engineering Department

Colorado State University

Fort Collins, CO 80523

Fall 2006

## ACKNOWLEDGMENTS

My foremost thank goes to my thesis advisor Dr. Louis Scharf. Without him, this dissertation would not have been possible. I gratefully acknowledge his constant and invaluable academic and personal support, his patience and encouragement that carried me on through difficult times, his enthusiasm, inspiration, suggestions and criticisms to shape my research skills and my English spoken and written ability.

I'm grateful to Dr. Edwin Chong, Dr. Donald Estep and Dr. Ronald Butler for their help and good suggestions for solving miscellaneous research problems I encountered during my research.

I would like to thank Dr. Peter Brockwell, Dr. Donald Estep, Dr. Edwin Chong and many other professors in Mathematics and Statistics departments who have taught me mathematics, statistics classes. Many of them let me sit in their classes for free. I also would like to thank Dr. Ali Pezeshki for many constructive discussion during the years. Without this good learning environment, it'll be much more difficult or even impossible for me to find the problem, formulate the problem and finish my thesis at this level.

I would also like to thank the support from the National Science Foundation, which has supported me financially during my 4 years of research.

Lastly, and most importantly, I wish to thank my husband, Hua Li, on whose constant encouragement and love I have relied throughout the years. And I wish to thank my parents, who are always there when I need them.

TABLE OF CONTENTS

Page

Page

Page

## LIST OF FIGURES

| Figure | | Page |
|---|---|---|

# CHAPTER 1

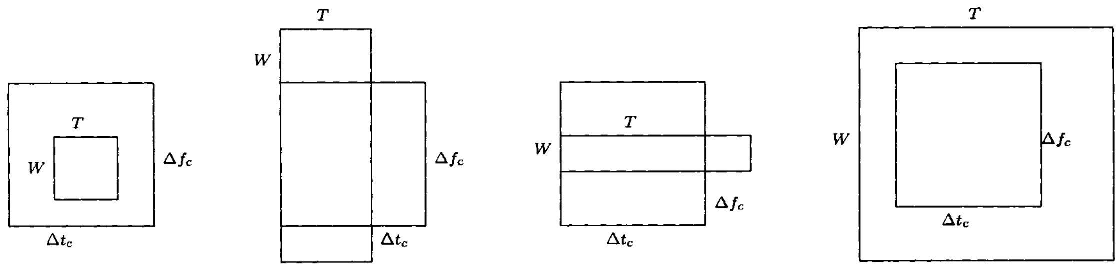
# Introduction

## 1.1 Degrees-of-Freedom in Wireless Communications

In wireless communications, the effect of the channel on the transmitted signal is a function of the signal bandwidth and signal duration. Figure 1 illustrates 4 different channel effects for signals with signal bandwidth $W$ and signal duration $T$. In the figure, $\Delta t_c$ is a coherent time that roughly characterizes time intervals over which the channel is time-invariant and $\Delta f_c$ characterizes the bandwidth of this time-invariant channel. If the signal duration $T$ is less than the channel coherence time $\Delta t_c$, and the signal bandwidth $W$ is less than the channel coherence bandwidth $\Delta f_c$, then the channel is time-invariant and frequency-nonselective. The signal is said to experience *flat fading* in time and frequency when transmitted through the channel. If the signal duration $T$ is less than the channel coherence time $\Delta t_c$, but the signal bandwidth $W$ is greater than the channel coherence bandwidth $\Delta f_c$, then the channel is time-invariant and frequency-selective. Sometimes this frequency selectivity may be associated with time-invariant multipath scattering. If the signal duration $T$ is greater than the channel coherence time $\Delta t_c$, but the signal bandwidth $W$ is less than the channel coherence bandwidth $\Delta f_c$, then the channel is time-variant and frequency-nonselective. The signal is said to experience time-varying flat fading in frequency. If the signal duration $T$ is greater than the channel coherence time $\Delta t_c$, and the signal bandwidth $W$ is greater than the channel coherence bandwidth $\Delta f_c$, then the channel is time-varying and frequency-selective. In other words, the channel is a time-varying filter. Sometimes these effects may be associated with Doppler spreading and multipath fading.

*Over- and Under-Spread Channels*

According to the product $p := \Delta t_c \Delta f_c$, the channel may be classified as un-




(a) Flat Fading (b) Time-Invariant Frequency-Selective (c) Time-Varying Frequency-Nonselective (d) Time-Varying Frequency-Selective

Figure 1. Channel effects for signals with different signal bandwidth $W$ and signal duration $T$. The first signal has $T < \Delta t_c$, and $W < \Delta f_c$, so it sees a channel that is time-invariant and frequency-nonselective. The second signal has $T < \Delta t_c$, and $W > \Delta f_c$, so it sees a channel that is time-invariant and frequency-selective. The third signal has $T > \Delta t_c$ and $W < \Delta f_c$, so it sees a channel that is time-varying and frequency-nonselective. The fourth signal has $T > \Delta t_c$ and $W > \Delta f_c$, so it sees a channel that is time-varying and frequency-selective.

derspread or overspread. When $p >> 1$, the channel is underspread. According to the Landau-Pollak Theorem, we can design a signal with $1 \leq TW < p$ to estimate the channel and then use it for communication. On the other hand, when $p < 1$, the channel is overspread. In this case, no signal can lie in a subspace of dimension less than 1, so the channel varies too fast to be estimated and used. In our work, we only consider the underspread channel. That is $p >> 1$. For intuition we note that $p \approx \frac{1}{B_d T_m} = \Delta t_c \Delta f_c$, where $B_d = \frac{1}{\Delta t_c}$ is the doppler spread and $T_m = \frac{1}{\Delta f_c}$ is the multipath spread, so that they cannot both be large.

*Usage of the Time-Frequency Signaling Space*

Given a frequency band $W$ which is much greater than the coherence bandwidth, $W >> \Delta f_c$, and given a signaling interval $T$ which is much greater than the coherence time, $T >> \Delta t_c$, then in this $TW$ time-frequency signaling space, there are $\frac{TW}{\Delta t_c \Delta f_c}$ independent fading paths. This is case ($d$) in Figure 1.

There are many ways to design the signaling in this $TW$ time-frequency signaling space. We can divide this time-frequency signaling space into non-overlapping time-frequency cells whose bandwidths are less than or equal to the coherence

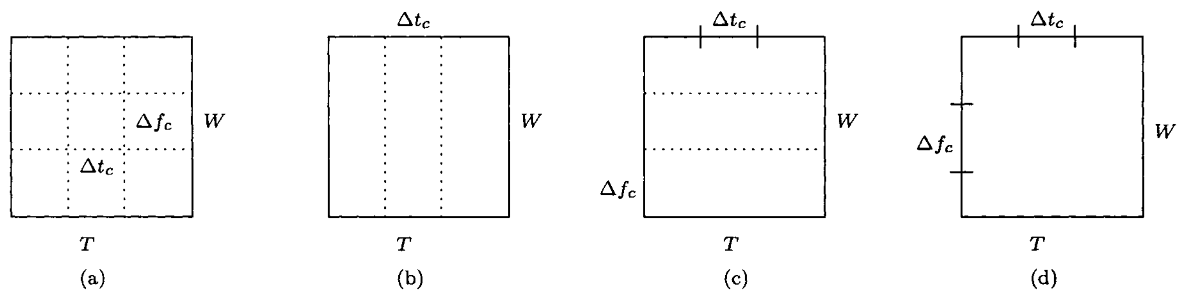



Figure 2. 4 different ways to design the signaling in the $TW$ time-frequency signaling space. In (a), we divide the $TW$ signaling space into non-overlapping time-frequency cells whose bandwidths are less than or equal to the coherence bandwidth and whose time durations are less than or equal to the coherence time. In (b), we divide the $TW$ signaling space into non-overlapping time cells whose time durations are less than or equal to the coherence time $\Delta t_c$. In (c), we partition the $TW$ signaling space into non-overlapping frequency cells whose bandwidths are less than or equal to the coherence bandwidth $\Delta f_c$. In (d), we transmit high time-frequency product signals in the space.

bandwidth and whose time durations are less than or equal to the coherence time. See Figure 2(a). Approximately, the signal sees $\frac{TW}{\Delta t_c \Delta f_c}$ independent *block fades.* Signal waveforms may be transmitted within each cell and each will see a time-invariant, frequency-nonselective flat fading channel.

Using time division, as in Figure 2(b), we can divide this $TW$ signaling space into non-overlapping time cells whose time durations are less than or equal to the coherence time $\Delta t_c$. Signal waveforms are transmitted in each cell and see a time-invariant, frequency-selective channel. Using frequency division, as in Figure 2(c), the $TW$ signaling space can be partitioned into non-overlapping frequency cells whose bandwidths are less than or equal to the coherence bandwidth $\Delta f_c$. Signal waveforms transmitted in each cell see a time-varying, frequency-nonselective channel.

We can transmit high time-frequency product signal waveforms, for example CDMA signals, in the time-frequency signaling space, as in Figure 2(d). The signal waveform sees a time-varying, frequency-selective channel.

Note that in each of these time-frequency signaling spaces, there are essentially $\frac{TW}{\Delta t_c \Delta f_c}$ independent fading paths, so, equivalently, there are $N := \frac{TW}{\Delta t_c \Delta f_c}$ independent parallel channels in each time-frequency signaling space, as illustrated in Figure 3. Since the probability of simultaneous deep fading of several independent parallel channels is far less than the probability of deep fading of one of them, this gives us a possibility to reduce error probability by repeated transmission of the same symbols carried by the information-bearing signals in different parallel channels, which on the other hand will reduce the data rate compared to transmission of each symbol once. In other words, the independent parallel channels supplied by this time-frequency signaling space gives us freedom to trade-off error probability and data rate. Note that there are many methods for linearly combining transmit symbols carried by information-bearing signals or modulation pulse shapes to transmit bits in the channel. But the underlying idea to get diversity is as described above.

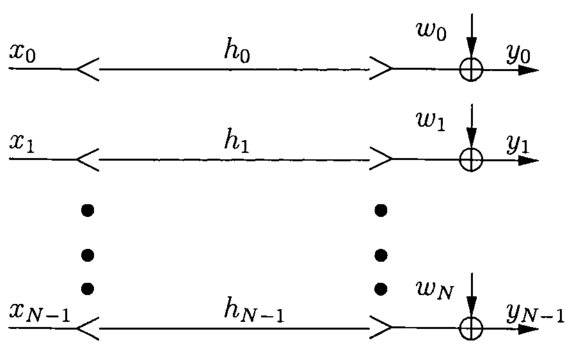



Figure 3. Equivalent $N$ independent parallel channels in the $TW$ time-frequency signaling space

In summary, to get diversity gain is to reduce the error probability. To get multiplexing gain is to increase the data rate. The trade-off between multiplexing gain and diversity gain is the trade-off between data rate and probability of error, at high SNR.

A natural question which come up is, what's the maximum diversity gain we can get for a given data rate? This question has been answered in [1] for MIMO channels. In [1] the authors explored the degrees of freedom in time and space. In our work, instead, we'll explore the degrees of freedom in time and frequency.

*Block Fading Sub-Channel*: In this dissertation, we consider the block fading sub-channel, i.e. the channel model in Figure 1 (a) and Figure 2 (a). We give an upper bound and a lower bound on the diversity gain at high SNR with fixed multiplexing gain in the case of symbolling in a time-frequency signaling space over a time-frequency-spread channel whose channel vector may be rank deficient and has arbitrary covariance structure. Then we give a simple linear programming optimization procedure to derive the exact diversity gain at high SNR with fixed multiplexing gain which doesn't depend on the detailed values of the channel covariance matrix. We derive the outage probability and error probability at high SNR for this channel.

*Frequency Selective Channel*: In this dissertation, we also consider the frequency-selective channel model, i.e. the channel model in Figure 1 (b) and Figure 2 (b). We derive the trade-off between multiplexing gain and diversity gain for the frequency-selective channel. Then we derive the scaling law of the outage probability of the frequency-selective channel (equivalent to the tapped delay line channel model in [2]) at high SNR, and give approximations of the outage probability at low and high SNR. We derive the error probability of the frequency-selective channel at high SNR. The outage probability of the tapped delay line channel with two taps is derived in [3], and the density of the mutual information of the MIMO tapped delay line channel with arbitrary taps is derived in [4] with special assumptions.

## 1.2 Signal Design in the Multiaccess Channel

A general approach to digital communication is orthogonal signaling in which transmitted symbols are modulated onto a set of orthonormal waveforms. One class of such schemes uses a Gabor system, which consists of time and frequency shifts of one pulse shape. Multiple carrier modulation, orthogonal frequency division multiplexing (OFDM), or discrete multitone modulation (DMT) are all important examples of Gabor systems. They divide the channel into many small time slots and frequency bands that are matched to the characteristics of a doubly dispersive channel.

To eliminate ISI and ICI in the dispersive channel, a cyclic prefix may be introduced, but this reduces the spectral efficiency. To combat both time-dispersion (multi-path fading) and frequency-dispersion (Doppler spreading), many Gabor systems with good time and frequency localized pulse shapes have been studied. These include coded OFDM [5], Gabor systems with time-frequency product greater than 1 [6], and OFDM with offset-QAM (OQAM) [7] [8], which also corresponds to the Wilson basis [9]. While these orthogonal signaling schemes address the signalling basis thoroughly, they leave the communication and information aspects unexplored.

A large class of linear precoder and equalizer designs based on a variety of criteria [10] [11] [12] [13], have been published for the discrete block transmission model [14], [15], [13], [16]. In this model the channel, together with the digital to analog converter in the transmit side and the analog to digital converter in the receive side, is assumed to be FIR with order less than the transmission block length. Zero padding or zero leading (including cyclic) prefixes are used to eliminate the inter-block interference and convert the system into a matrix map [14]. The implicit underlying assumptions when converting the continuous model into a matrix map are that the precoder and equalizer filter banks are time limited.

In this dissertation we bring the analog basis for precoder and equalizer design directly into the problem, as in [17]. We then optimize with respect to it and with respect to its linear combinations in order to design the analog precoders. We consider an analog channel, and design the precoder and equalizer in two steps. First, we match the subspace for the precoder filter-bank to the channel. Next, we design the precoder filter-bank as linear combinations in a basis for this subspace. In the under-spread linear time-variant channel, the importance of the two design steps becomes clear.

Next, we explore the canonical coordinate geometry of these designs [18]. We show that all the precoder and equalizer designs we consider are, in fact, decompositions of a virtual two-channel problem into a system of canonical coordinates, wherein variables in the canonical message channel are correlated only pairwise with corresponding variables in the canonical measurement channel. This finding clarifies the geometry of precoder and equalizer designs and illustrates that they decompose the two-channel communication problem into what might be called the Shannon channel.

## 1.3 Precoder Design in CDMA Multi-user Multi-path Systems

The linear minimum mean square error (LMMSE) receiver is a widely used linear equalizer in CDMA multiuser systems. To get the LMMSE equalizer, we need to compute a $K \times K$ matrix inversion $R_{yy}^{-1}$, where $K$ is the number of users. This is a time-consuming computation. There are many fast algorithms for the matrix inversion. The conjugate gradient method is one of them, which is guaranteed to converge in $K$ steps. When the number of users $K$ is large, this is still a big burden. It has been discovered that the convergence time of the conjugate gradient method depends on the number of distinct eigenvalues of the matrix $R_{yy}$. This initiates the thought that we might design the precoder to make $R_{yy}$ have a small number of distinct eigenvalues. In the flat fading CDMA system, precoding can be used in both the uplink and downlink. But in the multipath CDMA system, since each mobile receiver receives a signal experiencing different multipath fading, the precoder in the down link can't take care of everyone. We only consider the uplink multipath CDMA system. Suppose the base station can estimate the channel of each user, design the precoder, update the signature waveforms and transmit them back to each user. In the next symbol period, each mobile user uses the updated signature, and the MMSE equalizer is simplified.

### List of References

[1] L. Zheng and D. Tse, "Diversity and multiplexing: A fundanmental tradeoff in multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[2] J. G.Proakis, *Digital Communications, 4th Edition.* McGrawHill, 2000.

[3] L. H.Ozarow, S. Shamai, and A. D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicuar Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[4] A. Scaglione, "Statistical analysis of the capacity of mimo frquency selective rayleigh fading channels with arbitrary number of inputs and outputs," Lausanne, Switzerland, 2002.

[5] B. L. Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proceedings of the IEEE.*, vol. 83, no. 6, pp. 982–996, June 1995.

[6] R. Haas and J. C.Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Personal Communications*, vol. 5, pp. 1–18, 1997.

[7] R. W.Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *The Bell Syst. Tech. J.*, vol. 45, pp. 1775–1796, Dec. 1966.

[8] B. R.Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. on Comm. Tech.*, vol. COM-15, no. 6, pp. 805–811, Dec. 1967.

[9] I. Daubechies, S. Jaffard, and J. L.Journ, "A simple wilson orthonormal basis with exponential decay," *SIAM J. Math. Anal.*, vol. 22, pp. 554–572, 1991.

[10] A. Scaglione, G. B.Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part 1: unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[11] A. Scaglione, S. Barbarossa, and G. B.Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[12] Y. Ding, T. N.Davidson, Z. Luo, and K. M.Wong, "Minimum ber block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2410–2423, Sept. 2003.

[13] D. P.Palomar, J. M.Cioffi, and M. A.Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: a unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[14] Z. Wang and G. B.Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Magazine*, pp. 29–48, May 2000.

[15] A. Scaglione, P. Stoica, S. Barbarossa, G.B.Giannakis, and H.Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[16] G. G.Raleigh and V. K.Jones, "Multivariate modulation and coding for wireless communication," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 6, pp. 851–866, May 1999.

[17] Z. Fan, L. L.Scharf, and J. A.Gubner, "Analog precoder and equalizer design for multichannel communication," in *SPAWC*, Lisbon, Portugal, July 2004.

[18] Z. Fan, L. L.Scharf, and T. N.Davidson, "Canonical coordinate geometry of precoder and equalizer designs for multichannel communications," in *SPAWC*, Lisbon, Portugal, July 2004.

# CHAPTER 2

## Trade-off between Capacity and Diversity for Block Fading Sub-Channels

In this chapter we consider a "block fading sub-channel" communication problem, wherein a transmitter-receiver pair exchanges symbols in flat-fading subchannels to optimally trade off capacity and diversity. Capacity is used to increase data rate and diversity is used to decrease error rate. So there is a trade off to be made, which we characterize. The main result is to characterize admissible pairs of capacity and diversity. Roughly, the sum of diversity and capacity is the rank of a subchannel gain vector.

### 2.1 Channel Model

We consider the channel model in Figure 4, where we fix the time-frequency signal space of dimension $TW$. $\Delta t_c$ and $\Delta f_c$ are respectively coherence time and coherence bandwidth of the channel. We divide this $TW$ signaling space into non-overlapping $\Delta t_c \Delta f_c$ time-frequency cells. Define $K := \frac{T}{\Delta t_c}$, and $L := \frac{W}{\Delta f_c}$, and $N := KL$. In each cell, the signal sees a flat fading channel. The fading coefficients are assumed to be random variables with circularly symmetric complex Gaussian distribution. We assume that in the $(i, j)$th cell, the fading coefficient $h_{ij}$ is distributed as $h_{ij} \sim CN[0, \sigma^2]$. Define vector $\boldsymbol{h}_i := [h_{i,0}, \cdots, h_{i,L-1}]^T$, and the "vec" $\boldsymbol{h} := [\boldsymbol{h}_0^T, \cdots, \boldsymbol{h}_{K-1}^T]^T$, which is a vecing of columns. We assume that the correlation matrix of vector $\boldsymbol{h}$ is $E[\boldsymbol{h}\boldsymbol{h}^H] = \boldsymbol{R}_h$, the rank of $\boldsymbol{R}_h$ is $n \leq N$, and the eigenvalue decomposition of $\boldsymbol{R}_h$ is $\boldsymbol{R}_h = \boldsymbol{U}_h \boldsymbol{\Sigma}_h \boldsymbol{U}_h^H$. Recall that in the introduction, we described the case where $\boldsymbol{R}_h$ is diagonal with full rank $N$. Here, we consider the more general case where $\boldsymbol{R}_h$ has rank $n \leq N$. We call this channel model a *block fading sub-channel model.* In this problem, we only consider the high SNR case, and assume that the receiver knows the channel state information $h_{ij}, i = 0, 1, \cdots, K-1; j = 0, 1, \cdots, L-1$, but the transmitter only knows the distribution for these random variables.

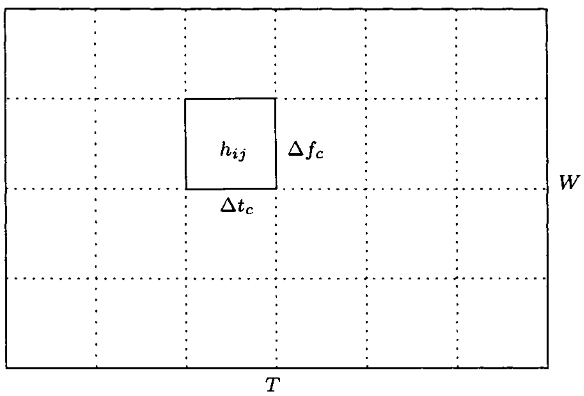



Figure 4. Block fading channel model. $TW$ time-frequency signaling space is partitioned into non-overlapping $\Delta t_c \Delta f_c$ time-frequency cells

Note that the two dimensional subscript does not affect the analysis. For simplicity, we change to a one dimensional "vec" description. The channel coefficients in the $N$ $\Delta t_c \Delta f_c$ blocks are $h_0, \cdots, h_{N-1}$. Redefine the channel vector $\boldsymbol{h} := [h_0, \cdots, h_{N-1}]^T$, $N = KL$.

In the signalling design for this $TW$ signalling space, suppose we separate the $N$ cells far enough apart in time-frequency that the basis for one $\Delta t_c \Delta f_c$ cell is orthogonal to the basis for another, so that there is no inter-cell interference. We transmit $p$ symbols $x_i[k], k = 0, \cdots, p-1$ in the $i$th cell. The $x_i[k]$ are modulated onto the transmitted signal $x_i(t) = \sum_{k=0}^{p-1} x_i[k]\eta_i(k,t)$, where $\eta_i(k,t)$ are an arbitrary orthonormal basis for the $i$th cell. The received signal in the $i$th cell is

$$y_i(t) = x_i(t) + n_i(t) = \sum_{k=0}^{p-1} x_i[k]\eta_i(k,t) + \sum_{k=0}^{p-1} n_i[k]\eta_i(k,t). \tag{1}$$

The $y_i[k] := x_i[k] + n_i[k]$ are the demodulated symbols. $n_i(t)$ is proper, complex white Gaussian noise with spectral density $N_0$.

Each $\Delta t_c \Delta f_c$ time frequency cell has approximately $2p+1$ dimensions according to the Landau-Pollak Theorem. In this time frequency cell, we can transmit $p$ complex symbols which can be decoded at the receiver. Define the transmit random vector in the $i$th block $\boldsymbol{x}_i := [x_i[0], \cdots, x_i[p-1]]^T$, and the received random vector $\boldsymbol{y}_i := [y_i[0], \cdots, y_i[p-1]]^T$, and define the transmit random "vec" $\boldsymbol{x} := [\boldsymbol{x}_0^T, \cdots, \boldsymbol{x}_{N-1}^T]^T$, and the receive random "vec" $\boldsymbol{y} := [\boldsymbol{y}_0^T, \cdots, \boldsymbol{y}_{N-1}^T]^T$.

We use the notation $\doteq$ in [1] to denote exponential equality. For example, $f(\rho) \doteq \rho^a$ means

$$\lim_{\rho \to \infty} \frac{\log f(\rho)}{\log \rho} = a$$

and $\dot{\geq}$, $\dot{\leq}$ are similarly defined.

## 2.2 Mutual Information

The mutual information between the transmit symbol, and the composition of the receive symbol and the channel is

$$I(\mathcal{X}; \mathcal{Y}, \mathcal{H}) = \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} \sum_{\boldsymbol{w}} p_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{v}, \boldsymbol{w})}{p_{\boldsymbol{x}}(\boldsymbol{u})}.$$

The conditional mutual information between the transmit symbols and the receive symbols, conditioned on the channel is

$$\begin{aligned} I(\mathcal{X}; \mathcal{Y}|\mathcal{H}) &= \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} \sum_{\boldsymbol{w}} p_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{v}, \boldsymbol{w})}{p_{\boldsymbol{x}|\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{w})} \\ &= \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} \sum_{\boldsymbol{w}} p_{\boldsymbol{h}}(\boldsymbol{w}) p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{h}}(\boldsymbol{u}, \boldsymbol{v}|\boldsymbol{w}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{v}, \boldsymbol{w})}{p_{\boldsymbol{x}|\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{w})} \\ &= \sum_{\boldsymbol{w}} p_{\boldsymbol{h}}(\boldsymbol{w}) \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{h}}(\boldsymbol{u}, \boldsymbol{v}|\boldsymbol{w}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{v}, \boldsymbol{w})}{p_{\boldsymbol{x}|\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{w})} = E\big[I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})\big] \end{aligned}$$

where the random variable $I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h}) = \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{h}}(\boldsymbol{u}, \boldsymbol{v}|\boldsymbol{h}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{h})}{p_{\boldsymbol{x}|\boldsymbol{h}}(\boldsymbol{u}|\boldsymbol{h})}$ is the mutual information between the random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, given the random channel $\boldsymbol{h}$.

It is easy to show that $I(\mathcal{X};\mathcal{Y},\mathcal{H}) = I(\mathcal{X};\mathcal{H}) + I(\mathcal{X};\mathcal{Y}|\mathcal{H}) = I(\mathcal{X};\mathcal{Y}|\mathcal{H})$. The last equality follows because, without knowledge of the channel realization, symbols are transmitted independently of the channel random variables.

Since in the communication system in Figure 4, given the channel and the transmitted symbols, the symbols received in subchannel $i$ are independent of those transmitted and received in subchannel $j$, or in other words, the channel is memoryless, $I(\mathcal{X};\mathcal{Y}|\boldsymbol{h}) \leq \sum_{i=0}^{N-1} I(\mathcal{X}_i;\mathcal{Y}_i|h_i)$ with equality if the transmitted symbols in different cells are independent, which is assumed hereafter. Thus $I(\mathcal{X};\mathcal{Y}|\boldsymbol{h}) = \sum_{i=0}^{N-1} I(\mathcal{X}_i;\mathcal{Y}_i|h_i)$.

## 2.3 Diversity Gain

In the $TW$ time-frequency signaling space, there are $N$ basic $\Delta t_c \Delta f_c$ time-frequency cells, which provide $n$ linearly independent fading channels, because the rank of $E[\boldsymbol{h}\boldsymbol{h}^H]$ is $n$. At high signal-to-noise ratio (SNR), without using diversity, we know that the probability of the symbol error $P_e$ is inversely proportional to SNR, $P_e \doteq \text{SNR}^{-1}$. If we use diversity, the probability of the symbol error is $P_e \doteq \text{SNR}^{-d}$, where $d$ is the diversity gain. So, $d = \lim_{\text{SNR}\to\infty} \frac{-\log P_e}{\log \text{SNR}}$. In the $TW$ time-frequency signaling space, with correlation between the channel gains $h_i$, there are $n \leq N$ linearly independent fading blocks, and the maximum diversity gain is $n \leq N$.

## 2.4 Multiplexing Gain

We define the data rate of a coding scheme in the $\Delta t_c \Delta f_c$ time-frequency cell as $R_{\Delta t_c \Delta f_c}$(bits/cell), and similarly the data rate of a coding scheme in a $TW$ time-frequency block as $R_{TW}$ (bits/block). Since we can not get diversity in a $\Delta t_c \Delta f_c$ time-frequency cell, but we can in the $TW$ signaling space, the spectral efficiency (in bits/sec per Hz) is $\frac{R_{TW}}{TW} = \gamma \frac{R_{\Delta t_c \Delta f_c}}{\Delta t_c \Delta f_c}$, or in bits $R_{TW} = \gamma N R_{\Delta t_c \Delta f_c}$

with $\gamma \in (0, 1]$. We shall call $r = \gamma N$ the multiplexing gain and note that this gain is never greater than $N$.

The ergodic channel capacity is the maximum data rate at which the source can transmit reliably. We show the ergodic capacity does exist in the communication system in Figure 4, and increases linearly with $\log \text{SNR}$: $C = \log \text{SNR} + \mathcal{O}(1)$, at high SNR.

Reliable communication at rates arbitrarily close to the ergodic capacity requires averaging across many independent realizations of the channel gains over time. Since at high SNR, the capacity increases linearly with $\log \text{SNR}$, we should consider schemes that support a data rate which also increases linearly with $\log \text{SNR}$. As we are considering coding over only a single $TW$ time-frequency signaling space, we must lower the data rate and take into account the randomness of the channel. At high SNR, we only consider coding schemes with data rate in each $\Delta t_c \Delta f_c$ time-frequency cell that increases linearly with $\log \text{SNR}$ approximately as $R_{\Delta t_c \Delta f_c} = p \log \text{SNR} + \mathcal{O}(1)$. Coding schemes whose rate does not increase linearly with $\log \text{SNR}$ have multiplexing gain 0.

### 2.5 Result Statement and Idea of Proof

We derive the trade-off between the multiplexing gain $r$ and the diversity gain $d$ in the $TW$ signaling space for $N$ block fading sub-channels, each of dimension $p$. The derivation follows the idea of the derivation in paper [1] where the authors derived the capacity versus diversity trade-off in multi-antenna channels. First, we prove that the ergodic capacity $C$ (bps/Hz) of the channel does exist and increases linearly with $\log \text{SNR}$ at high SNR: $C \doteq \text{SNR}$, or $C = \log \text{SNR} + \mathcal{O}(1)$ at high SNR. So, there exists a coding scheme whose data rate increases linearly with SNR.

At the data rate $R_{TW} = rp \log \text{SNR}$, with data rate $p \log \text{SNR}$ in each sub-channel, and multiplexing gain $r$, we derive the maximum diversity gain over all coding schemes. Instead of deriving the exact symbol error probability, we give an exponentially equivalent lower bound and an upper bound on $P_e$ [1],

$$P_{\text{out}}(r, \text{SNR}) \dot{\leq} P_e \leq P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage})(r, \text{SNR}),$$

where the last inequality comes from

$$\begin{aligned} P_e(r, \text{SNR}) &= P_{\text{out}}(r, \text{SNR}) P(\text{error}|\text{outage}) + P(\text{error, no outage}) \\ &\leq P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage}). \end{aligned}$$

Here, $P_{\text{out}}(r, \text{SNR})$ is the outage probability defined as the probability that the instantaneous mutual information of the channel is less than or equal to the given rate. Then we try to find the maximum diversity gain over all coding schemes for a given multiplexing gain $r$ by taking

$$\lim_{\text{SNR}\to\infty} \frac{\log P_{\text{out}}(r, \text{SNR})}{\log \text{SNR}} = -d$$

and finding some coding schemes such that

$$\lim_{\text{SNR}\to\infty} \frac{\log \left(P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage})\right)}{\log \text{SNR}} = -d$$

Then we are able to say the error probability is

$$\lim_{\text{SNR}\to\infty} \frac{\log P_e(r, \text{SNR})}{\log \text{SNR}} = -d. \tag{2}$$

This formula is used to explore the trade off between the time-frequency multiplexing gain and the diversity gain.

## 2.6 Ergodic Capacity

We consider the case where the channel changes slowly enough that the receiver can estimate the channel parameters but there is no feedback from the receiver to the transmitter. The channel is underspread, meaning $p >> 1$. So the

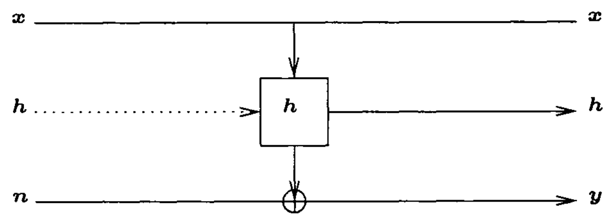



Figure 5. The realization of the channel is known to the receiver but not the transmitter. So equivalently, the channel output consists of the pair $(\boldsymbol{Y}, \boldsymbol{H})$.

channel output consists of the pair $(\boldsymbol{y}, \boldsymbol{h})$ and the distribution of $\boldsymbol{h}$ is known at the transmitter. See Figure 5.

In [2], it is proved that the discrete time-invariant memoryless channel has ergodic capacity equal to the maximum of the mutual information between the transmit symbols and receive symbols. We follow the arguments of [3] to construct a discrete time-invariant memoryless channel. Consider the communication system in Figure 4, which transmits $p$ independent complex symbols in each $\Delta t_c \Delta f_c$ time-frequency cell. We construct $p$ complex super symbols, where each super symbol is an $N := \frac{TW}{\Delta t_c \Delta f_c}$ tuple constructed by joining together one complex symbol from each of the $N$ time-frequency cells of area $\Delta t_c \Delta f_c$. For given values of the channel parameters $\{h_i\}$, we have a memoryless time-invariant super channel, to be defined in due course, for which the coding theorem [2] assures that rates close to $I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})$ can be reliably transmitted with the specific coding scheme above provided $p >> 1$, that is $p$ is very large such that transmission at rates close to $I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})$ with high reliability is possible [2]. As $I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})$ (for finite $T$) is a random variable, strict capacity in the Shannon sense does not exist. Thus the concept of capacity versus outage arises here, where the outage (or failure) probability for a given rate is interpreted as the probability that $I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})$ falls below that rate. When $T \to \infty$, then, under mild regularity conditions, the random variable $\lim_{T\to\infty} \frac{\Delta t_c}{T} I(\mathcal{X}; \mathcal{Y}|\boldsymbol{h})$ equals (due to the weak law of large numbers) to its average with probability 1.

In this sense, strict Shannon capacity does exist and it equals the average mutual information $I$. The mild regularity conditions require that the channel $\boldsymbol{h}$ satisfy the asymptotic mean stationarity property [3].

Consider the communication system in Figure 4. Suppose we repeatedly use the $TW$ signalling subspace to transmit symbols, and the fading is independent identically distributed from channel use to channel use. Obviously, this channel model is a kind of *block interference* (BI) channel defined in [4], and furthermore it satisfies the *compatibility assumption* in [4]. Since in every block the channel is memoryless, the ergodic capacity of this channel is the maximum of the average mutual information.

Given $\boldsymbol{h}$, channel (1) is a Gaussian channel, the optimal $\boldsymbol{x}$ is a complex Gaussian distributed random vector, and the mutual information between the transmit and receive vector is

$$I(\mathcal{X}_i; \mathcal{Y}_i | h_i) = \log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}|h_i|^2}{N_0})$$

where $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}$ is the covariance matrix of $\boldsymbol{x}_i$, and $\boldsymbol{x}_i$ is the symbol vector transmitted in subchannel $i$. We distribute power into each $\Delta t_c \Delta f_c$ time-frequency cell so that $\mathrm{tr}\{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}\} \leq P_i$ with $\sum_{i=0}^{N-1} P_i = PT$ as the power constraint.

The ergodic capacity of the communication system in Figure 4 is

$$C = \max_{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}, \forall i} \frac{1}{TW} E\left( \sum_{i=0}^{N-1} \log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}|h_i|^2}{N_0}) \right).$$

Suppose $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i} = \boldsymbol{U}_i \boldsymbol{\Delta}_i \boldsymbol{U}_i^H$ is the eigen decomposition of $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}$. Then recalling that $TW = Np$, the objective function may be written as

$$C = \max_{\delta_{ik}, \forall i,k} \frac{1}{Np} \sum_{i=0}^{N-1} \sum_{k=0}^{p-1} E \log(1 + \frac{\delta_{ik}\lambda_i}{N_0})$$

subject to

$$\sum_{i=0}^{N-1} \sum_{k=0}^{p-1} \delta_{ik} \leq PT.$$

Here, $\delta_{ik}$ is the $k$th diagonal element of $\Delta_i$ and $\lambda_i := |h_i|^2$.

Since $h_i \sim CN[0, \sigma^2]$, the random variable $x = |h_i|^2$ has density function

$$f(x) = \frac{1}{\sigma^2} \exp\{-\frac{x}{\sigma^2}\}.$$

Define $\rho_{ik} = \frac{\delta_{ik}}{N_0}$, so that

$$\begin{aligned} & E\big[\log(I + \rho_{ik}\lambda_i)\big] \\ = & \int_0^{\infty} \log(1 + \rho_{ik}x)\frac{1}{\sigma^2} \exp\{-\frac{x}{\sigma^2}\}dx \\ = & \frac{1}{\sigma^2} \int_0^{\infty} \exp\{-\frac{x}{\sigma^2}\} \log(1 + \rho_{ik}x)dx \\ = & -\exp\{\frac{1}{\rho_{ik}\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho_{ik}\sigma^2}), \end{aligned}$$

where Ei is the exponential integral function

$$\mathrm{Ei}(-z) = -\int_z^{\infty} e^{-t}t^{-1}dt.$$

This yields

$$C = -\max_{\rho_{ik}, \forall i,k} \frac{1}{Np} \sum_{i=0}^{N-1} \sum_{k=0}^{p-1} \exp\{\frac{1}{\rho_{ik}\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho_{ik}\sigma^2}),$$

subject to

$$\sum_{i=0}^{N-1} \sum_{k=0}^{p-1} \rho_{ik} \le \frac{PT}{N_0}.$$

It is easy to show that the optimal $\rho_{ik} = \rho$, for all $(i, k)$, where $\rho$ is defined to be $\rho := \frac{P}{N_0 W}$. From now on, we use $\rho$ to denote SNR. The capacity is then

$$C = -\exp\{\frac{1}{\rho\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho\sigma^2}).$$

An asymptotic expression for $C$ is

$$C = \left\{ \begin{array}{ll} \rho\sigma^2, & \rho\sigma^2 << 1 \\ \log(1 + \rho\sigma^2) - \mathcal{C}, & \rho\sigma^2 >> 1 \end{array} \right\}$$

where $\mathcal{C} \approx 0.577\cdots$ is the Euler constant. At high SNR,

$$C = \log\rho + \log\sigma^2 - \mathcal{C}.$$

So, at high SNR, $C = \log\rho + \mathcal{O}(1)$.

## 2.7 Outage Probability

Given a rate $R_{TW} = \gamma N R_{\Delta t_c \Delta f_c} = \gamma N p \log\rho = rp\log\rho$ per $TW$ time-frequency signaling space, we want to derive the maximum diversity gain of a coding scheme, or the maximum $d$ such that

$$P_e \doteq \frac{1}{\rho^d}, \text{ or, } \lim_{\rho\to\infty}\frac{\log P_e}{\log\rho} = -d.$$

We can follow [1] to prove that $P_{\text{out}}(r, \text{SNR}) \dot{\leq} P_e(r, \text{SNR})$. The proof is omitted since it is essentially unchanged from [1]. An upper bound for $P_e[r, \text{SNR}]$ is

$$P_e(r, \text{SNR}) \leq P_{\text{out}}(r, \text{SNR}) + P(\text{error}, \text{no outage}). \tag{3}$$

So, next we derive $P_{\text{out}}(r, \rho)$. Following the proof of [1] to choose the circularly Gaussian distributed random vector $\boldsymbol{x}$ and optimize over the input covariance matrix, the outage probability is

$$\begin{aligned} P_{\text{out}}(r,\rho) &= \inf_{p_{\boldsymbol{x}}(\boldsymbol{u})} P\big[I(\mathcal{X};\mathcal{Y}|\boldsymbol{h}) \leq rp\log\rho\big] \\ &= \inf_{(a)} P\big[\sum_{i=0}^{N-1}\log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}|h_i|^2}{N_0}) \leq rp\log\rho\big], \end{aligned}$$

where the area $(a)$ is $\sum_{i=0}^{N-1}\text{trace}(\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}) \leq PT$, and the probability is taken over the random channel vector $\boldsymbol{h}$. Recalling that $\boldsymbol{x}_i$ is a $p$ dimensional vector, and that $Np\frac{P}{W} = PT$, taking $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i} = \frac{P}{W}\boldsymbol{I}$ gives an upper bound on the outage probability,

$$P\big[p\sum_{i=0}^{N-1}\log(1+\rho|h_i|^2) \leq rp\log\rho\big]. \tag{4}$$

On the other hand, since $\sum_{i=0}^{N-1}\text{trace}(\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}) \leq PT$, each eigenvalue of $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}$, for all $i$ is less than or equal to $PT$. Then $PT\boldsymbol{I} - \boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i} \geq 0$, for all $i$, which means that it is a positive semi-definite matrix. If we replace each $\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}$ by $PT\boldsymbol{I}$, the mutual information is increased, since log det is an increasing function on the positive- semi-definite matrix:

$$\begin{aligned}\sum_{i=0}^{N-1} \log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{x}_i\boldsymbol{x}_i}|h_i|^2}{N_0}) &\leq \sum_{i=0}^{N-1} \log\det(\boldsymbol{I} + \frac{PT|h_i|^2\boldsymbol{I}}{N_0}) \\ &= p\sum_{i=0}^{N-1} \log(1 + Np\rho|h_i|^2). \end{aligned} \quad (5)$$

Hence, the outage probability satisfies

$$\begin{aligned} P\big[p\sum_{i=0}^{N-1} \log(1 + Np\rho|h_i|^2) \leq rp\log\rho\big] &\leq P_{\text{out}}(r,\rho) \\ &\leq P\big[p\sum_{i=0}^{N-1} \log(1 + \rho|h_i|^2) \leq rp\log\rho\big]. \end{aligned}$$

At high SNR,

$$\begin{aligned} &\lim_{\rho\to\infty} \frac{\log P\big[p\sum_{i=0}^{N-1} \log(1+\rho|h_i|^2) \leq rp\log\rho\big]}{\log\rho} \\ = &\lim_{\rho\to\infty} \frac{\log P\big[p\sum_{i=0}^{N-1} \log(1+Np\rho|h_i|^2) \leq rp\log\rho\big]}{\log(Np\rho)}. \end{aligned}$$

Therefore for high SNR or when $\rho \to \infty$, the bounds are tight, and we have

$$P_{\text{out}}[r,\rho] \doteq P\big[p\sum_{i=0}^{N-1} \log(1+\rho|h_i|^2) \leq rp\log\rho\big] = P\big[\sum_{i=0}^{N-1} \log(1+\rho|h_i|^2) \leq r\log\rho\big], \quad (6)$$

and we can fix the input distribution to be i.i.d. Gaussian without loss of generality.

### 2.7.1 A Lower Bound

Since $\boldsymbol{h} = [h_0, \cdots, h_{N-1}]^T$ is a circularly symmetric complex Gaussian random vector with covariance matrix $\boldsymbol{R_{hh}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$, we can find a vector $\boldsymbol{\psi} = [\psi_0, \cdots, \psi_{n-1}]^T$ whose elements are independent circularly symmetric complex Gaussian random variables $\psi_i \sim CN[0, \sigma_i^2]$, where $\sigma_i^2$ is the $i$th eigenvalue of $\boldsymbol{\Sigma}$, such that $\boldsymbol{h} = \boldsymbol{U}_n\boldsymbol{\psi}$, where $\boldsymbol{U}_n$ is defined to be the left $n$ columns of $\boldsymbol{U}$. Since the vector $\boldsymbol{h}$ has reduced rank with rank $n$, we can choose a vector of linearly independent elements with dimension $n$ from vector $\boldsymbol{h}$. W.l.o.g. we assume $\boldsymbol{h}_n :=$

$[h_0, \cdot, h_{n-1}]^T$ is this linear independent vector. Vector $\boldsymbol{h}_{N-n} := [h_n, \cdots, h_{N-1}]^T$ is a linear combination of vector $\boldsymbol{h}_n$ as

$$\boldsymbol{h}_{N-n} = \boldsymbol{W}_{(N-n)\times n}\boldsymbol{h}_n, \tag{7}$$

where $\boldsymbol{W}_{(N-n)\times n}$ is an $(N-n)\times n$ matrix whose $(i,j)$th element is defined to be $w_{i,j}$.

**Lemma 1.** *A lower bound on the outage probability in the block fading sub-channel is $P_{out}(r, \log\rho) \dot{\geq} \rho^{-n(1-\frac{r}{N})}$, or an upper bound on the maximum diversity gain is $n(1-\frac{r}{N})$, when the channel rank is n and the multiplexing gain is r, which is to say the spectral efficiency is fixed at $\frac{R_{TW}}{TW} = \frac{rp\log\rho}{pN} = \frac{r}{N}\log\rho$, where $rp = \gamma N$.*

*Proof.* Since the arithmetic average is greater than the geometric average, we get that

$$\begin{aligned}
\sum_{i=0}^{N-1}\log(1+\rho|h_i|^2) &= \log\prod_{i=0}^{N-1}(1+\rho|h_i|^2) \\
&\leq N\log\frac{1}{N}\sum_{i=0}^{N-1}(1+\rho|h_i|^2) \\
&= N\log\left(1+\rho\frac{\sum_{i=0}^{N-1}|h_i|^2}{N}\right) \\
&= N\log\left(1+\rho\frac{\sum_{i=0}^{n-1}|\phi_i|^2}{N}\right).
\end{aligned}$$

The last equality is easily obtained from $\boldsymbol{h} = \boldsymbol{U}_n\boldsymbol{\phi}$. Then, a lower bound on the outage probability is $P\left[N\log(1+\rho\frac{\sum_{i=0}^{n-1}|\phi_i|^2}{N}) \leq r\log\rho\right]$. The event $\left\{N\log(1+\rho\frac{\sum_{i=0}^{n-1}|\phi_i|^2}{N}) \leq r\log\rho\right\}$ is equivalent to event $\{\sum_{i=0}^{n-1}|\phi_i|^2 \leq N(\rho^{\frac{r}{N}-1}-\rho^{-1})\}$ and

$$\bigcap_{i=0}^{n-1}\left\{|\phi_i|^2 \leq \frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n}\right\} \subseteq \left\{\sum_{i=0}^{n-1}|\phi_i|^2 \leq N(\rho^{\frac{r}{N}-1}-\rho^{-1})\right\}.$$

Then, a new lower bound of the outage probability is

$$P\left[\bigcap_{i=0}^{n-1}\left\{|\phi_i|^2 \leq \frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n}\right\}\right] = \prod_{i=0}^{n-1} P\left\{|\phi_i|^2 \leq \frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n}\right\}. \tag{8}$$

This equality is because the $\phi_i$ are independent. Since

$$P\left\{|\phi_i|^2 \leq \frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n}\right\} = 1-\exp\left\{-\frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n\sigma_i^2}\right\},$$

Quantity (8) is equal to

$$\prod_{i=0}^{n-1}\left(1-\exp\left\{-\frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n\sigma_i^2}\right\}\right).$$

It follows that

$$-\lim_{\rho\to\infty}\frac{\log\left[\prod_{i=0}^{n-1}\left(1-\exp\left\{-\frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n\sigma_i^2}\right\}\right)\right]}{\log\rho} = n(1-\frac{r}{N})$$

or equivalently

$$\prod_{i=0}^{n-1} P\left\{|\phi_i|^2 \leq \frac{N(\rho^{\frac{r}{N}-1}-\rho^{-1})}{n}\right\} \doteq \rho^{-n(1-\frac{r}{N})}$$

from L'Hospital's method. This gives a lower bound on the error probability and an upper bound of the maximum diversity gain at a fixed multiplexing gain. □

### 2.7.2 An Upper Bound

**Lemma 2.** *An upper bound of the outage probability in the block fading sub-channel is* $P_{out}(r, \log\rho) \dot{\leq} \rho^{-(n-r)^+}$, *when the channel rank is* $n$ *and the rate is fixed at* $R_{TW} = rp\log\rho$. *Here* $x^+ = \max\{x, 0\}$.

*Proof.* Define $\lambda_i := |h_i|^2$, $\phi_i := \arg(h_i)$ and $\alpha_i := \frac{\log(1+\rho\lambda_i)}{\log\rho}$. The joint density of

$\boldsymbol{\alpha} := [\alpha_0, \cdots, \alpha_{n-1}]^T, \boldsymbol{\phi} := [\phi_0, \cdots, \phi_{n-1}]^T$ is

$$\begin{aligned} f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\boldsymbol{\alpha}, \boldsymbol{\phi}) &= (\log\rho)^n \rho^{(\sum_{l=0}^{n-1}\alpha_l - n)} \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi)^n} \\ &\times \exp\left( -\frac{1}{2}\left[ \sum_{l=0}^{n-1} A_{ll}(\rho^{\alpha_l - 1} - \rho^{-1}) + \sum_{i,j=0 i<j}^{n-1} (A_{ij}^2 + B_{ij}^2)^{\frac{1}{2}} \right.\right. \\ &\times \left.\left. \left((\rho^{\alpha_i - 1} - \rho^{-1})(\rho^{\alpha_j - 1} - \rho^{-1})\right)^{\frac{1}{2}} \cos(\phi_i - \phi_j + \theta_{ij}) \right] \right), \end{aligned}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are constant matrices depend on the correlations between $h_i$, for all $i \in [0, n-1]$, $A_{ij}$ and $B_{ij}$ are their $(i,j)$th entries, and $\theta_{ij}$ is a constant dependent on $\boldsymbol{A}$ and $\boldsymbol{B}$.

By changing variables, the outage probability becomes

$$P_{\text{out}}[r, \rho] = P\left[\sum_{i=0}^{N-1} \alpha_i \leq r\right] = \int_{\{\sum_{i=0}^{N-1}\alpha_i \leq r\}} f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\boldsymbol{\alpha}, \boldsymbol{\phi}) d\boldsymbol{\alpha} d\boldsymbol{\phi}. \tag{9}$$

We analyze $\exp(-\frac{1}{2}h(\boldsymbol{\alpha}, \boldsymbol{\phi}))$ in $f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\boldsymbol{\alpha}, \boldsymbol{\phi})$, where $h(\boldsymbol{\alpha}, \boldsymbol{\phi})$ is defined to be the part within the brackets. Note that $h(\boldsymbol{\alpha}, \boldsymbol{\phi}) \geq 0$, and for any $\alpha_l > 1$, $\exp\left(-\frac{1}{2}h(\boldsymbol{\alpha}, \boldsymbol{\phi})\right)$ decays with $\rho$ exponentially. At high SNR, we can therefore drop the integral over the range with any $\alpha_l > 1$, and replace the integral range from $\mathcal{A} := \left\{\sum_{l=0}^{N-1} \alpha_l \leq r\right\} \bigcap \{\alpha_i \geq 0, \text{ for all } i\} \bigcap \{-\pi \leq \phi_i \leq \pi\}$ to $\mathcal{A}' = \{\alpha_i \leq 1, i = 0, \cdots, n-1\} \bigcap \mathcal{A}$. In $\mathcal{A}'$, as $\rho \to \infty$, $\exp\left(-\frac{1}{2}h(\boldsymbol{\alpha}, \boldsymbol{\phi})\right) \to 1$ when $\alpha_l < 1$, and converges to

$$\exp\left(-\frac{1}{2}\sum_{l=0}^{n-1} A_{ll} + \sum_{i,j=0,i<j}^{n-1} (A_{ij}^2 + B_{ij}^2)^{\frac{1}{2}} \cos(\phi_i - \phi_j + \theta_{ij})\right), \tag{10}$$

when $\alpha_i = 1$, which does not affect the SNR exponent. So at high SNR, (9) is approximately

$$P\left[\sum_{l=0}^{N-1} \alpha_l \leq r\right] \doteq \rho^{-n} \int_{\mathcal{A}'} \rho^{(\sum_{l=0}^{n-1}\alpha_l)} d\boldsymbol{\alpha} d\boldsymbol{\phi}. \tag{11}$$

Since the integrand is constrained to be $\sum_{i=0}^{N-1} \alpha_i \leq r$, $\sum_{i=0}^{n-1} \alpha_i \leq r$. Recall that $\alpha_i \leq 1, i = 0, \cdots, n-1$ which implies that when $r > n$, $\sum_{i=0}^{n-1} \alpha_i \leq n$.

Therefore, an upper bound is

$$P\left[\sum_{l=0}^{N-1} \alpha_l \leq r\right] \doteq \rho^{-n} \int_{\mathcal{A}'} \rho^{\left(\sum_{l=0}^{n-1} \alpha_l\right)} d\boldsymbol{\alpha} d\boldsymbol{\phi} \dot{\leq} \rho^{-n}\rho^{r}, \tag{12}$$

which proves that $P_{\text{out}}(r, \log\rho) \dot{\leq} \rho^{-(n-r)^+}$. □

### 2.7.3 Methods to Derive the Exact Results

**Theorem 1.** *At high SNR, the outage probability is exponentially equivalent to* $P_{out}(r, \rho) \doteq \rho^{-(n-f(\boldsymbol{\alpha}^*))}$ *where* $f(\boldsymbol{\alpha}) := \sum_{l=0}^{n-1} \alpha_l$ *and* $\boldsymbol{\alpha}^*$ *is the solution of the following optimization problem.*

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ s.t. \quad & \sum_{i=0}^{n-1} \alpha_i + \sum_{i=n}^{N-1} \alpha_{\max}^k \leq r \\ & 0 \leq \alpha_i \leq 1, i \in [0, n-1], \end{aligned} \tag{13}$$

*where* $\alpha_{\max}^k$ *is defined to be* $\alpha_{\max}^k := \max(\alpha_i, i \in [0, n-1], s.t.\ w_{(k-n),i} \neq 0.)$

*Proof.* From (11), we know that outage probability is exponentially equal to

$$P_{\text{out}}(r, \rho) \doteq \rho^{-(n-\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*))}, \tag{14}$$

where $\hat{f}(\boldsymbol{\alpha}, \boldsymbol{\phi}) := \sum_{l=0}^{n-1} \alpha_l$ and $\hat{\boldsymbol{\alpha}}^*$, $\hat{\boldsymbol{\phi}}^*$ is the solution of the following optimization problem

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ \text{s.t.} \quad & \sum_{i=0}^{N-1} \alpha_i \leq r \\ & 0 \leq \alpha_i \leq 1, i \in [0, n-1] \\ & -\pi \leq \phi_i \leq \pi, i \in [0, n-1]. \end{aligned} \tag{15}$$

First, we prove that the exponential equivalence in equation (14) does hold. An exponentially equivalent upper bound is

$$P_{\text{out}}(r, \rho) \dot{\leq} (\log\rho)^n \frac{\det(\boldsymbol{A} + j\boldsymbol{B})}{(4\pi\rho)^n} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)} \doteq \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)-n}. \tag{16}$$

We follow the ideas of [1] to get a lower bound. Note that $\hat{f}(\boldsymbol{\alpha}, \boldsymbol{\phi})$ is continuous, therefore for any $\delta > 0$, there exists a neighborhood $\Omega$ of $\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*$, within which $\hat{f}(\boldsymbol{\alpha}, \boldsymbol{\phi}) \geq \hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*) - \delta$. It follows that

$$P_{\text{out}}(r,\rho) \dot{\geq} (\log\rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi\rho)^n} \text{vol}[\Omega \cap \mathcal{A}'] \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)-\delta} \doteq \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)-\delta-n}. \tag{17}$$

Since $P_{\text{out}}(r,\rho) \dot{\geq} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)-\delta-n}$ for any $\delta > 0$, we have $P_{\text{out}}(r,\rho) \dot{\geq} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\phi}}^*)-n}$ which together with the upper bound proves equation (14). This means the integral is dominated by the minimum SNR exponent.

By (7), we know that

$$\begin{aligned} \alpha_k &= \frac{\log\left(1+\rho|\sum_{i=0}^{n-1} w_{(k-n),i}\sqrt{\rho^{\alpha_i-1}-\rho^{-1}}e^{j\phi_i}|^2\right)}{\log\rho} \\ &= \frac{\log\left(1+(\rho^{\alpha_{\max}^k}-1)|\sum_{i=0}^{n-1} w_{(k-n),i}\sqrt{\frac{\rho^{\alpha_i-1}}{\rho^{\alpha_{\max}^k}-1}}e^{j\phi_i}|^2\right)}{\log\rho} \end{aligned}$$

for $k = n, \cdots, N$. It follows that $\lim_{\rho\to\infty} \alpha_k = \alpha_{\max}^k$ at high SNR, dominant convergence theorem can be used to change the optimization problem (15) into (13), and furthermore prove that $P_{\text{out}}(r,\rho) \doteq \rho^{-(n-f(\boldsymbol{\alpha}^*))}$, where $\boldsymbol{\alpha}^*$ is the optimal solution of (13), and $f(\boldsymbol{\alpha}) := \sum_{l=0}^{n-1} \alpha_l$. □

Note that given $\alpha_{\max}^i$, the optimization problem (13) is a linear programming problem and can be solved using the methods in [5]. But there are many ways to choose $\alpha_{\max}^i$ and we experiment to find the best choice. Theoretically, the maximum number of ways to try is $n^{N-n}$. Next, we convert this problem to a linear programming problem.

**Lemma 3.** *The optimization problem (13) can be solved by the following linear programming problem. In other words, the optimal feasible solution of the following* *linear programming problem is also an optimal feasible solution for problem (13):*

$$
\begin{aligned}
\max \quad & \sum_{i=0}^{n-1} \alpha_i \\
s.t. \quad & \sum_{i=0}^{N-1} \alpha_i \leq r \\
& 0 \leq \alpha_i \leq 1, i \in [0, n-1] \\
& \alpha_k \geq \alpha_{k_j}, \forall k \in [n, N-1]; \ \ and \ \ k_j \in \left\{ i \in [0, n-1], s.t. w_{(k-n),i} \neq 0 \right\}.
\end{aligned} \tag{18}
$$

*Proof.* It is obvious that the feasible solution of problem (13) is also a feasible solution of problem (18). For the other direction: define the feasible set of problem (13) as $\mathcal{S}_1$, which is a subset of the feasible set of problem (18) defined as $\mathcal{S}_2$. If the optimal solution of problem (18) lies in set $\mathcal{S}_1$, it is also an optimal solution of problem (13). Next, we prove that the optimal feasible solution of problem (18) does lie in set $\mathcal{S}_1$.

The KKT conditions [5] for problem (18) are

1. $\mu \geq 0; \eta_i \geq 0; \xi_i \geq 0; \rho_{k,k_j} \geq 0;$

   $\forall i \in [0, n-1]; k \in [n, N-1];$

2. $-1 + \mu + \eta_i - \xi_i + \sum_k \rho_{k,i} = 0, \forall i \in [0, n-1];$

   $\forall k, \text{s.t.} \exists k_j = i;$

3. $\mu - \sum_{k_j} \rho_{k,k_j} = 0, \forall k \in [n, N-1];$

4. $\mu(\sum_{i=0}^{N-1} \alpha_i - r) = 0;$

5. $\eta_i(\alpha_i - 1) = 0, \forall i \in [0, n-1];$

6. $\xi_i \alpha_i = 0; \forall i \in [0, n-1];$

7. $\rho_{k,k_j}(\alpha_k - \alpha_{k_j}) = 0; \forall k \in [n, N-1];$

8. $\sum_{i=0}^{N-1} \alpha_i \leq r;$

9. $0 \leq \alpha_i \leq 1, i \in [0, n-1]$

10. $\alpha_k \geq \alpha_{k_j}, \forall k \in [n, N-1].$

If the optimal feasible solution of problem (18) does not lie in set $\mathcal{S}_1$, then there exists an $m \in [n, N-1]$ such that $\alpha_m > \alpha_{m_j}$, for all $m_j$. From condition 7, we obtain $\rho_{m,m_j} = 0$, for all $m_j$. Then, from condition 3, we obtain $\mu = 0$. Then for all $k \in [n, N-1]$, $\sum_{k_j} \rho_{k,k_j} = 0$. Together with condition 1, we obtain $\rho_{k,k_j} = 0$, for all $k \in [n, N-1]$. Putting these into condition 2, we obtain $-1 + \eta_i - \xi_i = 0$, for all $i \in [0, n-1]$. Using this with the condition $\eta_i \geq 0$, for all $i$ and $\xi_i \geq 0$, for all $i$ in condition 1, we see that $\eta_i > 0$, for all $i$, which means $\alpha_i = 1$, for all $i \in [0, n-1]$ from condition 5, and furthermore that $\alpha_k \geq 1$, for all $k \in [n, N-1]$ from condition 10. This violates condition 8.

So, the assumption that the optimal feasible solution of problem (18) lies outside set $\mathcal{S}_1$ is impossible to satisfy. The optimal feasible solution of problem

(18) does lie in set $\mathcal{S}_1$, which means the optimal solution of problem (18) is also the optimal solution of problem (13). □

## 2.8 Error Probability with no Outage

Next we consider the other part of the error probability: the probability of symbol error in the case that no outage occurs, which affects the upper bound on the error probability.

In each $\Delta t_c \Delta f_c$ time-frequency cell, we can transmit $p$ symbols. In the $TW$ time-frequency signaling space, we can transmit $TW = pN$ symbols. Define $\boldsymbol{D}_h = \text{diag}[\boldsymbol{h}]$, which is a diagonal matrix with diagonal elements as the elements of vector $\boldsymbol{h}$ defined in Section 2.1. Then

$$\boldsymbol{y} = (\boldsymbol{D}_h \otimes \boldsymbol{I}_p)\boldsymbol{x},$$

where we omit the noise for notation simplicity, the definitions of $\boldsymbol{x}$, $\boldsymbol{y}$ are in Section 2.1, and $\otimes$ is the Kronecker product. We choose a random code from i.i.d Gaussian ensemble. Suppose $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are two $N$ dimensional complex symbols. If $\boldsymbol{x}_0$ is transmitted, the probability that a ML receiver makes a decision in favor of $\boldsymbol{x}_1$, conditioned on a realization of the channel, is [ [6] p. 318]

$$P\left[\boldsymbol{x}_0 \to \boldsymbol{x}_1 | \boldsymbol{h}\right] = Q\left(\sqrt{\frac{\| (\boldsymbol{D}_h \otimes \boldsymbol{I}_p)\boldsymbol{\Delta x} \|^2}{2N_0}}\right)$$

where $\boldsymbol{\Delta x} = \boldsymbol{x}_1 - \boldsymbol{x}_0$. By the Chernoff bound [7] p. 53, $Q(x) \le e^{-x^2/2}$, and

$$P\big[\boldsymbol{x}_0 \to \boldsymbol{x}_1 | \boldsymbol{h}\big] \le \exp\{-\frac{1}{4N_0} \| (\boldsymbol{D}_h \otimes \boldsymbol{I}_p)\boldsymbol{\Delta x} \|^2\}.$$

Averaging this bound over the ensemble of the random code, we have the average pairwise error probability given the channel realization:

$$P\big[\mathcal{X}_0 \to \mathcal{X}_1 | \boldsymbol{h}\big] \le \prod_{i=0}^{N-1} (1 + \frac{\rho}{2}|h_i|^2)^{-p}.$$

Now at a data rate $R_{TW} = rp\log\rho$ per $TW$ time-frequency signaling space, we have a total of $\rho^{rp}$ codewords. Apply the union bound, to get

$$P\left[\text{error, no outage}|\boldsymbol{h}\right] \leq \rho^{rp} \prod_{i=0}^{N-1} (1+\frac{\rho}{2}|h_i|^2)^{-p}.$$

Changing variables, $1+\rho|h_i|^2 = \rho^{\alpha_i}$, we obtain

$$P\left[\text{error, no outage}|\boldsymbol{h}\right] \dot{\leq} \rho^{-p(\sum_{i=0}^{N-1}\alpha_i - r)}.$$

Averaging with respect to the joint distribution of $\boldsymbol{\alpha}, \boldsymbol{\phi}$ (see Appendix 2.12), we have

$$P\left[\text{error, no outage}\right] \dot{\leq} \int_{\sum_{i=0}^{N-1}\alpha_i \geq r} \rho^{-p(\sum_{i=0}^{N-1}\alpha_i - r)} f(\boldsymbol{\alpha}, \boldsymbol{\phi}) d\boldsymbol{\alpha} d\boldsymbol{\phi}, \tag{19}$$

where $f(\boldsymbol{\alpha}, \boldsymbol{\phi})$ is defined in Appendix 2.12. Define the integral area as $\mathcal{B} := \{\sum_{i=0}^{N-1}\alpha_i \geq r\} \cap \{\alpha_i \geq 0, \text{ for all } i\} \cap \{-\pi \leq \phi_i \leq \pi, \text{ for all } i\}$. By a similar argument for the proof of Lemma 2, we replace the $\mathcal{B}$ by $\mathcal{B}' = \{\alpha_i \leq 1, \text{ for all } i\} \bigcap \mathcal{B}$, and get an exponential equivalent upper bound for $P\left[\text{error, no outage}\right]$,

$$P\left[\text{error, no outage}\right] \dot{\leq} (\log\rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi)^n \rho^{n-pr}} \int_{\mathcal{B}'} \rho^{(1-p)\sum_{i=0}^{n-1}\alpha_i} d\boldsymbol{\alpha} d\boldsymbol{\phi}. \tag{20}$$

### 2.8.1 An Upper Bound

**Lemma 4.** *An exponential equivalent upper bound on the symbol error probability without outage is*

$$P\left[\textit{error, no outage}\right] \dot{\leq} \rho^{-(n-r)}.$$

*Proof.* From (20),

$$\begin{aligned} & P\left[\text{error, no outage}\right] \\ \dot{\leq}\ & (\log\rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi)^n \rho^{n-pr}} \int_{\mathcal{B}'} \rho^{(1-p)\sum_{i=0}^{n-1}\alpha_i} d\boldsymbol{\alpha} d\boldsymbol{\phi} \\ \leq\ & (\log\rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{2^n \rho^{n-pr}} \rho^{(1-p)r} \\ \doteq\ & \rho^{-(n-r)}. \end{aligned}$$

□

### 2.8.2 Methods to Derive a Tighter Upper Bound

From (20), it is easy to show that

$$\begin{aligned} P\big[\text{error, no outage}\big] &\dot{\leq} (\log \rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi)^n \rho^{n-pr}} \int_{\mathcal{B}'} \rho^{(1-p)\sum_{i=0}^{n-1}\alpha_i} d\boldsymbol{\alpha} d\boldsymbol{\phi} \\ &\leq (\log \rho)^n \frac{\det(\boldsymbol{A}+j\boldsymbol{B})}{(4\pi\rho)^n} \rho^{\hat{g}(\tilde{\boldsymbol{\alpha}}^*,\tilde{\boldsymbol{\phi}}^*)} \doteq \rho^{-(n-\hat{g}(\tilde{\boldsymbol{\alpha}}^*,\tilde{\boldsymbol{\phi}}^*))}, \end{aligned}$$

where $\hat{g}(\boldsymbol{\alpha},\boldsymbol{\phi}) := \sum_{i=0}^{n-1}\alpha_i - p\sum_{i=0}^{N-1}\alpha_i + pr$ and $\tilde{\boldsymbol{\alpha}}^*, \tilde{\boldsymbol{\phi}}^*$, is the solution of the following optimization problem

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1}\alpha_i - p\sum_{i=0}^{N-1}\alpha_i + pr \\ \text{s.t.} \quad & \sum_{i=0}^{N-1}\alpha_i \geq r \\ & 0 \leq \alpha_i \leq 1, i \in [0, n-1] \\ & -\pi \leq \phi_i \leq \pi, i \in [0, n-1]. \end{aligned} \tag{21}$$

With the same arguments as those for the proof of Theorem 1, we show that at high SNR, the optimization problem above is

$$\begin{aligned} \max \quad & (1-p)\sum_{i=0}^{n-1}\alpha_i - p\sum_{k=n}^{N-1}\alpha_{\max}^k + pr \\ \text{s.t.} \quad & \sum_{i=0}^{n-1}\alpha_i + \sum_{k=n}^{N-1}\alpha_{\max}^k \geq r \\ & 0 \leq \alpha_i \leq 1, i \in [0, n-1], \end{aligned} \tag{22}$$

where $\alpha_{\max}^k$ is defined to be $\alpha_{\max}^k := \max(\alpha_i, i = 0, \cdots, n-1, \text{s.t.} w_{(k-n),i} \neq 0)$. Therefore,

$$P\big[\text{error, no outage}\big] \dot{\leq} \rho^{-(n-g(\boldsymbol{\alpha}^*))},$$

where $\boldsymbol{\alpha}^*$ is the solution of the above optimization problem (22), and

$$g(\boldsymbol{\alpha}) := \sum_{i=0}^{n-1}\alpha_i - p\sum_{i=0}^{N-1}\alpha_i + pr.$$

**Lemma 5.** *The optimization problems (22) and (13) are equivalent.*

*Proof.* First, we prove that given $\alpha^k_{\max}$, for all $k \in [n, N-1]$, the two optimization problems are equivalent. Given $\alpha^k_{\max}$, for all $k \in [n, N-1]$, problem (13) becomes

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ \text{s.t.} \quad & \sum_{i=0}^{n-1} c_i \alpha_i \leq r \\ & 0 \leq \alpha_i \leq 1, \forall i \in [0, n-1]. \end{aligned} \tag{23}$$

where $c_i \geq 1$ is the weight of $\alpha_i$ in the inequality constraint. Obviously, $c_i - 1$ $\alpha^k_{\max} = \alpha_i$.

Next we prove that the first inequality is actually equality by the KKT condition. The KKT condition of problem (23) is

1. $\mu \geq 0; \eta_i \geq 0; \xi_i \geq 0; \forall i \in [0, n-1];$
2. $-1 + c_i \mu + \eta_i - \xi_i = 0, \forall i \in [0, n-1];$
3. $\mu(\sum_{i=0}^{n-1} c_i \alpha_i - r) = 0;$
4. $\eta_i(\alpha_i - 1) = 0;$
5. $\xi_i \alpha_i = 0;$
6. $\sum_{i=0}^{n-1} c_i \alpha_i \leq r.$

Suppose the first inequality is strict, or $\sum_{i=0}^{n-1} c_i \alpha_i < r$. From condition 3, $\mu = 0$. Then from 2, $-1 + \eta_i - \xi_i = 0$. Together with conditions 1, 5, and 6, we get that $\alpha_i = 1$, for all $i \in [0, n-1]$, which violates the condition 6. Consequently, the first inequality should be equality, and the problem (23) becomes

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ \text{s.t.} \quad & \sum_{i=0}^{n-1} c_i \alpha_i = r \\ & 0 \leq \alpha_i \leq 1, \forall i \in [0, n-1]. \end{aligned} \tag{24}$$

Given $\alpha^k_{\max}$, for all $k \in [n, N-1]$, problem (22) becomes

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1}(1-pc_i)\alpha_i + pr \\ \text{s.t.} \quad & \sum_{i=0}^{n-1} c_i\alpha_i \geq r \\ & 0 \leq \alpha_i \leq 1, \forall i \in [0, n-1]. \end{aligned} \tag{25}$$

As above, we can prove that the first inequality is actually equality by KKT condition, and the problem (25) becomes

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1}(1-pc_i)\alpha_i + pr \\ \text{s.t.} \quad & \sum_{i=0}^{n-1} c_i\alpha_i = r \\ & 0 \leq \alpha_i \leq 1, \forall i \in [0, n-1]. \end{aligned} \tag{26}$$

Because of the first equality constraint, the objective function above is

$$\sum_{i=0}^{n-1}(1-pc_i)\alpha_i + pr = \sum_{i=0}^{n-1}\alpha_i - pr + pr = \sum_{i=0}^{n-1}\alpha_i,$$

which makes the problem (26) exactly the same as the problem (24), and furthermore proves the Lemma. □

From this Lemma, we know that $g(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha})$, and

$$P\big[\text{error, no outage}\big] \dot{\leq} \rho^{-(n-f(\boldsymbol{\alpha}^*))}.$$

## 2.9 Summary of the Results

From $P_{\text{out}}(r,\rho) \leq P_e(r,\rho)$, and Lemma 1, we get that $P_e(r,\rho) \dot{\geq} \rho^{-n(1-\frac{r}{N})}$. From inequality (3), Lemma 2 and Lemma 4, we get that $P_e(r,\rho) \dot{\leq} \rho^{-(n-r)^+}$. In summary,

$$\rho^{-n(1-\frac{r}{N})} \dot{\leq} P_e(r,\rho) \dot{\leq} \rho^{-(n-r)^+}.$$

Equivalently, we get an upper bound and a lower bound on the maximum diversity gain:

$$(n-r)^{+} \leq d \leq n(1-\frac{r}{N}).$$

Figure 6 illustrates the above inequality. The diversity gain $d$ can only take on values in the shaded area.

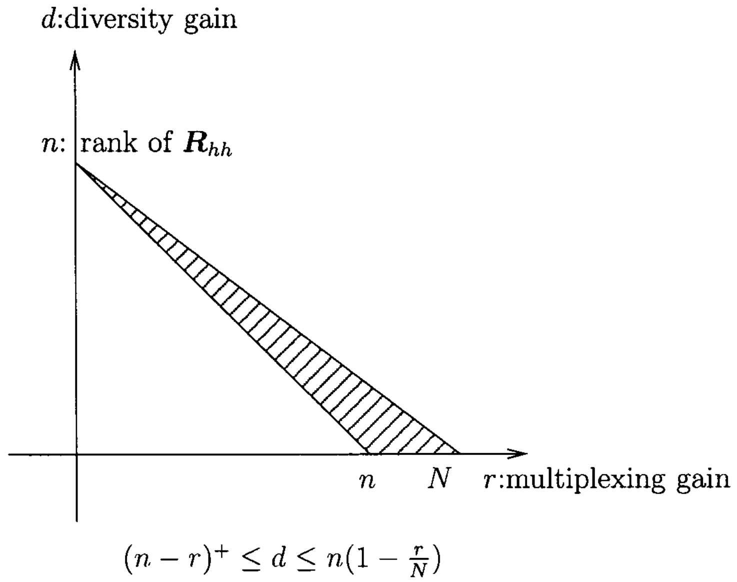



Figure 6. This figure gives the inequality $(n-r)^{+} \leq r \leq n(1-\frac{r}{N})$. The diversity gain $d$ can only be the values in the shaded area.

By solving the linear programming problem (18), we can get the exact exponential equivalent form for the error probability

$$P_e(r,\rho) \doteq \rho^{-(n-f(\boldsymbol{\alpha}^{*}))}$$

and the exact maximum diversity

$$d = n - f(\boldsymbol{\alpha}^{*}).$$

## 2.10 Examples

**Example 1.** *Suppose all the* $w_{ij}$ *in matrix* $\boldsymbol{W}_{(N-n)\times n}$ *(7) are non-zero. Then the optimization problem we need to solve is*

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ s.t. \quad & \sum_{i=0}^{n-1} \alpha_i + (N-n)\alpha_{\max} = r \\ & 0 \le \alpha_i \le 1, i \in [0, n-1], \end{aligned} \tag{27}$$

*where* $\alpha_{\max} := \max\{\alpha_0, \cdots, \alpha_{n-1}\}$. *Without loss of generality, we assume* $\alpha_{\max} = \alpha_0$. *Solving the modified optimization problem, we find that the solution is* $\alpha_i = \frac{r}{N}$, *for all* $i$ *and* $f(\boldsymbol{\alpha}^*) = \frac{rn}{N}$. *Then* $P_{out}(r, \rho) \doteq \rho^{-n(1-\frac{r}{N})}$. *Note that this solution matches the lower bound for outage probability and upper bound for the maximum diversity gain.*

*As a special case, consider the case where the channel vector* $\boldsymbol{h}$ *has full rank, or* $n = N$. *Then the upper bound of maximum diversity matches the lower bound, which is* $n(1-\frac{r}{N}) = N - r$. *So, in this case* $P_e(r, \rho) \doteq \rho^{-(N-r)}$, *and* $d = N - r$.

**Example 2.** *Suppose* $N - n \le n$, *and in each row of* $W_{(N-n)\times n}$ *(7), only one* $w_{ij} \neq 0$, *and in different rows the* $w_{ij}$ *that are nonzero are in different columns. Then, w.l.o.g. we get the following optimization problem*

$$\begin{aligned} \max \quad & \sum_{i=0}^{n-1} \alpha_i \\ s.t. \quad & \sum_{i=0}^{n-1} \alpha_i + \sum_{i=0}^{N-n-1} \alpha_i = r \\ & 0 \le \alpha_i \le 1, i \in [0, n-1]. \end{aligned} \tag{28}$$

*The optimal solution is as follows: when* $r \le n - (N-n)$, $\alpha_i = 1, i \in [N-n, N-n+r-1]$ *and* $\alpha_i = 0$, *for all other* $i$; *when* $r > n - (N-n)$, $\alpha_i = 1, i \in [N-n, n-1]$ *and* $\alpha_i = \frac{r-(2n-N)}{2(N-n)}$, *for all other* $i$. *When* $r \le n - (N-n)$, $f(\boldsymbol{\alpha}^*) = r$,

$P_{out}(r,\rho) \doteq \rho^{-(n-r)}$. *Note that in this case, the solution reaches the upper bound for the outage probability, and a potential lower bound for the maximum diversity. When* $r > n-(N-n)$, $f(\boldsymbol{\alpha}^*) = n - \frac{N-r}{2}$. $P_{out}(r,\rho) \doteq \rho^{\frac{N-r}{2}}$. *Note that this outage probability is between the lower and upper bounds.*

The above two examples are extreme cases where there is only one choice of $\alpha^k_{\max}$, for all$k \in [n, N-1]$. Then the optimization problems are linear programming problems, which are easy to solve. We do not need to solve them using tricks in Lemma 3. We give a more complicated example to illustrate the optimization procedure.

**Example 3.** *Consider the channel model* $[h_0, \cdots, h_3]$ *whose rank is* 2. $[h_0, h_1]$ *is a full rank sub-vector.* $h_2$ *depends on* $h_0, h_1$. $h_3$ *depends on* $h_2$ *Then according to Theorem 1 and Lemma 3, we formulate the optimization problem as follows:*

$$\begin{aligned} \max \quad & \alpha_0 + \alpha_1 \\ s.t. \quad & \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 \le r \\ & 0 \le \alpha_i \le 1, i = 0,1; \alpha_2 \ge \alpha_i, i = 0,1; \alpha_3 \ge \alpha_2. \end{aligned}$$

*Following the simplex method in [5], we can solve this linear programming problem. The optimal solution is* $\alpha_0 = \alpha_1 = \frac{r}{4}$, *and the optimal objective function is* $\frac{r}{2}$. *To verify, we see that* $\alpha_2 = \alpha_3 = \frac{r}{4}$, *which satisfy the inequality constraints, but not the strict inequality.*

*We can check this solution by choosing the* $\alpha^k_{\max}$, *for all* $k = 2, 3$, *and this shows the optimal solution is correct.*

## 2.11 Discussion and Conclusion

In a block fading sub-channel model, when the covariance of channel coefficients has rank $n \le N$, and the multiplexing gain is $r$, the maximum diversity gain is between $(n-r)^+$ and $n(1-\frac{r}{N})$. We have found a simplified optimization problem to get the maximum diversity gain without knowing the values of the entries of the covariance matrix.

From the examples, we see that when the channel has full rank, the maximum diversity gain is $N-r$, which fits the intuition that the more independent channels, the more diversity we can gain, and the smaller error probability we can have.

In the Introduction and Section 2.1, we set the block size for a sub-channel to be $p = \Delta t_c \Delta f_c$. Actually, we do not have to assume that the block size is equal to $\Delta t_c \Delta f_c$ as long as $p >> 1$, which guarantees the existence of the ergodic capacity. Note that from the derivation of the outage probability, we see that the block size $p >> 1$ does not affect the result. Therefore, we have given a general method to derive the outage probability at high SNR for the block fading sub-channel model regardless of the size of the sub-channel block. That is to say, the model we study is a construct for deriving a more general result where the sub-channel block size need not match the channel coherence. The fading correlations are evidently enough to capture all channel effects.

## 2.12 Appendix:Joint Density

In this appendix, we derive the joint density of $\alpha_0, \cdots, \alpha_{n-1}$ in Section 2.7. $\boldsymbol{h}_0$ is a complex circularly symmetric Gaussian distributed random vector. As in [8], we define $\boldsymbol{X}_c := \left[X_{c_0}, \cdots, X_{c_{n-1}}\right]^T = \mathrm{Re}(\boldsymbol{h}_0)$ and $\boldsymbol{X}_s := \left[X_{s_0}, \cdots, X_{s_{n-1}}\right]^T = \mathrm{Im}(\boldsymbol{h}_0)$ with covariance matrices $\boldsymbol{K}_{cc}$ and $\boldsymbol{K}_{ss}$ and cross-covariance matrix $\boldsymbol{K}_{cs}$. Then the joint pdf of $\boldsymbol{X}_c$ and $\boldsymbol{X}_s$ can be written as

$$f_{\boldsymbol{X}_c,\boldsymbol{X}_s}(\boldsymbol{x}_c, \boldsymbol{x}_s) = \frac{1}{(2\pi)^n(\det(\boldsymbol{K}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left[\boldsymbol{x}_c^T \ \boldsymbol{x}_s^T\right] \boldsymbol{K}^{-1} \left[\begin{array}{c} \boldsymbol{x}_c \\ \boldsymbol{x}_s \end{array}\right]\right),$$

where

$$\boldsymbol{K} = \left[\begin{array}{cc} \boldsymbol{K}_{cc} & \boldsymbol{K}_{cs} \\ \boldsymbol{K}_{cs}^T & \boldsymbol{K}_{ss} \end{array}\right]$$

is a nonsingular symmetric matrix. Note that $[\boldsymbol{x}_c^T\ \boldsymbol{x}_s^T]\,\boldsymbol{K}^{-1}\begin{bmatrix}\boldsymbol{x}_c\\ \boldsymbol{x}_s\end{bmatrix} \geq 0$. Denote $\boldsymbol{K}^{-1} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{B}^T & \boldsymbol{D}\end{bmatrix}$. So $(\boldsymbol{K}_{cc} + j\boldsymbol{K}_{cs})^{-1} = \boldsymbol{A} + j\boldsymbol{B}$.

Define random variables $\gamma_0, \cdots, \gamma_{n-1}$, $\phi_0, \cdots, \phi_{n-1}$ in terms of $\boldsymbol{X}_c$, and $\boldsymbol{X}_s$ as $\gamma_i := (X_{c_i}^2 + X_{s_i}^2)^{1/2}$, $\phi_i := \tan^{-1}\left(\frac{X_{s_i}}{X_{c_i}}\right), i = 0, \cdots, n-1$. Define $\lambda_m := |h_m|^2$, and $\alpha_i := \frac{\log(1+\rho\lambda_i)}{\log\rho}$. Note that since $\lambda_l \geq 0$, $\alpha_l \geq 0$, for all $i$. Define $\boldsymbol{\alpha} := [\alpha_0, \cdots, \alpha_{n-1}]^T$. The joint probability density of $\boldsymbol{\alpha}$, and $\boldsymbol{\phi}$ is

$$f_{\boldsymbol{\alpha},\boldsymbol{\phi}}(\boldsymbol{\alpha}, \boldsymbol{\phi}) = (\log\rho)^n \rho^{(\sum_{l=0}^{n-1}\alpha_l - n)} \frac{\det(\boldsymbol{A} + j\boldsymbol{B})}{(4\pi)^n} \exp\left(-\frac{1}{2}h(\boldsymbol{\alpha}, \boldsymbol{\phi})\right), \tag{29}$$

where

$$\begin{aligned} h(\boldsymbol{\alpha}, \boldsymbol{\phi}) \ &:= \ \sum_{l=0}^{n-1} A_{ll}(\rho^{\alpha_l - 1} - \rho^{-1}) + \sum_{i,j=0 i<j}^{n-1} (A_{ij}^2 + B_{ij}^2)^{\frac{1}{2}} \\ &\times \left((\rho^{\alpha_i - 1} - \rho^{-1})(\rho^{\alpha_j - 1} - \rho^{-1})\right)^{\frac{1}{2}} \cos(\phi_i - \phi_j + \theta_{ij}). \end{aligned}$$

And the joint probability density of $\boldsymbol{\alpha}$ is

$$\begin{aligned} f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \ &= \ (\log\rho)^n \rho^{(\sum_{l=0}^{n-1}\alpha_l - n)} \frac{\det(\boldsymbol{A} + j\boldsymbol{B})}{(4\pi)^n} \\ &\times \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp\left(-\frac{1}{2}h(\boldsymbol{\alpha}, \boldsymbol{\phi})\right) d\phi_0 \cdots d\phi_{n-1}. \end{aligned} \tag{30}$$

### List of References


[1] L. Zheng and D. Tse, "Diversity and multiplexing: A fundanmental tradeoff in multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[2] R. G.Gallager, *Information Theory and Reliable Communication.* New York, United States of America: Wiley, 1968.

[3] L. H.Ozarow, S. Shamai, and A. D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicuar Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[4] R. J.Mceliece and W. E.Stark, "Channels with block interference," *IEEE Transactions on Information Theory*, vol. IT-30, no. 1, pp. 44–53, Jan. 1984.

[5] E. K.P.Chong and S. H.Zak, *An Introduction To Optimization, 2nd Edition.* John Wiley and Sons, INC: A Wiley-Interscience Publication, 2001.

[6] E. A.Lee and D. G.Messerschmitt, *Digital Communication, 2nd Edition.* Kluwer Academic Publishers: Boston/Dordrecht/London, 1994.

[7] J. G.Proakis, *Digital Communications, 4th Edition.* McGrawHill, 2000.

[8] R. K.Mallik, "On multivariate rayleigh and exponential distributions," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1499–1515, June 2003.

# CHAPTER 3

## The Approximation of Outage Probability and the Trade-off between Capacity and Diversity for the Frequency-Selective Channel

In this chapter, we derive the capacity and diversity trade-off for the time-invariant, frequency-selective, channel model. The result is $d + r = L$, where $L$ is the number of independent channel taps in an $L-$path model for the channel.

### 3.1 Channel Model

In this chapter, we consider the time-invariant, frequency-selective channel model in case ($b$) of Figures 1 and 2. The signal resolves $L = \frac{W}{\Delta f_c}$ multipaths, approximately. Here, we assume that $L$ is the smallest integer greater than or equal to $\frac{W}{\Delta f_c}$. Then we get a tapped-delay-line channel model [1]

$$h(\tau) = \sum_{m=0}^{L-1} h_m \delta(\tau - \frac{m}{W}),$$

We assume $h_m$ are i.i.d. complex Gaussian, circularly symmetric random variables $h_m \sim CN[0, \frac{\sigma^2}{L}]$, and we assume that the channel independent identically changes from one $\Delta t_c W$ block to another. We assume that the receiver knows the channel state information $h_m$, $m = 0, \cdots, L-1$, but the transmitter only knows the distribution for these random variables. Define the channel vector $\boldsymbol{h} := [h_0, \cdots, h_{L-1}]'$.

In the design for the signalling space, suppose we separate the $\Delta t_c W$ blocks (the number of resolvable paths in a coherence time) far enough in time-frequency that the basis for one $\Delta t_c W$ block is orthogonal to the basis for another, such that there is no inter-block interference. We transmit the signal $u(t)$ in one of the blocks. Then the received signal $v(t)$ is

$$v(t) = \sum_{m=0}^{L-1} h_m u(t - \frac{m}{W}) + n(t), 0 \leq t \leq \Delta t_c,$$

where $n(t)$ is proper, complex white Gaussian noise with density $N_0$. Suppose the transmit signal and the receive signal are all bandlimited to $[-\frac{W}{2}, \frac{W}{2}]$. Then by sampling both the transmit signal and the receive signal at Nyquist frequency $W$, we can get the discrete time model

$$v[\frac{n}{W}] = \sum_{m=0}^{L-1} h_m u[\frac{n}{W} - \frac{m}{W}] + n[\frac{n}{W}].$$

The discrete-time notation is

$$v_n = \sum_{m=0}^{L-1} h_m u_{n-m} + n_n.$$

Put $N$ symbols into one block to get the block system of equations

$$\begin{bmatrix} v_0 \\ \\ \vdots \\ v_{L-1} \\ \\ \vdots \\ v_{N-1} \end{bmatrix} = \begin{bmatrix} h_0 & 0 & \cdots & \cdots & & 0 \\ h_1 & h_0 & & & & 0 \\ \vdots & & \ddots & & & \vdots \\ h_{L-1} & h_{L-2} & \cdots & h_0 & \cdots & 0 \\ 0 & h_{L-1} & \cdots & & & \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & & h_{L-1} & \cdots & h_0 \end{bmatrix} \begin{bmatrix} u_0 \\ \\ \vdots \\ v_{L-1} \\ \\ \vdots \\ u_{N-1} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & \cdots & 0 & h_{L-1} & h_{L-2} & \cdots & h_1 \\ 0 & & 0 & 0 & h_{L-1} & \cdots & h_2 \\ \vdots & & \vdots & \vdots & & \ddots & \vdots \\ & & & & & & h_{L-1} \\ & & & & & & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} u_{-N} \\ \\ \vdots \\ u_{-N-(L-1)} \\ \\ \vdots \\ u_{-1} \end{bmatrix}.$$

Here we omit the noise for simplicity of notation. By forcing the first $L$ [1] samples of the received vector $\boldsymbol{v}$ to be zeros, and forcing the first $L$ samples of the transmit vector $\boldsymbol{u}$ to be same as the last $L$ samples, or in other words applying the cyclic

[1]Eliminating $L-1$ samples is enough to get rid of the inter-block interference. Here we remove 1 more symbol for the convenience of the outage probability derivation.

prefix in the transmit vector, we get the equivalent block system of equations

$$\begin{bmatrix} v_L \\ \vdots \\ \\ \vdots \\ \\ \\ \vdots \\ v_{N-1} \end{bmatrix} = \begin{bmatrix} h_0 & 0 & & \cdots & 0 & h_{L-1} & h_{L-2} & \cdots & h_1 \\ h_1 & h_0 & 0 & & & 0 & h_{L-1} & \cdots & h_2 \\ \vdots & & & \ddots & & & & & \vdots \\ & & & & 0 & & & 0 & h_{L-1} \\ \vdots & & & \ddots & & & & & \vdots \\ h_{L-1} & h_{L-2} & & & & & & & \\ 0 & h_{L-1} & & & & & & & \\ \vdots & \vdots & \ddots & & & & & \ddots & \vdots \\ 0 & 0 & & h_{L-1} & \cdots & & & & h_0 \end{bmatrix} \begin{bmatrix} u_L \\ \vdots \\ \\ \vdots \\ \\ \\ \vdots \\ u_{N-1} \end{bmatrix}$$

Since the equivalent channel matrix is circulant, applying the $M \times M$ Fourier matrix $\mathcal{F}_M$ and inverse Fourier matrix $\mathcal{F}_M^H$ before and after the matrix, where $M := N - L$, we get

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{31}$$

where

$$\boldsymbol{x} := \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix} = \mathcal{F}_M^H \begin{bmatrix} u_L \\ \vdots \\ u_{N-1} \end{bmatrix},$$

$$\boldsymbol{y} := \begin{bmatrix} y_0 \\ \vdots \\ y_{M-1} \end{bmatrix} = \mathcal{F}_M^H \begin{bmatrix} v_L \\ \vdots \\ v_{N-1} \end{bmatrix},$$

and $\boldsymbol{H}$ is the channel matrix $\boldsymbol{H} := \text{diag}[H_0, \cdots, H_{M-1}]$. The complex frequency responses $H_m = \sum_{l=0}^{L-1} h_l e^{\frac{-j2\pi lm}{M}}$ are complex circularly symmetric Gaussian random variables $H_m \sim CN[0, \sigma^2]$. The correlation between $H_m$ and $H_n$ is

$$E[H_m \bar{H}_n] = \frac{\sigma^2}{L} \frac{1 - e^{\frac{-j2\pi L(m-n)}{M}}}{1 - e^{\frac{-j2\pi(m-n)}{M}}}.$$

In summary, this models the OFDM system in Figure 7.

Since we assume that the channel changes independently and identically from block to block with block size $\Delta t_c W$, we have to choose $N$ less than or equal to the block size $\Delta t_c W$. Making $N$ to be the block size of $\Delta t_c W$, we have $L$ redundant

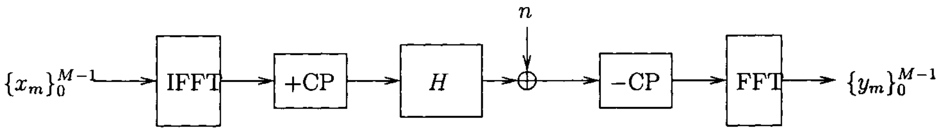



Figure 7. The OFDM system whose equivalent system function is (31), where $\{x_m\}_0^{M-1}$ and $\{y_m\}_0^{M-1}$ represent the transmit and receive $M$ symbol blocks respectively. +CP and −CP are the operation of adding and subtracting the cyclic prefix. $H$ is the $L$ tap FIR channel. In this figure, we omit the transmit and receive filters which can be included in the channel $H$.

symbols in one block. Remember that $L = \frac{W}{\Delta f_c}$. So $N = \Delta t_c \Delta f_c L = (p+1)L$. Since we assume $p + 1 := \Delta t_c \Delta f_c >> 1$, we have $N >> L$. So, $M = N - L = pL$. Then the block size of the equivalent channel model (31) is a multiple of the channel length, which for simplicity of notation in the following derivation explains why we set the block size $\Delta t_c \Delta f_c$ to $p + 1$.

From the equivalent channel model, we see that in each transmission block we use a cyclic prefix to add $L$ redundant symbols. This operation sacrifices the information rate to eliminate the inter-block interference. To make the loss of the information rate as small as possible, we need to make the block size $N >> L$, which is what we assume. So, in the following derivation, we omit the loss of the information rate, and use $M = pL$ to represent the block size of the time-frequency block $\Delta t_c W$.

Recall we use the notation $\doteq$ established in [2] to denote exponential equality. For example, $f(\rho) \doteq \rho^a$ means

$$\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = a$$

and $\dot{\geq}$, $\dot{\leq}$ are similarly defined.

## 3.2 Mutual Information

The mutual information between the transmit symbol, and the composition of the receive symbol and the channel is

$$I(\mathcal{X};\mathcal{Y},\mathcal{H}) = \sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}}\sum_{\boldsymbol{W}} p_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{W}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{W})}{p_{\boldsymbol{x}}(\boldsymbol{u})},$$

where $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{W}$ are realizations of random vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and random matrix $\boldsymbol{H}$.

The conditional mutual information between the transmit symbols and the receive symbols, conditioned on the channel is

$$\begin{aligned}
I(\mathcal{X};\mathcal{Y}|\mathcal{H}) &= \sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}}\sum_{\boldsymbol{W}} p_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{W}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{W})}{p_{\boldsymbol{x}|\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{W})} \\
&= \sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}}\sum_{\boldsymbol{W}} p_{\boldsymbol{H}}(\boldsymbol{W}) p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{H}}(\boldsymbol{u},\boldsymbol{v}|\boldsymbol{W}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{W})}{p_{\boldsymbol{x}|\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{W})} \\
&= \sum_{\boldsymbol{W}} p_{\boldsymbol{H}}(\boldsymbol{W}) \sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}} p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{H}}(\boldsymbol{u},\boldsymbol{v}|\boldsymbol{W}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{W})}{p_{\boldsymbol{x}|\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{W})} \\
&= E\big[I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})\big],
\end{aligned}$$

where

$$I(\mathcal{X};\mathcal{Y}|\boldsymbol{H}) = \sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}} p_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{H}}(\boldsymbol{u},\boldsymbol{v}|\boldsymbol{H}) \log \frac{p_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{H})}{p_{\boldsymbol{x}|\boldsymbol{H}}(\boldsymbol{u}|\boldsymbol{H})}$$

is the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$, given $\boldsymbol{H}$.

It is easy to show that

$$\begin{aligned}
I(\mathcal{X};\mathcal{Y},\mathcal{H}) &= I(\mathcal{X};\mathcal{H}) + I(\mathcal{X};\mathcal{Y}|\mathcal{H}) \\
&= I(\mathcal{X};\mathcal{Y}|\mathcal{H})
\end{aligned}$$

The last equality follows because, without knowledge of the channel realization, symbols are transmitted independently of the channel random variables.

## 3.3 Diversity Gain

In the $\Delta t_c W$ time-frequency signaling space of dimension $M$, there are $L$ basic $\Delta t_c \Delta f_c$ time-frequency cells, which provide $L$ linearly independent fading channels. We can reduce the probability of error by repeatedly transmitting the same symbols carried by the information-bearing signals in different independent channels, for diversity. At high signal-to-noise ratio (SNR), without using diversity, we know that the probability of the symbol error $P_e$ is inversely proportional to SNR, $P_e \doteq \text{SNR}^{-1}$. If we use diversity, the probability of the symbol error is $P_e \doteq \text{SNR}^{-d}$. Here $d$ is defined to be the diversity gain. So, $d = \lim_{\text{SNR}\to\infty} \frac{-\log P_e}{\log \text{SNR}}$. In the $\Delta t_c W$ time-frequency signaling space, since there are $L$ linearly independent fading blocks, the maximum diversity gain is $L$.

## 3.4 Multiplexing Gain

We define the data rate of a coding scheme in the $\Delta t_c \Delta f_c$ time-frequency cell as $R_{\Delta t_c \Delta f_c}$(bits/cell), and similarly the data rate of a coding scheme in a $\Delta t_c W$ time-frequency block as $R_{\Delta t_c W}$ (bits/block). Since we can not get diversity in a $\Delta t_c \Delta f_c$ time-frequency cell, but we can in the $\Delta t_c W$ signaling space, we assume that we have spectral efficiency (in bits/dim) $\frac{R_{\Delta t_c W}}{\Delta t_c W} = \gamma \frac{R_{\Delta t_c \Delta f_c}}{\Delta t_c \Delta f_c}$, or (in bits) $R_{\Delta t_c W} = \gamma L R_{\Delta t_c \Delta f_c}$ with $\gamma \in (0, 1]$ and $L = \frac{W}{\Delta f_c}$. We call $0 < r = \gamma L \leq L$ the multiplexing gain.

The ergodic channel capacity is the maximum data rate at which the source can transmit reliably. We show the ergodic capacity does exist in this communication system, and increases linearly with $\log \text{SNR}$: $C = \log \text{SNR} + \mathcal{O}(1)$ at high SNR.

Reliable communication at rates arbitrarily close to the ergodic capacity requires averaging across many independent realizations of the channel gains over time. Since we are considering coding over only a single $\Delta t_c W$ time-frequency block, we must lower the data rate and take into account the randomness of the channel. Since at high SNR, the capacity increases linearly with $\log \text{SNR}$, we consider schemes that support a data rate that also increases linearly with $\log \text{SNR}$.

At high SNR, we only consider a coding scheme with data rate per $\Delta t_c \Delta f_c$ time-frequency cell that increases linearly with $\log \text{SNR}$ approximately as

$$R_{\Delta t_c \Delta f_c} = (p+1)\log \text{SNR} + \mathcal{O}(1) \approx p \log \text{SNR} + \mathcal{O}(1).$$

Note that coding schemes whose rate does not increase linearly with $\log \text{SNR}$ have multiplexing gain 0.

### 3.5 Result Statement and Idea of Proof

We find that the trade-off between the multiplexing gain $r$ and the diversity gain $d$ in the $\Delta t_c W$ signaling space for the frequency-selective channel is

$$r + d = L.$$

The proof follows the idea of proof in [2], where the authors derived the capacity versus diversity trade-off in multi-antenna channels. First, we prove that the ergodic capacity $C$ (bps/Hz) of the channel does exist and increases linearly with $\log \text{SNR}$ at high SNR: $C \doteq \text{SNR}$, or $C = \log \text{SNR} + \mathcal{O}(1)$. So, there exists a coding scheme whose data rate increases linearly with SNR.

At the data rate $R_{\Delta t_c W} = \gamma L R_{\Delta t_c \Delta f_c} = rp \log \text{SNR}$, with multiplexing gain $r = \gamma L$, we derive the maximum diversity gain for a coding scheme. Instead of deriving the exact symbol error probability, we give an exponentially equivalent lower bound and an upper bound on $P_e$ [2]:

$$P_{\text{out}}(r, \text{SNR}) \dot{\leq} P_e \leq P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage})(r, \text{SNR}),$$

where the last inequality comes from

$$\begin{aligned} P_e(r, \text{SNR}) &= P_{\text{out}}(r, \text{SNR}) P(\text{error}|\text{outage}) + P(\text{error, no outage}) \\ &\leq P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage}). \end{aligned}$$

Here, $P_{\text{out}}(r, \text{SNR})$ is the outage probability defined as the probability that the instantaneous mutual information of the channel is less than or equal to the given rate. Then we find the maximum diversity gain over all coding schemes for a given multiplexing gain $r$ by taking

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log P_{\mathrm{out}}(r, \mathrm{SNR})}{\log \mathrm{SNR}} = -d,$$

and finding some coding schemes such that

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log\left(P_{\mathrm{out}}(r, \mathrm{SNR}) + P(\text{error, no outage})\right)}{\log \mathrm{SNR}} = -d.$$

Then allows us to say the error probability is

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(r, \mathrm{SNR})}{\log \mathrm{SNR}} = -d. \tag{32}$$

This formula is used to explore the trade off between the time-frequency multiplexing gain and the diversity gain.

## 3.6 Ergodic Capacity

We consider the case where the channel changes sufficiently slowly that the receiver can estimate the channel parameters, but there is no feedback from the receiver to the transmitter. The channel is underspread, meaning $p >> 1$. So the channel output consists of the pair $(\boldsymbol{y}, \boldsymbol{H})$ and only the distribution of $\boldsymbol{H}$ is known at the transmitter. See Figure 8, where $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$.

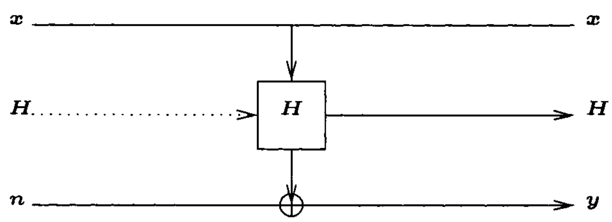



Figure 8. The realization of the channel is known to the receiver but not the transmitter. So equivalently, the channel output consists of the pair $(\boldsymbol{y}, \boldsymbol{H})$.

In [3], it is proved that the discrete time-invariant memoryless channel has ergodic capacity which is the maximum of the mutual information between the transmit symbols and receive symbols. Suppose we transmit symbols in $\mathcal{N}$ consecutive $\Delta t_c W$ blocks, and the fading is independent identically distributed from block to block. Ergodic capacity does have meaning when $\mathcal{N} \to \infty$. We follow the arguments of [4] to construct a discrete time-invariant memoryless channel. Consider the equivalent communication system (31) which transmits $M$ complex symbols in each $\Delta t_c W$ time-frequency block. We may construct $M$ complex super symbols where each super symbol is an $\mathcal{N}$-tuple constructed by joining together one complex symbol from each of the consecutive $\mathcal{N}$ time-frequency cells of area $\Delta t_c W$. For given values of the channel parameters $\{h_i\}$ we have a memoryless time-invariant super channel (31) for which the coding theorem [3] assures that rates close to $I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})$ can be reliably transmitted with the specific coding scheme provided $M >> 1$. Then transmission at rates close to $I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})$ with high reliability is possible [3].

As $I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})$ (for finite $T$) is a random variable, strict capacity in the Shannon sense doesn't exist. Thus the concept of capacity versus outage arises, where the outage (or failure) probability for a given rate is interpreted as the probability that $I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})$ falls below that rate.

Under mild regularity conditions, when $\mathcal{N} \to \infty$, the random variable $\lim_{\mathcal{N}\to\infty} \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} I(\mathcal{X}_i;\mathcal{Y}_i|\boldsymbol{H}_i)$ converges to its average in probability (the weak law of large numbers). In this sense, strict Shannon capacity does exist and it equals the average mutual information $I$. The mild regularity conditions require that the channel $\boldsymbol{H}$ should satisfy the asymptotic mean stationarity property [4].

Consider the communication system. We assume that we transmit symbols in succeeding $\Delta t_c W$ time-frequency blocks, and the fading is independent identically distributed from channel use to channel use. Obviously, the channel model is a kind of *block interference* (BI) channel defined in [5], and furthermore it satisfies the *compatibility assumption* in [5]. Since in every block the channel is memoryless, the ergodic capacity of this channel is the maximum of the average mutual information.

Given $\boldsymbol{H}$, channel (31) is a Gaussian channel. The optimal distribution of $\boldsymbol{x}$ is circularly symmetric complex Gaussian, and therefore the mutual information between the transmit and receive vector is

$$I(\mathcal{X};\mathcal{Y}|\boldsymbol{H}) = \log\det(\boldsymbol{I} + \frac{\boldsymbol{R_{xx}}\boldsymbol{H}^H\boldsymbol{H}}{N_0}),$$

where $\boldsymbol{R_{xx}}$ is the covariance matrix of $\boldsymbol{x}$. The power constraint is $\text{tr}\{\boldsymbol{R_{xx}}\} \leq P\Delta t_c$. The ergodic capacity is

$$\begin{aligned} C &= \max_{\boldsymbol{R_{xx}}} \frac{1}{M} I(\mathcal{X};\mathcal{Y}|\mathcal{H}) \\ &= \max_{\boldsymbol{R_{xx}}} \frac{1}{M} E\big[I(\mathcal{X};\mathcal{Y}|\boldsymbol{H})\big] \\ &= \max_{\boldsymbol{R_{xx}}} \frac{1}{M} E\big[\log\det(\boldsymbol{I} + \frac{\boldsymbol{R_{xx}}\boldsymbol{H}^H\boldsymbol{H}}{N_0})\big] \\ &\leq \max_{\boldsymbol{R_{xx}}} \frac{1}{M} E\big[\sum_{i=0}^{M-1} \log(1 + \frac{\delta_i |H_i|^2}{N_0})\big], \end{aligned}$$

where the $\delta_i$ are diagonal elements of $\boldsymbol{R_{xx}}$. The equality holds when $\boldsymbol{R_{xx}}$ is diagonal. The objective function is

$$C = \max_{\delta_i,\forall i} \frac{1}{M} \sum_{i=0}^{M-1} E\log(1 + \frac{\delta_i\lambda_i}{N_0})$$

subject to

$$\sum_{i=0}^{M-1} \delta_i \leq P\Delta t_c,$$

where $\lambda_i := |H_i|^2$. Since $H_i \sim CN[0,\sigma^2]$, then $|H_i|^2$ has density function

$$f(x) = \frac{1}{\sigma^2}\exp\{-\frac{x}{\sigma^2}\}.$$

Define $\rho_i = \frac{\delta_i}{N_0}$ and compute the expectation

$$\begin{aligned} E\big[\log(I + \rho_i\lambda_i)\big] &= \int_0^\infty \log(1+\rho_i x)\frac{1}{\sigma^2}\exp\{-\frac{x}{\sigma^2}\}dx \\ &= \frac{1}{\sigma^2}\int_0^\infty \exp\{-\frac{x}{\sigma^2}\}\log(1+\rho_i x)dx \\ &= -\exp\{\frac{1}{\rho_i\sigma^2}\}\text{Ei}(-\frac{1}{\rho_i\sigma^2}) \end{aligned}$$

where Ei is the exponential integral function defined to be

$$\mathrm{Ei}(-z) = -\int_z^{\infty} e^{-t}t^{-1}dt.$$

This yields

$$C = -\max_{\rho_i,\forall i} \frac{1}{M}\sum_{i=0}^{M-1} \exp\{\frac{1}{\rho_i\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho_i\sigma^2})$$

subject to

$$\sum_{i=0}^{M-1} \rho_i \leq \frac{P\Delta t_c}{N_0}.$$

It is easy to show that the optimal $\rho_i = \rho$, where $\rho$ is defined to be $\rho := \frac{P}{N_0 W}$. From now on, we use $\rho$ instead of SNR. The capacity is then

$$C = -\exp\{\frac{1}{\rho\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho\sigma^2}).$$

An asymptotic expression for $C$ is

$$C = -\exp\{\frac{1}{\rho\sigma^2}\}\mathrm{Ei}(-\frac{1}{\rho\sigma^2}) = \left\{ \begin{array}{ll} \rho\sigma^2, & \rho\sigma^2 << 1 \\ \log(1+\rho\sigma^2) - \mathcal{C}, & \rho\sigma^2 >> 1 \end{array} \right\},$$

where $\mathcal{C} \approx 0.577\cdots$ is the Euler constant. At high SNR,

$$C = \log\rho + \log\sigma^2 - \mathcal{C}.$$

So, at high SNR, $C \doteq \rho$, or $\lim_{\rho\to\infty} \frac{C}{\log\rho} = 1$. This proves that ergodic capacity increases linearly with $\log\frac{P}{N_0W}$ for large $\frac{P}{N_0W}$.

### 3.7 Outage Probability

Given a rate $R_{\Delta t_c W} = \gamma L R_{\Delta t_c \Delta f_c} = \gamma L p \log\rho = rp\log\rho$ per $\Delta t_c W$ time-frequency block, we want to derive the maximum diversity gain $d$ of a coding scheme with multiplexing gain $r = \gamma L$. Or the maximum $d$ such that

$$P_e \doteq \frac{1}{\rho^d}, \text{ or, } \lim_{\rho\to\infty}\frac{\log P_e}{\log\rho} = -d.$$

We can easily follow [2] to prove that $P_{\text{out}}(r, \text{SNR}) \dot{\leq} P_e(r, \text{SNR})$. The proof is omitted since it is essentially unchanged from [2]. An upper bound for $P_e[r, \text{SNR}]$ is

$$P_e(r, \text{SNR}) \leq P_{\text{out}}(r, \text{SNR}) + P(\text{error, no outage}). \tag{33}$$

So, next we derive the $P_{\text{out}}(r, \rho)$.

Following the proof of [2] to choose the circularly Gaussian distributed random vector $\boldsymbol{x}$ and optimize over the input covariance matrix, the outage probability is

$$\begin{aligned} P_{\text{out}}(r, \rho) &= \inf_{p_{\boldsymbol{x}}(\boldsymbol{u})} P\big[I(\mathcal{X}; \mathcal{Y}|\boldsymbol{H}) \leq rp \log \rho\big] \\ &= \inf_{\text{trace}(\boldsymbol{R}_{\boldsymbol{xx}}) \leq P\Delta t_c} P\big[\log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{xx}}\boldsymbol{H}^H\boldsymbol{H}}{N_0}) \leq rp \log \rho\big], \end{aligned}$$

where the probability is taken over the random channel matrix $\boldsymbol{H}$. Recalling that $\boldsymbol{x}$ is an $M$ dimensional vector, and since $Mp\frac{P}{W} = P\Delta t_c$, taking $\boldsymbol{R}_{\boldsymbol{xx}} = \frac{P}{W}\boldsymbol{I}$ gives an upper bound on the outage probability

$$P\left[\sum_{i=0}^{M-1} \log(1 + \rho|H_i|^2)\right] \leq rp \log \rho.$$

On the other hand, since $\text{trace}(\boldsymbol{R}_{\boldsymbol{xx}}) \leq P\Delta t_c$, each eigenvalue of $\boldsymbol{R}_{\boldsymbol{xx}}$ is less than or equal to $P\Delta t_c$. Then $P\Delta t_c\boldsymbol{I} - \boldsymbol{R}_{\boldsymbol{xx}} \geq 0$, which means that it is a positive semi-definite matrix. If we replace $\boldsymbol{R}_{\boldsymbol{xx}}$ by $P\Delta t_c\boldsymbol{I}$, the mutual information is increased, since log det is an increasing function on positive- semi-definite matrices:

$$\log\det(\boldsymbol{I} + \frac{\boldsymbol{R}_{\boldsymbol{xx}}\boldsymbol{H}^H\boldsymbol{H}}{N_0}) \leq \sum_{i=0}^{M-1} \log(1 + M\rho|h_i|^2).$$

Hence the outage probability satisfies

$$\begin{aligned} P\big[\sum_{i=0}^{M-1} \log(1 + M\rho|H_i|^2) \leq rp \log \rho\big] &\leq P_{\text{out}}(r, \rho) \\ &\leq P\big[\sum_{i=0}^{M-1} \log(1 + \rho|H_i|^2) \leq rp \log \rho\big]. \end{aligned}$$

At high SNR,

$$\lim_{\rho\to\infty} \frac{\log P\big[\sum_{i=0}^{M-1}\log(1+\rho|H_i|^2)\le rp\log\rho\big]}{\log\rho}$$
$$= \lim_{\rho\to\infty} \frac{\log P\big[\sum_{i=0}^{M-1}\log(1+M\rho|H_i|^2)\le rp\log\rho\big]}{\log(M\rho)}.$$

Therefore for high SNR or when $\rho\to\infty$, the bounds are tight, and we have this probability statement for outage probability:

$$P_{\text{out}}[r,\rho] \doteq P\big[\sum_{i=0}^{M-1}\log(1+\rho|H_i|^2)\le rp\log\rho\big]. \tag{34}$$

### 3.7.1 Preliminary

From the channel model (31), we have that the frequency responses $H_m = \sum_{l=0}^{L-1} h_l e^{\frac{-j2\pi lm}{M}}$, for all $m$ are circularly symmetric complex Gaussian random variables, with correlations

$$E[H_m\bar{H}_n] = \frac{\sigma^2}{L}\frac{1-e^{\frac{-j2\pi L(m-n)}{M}}}{1-e^{\frac{-j2\pi(m-n)}{M}}} = \frac{\sigma^2 P}{M}\frac{1-e^{\frac{-j2\pi(m-n)}{p}}}{1-e^{\frac{-j2\pi(m-n)}{M}}}.$$

The last equality uses $M=pL$.

We divide $H_0,\cdots,H_{M-1}$ into $p$ groups, each of which has $L$ elements, as follows

$$\begin{bmatrix} H_0\\ H_p\\ \vdots\\ H_{(L-1)p}\end{bmatrix}\begin{bmatrix} H_1\\ H_{p+1}\\ \vdots\\ H_{(L-1)p+1}\end{bmatrix}\cdots\begin{bmatrix} H_{p-1}\\ H_{2p-1}\\ \vdots\\ H_{Lp-1}\end{bmatrix}. \tag{35}$$

According to the DFT formula $H_m = \sum_{l=0}^{L-1} h_l e^{\frac{-j2\pi lm}{M}}$, we see that the first column is the standard DFT

$$\begin{bmatrix} H_0\\ H_p\\ \vdots\\ H_{(L-1)p}\end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1\\ 1 & e^{\frac{-j2\pi}{L}} & \cdots & e^{\frac{-j2\pi(L-1)}{L}}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & e^{\frac{-j2\pi(L-1)}{L}} & \cdots & e^{\frac{-j2\pi(L-1)(L-1)}{L}}\end{bmatrix}\begin{bmatrix} h_0\\ h_1\\ \vdots\\ h_{L-1}\end{bmatrix}.$$

Define

$$\boldsymbol{h}_0 = \boldsymbol{\mathcal{F}}_L\boldsymbol{h}$$

$$\boldsymbol{h}_m := \begin{bmatrix} H_m \\ H_{p+m} \\ \vdots \\ H_{(L-1)p+m} \end{bmatrix} = \boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{h} = \frac{1}{L}\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{\mathcal{F}}_L^H \boldsymbol{h}_0, m = 0, 1, \cdots, p-1,$$

where $\frac{1}{L}\boldsymbol{\mathcal{F}}_L^H$ is the inverse $L$ point DFT matrix and

$$\boldsymbol{D}_m = \operatorname{diag}\left[1, e^{\frac{-j2\pi m}{M}}, \cdots, e^{\frac{-j2\pi m(L-1)}{M}}\right].$$

Obviously, for arbitrary $n$ and $m$,

$$\boldsymbol{h}_n = \frac{1}{L}\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_{n-m} \boldsymbol{\mathcal{F}}_L^H \boldsymbol{h}_m. \tag{36}$$

$\frac{1}{L}\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{\mathcal{F}}_L^H$ is a unitary matrix. It is easy to prove that all $\boldsymbol{h}_m$ are identically distributed.

Define the polar coordinates $r_m$ and $\theta_m$ of $H_m$:

$$r_m \cos(\theta_m) = \mathrm{Re} H_m, \qquad r_m \sin(\theta_m) = \mathrm{Im} H_m,$$

random vector $\boldsymbol{r}_m := \left[r_m, \cdots, r_{(L-1)p+m}\right]', \boldsymbol{\theta}_m := \left[\theta_m, \cdots, \theta_{(L-1)p+m}\right]'$, $\lambda_m = r_m^2$, and random vector $\boldsymbol{\lambda}_m := \left[\lambda_m, \cdots, \lambda_{(L-1)p+m}\right]'$.

Using this notation, the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$, given channel realization $\boldsymbol{H}$ is

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}|\boldsymbol{H}) &= \sum_{m=0}^{M-1} \log(1+\rho|H_m|^2) = \sum_{m=0}^{M-1} \log(1+\rho r_m^2) = \sum_{m=0}^{M-1} \log(1+\rho\lambda_m) \\ &= \sum_{m=0}^{p-1}\sum_{l=0}^{L-1} \log(1+\rho|H_{lp+m}|^2) = \sum_{m=0}^{p-1}\sum_{l=0}^{L-1} \log(1+\rho r_{lp+m}^2) \\ &= \sum_{m=0}^{p-1}\sum_{l=0}^{L-1} \log(1+\rho\lambda_{lp+m}) \\ &= \sum_{m=0}^{p-1} I_m, \end{aligned} \tag{37}$$

where $I_m$ is the rate

$$I_m := \sum_{l=0}^{L-1} \log(1+\rho|H_{lp+m}|^2). \tag{38}$$

Then, the outage event is

$$\left\{\frac{1}{p}\sum_{m=0}^{M-1}\log\left(1+\rho|H_m|^2\right)\le r\log\rho\right\} = \left\{\frac{1}{p}\sum_{m=0}^{p-1}I_m\le r\log\rho\right\}$$
$$= \left\{\frac{1}{p}\sum_{m=0}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m}\le r\right\}. \quad (39)$$

The fact that

$$\bigcap_{m=0}^{p-1}\{I_m\le r\log\rho\}\subseteq\left\{\frac{1}{p}\sum_{m=0}^{p-1}I_m\le r\log\rho\right\}\subseteq\bigcup_{m=0}^{p-1}\{I_m\le r\log\rho\}$$

yields

$$P\left[\bigcap_{m=0}^{p-1}\{I_m\le r\log\rho\}\right]\le P\left[\frac{1}{p}\sum_{m=0}^{p-1}I_m\le r\log\rho\right]\le P\left[\bigcup_{m=0}^{p-1}\{I_m\le r\log\rho\}\right]. \quad (40)$$

### 3.7.2 Joint Density

Since the elements of $\boldsymbol{h}_0$, $H_{lp}, l\in[0,L-1]$, are i.i.d. complex Gaussian distributed random variables, it is easy to derive the joint Rayleigh distribution

$$f(r_0,\cdots,r_{(L-1)p},\theta_0,\cdots,\theta_{(L-1)p})=\frac{\prod_{l=0}^{L-1}r_l}{(\pi\sigma^2)^L}\exp\left\{-\frac{\sum_{l=0}^{L-1}r_l^2}{\sigma^2}\right\}$$

and

$$f(r_0,\cdots,r_{(L-1)p})=\frac{\prod_{l=0}^{L-1}2r_l}{(\sigma^2)^L}\exp\left\{-\frac{\sum_{l=0}^{L-1}r_l^2}{\sigma^2}\right\}.$$

Since $\boldsymbol{h}_m$, for all $m\in[0,p-1]$, are identically distributed, $\boldsymbol{r}_m := \left[r_m,\cdots,r_{(L-1)p+m}\right]'$, $\boldsymbol{\theta}_m := \left[\theta_m,\cdots,\theta_{(L-1)p+m}\right]'$ consists of i.i.d. random variables, and all $\boldsymbol{r}_m$ and $\boldsymbol{\theta}_m$ are identically distributed, and independent.

The joint distribution of $\boldsymbol{\lambda}_m := \left[\lambda_m,\cdots,\lambda_{(L-1)p+m}\right]'$, and $\boldsymbol{\theta}_m$ is

$$f(\boldsymbol{\lambda}_m,\boldsymbol{\theta}_m)=\frac{2^L}{(\pi\sigma^2)^L}\exp\left\{-\frac{\sum_{l=0}^{L-1}\lambda_{lp+m}}{\sigma^2}\right\}\forall m\in[0,p-1],$$

and furthermore

$$f(\boldsymbol{\lambda}_m)=\frac{1}{(\sigma^2)^L}\exp\left\{-\frac{\sum_{l=0}^{L-1}\lambda_{lp+m}}{\sigma^2}\right\}\forall m\in[0,p-1]. \quad (41)$$

Define $\alpha_m = \frac{\log(1+\rho\lambda_m)}{\log\rho}$. Then, the joint distribution of $\boldsymbol{\alpha}_m := \left[\alpha_m, \cdots, \alpha_{(L-1)p+m}\right]'$, and $\boldsymbol{\theta}_m$ is

$$f(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) = \frac{(2\log\rho)^L}{\rho^L(\pi\sigma^2)^L}\rho^{\sum_{l=0}^{L-1}\alpha_{(lp+m)}} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{(lp+m)}-1} - \frac{1}{\rho}\right)}{\sigma^2}\right\} \forall m \in [0, p-1] \tag{42}$$

and

$$f(\boldsymbol{\alpha}_m) = \frac{(\log\rho)^L}{\rho^L(\sigma^2)^L}\rho^{\sum_{l=0}^{L-1}\alpha_{(lp+m)}} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{(lp+m)}-1} - \frac{1}{\rho}\right)}{\sigma^2}\right\} \forall m \in [0, p-1]. \tag{43}$$

Since $\boldsymbol{h}_m = \frac{1}{L}\boldsymbol{\mathcal{F}}_L\boldsymbol{D}_{m-n}\boldsymbol{\mathcal{F}}_L^H\boldsymbol{h}_n$, by changing the variables as above, we can get the relationship between the $\boldsymbol{\beta}_m := [\boldsymbol{\alpha}_m', \boldsymbol{\theta}_m']'$, for all $m \in [0, p-1]$, as

$$\boldsymbol{\beta}_m = \mathcal{K}_{m-n}(\boldsymbol{\beta}_n).$$

Here $\mathcal{K}_{m-n}$ depends on the difference $m - n$. Then defining $\boldsymbol{\beta} := [\boldsymbol{\beta}_0' \cdots, \boldsymbol{\beta}_{p-1}]'$

$$\begin{aligned} f(\boldsymbol{\beta}) &= f(\boldsymbol{\beta}_0)\prod_{m=1}^{p-1}\delta\left(\boldsymbol{\beta}_m - \mathcal{K}_m(\boldsymbol{\beta}_0)\right) \\ &= f(\boldsymbol{\beta}_n)\prod_{m=0, m\neq n}^{p-1}\delta\left(\boldsymbol{\beta}_m - \mathcal{K}_{m-n}(\boldsymbol{\beta}_n)\right). \end{aligned} \tag{44}$$

### 3.7.3 Lower Bound for the Outage Probability

**Lemma 6.** *An exponentially equivalent lower bound on the outage probability is*

$$P_{out}(r, \rho) \dot{\geq} \rho^{-(L-r)}.$$

*Proof.* From equation (37) and (39), we get

$$\begin{aligned}
\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m) &= \frac{1}{p}\sum_{m=0}^{p-1}\sum_{l=0}^{L-1}\log(1+\rho\lambda_{lp+m}) \\
&= \frac{1}{p}\sum_{m=0}^{p-1}\log\left(\prod_{l=0}^{L-1}(1+\rho\lambda_{lp+m})\right) \\
&\overset{(a)}{\leq} \frac{1}{p}\sum_{m=0}^{p-1}\log\left(\frac{1}{L}(\sum_{l=0}^{L-1}(1+\rho\lambda_{lp+m}))\right)^L \\
&= \frac{1}{p}\sum_{m=0}^{p-1}L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp+m}) \\
&\overset{(b)}{=} L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp}).
\end{aligned} \tag{45}$$

The inequality $(a)$ is the inequality that the geometric average is less than or equal to the arithmetic average. The equality $(b)$ follows from

$$\begin{aligned}
\sum_{l=0}^{L-1}\lambda_{lp+m} = \sum_{l=0}^{L-1}|H_{lp+m}|^2 &= \text{trace}\left\{\boldsymbol{h}_m\boldsymbol{h}_m^H\right\} \\
&= \text{trace}\left\{\frac{1}{L}\boldsymbol{\mathcal{F}}_L\boldsymbol{D}_m\boldsymbol{\mathcal{F}}_L^H\boldsymbol{h}_0\boldsymbol{h}_0^H\left(\frac{1}{L}\boldsymbol{\mathcal{F}}_L\boldsymbol{D}_m\boldsymbol{\mathcal{F}}_L^H\right)^H\right\} \\
&= \text{trace}\left\{\boldsymbol{h}_0\boldsymbol{h}_0^H\right\} \\
&= \sum_{l=0}^{L-1}|H_{lp}|^2 = \sum_{l=0}^{L-1}\lambda_{lp}.
\end{aligned}$$

So, a lower bound for the outage probability is

$$P\left[L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp}) \leq r\log\rho\right] \leq P\left[\frac{1}{p}\sum_{m=0}^{p-1}I_m \leq r\log\rho\right].$$

The event $\left\{L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp}) \leq r\log\rho\right\} = \left\{\sum_{l=0}^{L-1}\lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$. We further give an event which is contained in this event, and thus produces a weaker, but more tractable, lower bound for the outage probability.

Since

$$\bigcap_{l=0}^{L-1}\left\{\lambda_{lp} \leq \frac{\rho^{\frac{r}{L}}-1}{\rho}\right\} \subseteq \left\{\sum_{l=0}^{L-1}\lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\},$$

then a weaker lower bound is

$$P\left[\bigcap_{l=0}^{L-1}\left\{\lambda_{lp} \leq \frac{\rho^{\frac{r}{L}}-1}{\rho}\right\}\right] = \int \cdots \int_0^{\cap_{l=0}^{L-1}\left[\lambda_{lp}\leq\frac{\rho^{\frac{r}{L}}-1}{\rho}\right]} f(\boldsymbol{\lambda}_0)d\boldsymbol{\lambda}_0$$
$$= \int \cdots \int_0^{\cap_{l=0}^{L-1}\left[\lambda_{lp}\leq\frac{\rho^{\frac{r}{L}}-1}{\rho}\right]} \frac{1}{\sigma^{2L}}\exp\left\{-\frac{\sum_{l=0}^{L-1}(\lambda_{lp})}{\sigma^2}\right\}d\boldsymbol{\lambda}_0$$
$$= \left[\int_0^{\lambda_{lp}\leq\frac{\rho^{\frac{r}{L}}-1}{\rho}} \frac{1}{\sigma^2}\exp\left\{-\frac{\lambda_{lp}}{\sigma^2}\right\}d\lambda_{lp}\right]^L$$
$$= \left(1-\exp\left\{-\frac{\rho^{\frac{r}{L}}-1}{\rho\sigma^2}\right\}\right)^L.$$

We can show that

$$\lim_{\rho\to\infty} -\frac{\log\left(1-\exp\left\{-\frac{\rho^{\frac{r}{L}}-1}{\rho\sigma^2}\right\}\right)^L}{\log\rho} = L-r,$$

by repeatedly using L'Hospital's rule, which gives a lower bound for the outage probability

$$P_{\text{out}}(r,\rho) \dot{\geq} \rho^{-(L-r)}.$$

□

*A Geometric Explanation.* Consider the set

$$\{I_0 \leq r\log\rho\} = \left\{\sum_{l=0}^{L-1}\log\left(1+\rho|H_{lp}|^2\right) \leq r\log\rho\right\}.$$

Suppose we have $L$ dimensional complex Euclidean space with coordinates $H_{lp}, l \in [0, L-1]$. It is easy to see that the set above is symmetric with respect to every coordinate in this $L$ dimensional space and invariant to permutation $\prod \boldsymbol{h}_0$, as explained in Figure 9 in 1 dimensional space.

Since $\boldsymbol{h}_m = \frac{1}{L}\boldsymbol{\mathcal{F}}_L\boldsymbol{D}_m\boldsymbol{\mathcal{F}}_L^H\boldsymbol{h}_0$, where $\frac{1}{L}\boldsymbol{\mathcal{F}}_L\boldsymbol{D}_m\boldsymbol{\mathcal{F}}_L^H$ is a unitary matrix, we can get coordinates $H_{lp+m}, l \in [0, L-1]$ by rotating the coordinates $H_{lp}, l \in [0, L-1]$ in the same $L$ dimensional space as explained in Figure 10. In the new co-

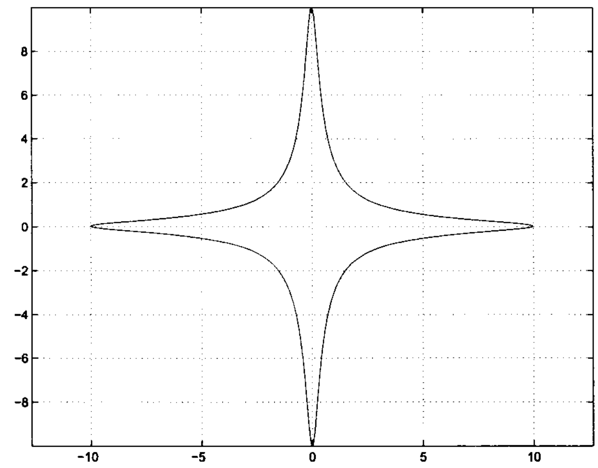



Figure 9. Set $\{I_0 \le r\log\rho\} = \left\{\sum_{l=0}^{L-1}\log\left(1+\rho|H_{lp}|^2\right) \le r\log\rho\right\}$ in one dimensional space

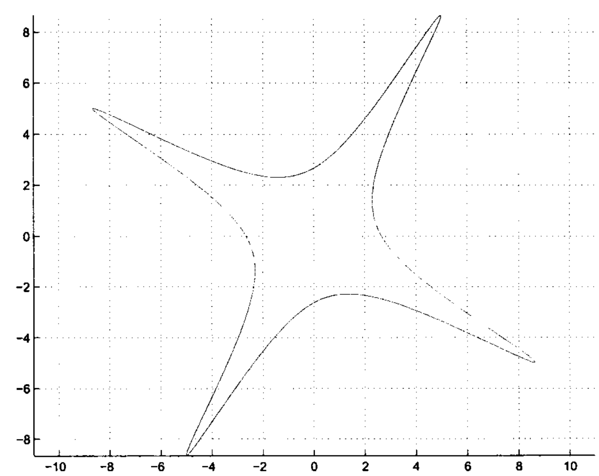



Figure 10. Set $\{I_m \le r\log\rho\} = \left\{\sum_{l=0}^{L-1}\log\left(1+\rho|H_{lp+m}|^2\right) \le r\log\rho\right\}$ in one dimensional space

ordinate system, $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp+m}|^2) \le r\log\rho\right\}$ has the same shape as $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp}|^2) \le r\log\rho\right\}$ in the old coordinate system. We can get $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp+m}|^2) \le r\log\rho\right\}$ for all $m$, by coordinate rotations. Since all the sets are symmetric in their own coordinates, we may search for a convex set that belongs to all the sets.

We stick to coordinates $H_{lp}$ and the set $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp}|^2) \le r\log\rho\right\}$. From the inequality $(a)$ of equation (45), we see that $\left\{\sum_{l=0}^{L-1} \lambda_{lp} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\} \subseteq \left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp}|^2) \le r\log\rho\right\}$, as explained in Figure 11. Geometrically,

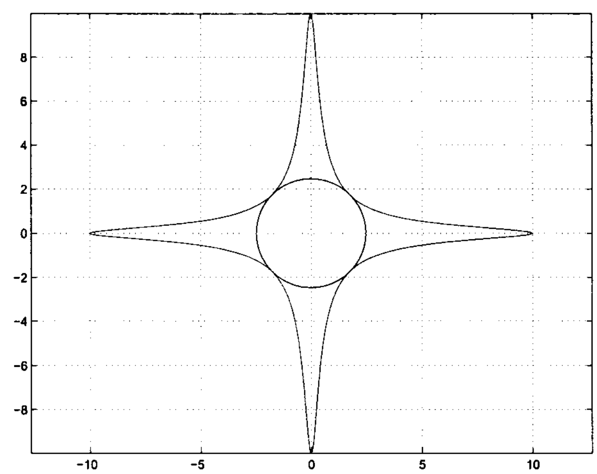



Figure 11. This figure shows that $\left\{\sum_{l=0}^{L-1} \lambda_{lp} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\} \subseteq \left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp}|^2) \le r\log\rho\right\}$

the set $\left\{\sum_{l=0}^{L-1} \lambda_{lp} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$ is a multi-dimensional ball invariant to coordinate rotations. It is easy to prove geometrically that this ball is contained in the set $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp}|^2) \le r\log\rho\right\}$. Similarly in coordinates $H_{lp+m}$, there exists a multidimensional ball $\left\{\sum_{l=0}^{L-1} \lambda_{lp+m} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$ which is inside of the set $\left\{\sum_{l=0}^{L-1} \log(1+\rho|H_{lp+m}|^2) \le r\log\rho\right\}$, as explained in Figure 12. Geometrically the multidimensional ball $\left\{\sum_{l=0}^{L-1} \lambda_{lp+m} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$ is the rotation of $\left\{\sum_{l=0}^{L-1} \lambda_{lp} \le \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$. So they are the same as we proved before. Then this

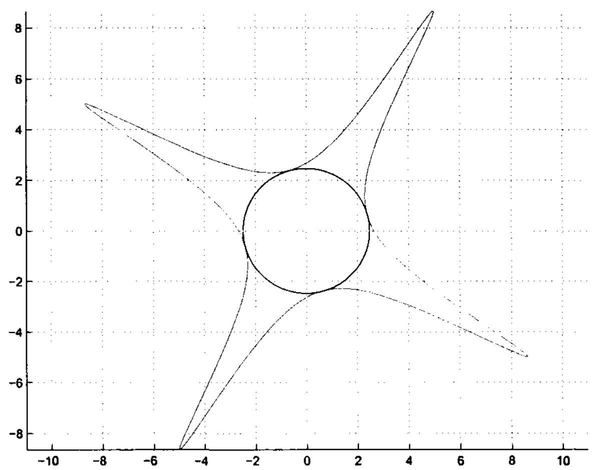



Figure 12. This figure explains that $\left\{\sum_{l=0}^{L-1} \lambda_{lp+m} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\} \subseteq \left\{\sum_{l=0}^{L-1} \log\left(1+\rho|H_{lp+m}|^2\right) \leq r\log\rho\right\}$

multidimensional ball is a convex set contained in all the sets, and therefore the probability of this set is a lower bound on the outage probability, as explained in Figure 13.

### 3.7.4 An Upper Bound

**Lemma 7.** *An exponentially equivalent upper bound for the outage probability is*

$$P_{out}(r,\rho) \dot{\leq} \rho^{-(L-r)}.$$

*Proof.* From (40), we know an upper bound on the outage probability is

$$P\left[\bigcup_{m=0}^{p-1} \{I_m \leq r\log\rho\}\right] \leq \sum_{m=0}^{p-1} P\{I_m \leq r\log\rho\}.$$

By (39),

$$P\{I_0 \leq r\log\rho\} = P\Big[\sum_{l=0}^{L-1} \alpha_{lp} \leq r\Big] = \int_0^{\sum_{l=0}^{L-1}\alpha_{lp}\leq r} f(\boldsymbol{\alpha}_0)d\boldsymbol{\alpha}_0$$

$$= \frac{(\log\rho)^L}{\rho^L\sigma^{2L}} \int_0^{\sum_{l=0}^{L-1}\alpha_{lp}\leq r} \rho^{\sum_{l=0}^{L-1}\alpha_{lp}} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{lp}-1}-\frac{1}{\rho}\right)}{\sigma^2}\right\} d\boldsymbol{\alpha}_0. \quad (46)$$

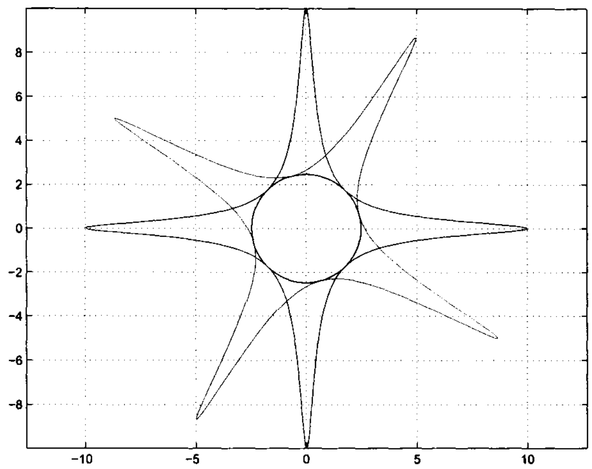



Figure 13. This figure shows that set $\left\{\sum_{l=0}^{L-1}\lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$ is a subset of the intersection of $\left\{\sum_{l=0}^{L-1}\log\left(1+\rho|H_{lp}|^2\right) \leq r\log\rho\right\}$ and $\left\{\sum_{l=0}^{L-1}\log\left(1+\rho|H_{lp+m}|^2\right) \leq r\log\rho\right\}$

We analyze $\exp\left\{-\frac{\sum_{l=0}^{L-1}(\rho^{\alpha_{lp}-1}-\frac{1}{\rho})}{\sigma^2}\right\}$. Note that for any $\alpha_{lp} > 1$, $\exp\{-\frac{\rho^{\alpha_{lp}-1}-\frac{1}{\rho}}{\sigma^2}\}$ decays with $\rho$ exponentially. At high SNR, we therefore ignore the integral over the range with any $\alpha_{lp} > 1$, and replace the integral range from $\mathcal{A} := \{\alpha_{lp} \geq 0, \text{ for all } l\}\bigcap\{\sum_{l=0}^{L-1}\alpha_{lp} \leq r\}$, to $\mathcal{A}' = \{\alpha_{lp} \leq 1, \text{ for all } l\}\bigcap\mathcal{A}$. In $\mathcal{A}'$, as $\rho \to \infty$, $\exp\{-\frac{\rho^{\alpha_{lp}-1}-\frac{1}{\rho}}{\sigma^2}\} \to 1$. So at high SNR, (46) is approximately

$$P\left[\sum_{l=0}^{L-1}\alpha_{lp} \leq r\right] \doteq \frac{(\log\rho)^L}{\rho^L\sigma^{2L}}\int_{\mathcal{A}'}\rho^{\sum_{l=0}^{L-1}\alpha_{lp}}d\boldsymbol{\alpha}_0. \tag{47}$$

We next derive an upper bound for (47). Let $\mathcal{V} = \{\alpha_{lp} \leq 1, \text{ for all } l\}\bigcap\{\alpha_{lp} \geq 0, \text{ for all } l\}$. So, $\mathcal{A}' = \mathcal{V}\bigcap\{\sum_{l=0}^{L-1}\alpha_{lp} \leq r\}$. The upper bound of (47) is derived as follows

$$\begin{aligned}
P[\sum_{l=0}^{L-1}\alpha_{lp} \le r] &\doteq \frac{(\log\rho)^L}{\rho^L\sigma^{2L}}\int_{\mathcal{A}'}\rho^{\sum_{l=0}^{L-1}\alpha_{lp}}d\boldsymbol{\alpha}_0 \\
&\le \frac{(\log\rho)^L}{\rho^L\sigma^{2L}}\int_{\mathcal{V}}\rho^{\sum_{l=0}^{L-1}\alpha_{lp}}d\boldsymbol{\alpha}_0 \\
&\le \frac{(\log\rho)^L}{\rho^L\sigma^{2L}}\int_{\mathcal{V}}\rho^{r}d\boldsymbol{\alpha}_0 \\
&= \frac{(\log\rho)^L}{\rho^L\sigma^{2L}}\rho^{r}.
\end{aligned}$$

Since all $I_m$ are identically distributed, it is easy to prove Lemma 7:

$$P_{\text{out}}(r,\rho) \dot{\le} \rho^{-(L-r)}.$$

□

Combining Lemma 6 and 7, we get that

$$P_{\text{out}}(r,\rho) \doteq \rho^{-(L-r)}$$

and

$$\lim_{\rho\to\infty} -\frac{\log P_{\text{out}}(r,\rho)}{\log\rho} = L - r.$$

### 3.8 Error Probability with no Outage

Next we consider the other part of the error probability: the probability of symbol error in the case that no outage occurs, which affects the upper bound on the error probability.

Recall the channel model in (31). We choose a random code from the i.i.d Gaussian ensemble. Suppose $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are two $M$ dimensional complex symbols.

If $\boldsymbol{X}_0$ is transmitted, the probability that a ML receiver makes a decision in favor of $\boldsymbol{X}_1$, conditioned on a realization of the channel, is in ( [6] p. 318)

$$P[\boldsymbol{X}_0 \to \boldsymbol{X}_1|\boldsymbol{H}] = Q(\sqrt{\frac{\| (\Delta\boldsymbol{X})\boldsymbol{H} \|^2}{2N_0}}),$$

where $\boldsymbol{\Delta X} = \boldsymbol{X}_1 - \boldsymbol{X}_0$. By the Chernoff bound ( [1] p. 53), $Q(x) \leq e^{-x^2/2}$,

$$P[\boldsymbol{X}_0 \to \boldsymbol{X}_1|\boldsymbol{H}] \leq \exp\{-\frac{1}{4N_0} \parallel (\boldsymbol{\Delta X})\boldsymbol{H} \parallel^2\}.$$

Averaging this bound over the ensemble of the random code with density function

$$f_{\boldsymbol{X}_0}(\boldsymbol{X}) = \frac{1}{(2\pi\frac{P}{2W})^M} \exp\{-\frac{\parallel \boldsymbol{X} \parallel^2}{2\frac{P}{2W}}\}$$

we have the average pairwise error probability given the channel realization:

$$P[\mathcal{X}_0 \to \mathcal{X}_1|\boldsymbol{H}] \leq \prod_{i=0}^{M-1}(1+\frac{\rho}{2}|H_i|^2)^{-1}.$$

Now at a data rate $R = \gamma Lp \log\rho = rp\log\rho$ per $\Delta t_c W$ time-frequency signaling space, we have a total of $\rho^{rp}$ codewords. Apply the union bound, to get

$$P[\text{error, no outage}|\boldsymbol{H}] \leq \rho^{rp}\prod_{i=0}^{M-1}(1+\frac{\rho}{2}|H_i|^2)^{-1}.$$

Changing variables, $|H_i|^2 = \rho^{\alpha_i - 1} - \frac{1}{\rho}$, we have

$$P[\text{error, no outage}|\boldsymbol{H}] \dot{\leq} \rho^{-(\sum_{i=0}^{M-1}\alpha_i - pr)}.$$

Averaging with respect to the joint distribution of $\boldsymbol{\beta}$ defined in (44), we obtain

$$P[\text{error, no outage}] \dot{\leq} \int_{\{\sum_{m=0}^{M-1}\alpha_m > rp\}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta})d\boldsymbol{\beta}. \tag{48}$$

**Lemma 8.** *An exponential equivalent upper bound for the error probability with no outage is*

$$P[\textit{error, no outage}] \dot{\leq} \rho^{-(L-r)}.$$

*Proof.* Please see Appendix 3.14-A for the proof and Appendix 3.14-B for an alternative proof. □

Lemmas 6, 7 and 8 prove the following Theorem.

**Theorem 2.** *Fixing the data rate $R = rp\log\rho$, the error probability of the DFT channel (31) is*

$$P_e(r,\rho) \doteq \rho^{-(L-r)},$$

*or in other words, fixing the multiplexing gain $r$, the maximum diversity gain is $d = L - r$.*

## 3.9 An Alternative Way to Derive the Maximum Diversity

In this section, we follow the ideas in [2] to give another prove of Theorem 2, which gives more intuition for this problem. Please see Appendix 3.14-B for the detailed proof.

Recall the process by which we derived the channel model (31). Using the Fourier transform, we transform $h_l, l = 0, \cdots, L-1$ to $H_m, m = 0, \cdots, M-1$ in the frequency domain. $M = pL$ is far greater than the $L$ which is the degrees of freedom of the channel. The $H_m$ are highly correlated. From the proofs, we see that the lower bound and upper bound on the error probability do not depend at all on channel spreading factor $p$. They only depend on the filter order $L$.

## 3.10 Approximations of the Outage Probability at Low and High SNR

In the previous sections, we derive the *scaling* law for the outage probability at high SNR. Motivated by the methods we used in those sections, we give approximations of the outage probability at low and high SNR in this section. Although these results don't contribute to the main subject, they are important in their own right.

From equation (45) in Subsection 3.7.3, we get that

$$\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m) = \frac{1}{p}\sum_{m=0}^{p-1}\log\left(\prod_{l=0}^{L-1}(1+\rho\lambda_{lp+m})\right).$$

At low SNR, approximately

$$\prod_{l=0}^{L-1}(1+\rho\lambda_{lp+m}) \approx 1+\rho\sum_{l=0}^{L-1}\lambda_{lp+m}.$$

So,

$$\begin{aligned}
\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m) &= \frac{1}{p}\sum_{m=0}^{p-1}\log\left(\prod_{l=0}^{L-1}(1+\rho\lambda_{lp+m})\right)\\
&\approx \frac{1}{p}\sum_{m=0}^{p-1}\log\left(1+\rho\sum_{l=0}^{L-1}\lambda_{lp+m}\right)\\
&\overset{(a)}{=} \log\left(1+\rho\sum_{l=0}^{L-1}\lambda_{lp}\right)\\
&\approx \rho\sum_{l=0}^{L-1}\lambda_{lp}.
\end{aligned} \tag{49}$$

The equality $(a)$ is because $\sum_{l=0}^{L-1}\lambda_{lp+m} = \sum_{l=0}^{L-1}\lambda_{lp+n}$, for all $m \neq n$. From equation (45) at low SNR,

$$\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m) \leq L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp}) \approx \rho\sum_{l=0}^{L-1}\lambda_{lp}.$$

Summarizing the above two equations, we can use $L\log(1+\frac{\rho}{L}\sum_{l=0}^{L-1}\lambda_{lp})$ to approximate $\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m)$. Therefore, the outage probability can be approximated by

$$P\left[\sum_{l=0}^{L-1}\lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right],$$

which is the lower bound of the outage probability we derived before. We explain the low SNR case in Figure 14 and 15 that when SNR gets lower and lower, the lower bound of the outage probability gets closer and closer to the outage probability. In paper [7], the author used a similar idea to get the characteristic function of the mutual information of the MIMO-OFDM system at low SNR.

From Subsection 3.7.3, we know that the outage event $\left\{\frac{1}{p}\sum_{m=0}^{M-1}\log(1+\rho\lambda_m) \leq r\right\}$ has a convex set which is $\left\{\sum_{l=0}^{L-1}\lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}\right\}$.

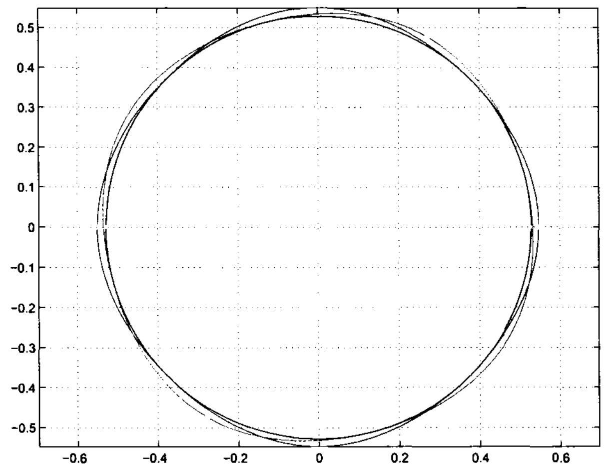



Figure 14. When SNR gets smaller, the probability of the set $\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}$ gets closer to the outage probability.

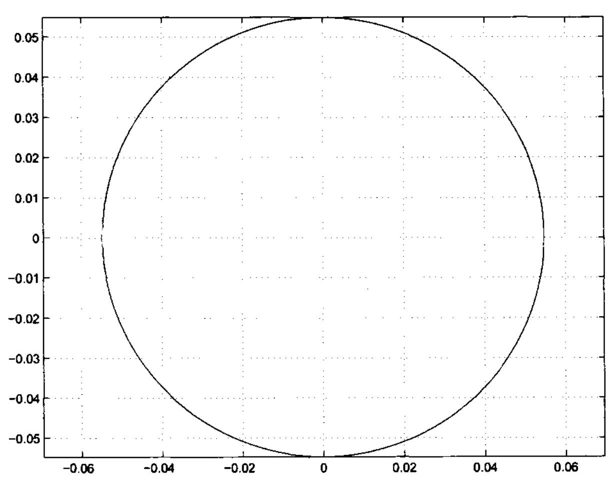



Figure 15. When SNR gets even smaller, the probability of the set $\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}$ gets even closer to the outage probability.

The points $\lambda_{lp} = \frac{\rho^{\frac{r}{L}}-1}{\rho}$, for all $l$ or equivalently $\alpha_{lp} = \frac{r}{L}$ are the intersection points of the boundaries of two sets. In Section 3.9, we show that at high SNR, the outage event is equivalent to

$$\left\{ \sum_{l=0}^{L-1} \alpha_{lp} + L(p-1)\alpha_{\max} \leq r \right\},$$

where $\alpha_{\max} = \max\{\alpha_{lp}, \forall l\}$. The boundary $\sum_{l-0}^{L-1} \alpha_{lp} + L(p-1)\alpha_{\max} = r$ is symmetric for every $\alpha_{lp}$.

We consider the area where $\alpha_0 \geq \alpha_{lp}$, for all $l \neq 0$. It is easy to check that the highest value is where

$$\alpha_0 = \frac{r}{L - \frac{L-1}{p}}, \alpha_{lp} = 0, \forall l \neq 0.$$

When $Lp >> L$,

$$\alpha_0 \approx \frac{r}{L}, \alpha_{lp} = 0, \forall l \neq 0,$$

or equivalently,

$$\lambda_0 \approx \frac{\rho^{\frac{r}{L}} - 1}{\rho}, \lambda_{lp} = 0, \forall l \neq 0.$$

This is approximately equal to

$$\lambda_0 = \frac{L(\rho^{\frac{r}{L}} - 1)}{\rho}, \lambda_{lp} = 0, \forall l \neq 0,$$

at high SNR, a point on the boundary of the convex set $\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}$. So at high SNR, we can also use the convex set to approximate the outage event, and get a unified approximation of the outage probability both at low and high SNR as

$$P\left[\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}} - 1)}{\rho}\right].$$

This outage event is illustrated in Figure 16 and 17.

The condition $Lp >> L$ is equivalent to $M >> L$ which is what is required to minimize the information loss. When the channel is underspread, then $p >> 1$, $M = \Delta t_c W = pL >> L$, and the condition is justified.

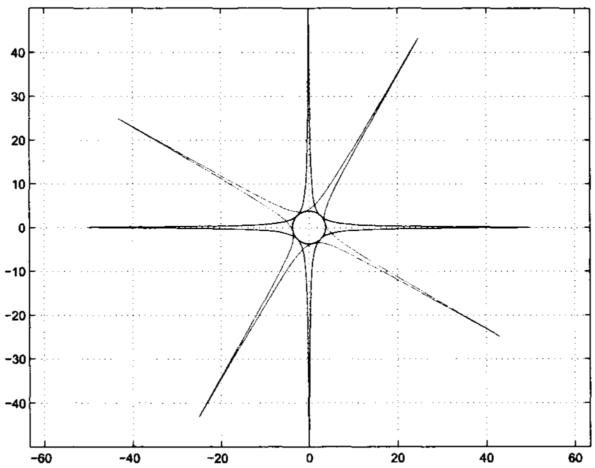



Figure 16. When SNR gets larger, the probability of the set $\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}$ gets closer to the outage probability event.

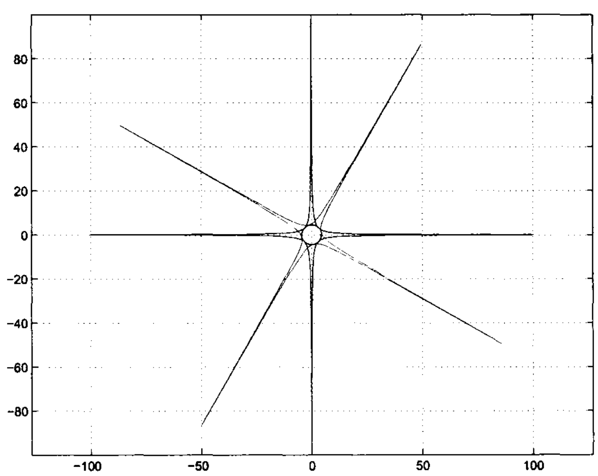



Figure 17. When SNR gets even larger, the probability of the set $\sum_{l=0}^{L-1} \lambda_{lp} \leq \frac{L(\rho^{\frac{r}{L}}-1)}{\rho}$ gets even closer to the outage probability event.

### 3.11 More General Channel Models

We have assumed that the channel changes independently and identically from block to block, with block size $N$. Or in other words, the channel is time-invariant in one block and changes independently from block to block. In this case, we choose the block length for the block input-output model as $N = \Delta t_c W$, which actually has to be less than or equal to $N$, and we consider the outage probability for one block. For a channel that is time-invariant, the block size $N$ in the model may increase without bound, in which case rate loss $\frac{L}{N}$ may be reduced to zero.

We have supposed that the original tapped-delay-line channel has independent identically distributed channel coefficients. It is easy to find that the i.i.d. constraint can be relaxed to linear independence, so that the channel coefficients have full rank covariance, in which case the proofs are essentially the same.

Another constraint is the block length $M = pL$ is a multiple of the number of channel taps. If it is not, all of the results above still hold, but not all the proofs. The ergodic capacity arguments do not change. We can still use the method in Subsection 3.7.3 to derive the lower bound for the outage probability, but the geometric explanation doesn't apply. Both the lower bound and the upper bound on the outage probability and the error probability without outage can be proved using alternative methods in Section 3.9.

### 3.12 Outage Probability at High SNR for a MIMO-OFDM System

We assume a MIMO system with $N_t$ transmit antennas and $N_r$ receive antennas. In each channel, we use signals with time duration $T < \Delta t_c$ and bandwidth $W > \Delta f_c$. Then the signal in each channel resolves a time-invariant frequency-selective channel with $L := \frac{W}{\Delta f_c}$ channel taps. Following the ideas in Section 3.1, we pose the channel model from the $n_t$th transmit antenna to the $n_r$th receive antenna as

$$\begin{bmatrix} y_{n_r}[0] \\ \vdots \\ y_{n_r}[M-1] \end{bmatrix} = \begin{bmatrix} H_{n_r,n_t}[0] & & & \\ & H_{n_r,n_t}[1] & & \\ & & \ddots & \\ & & & H_{n_r,n_t}[M-1] \end{bmatrix} \begin{bmatrix} x_{n_t}[0] \\ \vdots \\ x_{n_t}[M-1] \end{bmatrix},$$

where $H_{n_r,n_t}[n] = \sum_{l=0}^{L-1} h_{n_r,n_t}[l]e^{\frac{-j2\pi nl}{M}}$ is a circularly symmetric complex Gaussian random variable $H_{n_r,n_t}[n] \sim CN[0,\sigma^2]$. The correlation between $H_{n_r,n_t}[m]$ and $H_{n_r,n_t}[n]$ is

$$E[H_{n_r,n_t}[m]\bar{H}_{n_r,n_t}[n]] = \frac{\sigma^2}{L}\frac{1-e^{\frac{-j2\pi L(i-k)}{M}}}{1-e^{\frac{-j2\pi(i-k)}{M}}}.$$

Considering all the channels from $N_t$ transmit antennas to $N_r$ receive antennas, we get the following channel model:

$$\begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{M-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_0 & & & \\ & \boldsymbol{H}_1 & & \\ & & \ddots & \\ & & & \boldsymbol{H}_{M-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_{M-1} \end{bmatrix}, \qquad (50)$$

where $\boldsymbol{H}_m$ is an $N_r \times N_t$ matrix with i.i.d. circularly symmetric complex Gaussian distributed random elements, whose $(i,j)$th element is the $m$th frequency response of the channel from the $j$th transmit antenna to the $i$th receive antenna, $\boldsymbol{x}_m := [x_0[m], \cdots, x_{n_t-1}[m]]'$ and $\boldsymbol{y}_m := [y_0[m], \cdots, y_{n_r-1}[m]]'$.

As in the SISO case, we divide the channel matrices into $p$ groups, each of which has $L$ elements as follows,

$$\begin{bmatrix} \boldsymbol{H}_0 \\ \boldsymbol{H}_p \\ \vdots \\ \boldsymbol{H}_{(L-1)p} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_{p+1} \\ \vdots \\ \boldsymbol{H}_{(L-1)p+1} \end{bmatrix} \cdots \begin{bmatrix} \boldsymbol{H}_{p-1} \\ \boldsymbol{H}_{2p-1} \\ \vdots \\ \boldsymbol{H}_{Lp-1} \end{bmatrix}. \qquad (51)$$

According to the DFT formula $H_{n_r,n_t}[n] = \sum_{l=0}^{L-1} h_{n_r,n_t}[l]e^{\frac{-j2\pi nl}{M}}$, we see that the first column is a standard DFT

$$\begin{bmatrix} \boldsymbol{H}_0 \\ \boldsymbol{H}_p \\ \vdots \\ \boldsymbol{H}_{(L-1)p} \end{bmatrix} = \left( \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\frac{-j2\pi}{L}} & \cdots & e^{\frac{-j2\pi(L-1)}{L}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{\frac{-j2\pi(L-1)}{L}} & \cdots & e^{\frac{-j2\pi(L-1)(L-1)}{L}} \end{bmatrix} \otimes \boldsymbol{I}_{L\times L} \right) \begin{bmatrix} \boldsymbol{h}_0 \\ \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_{L-1} \end{bmatrix}, \qquad (52)$$

where $\otimes$ is the Kronecker product, and $\boldsymbol{h}_n$ is an $n_r \times n_t$ matrix whose $(i,j)$th element is the $n$th tap of the channel model from the $j$th transmitter to the $i$th receiver. With $\boldsymbol{h} := [\boldsymbol{h}_0 \cdots \boldsymbol{h}_{L-1}]'$ and $\boldsymbol{\mathcal{H}}_m := [\boldsymbol{H}_m, \boldsymbol{H}_{p+m} \cdots, \boldsymbol{H}_{(L-1)p+m}]'$, then

$$\boldsymbol{\mathcal{H}}_0 = (\boldsymbol{\mathcal{F}}_L \otimes \boldsymbol{I}_{L\times L})\boldsymbol{h}.$$

The relationships between $\boldsymbol{\mathcal{H}}_m$, $\boldsymbol{h}$ and $\boldsymbol{\mathcal{H}}_n$ are

$$\boldsymbol{\mathcal{H}}_m = ((\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m) \otimes \boldsymbol{I}_{L\times L})\,\boldsymbol{h} = \left(\frac{1}{L}(\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{\mathcal{F}}_L^H) \otimes \boldsymbol{I}_{L\times L}\right)\boldsymbol{\mathcal{H}}_0$$

and

$$\boldsymbol{\mathcal{H}}_m = \left(\frac{1}{L}(\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_{m-n} \boldsymbol{\mathcal{F}}_L^H) \otimes \boldsymbol{I}_{L\times L}\right)\boldsymbol{\mathcal{H}}_n. \tag{53}$$

$\left(\frac{1}{L}(\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{\mathcal{F}}_L^H) \otimes \boldsymbol{I}_{L\times L}\right)$ is a unitary matrix. It is easy to prove that all $\boldsymbol{\mathcal{H}}_m$ are identically distributed. We assume $p > (N_t + N_r - 1)$.

### 3.12.1 Outage Probability

Following the arguments in the derivation of the SISO case and in [2], we can choose the Gaussian distributed input vector and obtain the exponential equivalent outage probability

$$P_{\text{out}}(r,\rho) \doteq P\left[\sum_{m=0}^{M-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_m \boldsymbol{H}_m^H) \leq rp\log\rho\right].$$

#### Block-Fading Sub-Channel Model

In this Subsection, we consider a special case of the frequency-selective channel model, the block-fading sub-channel model, where the columns of the channels in (51) are the same, or the channel does not change during $p$ symbols. The outage probability for this block fading channel is

$$P_{\text{out}}(r,\rho) = P\left[I(\mathcal{X};\mathcal{Y}|\boldsymbol{H}) \leq rp\log\rho\right] = P\left[\sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp} \boldsymbol{H}_{lp}^H) \leq r\log\rho\right].$$

Suppose the eigenvalues of $\boldsymbol{H}_{lp}\boldsymbol{H}_{lp}^H$ are $\boldsymbol{\lambda}_{lp} = [\lambda_{lp,0}, \cdots, \lambda_{lp,K-1}]'$, where $K = \min(N_t, N_r)$. The eigenvalues are supposed to be in increasing order. So, the joint distribution of these eigenvalues is

$$p(\boldsymbol{\lambda}_0, \cdots, \boldsymbol{\lambda}_{(L-1)p}) = C_{N_t,N_r}^{-K} \prod_{l=0}^{L-1} \left( \prod_{i=0}^{K-1} \lambda_{lp,i}^{|N_t-N_r|} \prod_{i<j} (\lambda_{lp,i} - \lambda_{lp,j})^2 \right) e^{-\sum_{l,i} \lambda_{lp,i}}. \quad (54)$$

Where $C_{N_t,N_r}$ is a normalizing constant.

Define $\alpha_{lp,i} = \frac{\log(1+\rho\lambda_{lp,i})}{\log \rho}$. Note that since $\lambda_{lp,i} \geq 0$, $\alpha_{lp,i} \geq 0$, for all $l, i$. The joint density of $\boldsymbol{\alpha}_{mp} = [\alpha_{mp,0}, \cdots, \alpha_{mp,L-1}]'$ is

$$\begin{aligned} p(\boldsymbol{\alpha}_0, \cdots, \boldsymbol{\alpha}_{(L-1)p}) &= C_{N_t,N_r}^{-K} (\log \rho)^{LK} \rho^{\sum_{l=0}^{L-1} \sum_{i=0}^{K-1} \alpha_{lp,i}} \\ &\times \prod_{l=0}^{L-1} \left( \prod_{i=0}^{K-1} (\rho^{\alpha_{lp,i}-1} - \rho^{-1})^{|N_t-N_r|} \prod_{i<j} (\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 \right) \\ &\times e^{-\sum_{l,i}(\rho^{\alpha_{lp,i}-1} - \rho^{-1})}. \end{aligned} \quad (55)$$

At high SNR, the density is approximately

$$\begin{aligned} p(\boldsymbol{\alpha}) &= C_{N_t,N_r}^{-K} (\log \rho)^{LK} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \\ &\times \prod_{l=0}^{L-1} \prod_{i<j} (\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 e^{-\sum_{l,i}(\rho^{\alpha_{lp,i}-1} - \rho^{-1})}, \end{aligned}$$

where $\boldsymbol{\alpha} := [\boldsymbol{\alpha}_0', \cdots, \boldsymbol{\alpha}_{(L-1)p}']'$. After changing variables, the outage probability becomes

$$\begin{aligned} P(\sum_{l,i} \alpha_{lp,i} \leq r) &= \int_{\mathcal{A}} C_{N_t,N_r}^{-K} (\log \rho)^{LK} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \\ &\times \prod_{l=0}^{L-1} \prod_{i<j} (\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 e^{-\sum_{l,i}(\rho^{\alpha_{lp,i}-1} - \rho^{-1})} d\boldsymbol{\alpha}. \end{aligned} \quad (56)$$

Here, we define $\mathcal{A} = \{\sum_{l,i} \alpha_{l,i} \leq r\} \cap \{\alpha_{lp,i} \geq 0,\ \text{for all } l, i\} \cap \{\alpha_{lp,0} \leq \alpha_{lp,1} \leq \cdots \leq \alpha_{lp,K-1},\ \text{for all } l\}$.

Note that when $\alpha_{lp,i} > 1$ for any $l$ or $i$, $e^{-\sum_{l,i}(\rho^{\alpha_{lp,i}-1} - \rho^{-1})}$ decays exponentially in $\rho$. At high SNR, we therefore drop the integral over the range with any $\alpha_{lp,i} > 1$, and replace the integral range from $\mathcal{A}$ to $\mathcal{A}' = \{\alpha_{lp,i} \leq 1, \text{ for all } l,i\} \cap \mathcal{A}$. In $\mathcal{A}'$, as $\rho \to \infty$, $\exp\{-\frac{\rho^{\alpha_{lp,i}-1}-\rho^{-1}}{\sigma^2}\} \to 1$. So at high SNR, outage probability is approximately

$$\begin{aligned} P(\sum_{l,i} \alpha_{lp,i} \leq r) &\doteq \int_{\mathcal{A}'} C_{N_t,N_r}^{-K} (\log \rho)^{LK} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \\ &\times \prod_{l=0}^{L-1} \prod_{i<j} (\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha}. \end{aligned} \tag{57}$$

Suppose $r = sL + q$ where $s$ is an integer, $q < L$.

**Lemma 9.** *The outage probability of this block fading channel is exponentially equivalent to* $\rho^{-[q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]]}$.

*Proof.* Please see the Appendix 3.14-C for the proof. □

**Frequency-Selective Channel Model**

We return to the MIMO-OFDM channel in this Subsection.

**An Upper Bound** Since the outage event

$$\begin{aligned} &\left\{ \sum_{m=0}^{p-1} \sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp+m} \boldsymbol{H}_{lp+m}^H) \leq rp \log \rho \right\} \\ \subseteq &\bigcup_{m=0}^{p-1} \left\{ \sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp+m} \boldsymbol{H}_{lp+m}^H) \leq r \log \rho \right\}, \end{aligned}$$

then

$$\begin{aligned} P_{\text{out}}(r,\rho) &\leq \sum_{m=0}^{p-1} P\left[ \sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp+m} \boldsymbol{H}_{lp+m}^H) \leq r \log \rho \right] \\ &= pP\left[ \sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp} \boldsymbol{H}_{lp}^H) \leq r \log \rho \right]. \end{aligned} \tag{58}$$

The last equality is because the $\{\boldsymbol{H}_m\}$ are identically distributed. Since

$$P\left[ \sum_{l=0}^{L-1} \log\det(\boldsymbol{I} + \rho \boldsymbol{H}_{lp} \boldsymbol{H}_{lp}^H) \leq r \log \rho \right],$$

is the same with that in (54), we can get the upper bound using the results in Subsection 3.12.1. Given $r = sL + q$,

$$P_{\text{out}}(r,\rho) \dot{\leq} \rho^{-\left(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\right)}.$$

### Exact Result

**Lemma 10.** *The outage probability of this block fading channel is exponentially equivalent to* $\rho^{-[q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]]}$.

*Proof.* Please see the Appendix 3.14-D for the proof. □

## 3.13 Conclusion and Discussion

We have proved that in the frequency-selective channel with $L$ independent identically distributed taps, the maximum diversity gain is $d = L - r$, with multiplexing gain $r$.

We use the OFDM technique to transfer the tapped-delay-line channel into the frequency domain to get a diagonalized, highly correlated channel model, where the channel vector in the frequency domain is reduced rank. From the proof, we find that regardless of the block size and the size of the FFT for the channel, the diversity gain depends only on the number of the independent channel taps of the original tapped-delay-line channel model, and so does the scaling law of the outage probability and the error probability without outage at high SNR.

We apply various methods to derive the scaling law of the outage probability and the error probability without outage. We use a geometric picture to derive a lower bound for the outage probability, which is also a key step to derive the approximations at low and high SNR.

## 3.14 Appendices

### 3.14-A Proof of Lemma 8

*Proof.* We analyze the integral area in equation (48)

$$\begin{aligned}
\left\{\sum_{m=0}^{M-1} \alpha_m > rp\right\} &= \left\{\sum_{l=0}^{L-1} \alpha_{lp} + \sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} > rp\right\} \\
&\subseteq \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} > r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} > r(p-1)\right\}\right\} \\
&\quad \bigcup \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} \le r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} > r(p-1)\right\}\right\} \\
&\quad \bigcup \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} > r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} \le r(p-1)\right\}\right\}.
\end{aligned}$$

Define

$$\mathcal{G} := \left\{\sum_{m=0}^{M-1} \alpha_m > rp\right\}$$

$$\mathcal{B} := \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} > r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} > r(p-1)\right\}\right\}$$

$$\mathcal{C} := \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} \le r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} > r(p-1)\right\}\right\}$$

and

$$\mathcal{D} := \left\{\left\{\sum_{l=0}^{L-1} \alpha_{lp} > r\right\} \cap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} \le r(p-1)\right\}\right\}$$

Since $\mathcal{G} \subseteq \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}$,

$$\mathcal{G} = (\mathcal{B} \cup \mathcal{C} \cup \mathcal{D}) \cap \mathcal{G} = (\mathcal{B} \cap \mathcal{G}) \cup (\mathcal{C} \cap \mathcal{G}) \cup (\mathcal{D} \cap \mathcal{G}) \subseteq \mathcal{B} \cup \mathcal{C} \cup (\mathcal{D} \cap \mathcal{G}).$$

We analyze $\mathcal{D} \cap \mathcal{G}$,

$$\mathcal{D} \cap \mathcal{G} = \left\{\sum_{l=0}^{L-1} \alpha_{lp} > r\right\} \bigcap \left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} \le r(p-1)\right\} \bigcap \left\{\sum_{m=0}^{M-1} \alpha_m > rp\right\}.$$

Note that the middle term

$$\left\{\sum_{m=1}^{p-1}\sum_{l=0}^{L-1} \alpha_{lp+m} \le r(p-1)\right\} \subseteq \bigcup_{m=1}^{p-1} \left\{\sum_{l=0}^{L-1} \alpha_{lp+m} \le r\right\}.$$

So,

$$\begin{aligned}
\mathcal{D}\cap\mathcal{G} &\subseteq \left\{\sum_{l=0}^{L-1}\alpha_{lp} > r\right\}\bigcap\left\{\bigcup_{m=1}^{p-1}\left\{\sum_{l=0}^{L-1}\alpha_{lp+m}\leq r\right\}\right\}\bigcap\left\{\sum_{n=0}^{M-1}\alpha_n > rp\right\}\\
&\subseteq \left\{\bigcup_{m=1}^{p-1}\{\sum_{l=0}^{L-1}\alpha_{lp+m}\leq r\}\right\}\bigcap\left\{\sum_{n=0}^{M-1}\alpha_n > rp\right\}\\
&= \bigcup_{m=1}^{p-1}\left\{\left\{\sum_{l=0}^{L-1}\alpha_{lp+m}\leq r\right\}\bigcap\left\{\sum_{n=0}^{M-1}\alpha_n > rp\right\}\right\}\\
&= \bigcup_{m=1}^{p-1}\left\{\left\{\sum_{l=0}^{L-1}\alpha_{lp+m}\leq r\right\}\bigcap\left\{\sum_{n=0,n\neq m}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+n} > r(p-1)\right\}\right\}.
\end{aligned}$$

Define

$$\mathcal{D}_m := \left\{\sum_{l=0}^{L-1}\alpha_{lp+m}\leq r\right\}\bigcap\left\{\sum_{n=0,n\neq m}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+n} > r(p-1)\right\}.$$

So, finally,

$$\mathcal{G}\subseteq\mathcal{B}\cup\mathcal{C}\cup(\mathcal{D}\cap\mathcal{G})\subseteq\mathcal{B}\cup\mathcal{C}\cup(\cup_{m=1}^{p-1}\mathcal{D}_m).$$

Since the integrand in (48) is greater than 0, then

$$\begin{aligned}
P\left[\text{error, no outage}\right] &\leq \int_{\mathcal{G}}\frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}}f(\boldsymbol{\beta})d\boldsymbol{\beta}\\
&\leq \int_{\mathcal{B}\cup\mathcal{C}\cup(\cup_{m=1}^{p-1}\mathcal{D}_m)}\frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}}f(\boldsymbol{\beta})d\boldsymbol{\beta}\\
&\leq \int_{\mathcal{B}}\frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}}f(\boldsymbol{\beta})d\boldsymbol{\beta}+\int_{\mathcal{C}}\frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}}f(\boldsymbol{\beta})d\boldsymbol{\beta}\\
&+ \sum_{m=1}^{p-1}\int_{\mathcal{D}_m}\frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}}f(\boldsymbol{\beta})d\boldsymbol{\beta}.
\end{aligned}$$

We analyze these terms one by one.

$$\int_{\mathcal{B}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta})d\boldsymbol{\beta} = \int_{\mathcal{B}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta}_0) \prod_{m=1}^{p-1} \delta\left(\boldsymbol{\beta}_m - \mathcal{K}_m(\boldsymbol{\beta}_0)\right) d\boldsymbol{\beta}_0$$

$$= \int_{\mathcal{B}} \frac{\rho^{rp-L}(\log\rho)^L}{(\pi\sigma^2)^L} \rho^{-\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{(lp+m)}} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{lp}-1}-\frac{1}{\rho}\right)}{\sigma^2}\right\}$$

$$\times \prod_{m=1}^{p-1} \delta\left(\boldsymbol{\beta}_m - \mathcal{K}_m(\boldsymbol{\beta}_0)\right) d\boldsymbol{\beta}_0$$

$$\overset{\text{a}}{\leq} \int_{\left\{\sum_{l=0}^{L-1}\alpha_{lp}>r\right\}} \frac{\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{lp}-1}-\frac{1}{\rho}\right)}{\sigma^2}\right\} d\boldsymbol{\alpha}_0.$$

Inequality (a) is because in area $\mathcal{B}$,

$$\rho^{-\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m}} < \rho^{-r(p-1)}.$$

Using the techniques, for deriving the upper bound of the outage probability, we can easily get that at high SNR,

$$\int_{\mathcal{B}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f_{\underline{\beta}}(\underline{\beta})d\underline{\beta} \dot{\leq} \frac{\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L}. \tag{59}$$

Similarly, we can get that

$$\int_{\mathcal{C}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f_{\underline{\beta}}(\underline{\beta})d\underline{\beta} \dot{\leq} \frac{\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L}. \tag{60}$$

Next we derive the integral under area $D_m$.

$$\int_{\mathcal{D}_m} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta})d\boldsymbol{\beta} = \int_{\mathcal{D}_m} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta}_m) \prod_{n=1}^{p-1} \delta\left(\boldsymbol{\beta}_n - \mathcal{K}_{n-m}(\boldsymbol{\beta}_m)\right) d\boldsymbol{\beta}_m$$

$$= \int_{\mathcal{D}_m} \frac{\rho^{rp-L}(\log\rho)^L}{(\pi\sigma^2)^L} \rho^{-\sum_{n=1,n\neq m}^{p-1}\sum_{l=0}^{L-1}\alpha_{(lp+n)}} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{lp+m}-1}-\frac{1}{\rho}\right)}{\sigma^2}\right\}$$

$$\times \prod_{n=1}^{p-1} \delta\left(\boldsymbol{\beta}_n - \mathcal{K}_{n-m}(\boldsymbol{\beta}_m)\right) d\boldsymbol{\beta}_m$$

$$\leq \int_{\left\{\sum_{l=0}^{L-1}\alpha_{lp+m}>r\right\}} \frac{\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L} \exp\left\{-\frac{\sum_{l=0}^{L-1}\left(\rho^{\alpha_{lp+m}-1}-\frac{1}{\rho}\right)}{\sigma^2}\right\} d\boldsymbol{\alpha}_m.$$

Since $\boldsymbol{\alpha}_m$ are all identically distributed, we can get for all $m \in [1, p-1]$,

$$\int_{\mathcal{D}_m} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\beta}) d\boldsymbol{\beta} \dot{\leq} \frac{\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L}. \tag{61}$$

From (59), (60) and (61) we obtain

$$P\left[\text{error, no outage}\right] \dot{\leq} \frac{(p+1)\rho^{r-L}(\log\rho)^L}{(\sigma^2)^L}.$$

□

### 3.14-B Alternative Proof of Theorem 2

*Proof.* From equation (39), and the density functions (44), (43) and (42) we can represent outage probability as follows

$$\begin{aligned} P_{\text{out}}(r,\rho) &= P\left[\frac{1}{p}\sum_{m=0}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m} \leq r\right] = \int_{\mathcal{A}} f(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0) d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &= \int_{\mathcal{A}} \frac{(2\log\rho)^L \rho^{\sum_{l=0}^{L-1}\alpha_{lp}-L}}{(\pi\sigma^2)^L} \exp\left\{-\frac{\sum_{l=0}^{L-1}(\rho^{\alpha_{lp}-1}-\rho^{-1})}{\sigma^2}\right\} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &\overset{(a)}{\doteq} \int_{\mathcal{A}'} \rho^{\sum_{l=0}^{L-1}\alpha_{lp}-L} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &\overset{(b)}{=} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*)-L}. \end{aligned} \tag{62}$$

Here the integral area $\mathcal{A} := \left\{\frac{1}{p}\sum_{m=0}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m} \leq r\right\} \cap \{0 \leq \alpha_{lp},\ \text{for all } l\} \cap \{-\pi \leq \theta_{lp} \leq \pi,\ \text{for all } l\}$, $\mathcal{A}' = \mathcal{A} \cap \{\alpha_{lp} \leq 1,\ \text{for all } l\}$. Define $\mathcal{V} := \{0 \leq \alpha_{lp} \leq 1,\ \text{for all } l\}$. We see that $\mathcal{A}' \subseteq \mathcal{V}$. We get the exponential equality $(a)$ using the same arguments as in the proof of Lemma 7 $\hat{f}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) := \sum_{l=0}^{L-1}\alpha_{lp}$. $\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*$ is the optimal solution of the following optimization problem

$$\begin{aligned} \max \quad & \sum_{l=0}^{L-1}\alpha_{lp} \\ \text{s.t.} \quad & \frac{1}{p}\sum_{m=0}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m} \leq r \\ & 0 \leq \alpha_{lp} \leq 1, \forall l \\ & -\pi \leq \theta_l \leq \pi, \forall l. \end{aligned} \tag{63}$$

We prove the exponential equality $(b)$ in equation (62) next. An upper bound is derived in the following

$$\begin{aligned} P_{\text{out}}(r,\rho) &\doteq \int_{\mathcal{A}'} \rho^{\sum_{l=0}^{L-1}\alpha_{lp}-L} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &\stackrel{(b)}{\leq} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-L}\int_{\mathcal{V}} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &\leq \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-L}\text{vol}(\mathcal{V}) \\ &\doteq \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-L}. \end{aligned} \tag{64}$$

The inequality $(b)$ is obvious since $\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)$ is the maxima of $\sum_{l=0}^{L-1}\alpha_{lp}$ in the integral area $\mathcal{A}'$, and the integral area $\mathcal{A}' \subseteq \mathcal{V}$. The vol$(\mathcal{V})$ is fixed yields the last exponential equality.

We follow the ideas of [2] to get a lower bound. Note that $\hat{f}(\boldsymbol{\alpha},\boldsymbol{\theta})$ is continuous, therefore for any $\delta > 0$, there exists a neighborhood $\Omega$ of $\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*$, within which $\hat{f}(\boldsymbol{\alpha},\boldsymbol{\theta}) \geq \hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*) - \delta$. It follows that

$$\begin{aligned} P_{\text{out}}(r,\rho) &\doteq \int_{\mathcal{A}'} \rho^{\sum_{l=0}^{L-1}\alpha_{lp}-L} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\ &\geq \text{vol}[\Omega \cap \mathcal{A}']\rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-\delta-L} \\ &\doteq \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-\delta-L}. \end{aligned} \tag{65}$$

Since $P_{\text{out}}(r,\rho) \dot{\geq} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-\delta-L}$ for any $\delta > 0$, we have $P_{\text{out}}(r,\rho) \dot{\geq} \rho^{\hat{f}(\hat{\boldsymbol{\alpha}}^*,\hat{\boldsymbol{\theta}}^*)-L}$ which together with the upper bound proves the Lemma, which means the integral is dominated by the minimum SNR exponent.

From equation (36), defining

$$\boldsymbol{U}_m := \frac{1}{L}\boldsymbol{\mathcal{F}}_L \boldsymbol{D}_m \boldsymbol{\mathcal{F}}_L^H,$$

which is a unitary matrix with $(i,l)$th element $u_{i,l}^m$, we can get that

$$\alpha_{ip+m} = \frac{\log\left(1+\rho|\sum_{l=0}^{L-1} u_{i,l}^m \sqrt{\rho^{\alpha_{lp}-1}-\rho^{-1}}e^{j\theta_{lp}}|^2\right)}{\log\rho}, i = 1,\cdots,p-1.$$

Define $\alpha_{\max} = \max\{\alpha_{lp}, \text{ for all } l\}$. At high SNR, because $u_{i,l}^m \neq 0$, for all $i, l$ and$m \neq 0$

$$\lim_{\rho\to\infty} \alpha_{ip+m} = \lim_{\rho\to\infty} \frac{\log\left(1 + (\rho^{\alpha_{\max}} - 1)|\sum_{l=0}^{L-1} u_{i,k}^m \sqrt{\frac{\rho^{\alpha_l - 1} - \rho^{-1}}{\rho^{\alpha_{\max}-1} - \rho^{-1}}} e^{j\theta_{lp}}|^2\right)}{\log\rho}$$
$$= \alpha_{\max}. \quad (66)$$

W.l.o.g. assuming $\alpha_{\max} = \alpha_0$, because of the dominated convergence Theorem at high SNR the optimization problem (63) becomes

$$\begin{aligned} \max \quad & \sum_{l=0}^{L-1} \alpha_{lp} \\ \text{s.t.} \quad & \frac{1}{p}\left(\sum_{l=0}^{L-1} \alpha_{lp} + L(p-1)\alpha_0\right) \leq r \\ & \alpha_{lp} \leq \alpha_0, \forall l \in [1, p-1] \\ & 0 \leq \alpha_{lp} \leq 1, \forall l. \end{aligned}$$

By KKT condition in [8], it is easy to check that the first inequality constraint can be changed to be equality, which yields the following optimization problem

$$\begin{aligned} \max \quad & \sum_{l=0}^{L-1} \alpha_{lp} \\ \text{s.t.} \quad & \frac{1}{p}\left(\sum_{l=0}^{L-1} \alpha_{lp} + L(p-1)\alpha_0\right) = r \\ & \alpha_{lp} \leq \alpha_0, \forall l \in [1, p-1] \\ & 0 \leq \alpha_{lp} \leq 1, \forall l. \end{aligned} \quad (67)$$

Solving this problem, we get that the optimal $\alpha_{lp} = \frac{r}{L}$, for all $l$, and the optimal objective function is $f(\boldsymbol{\alpha}^*) := \sum_{l=0}^{L-1} \alpha_{lp} = r$, which proves that

$$P_{\text{out}} \doteq \rho^{-(L-r)}.$$

We consider the error probability without outage now. From equation (48), we get

$$\begin{aligned}
& P\left[\text{error, no outage}\right] \\
\leq\ & \int_{\bar{\mathcal{A}}} \frac{\rho^{rp}}{\rho^{\sum_{m=0}^{M-1}\alpha_m}} f(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0) d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\
=\ & \int_{\bar{\mathcal{A}}} \frac{(2\log\rho)^L \rho^{-\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m}+rp-L}}{(\pi\sigma^2)^L} \exp\left\{-\frac{\sum_{l=0}^{L-1}(\rho^{\alpha_{lp}-1}-\rho^{-1})}{\sigma^2}\right\} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\
\stackrel{(a)}{\doteq}\ & \int_{\mathcal{B}} \rho^{-\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m}+rp-L} d\boldsymbol{\alpha}_0 d\boldsymbol{\theta}_0 \\
\stackrel{(b)}{\dot{\leq}}\ & \rho^{g(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*)-L},
\end{aligned} \tag{68}$$

where $\bar{\mathcal{A}} := \left\{\sum_{m=0}^{M-1}\alpha_m > rp\right\} \cap \{\alpha_{lp} \geq 0, \text{ for all } l\} \cap \{-\pi \leq \theta_{lp} \leq \pi, \text{ forall } l\}$. The exponential equality $(a)$ is got with the same arguments as we get the exponential equality $(a)$ in (62). $\mathcal{B} := \bar{\mathcal{A}} \cap \mathcal{V}$. Define $g(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*) := -\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m} + rp$. We prove the exponential equality $(b)$ next.

With the same arguments as before, we get that

$$P\left[\text{error, no outage}\right] \dot{\leq} \rho^{g(\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*)-L},$$

where $\hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\theta}}^*$ is the solution of the following optimization problem

$$\begin{aligned}
\max \quad & -\sum_{m=1}^{p-1}\sum_{l=0}^{L-1}\alpha_{lp+m} + rp \\
\text{s.t.} \quad & \sum_{m=0}^{M-1}\alpha_m > rp \\
& 0 \leq \alpha_{lp} \leq 1, \forall l \\
& -\pi \leq \theta_{lp} \leq \pi, \forall l.
\end{aligned}$$

Next, we prove that this optimization problem is equivalent to problem (67) at high SNR. From equation (66), the optimization problem above becomes

$$\begin{aligned} \max \quad & rp - pL\alpha_0 \\ \text{s.t.} \quad & \sum_{l=0}^{L-1} \alpha_{lp} + L(p-1)\alpha_0 > rp \\ & \alpha_{lp} \leq \alpha_0, \forall l \in [1, p-1] \\ & 0 \leq \alpha_{lp} \leq 1, \forall l. \end{aligned}$$

By the KKT condition in [8], the first inequality constraint can be changed to be equality, which makes the above optimization problem equivalent to problem (67). Therefore the solution is the same, which proves that

$$P\big[\text{error, no outage}\big] \dot{\leq} \rho^{-(L-r)}.$$

□

### 3.14-C Proof of Lemma 9

*Proof.* Following the proof of [2], we give an upper bound $\bar{P}_{\text{out}}(r,\rho)$ and a lower bound of the outage probability.

First we derive an upper bound. Since we are only interested in the SNR exponent, we simplify the integral by ignoring all the parts which has no effect on the SNR exponent. Consider

$$\begin{aligned} P_{\text{out}}(r,\rho) &:= P(\sum_{l,i} \alpha_{lp,i} \leq r) \\ &\doteq \int_{\mathcal{A}'} \rho^{\sum_{lp,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \prod_{l=0}^{L-1} \prod_{i<j} (\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha} \\ &\dot{\leq} \int_{\mathcal{A}'} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \prod_{l=0}^{L-1} \prod_{i<j} (0 - \rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha} \\ &= \int_{\mathcal{A}'} \rho^{\sum_{l=0}^{L-1} \sum_{i=0}^{K-1} (\alpha_{lp,i}-1)(|N_t-N_r|+2i+1)} d\boldsymbol{\alpha} = \bar{P}_{\text{out}}(r,\rho). \end{aligned}$$

Define $f(\boldsymbol{\alpha}) := \sum_{l=0}^{L-1} \sum_{i=0}^{K-1} (\alpha_{lp,i} - 1)(|N_t - N_r| + 2i + 1)$, and denote $\boldsymbol{\alpha}^* = \arg\sup_{\mathcal{A}'} f(\boldsymbol{\alpha})$. Define $\mathcal{V} = \{\alpha_{lp,i} \geq 0, \forall l, i\} \cap \{\alpha_{lp,i} \leq 1, \forall l, i\}$. We see that $\mathcal{A}' \subseteq \mathcal{V}$.

Then

$$\begin{aligned}
\bar{P}_{\text{out}}(r,\rho) &= \int_{\mathcal{A}'} \rho^{\sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1)} d\boldsymbol{\alpha} \\
&= \int_{\mathcal{A}'\cap\mathcal{V}} \rho^{\sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1)} d\boldsymbol{\alpha} \\
&\leq \int_{\mathcal{V}} \rho^{f(\boldsymbol{\alpha}^*)} d\boldsymbol{\alpha} \\
&\doteq \rho^{f(\boldsymbol{\alpha}^*)}.
\end{aligned}$$

A lower bound of $\bar{P}_{\text{out}}(r,\rho)$ can be found by noting that $f(\boldsymbol{\alpha})$ is continuous, therefore for any $\delta > 0$, there exists a neighborhood $\Omega$ of $\boldsymbol{\alpha}^*$, within which $f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\alpha}^*) - \delta$. Now

$$\begin{aligned}
\bar{P}_{\text{out}}(r,\rho) &\geq \int_{\Omega\cap\mathcal{A}'} \rho^{f(\boldsymbol{\alpha}^*)-\delta} d\boldsymbol{\alpha} \\
&= \text{vol}[\Omega\cap\mathcal{A}']\rho^{f(\boldsymbol{\alpha}^*)-\delta}.
\end{aligned}$$

Since at high SNR, $\bar{P}_{\text{out}}(r,\rho) \dot{\geq} \rho^{f(\boldsymbol{\alpha}^*)-\delta}$ for any $\delta > 0$, we have $\bar{P}_{\text{out}}(r,\rho) \dot{\geq} \rho^{f(\boldsymbol{\alpha}^*)}$, which proves that

$$\lim_{\rho\to\infty} \frac{\log \bar{P}_{\text{out}}(r,\rho)}{\log \rho} = f(\boldsymbol{\alpha}^*).$$

Now to derive a lower bound of $P_{\text{outage}}(\rho)$, for any $\delta > 0$, define the set

$$S_\delta = \{\boldsymbol{\alpha} : |\alpha_{lp,j} - \alpha_{lp,i}| > \delta, \forall i \neq j\}.$$

Now

$$\begin{aligned}
P_{\text{out}}(r,\rho) &= \int_{\mathcal{A}'} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \prod_{l=0}^{L-1}\prod_{i<j}(\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha} \\
&\geq \int_{\mathcal{A}'\cap S_\delta} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \prod_{l=0}^{L-1}\prod_{i<j}(\rho^{\alpha_{lp,i}-1} - \rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha} \\
&\geq \int_{\mathcal{A}'\cap S_\delta} \rho^{\sum_{l,i}(\alpha_{lp,i}-1)(|N_t-N_r|+1)} \prod_{l=0}^{L-1}\prod_{i<j}((1-\rho^{-\delta})\rho^{\alpha_{lp,j}-1})^2 d\boldsymbol{\alpha} \\
&= (1-\rho^{-\delta})^{K^2} \int_{\mathcal{A}'\cap S_\delta} \rho^{f(\boldsymbol{\alpha})} d\boldsymbol{\alpha}.
\end{aligned}$$

Following the above arguments $\int_{\mathcal{A}'\cap S_\delta} \rho^{f(\boldsymbol{\alpha})} d\boldsymbol{\alpha}$ has SNR exponent

$$\sup_{\mathcal{A}'\cap S_\delta} f(\boldsymbol{\alpha})$$

which, by the continuity of $f$, approaches $f(\boldsymbol{\alpha}^*)$ as $\delta \to 0$. Combining the upper and lower bound, we have that

$$\lim_{\rho\to\infty} \frac{\log P_{\text{out}}(r,\rho)}{\log \rho} = f(\boldsymbol{\alpha}^*).$$

Next, we derive $f(\boldsymbol{\alpha}^*)$. First we get $\boldsymbol{\alpha}^*$ by solving the following optimization problem.

$$\begin{aligned} \max \quad & \sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1) \\ \text{st.} \quad & \sum_{l=0}^{L-1}\sum_{i=0}^{K-1}\alpha_{lp,i} \le r \\ & 0 \le \alpha_{lp,i} \le 1, \forall l, i \\ & \alpha_{lp,0} \le \alpha_{lp,1} \le \cdots \le \alpha_{lp,K-1}, \forall l. \end{aligned} \tag{69}$$

It is easy to get the optimal $\boldsymbol{\alpha}^*$ as

$$\begin{aligned} & \alpha_{lp,K-s} = \cdots = \alpha_{lp,K-1} = 1, \qquad \forall l \\ & \sum_{l=0}^{L-1}\alpha_{lp,K-s-1} = q, \text{and } \alpha_{lp,K-s-1} \le 1 \qquad \forall l \\ & \alpha_{lp,0} = \cdots = \alpha_{lp,K-s-2} = 0, \qquad \forall l. \end{aligned} \tag{70}$$

So,

$$f(\boldsymbol{\alpha}^*) = -\left[q[(N_t-s-1)(N_r-s-1)] + (L-q)[(N_t-s)(N_r-s)]\right].$$

□

### 3.14-D Proof of Lemma 10

*Proof.* Applying the SVD to each $\boldsymbol{H}_{lp+m}$, we get $\boldsymbol{H}_{lp+m} = \boldsymbol{U}_{lp+m}\boldsymbol{\Sigma}_{lp+m}\boldsymbol{V}_{lp+m}^H$, where $\boldsymbol{\Sigma}_{lp+m}$ is the diagonal matrix with diagonal elements as the singular values of $\boldsymbol{H}_{lp+m}$, and $\boldsymbol{U}_{lp+m}$ and $\boldsymbol{V}_{lp+m}$ are unitary matrices having the invariant Haar measure [9]. Knowing that the singular values of $\boldsymbol{H}_{lp+m}$ and the unitary matrices $\boldsymbol{U}_{lp+m}$ and $\boldsymbol{V}_{lp+m}$ are independent, together with the density in equation (54), we get the joint distribution of eigenvalues $\boldsymbol{\lambda}_{lp} = [\lambda_{lp,0}\cdots,\lambda_{lp,K-1}]', \forall l$, $\boldsymbol{U}_{lp}, \forall l$ and $V_{lp}, \forall l$ as

$$\begin{aligned} & p\left(\boldsymbol{\lambda}_0,\cdots,\boldsymbol{\lambda}_{(L-1)p},\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p},\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right) \\ = \; & C_{N_t,N_r}^{-K}\prod_{l=0}^{L-1}\left(\prod_{i=0}^{K-1}\lambda_{lp,i}^{|N_t-N_r|}\prod_{i<j}(\lambda_{lp,i}-\lambda_{lp,j})^2\right)e^{-\sum_{l,i}\lambda_{lp,i}} \\ \times \; & p\left(\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p}\right)p\left(\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right), \end{aligned} \tag{71}$$

where $p\left(\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p}\right)$ and $p\left(\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right)$ are joint distributions of $\boldsymbol{U}_{lp}, \forall l$ and $\boldsymbol{V}_{lp}, \forall l$ respectively. Therefore, the joint distribution of $\boldsymbol{\alpha}_{lp}, \forall l$, $\boldsymbol{U}_{lp}, \forall l$ and $\boldsymbol{V}_{lp}, \forall l$ is

$$\begin{aligned} & p\left(\boldsymbol{\alpha}_0,\cdots,\boldsymbol{\alpha}_{(L-1)p},\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p},\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right) \\ = \; & C_{N_t,N_r}^{-K}(\log\rho)^{LK}\rho^{\sum_{l=0}^{L-1}\sum_{i=0}^{K-1}\alpha_{lp,i}} \\ \times \; & \prod_{l=0}^{L-1}\left(\prod_{i=0}^{K-1}(\rho^{\alpha_{lp,i}-1}-\rho^{-1})^{|N_t-N_r|}\prod_{i<j}(\rho^{\alpha_{lp,i}-1}-\rho^{\alpha_{lp,j}-1})^2\right)e^{-\sum_{l,i}\left(\rho^{\alpha_{lp,i}-1}-\rho^{-1}\right)} \\ \times \; & p\left(\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p}\right)p\left(\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right). \end{aligned} \tag{72}$$

Then the outage probability is

$$\begin{aligned} P_{\text{out}}(r,\rho) &= P\left[\sum_{l=0}^{L-1}\sum_{m=0}^{p-1}\sum_{i=0}^{K-1}\alpha_{lp+m,i}\le rp\right] \\ &= \int_{\mathcal{A}} p\left(\boldsymbol{\alpha}_0,\cdots,\boldsymbol{\alpha}_{(L-1)p},\boldsymbol{U}_0,\cdots,\boldsymbol{U}_{(L-1)p},\boldsymbol{V}_0,\cdots,\boldsymbol{V}_{(L-1)p}\right) \\ &\qquad d\boldsymbol{\alpha}_0,\cdots,d\boldsymbol{\alpha}_{(L-1)p},d\boldsymbol{U}_0,\cdots,d\boldsymbol{U}_{(L-1)p},d\boldsymbol{V}_0,\cdots,d\boldsymbol{V}_{(L-1)p}, \end{aligned}$$

where area $\mathcal{A} := \sum_{l=0}^{L-1}\sum_{m=0}^{p-1}\sum_{i=0}^{K-1}\alpha_{lp+m,i}\le rp$.

Following the arguments in Apendix 3.14-C, we can get the exponential equivalent form for the outage probability by solving the following optimization problem

$$\begin{aligned}
\max \quad & \sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1) \\
\text{st.} \quad & \sum_{l=0}^{L-1}\sum_{m=0}^{p-1}\sum_{i=0}^{K-1}\alpha_{lp+m,i} \le rp \\
& 0 \le \alpha_{lp,i} \le 1, \forall l, i \\
& \alpha_{lp,0} \le \alpha_{lp,1} \le \cdots \le \alpha_{lp,K-1}, \forall l \\
& \boldsymbol{U}_{lp}, \boldsymbol{V}_{lp} \text{ are unitary } \forall l.
\end{aligned} \tag{73}$$

Solving this problem is difficult. But we derive an upper bound for the outage probability in the last section 3.12.1, which means that

$$\begin{aligned}
f(\boldsymbol{\alpha},\boldsymbol{U},\boldsymbol{V}) &:= \sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1) \\
&\le -\big(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\big)
\end{aligned}$$

for all $\{\boldsymbol{\alpha}, \boldsymbol{U} := [\boldsymbol{U}_0', \cdots, \boldsymbol{U}_{lp}']', \boldsymbol{V} := [\boldsymbol{V}_0', \cdots, \boldsymbol{V}_{lp}']'\}$ satisfying the constraints. If we find a solution $\{\boldsymbol{\alpha}^*, \boldsymbol{U}^*, \boldsymbol{V}^*\}$ such that

$$f(\boldsymbol{\alpha}^*,\boldsymbol{U}^*,\boldsymbol{V}^*) = -\big(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\big),$$

then this solution is one of the best solutions, and the exponential equivalent form of the outage probability is $\rho^{-\left(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\right)}$.

From the solution (70) of the optimization problem (69), we know that

$$\begin{aligned}
&\alpha_{lp,K-s} = \cdots = \alpha_{lp,K-1} = 1, \quad \forall l \\
&\alpha_{lp,K-s-1} = \frac{q}{L}, \quad \forall l \\
&\alpha_{lp,0} = \cdots = \alpha_{lp,K-s-2} = 0, \quad \forall l
\end{aligned} \tag{74}$$

makes

$$\begin{aligned}
&\sum_{l=0}^{L-1}\sum_{i=0}^{K-1}(\alpha_{lp,i}-1)(|N_t-N_r|+2i+1) \\
=\ & -\big(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\big).
\end{aligned}$$

From this and the relationship $\mathcal{H}_m = \left(\frac{1}{L}(\mathcal{F}_L \boldsymbol{D}_m \mathcal{F}_L^H) \otimes \boldsymbol{I}_{L\times L}\right)\mathcal{H}_0$, we can easily get that (74) together with $\boldsymbol{U}_{lp} = I_{N_r\times N_r}, \forall l$ and $\boldsymbol{V}_{lp} = I_{N_t\times N_r}, \forall l$ satisfy the constraints in (73) and make the objective function

$$f(\boldsymbol{\alpha}, \boldsymbol{U}, \boldsymbol{V}) = -\big(q[(N_t - s - 1)(N_r - s - 1)] + (L - q)[(N_t - s)(N_r - s)]\big),$$

which proves that

$$P_{\text{out}}(r, \rho) \doteq \rho^{-\left(q[(N_t-s-1)(N_r-s-1)]+(L-q)[(N_t-s)(N_r-s)]\right)}.$$

□

## List of References


[1] J. G.Proakis, *Digital Communications, 4th Edition.* McGrawHill, 2000.

[2] L. Zheng and D. Tse, "Diversity and multiplexing: A fundanmental tradeoff in multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[3] R. G.Gallager, *Information Theory and Reliable Communication.* New York, United States of America: Wiley, 1968.

[4] L. H.Ozarow, S. Shamai, and A. D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicuar Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[5] R. J.Mceliece and W. E.Stark, "Channels with block interference," *IEEE Transactions on Information Theory*, vol. IT-30, no. 1, pp. 44–53, Jan. 1984.

[6] E. A.Lee and D. G.Messerschmitt, *Digital Communication, 2nd Edition.* Kluwer Academic Publishers: Boston/Dordrecht/London, 1994.

[7] A. Scaglione, "Statistical analysis of the capacity of mimo frquency selective rayleigh fading channels with arbitrary number of inputs and outputs," Lausanne, Switzerland, 2002.

[8] E. K.P.Chong and S. H.Zak, *An Introduction To Optimization, 2nd Edition.* John Wiley and Sons, INC: A Wiley-Interscience Publication, 2001.

[9] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, May 1989.

# CHAPTER 4

## Analog Precoder and Equalizer Designs and Their Geometry for Multichannel Communication

The problem of precoder and equalizer design is the problem of jointly designing a transmitter and a receiver so that some property is optimized. Typically the property is information rate, mean-squared error, or error probability. In this chapter we design transmitter and receiver pairs directly in their analog domain.

### 4.1 Preliminaries

Consider the complex analytic received waveform $y(t) = (Au)(t) + n(t)$ where the transmitted waveform is $(Au)(t) := \sum_{k=1}^{m} u_k a_k(t)$. We assume that the waveforms $a_k(t)$ are linearly independent and that the message vector $u := [u_1, \cdots, u_m]'$ is complex proper $N(0, R_{uu})$ and independent of the proper zero-mean, white Gaussian noise $n(t)$ with power spectral density $S(f) = \sigma^2$. Direct evaluations of mutual information, mean squared error, or bit error rate are complicated by the fact that $u$ is a column vector and $y$ is a continuous-time random process. We follow the arguments of [1], [2] to show that there exists a column vector $v$ which is a sufficient statistic for the random process $y$. We use notation for constructing the $m$-dimensional transmission $Au$ that allows $m$ users to be accommodated, but of course $k < m$ users can share the subspace and use the extra dimensions for diversity gain.

Let $\mathcal{A} := \text{span}\{a_1, \cdots, a_m\}$, and let $P_{\mathcal{A}}$ denote the projection operator [1] onto

[1] A projection operator on a vector space is an idempotent linear transformation. Such transformations project any point in the vector space to a point in the subspace that is the image of the transformation. For example, $P_{\mathcal{A}}$ is $P_{\mathcal{A}}y := (A(A^*A)^{-1}A^*y)(t)$ in the space $\mathcal{A}$ we defined, where $A^*$ is the adjoint operator of $A$ which will be defined later. In an inner product space, such an operator is an orthogonal projection if and only if it is self-adjoint. In finite-dimensional inner product spaces, an orthogonal projection matrix is one whose matrix $M$ satisfies $M^2 = M$ and $M^* = M$ where $M^*$ is the conjugate transpose of $M$.

$\mathcal{A}$. Since $y = Au + n$, where $Au \in \mathcal{A}$, we have $\hat{y} := P_{\mathcal{A}}y = Au + P_{\mathcal{A}}n$ and $\tilde{y} := y - \hat{y} = n - P_{\mathcal{A}}n = P_{\mathcal{A}}^{\perp}n$. Here $P_{\mathcal{A}}^{\perp}$ denotes the projection onto $\mathcal{A}^{\perp}$, the orthogonal complement of $\mathcal{A}$. Since $\hat{y}$ and $\tilde{y}$ are uncorrelated and jointly Gaussian, they are independent. Moreover the mapping $y \longmapsto (\hat{y}, \tilde{y})$ is invertible, and only $\hat{y}$ carries information about $u$, Thus $\hat{y}$ is a sufficient statistic for $y$, which means $\hat{y}$ contains all the relevant information in $y$ for purposes of detecting or estimating the column vector $u$, with respect to any criterion of optimality.

We next show that the analog waveform $\hat{y}$ is equivalent to the digital column vector

$$A^*y = \begin{bmatrix} < y, a_1 > \\ \vdots \\ < y, a_m > \end{bmatrix}$$

where $A^*$ denotes the adjoint of operator $A$ and $< \cdot, \cdot >$ denotes the inner product, $< y, a > := \int y(t)\bar{a}(t)dt$. Here $\bar{a}(t)$ is the complex conjugate of $a(t)$. By equivalent we mean that $\hat{y}$ can be obtained as a function of $A^*y$ and $A^*y$ can be obtained as a function of $\hat{y}$. To see this, first recall that $P_{\mathcal{A}} = A(A^*A)^{-1}A^*$. Thus, $\hat{y} = P_{\mathcal{A}}y$ is a function of $A^*y$. Conversely, $A^*y$ is a function of $\hat{y}$: $A^*y = A^*A(A^*A)^{-1}A^*y = A^*\hat{y}$. Since $\hat{y}$ and $A^*y$ are equivalent, $A^*y$ is also sufficient for the message $u$. (Note that the adjoint operation $A^*y$ corresponds to putting the received waveform into a bank of matched filters.) Define $v$ in terms of $A^*y$ by the invertible transformation $v := (A^*A)^{-1/2}A^*y$, where each element of the $m \times m$ Hermitian matrix $A^*A$ is an inner product of the form $< a_m, a_n >$. Hence we get a column vector $v$ which is sufficient for the message $u$. We assume here and throughout that $A^*A$ is non-singular, which defines what we mean by the waveforms $\{a_k(t)\}$ being independent.

Specifically and as an example, we analyze the average mutual information between $u$ and $y$ which is denoted by $I(u, y)$. Using the fact that the mapping $y \longmapsto (\hat{y}, \tilde{y})$ is invertible, along with a standard identity, we have

$$I(u; y) = I(u; \hat{y}, \tilde{y}) = I(u; \hat{y}) + I(u; \tilde{y}|\hat{y}).$$

To see that this last term is zero, observe that

$$\begin{aligned} I(u;\tilde{y}|\hat{y}) &= H(\tilde{y}|\hat{y}) - H(\tilde{y}|\hat{y},u) \\ &= H(\tilde{y}) - H(\tilde{y}), \text{ by independence of } \tilde{y},\hat{y} \\ &= 0. \end{aligned}$$

Thus, $I(u;y) = I(u;\hat{y}) = I(u;A^*y) = I(u,v)$, as $v$ and $A^*y$ are related by an invertible transformation.

### 4.2 Signaling Waveforms for Maximum Mutual Information

We derive the optimal precoder and equalizer to maximize the mutual information $I(u;y)$ between the received signal $y$ and the transmitter vector $u$. Observe that

$$v = (A^*A)^{-1/2}A^*y = (A^*A)^{1/2}u + w \tag{75}$$

where $w := (A^*A)^{-1/2}A^*n$ is an $m$-dimensional complex $N(0,\sigma^2 I)$ random vector. It follows that $u$ and $v$ are jointly Gaussian with cross-covariance matrix $R_{uv} = Euv^H = R_{uu}(A^*A)^{H/2}$ and auto-covariance matrices $R_{uu}$ and $R_{vv}$:

$$\begin{aligned} R_{vv} &= (A^*A)^{1/2}R_{uu}(A^*A)^{H/2} + \sigma^2 I \\ &= (A^*A)^{1/2}R_{uu}^{1/2}[I + \sigma^2 R_{uu}^{-1/2}(A^*A)^{-1}R_{uu}^{-H/2}]R_{uu}^{H/2}(A^*A)^{H/2} \end{aligned}.$$

The mutual information between $u$ and $v$ is

$$\begin{aligned} I(u;v) &= H(u) + H(v) - H(u,v) \\ &= \log\det\left(I + \sigma^{-2}R_{uu}^{H/2}(A^*A)R_{uu}^{1/2}\right). \end{aligned} \tag{76}$$

#### 4.2.1 Precoding

We now introduce a precoder $G$ and channel filter $h$, and define the digital-to-analog operator,

$$A = hG, \tag{77}$$

where $h$ is determined by the time-varying channel impulse response $h(t,\tau)$, and $G$ by the precoders $g_k(t)$. So, now $A$ models the action of the precoder and channel on the transmit symbols $u$ :

$$\begin{aligned}(Au)(t) &:= \sum_{k=1}^{m} u_k a_k(t) = \int h(t,t-\tau)\sum_{k=1}^{m} u_k g_k(\tau)d\tau \\ &= \sum_{k=1}^{m} u_k \int h(t,t-\tau)g_k(\tau)d\tau := (hGu)(t).\end{aligned}$$

By $A = hG$, we mean $a_k(t)$ is the response of the time-varying channel to the precoder waveform $g_k(t)$: $a_k(t) = \int h(t,t-\tau)g_k(\tau)d\tau = (h*g_k)(t)$. We assume the functions $(h*g_k)(t)$ are linearly independent.

The adjoint operator $A^*$ is $A^* = G^*h^*$, where

$$\begin{aligned}(A^*y)_k &:= <y,a_k>= \int y(t)\bar{a}_k(t)dt = \int y(t)dt\int \bar{h}(t,t-\tau)\bar{g}_k(\tau)d\tau \\ &= \int \bar{g}_k(\tau)d\tau\int \bar{h}(t,t-\tau)y(t)dt = (G^*h^*y)_k.\end{aligned}$$

It is important to note that the analog - analog operators $h$ and $h^*$ work like this: $(hy)(t) = \int h(t,t-\tau)y(\tau)d\tau$ and $(h^*y)(\tau) = \int \bar{h}(t,t-\tau)y(t)dt$. The matrix $A^*A = G^*h^*hG$ is developed by first writing $(hg_j)(t) = \int h(t,t-\tau)g_j(\tau)d\tau$ and $(h^*hg_j)(t) = \int \bar{h}(s,s-t)\int h(s,s-\tau)g_j(\tau)d\tau ds$. The $i,j$th element of $A^*A = G^*h^*hG$ is then

$$\begin{aligned}(A^*A)_{ij} &= g_i^*h^*hg_j \\ &= \int \bar{g}_i(\tau')d\tau'\int \bar{h}(s,s-\tau')\int h(s,s-\tau)g_j(\tau)d\tau ds \\ &= \int \bar{g}_i(\tau')\int g_j(\tau)\int \bar{h}(s,s-\tau')h(s,s-\tau)dsd\tau d\tau'.\end{aligned}$$

With $A^*A$ determined, we may re-write the mutual information $I(u;v)$ in (76) as

$$I(u;v) = \log\det\left(I + \sigma^{-2}R_{uu}^{H/2}(G^*h^*hG)R_{uu}^{1/2}\right). \tag{78}$$

Note that the only analog channel characterization that matters is this deterministic second-order characterization of $h$:

$$(h^*h)(\tau',\tau) = \int \bar{h}(s, s-\tau')h(s, s-\tau)ds,$$

which is a correlation function in the local delay variables $(\tau', \tau)$.

We assume the precoder subspace $\mathcal{G} := \text{span}\{g_1, \cdots, g_m\} \subset \mathcal{S}$, which indicates that $\mathcal{G}$ is a subspace of a signal space $\mathcal{S}$, and further assume $\{\psi_1, \cdots, \psi_n\}$ is an orthonormal basis of $\mathcal{S}$, assumed to be $n$-dimensional. Thus $g_k(t) = (\Psi\gamma_k)(t) = \sum_{i=1}^n \gamma_{k,i}\psi_i(t)$, where the $\gamma_{k,i}$ are expansion coefficients for $g_k$ in the basis $\{\psi_1, \cdots, \psi_n\}$. Define the $n \times m$ coefficient matrix $\Gamma := [\gamma_1, \cdots, \gamma_m]$, where the column vectors $\gamma_j$ are $\gamma_j = [\gamma_{j1}, \cdots, \gamma_{jn}]^T$. Then,

$$\begin{aligned}(A^*A)_{ij} &= \sum_{k=1}^{n}\sum_{l=1}^{n} \bar{\gamma}_{i,k} \int\int \bar{\psi}_k(\tau')(h^*h)(\tau',\tau)\psi_l(\tau)d\tau' d\tau \gamma_{j,l} \\ &= \sum_{k=1}^{n}\sum_{l=1}^{n} \bar{\gamma}_{i,k} Q_{k,l}\gamma_{j,l},\end{aligned}$$

where

$$Q_{k,l} := \int\int \bar{\psi}_k(\tau')(h^*h)(\tau',\tau)\psi_l(\tau)d\tau' d\tau. \tag{79}$$

So, $A^*A = \Gamma^H(\Psi^*h^*h\Psi)\Gamma = \Gamma^H Q \Gamma$ and the mutual information in (78) becomes

$$I(u;v) = \log\det\left(I + \sigma^{-2}R_{uu}^{1/2}\Gamma^H Q \Gamma R_{uu}^{1/2}\right). \tag{80}$$

The matrix $Q$ is a complete deterministic second-order characterization of the action of the channel $h$ on the basis $\{\psi_1, \cdots, \psi_n\}$. It is positive definite Hermitian with unitary decomposition $Q = V\Lambda V^H$, where $V$ is an $n \times n$ unitary matrix, and $\Lambda$ is an $n \times n$ diagonal matrix with non-negative diagonal elements in decreasing order.

W.l.o.g., we parameterize the coefficient matrix $\Gamma$ as $\Gamma = V\Phi R_{uu}^{-1/2}$, where $\Phi$ is an arbitrary $n \times m$ matrix. Then the information rate in (80) becomes

$$I(u;v) = \log\det(I + \sigma^{-2}\Phi^H\Lambda\Phi). \tag{81}$$

The design problem for the precoder $G$ is to maximize the mutual information between $u$ and $v$ (or $u$ and $y$):

$$\begin{aligned} \max_{\Phi} I(u;v) &= \max_{\Phi} \log\det(I + \sigma^{-2}\Phi^H \Lambda \Phi), \\ \text{s.t.} & \quad \operatorname{tr}(\Phi\Phi^H) \leq \mathcal{P}, \end{aligned}$$

where the power constraint comes from the constraint $\operatorname{tr}(G^* G R_{uu}) = \operatorname{tr}(\Gamma^H \Gamma R_{uu}) = \operatorname{tr}(\Gamma R_{uu} \Gamma^H) = \operatorname{tr}(\Phi\Phi^H) \leq \mathcal{P}$. But we have to show that $\operatorname{tr}(G^* G R_{uu}) \leq \mathcal{P}$. Because the expected transmit power is

$$\begin{aligned} E < x(t), x(t) >&= E \int \sum_{k=1}^{m} u_k g_k(t) \sum_{i=1}^{m} \bar{u}_i \bar{g}_i(t) dt \\ &= \sum_{k=1}^{m} \sum_{i=1}^{m} r_{ki} \int g_k(t) \bar{g}_i(t) dt = \sum_{k=1}^{m} \sum_{i=1}^{m} r_{ki} c_{ki} = \operatorname{tr}(G^* G R_{uu}), \end{aligned}$$

where $r_{ki} = E u_k \bar{u}_i$, and $c_{ki} = \int g_k(t) \bar{g}_i(t) dt$.

We know from [3] [4] and [5] that the solution for $\Phi$ is an $n \times m$ diagonal matrix with nonzero elements $\phi_i$ only on its main diagonal, with

$$\begin{aligned} |\phi_i|^2 &= \left(\frac{1}{\mu} - \frac{\sigma^2}{\lambda_i}\right)^+ \\ \mu &= \frac{1}{\frac{\mathcal{P}}{\bar{m}} + \frac{1}{\bar{m}} \sum_{i=1}^{\bar{m}} \frac{\sigma^2}{\lambda_i}}. \end{aligned} \tag{82}$$

Here, $(x)^+ = \max(x, 0)$, $\bar{m}$ is the number of the subchannels where $|\phi_i|^2 > 0$, and $\lambda_i$ are diagonal elements of $\Lambda$. Large power $\mathcal{P}$ and large "per-channel" SNR $\frac{\lambda_i}{\sigma^2}$ is favored, so that many channels may be used. The maximum mutual information is then

$$I(u;v) = \sum_{i=1}^{m} \max(\log(\frac{\lambda_i}{\sigma^2 \mu}), 0), \tag{83}$$

which is a sum of "per-channel information rates" of the form $\log(\frac{1}{\mu}\frac{\lambda_i}{\sigma^2})$. So information rate increases with increasing SNRs $\frac{\lambda_i}{\sigma^2}$ and with increasing power.

If the correlation of the symbols $R_{uu}$, may also be designed without constraint, then $\Gamma R_{uu}^{1/2}$ is replaced by $\hat{\Gamma} = V\Phi$, and the design proceeds unchanged.

### 4.2.2 Design Rules

We summarize the design rules.

1. Choose signal space $\mathcal{S}$ and the corresponding basis $\{\psi_k\}_{k=1}^n$ to match the channel $h(t,\tau)$.

2. Compute $Q$ whose $(k,l)$th element $Q_{k,l} := \int\int \bar{\psi}_k(\tau')(h^*h)(\tau',\tau)\psi_l(\tau)d\tau' d\tau$ is the second-order characteristic of the channel.

3. Decompose $Q$ as $Q = V\Lambda V^H$, and construct the diagonal matrix $\Phi$ with the diagonal elements computed according to the water filling solution of equation (82).

4. Compute $\Gamma = V\Phi R_{uu}^{-1/2}$, and design precoders $g_k(t) = \sum_{i=1}^n \gamma_{ki}\psi_i(t)$.

In this design, we note that the channel is determined by $h(t,\tau)$, and the design is constrained by the basis $\{\psi_k\}_{k=1}^n$. The best we can do is to design the best waveforms $g_i$ in the space spanned by these basis.

### 4.2.3 Precoder Subspaces

Since $(h^*h)(\tau',\tau)$ is nonnegative definite and we assume $(h^*h)(\tau',\tau)$ is continuous, and square integrable on $I \times I$ ($I$ is an interval on $\mathcal{R}$), according to Mercer's theorem, $(h^*h)(\tau',\tau)$ can be decomposed into $(h^*h)(\tau',\tau) = \sum_{i=1}^\infty \mu_i\bar{\varphi}_i(\tau')\varphi_i(\tau)$, where the series converges absolutely and uniformly on $I \times I$, and the continuous functions $\varphi_i(\tau)$ are eigenfunctions of $(h^*h)(\tau',\tau)$ corresponding to eigenvalues $\mu_i$: $\int(h^*h)(\tau',\tau)\varphi_i(\tau')d\tau' = \mu_i\varphi_i(\tau)$. The eigenfunctions $\{\varphi_i(\tau)\}$ are orthonormal on $I$: $\int \varphi_i(\tau)\bar{\varphi}_j(\tau)dt = \delta_{ij}$.

Assume the $\mu_i$ are in decreasing order. Given any $\epsilon > 0$, there exists an integer $N_\epsilon$, such that for all $n > N_\epsilon$, $\sum_n^\infty \mu_n < \epsilon$. We shall assume, as a practical solution, that $n$, the dimension of $\mathcal{S}$, approximates $N_\epsilon$, the $\epsilon$-dimension of $(h^*h)(\tau',\tau)$. Then the $\psi_j(\tau)$ may be assumed to equal the $\varphi_i(\tau)$, and $\lambda_i = \mu_i$. The maximized mutual information is given in equations (83) and (82).

From the solution, we see that the actual dimension of the subspace the precoders lies in is $\bar{m}$, which depends on the eigenvalues $\mu_i$, noise variance $\sigma^2$ and power constraint $\mathcal{P}$. As long as $\mathcal{P}$ is finite, the dimensions $N_\epsilon$ and $\bar{m}$ should be finite. We can choose $n = N_\epsilon$ large enough such that the optimal solution $\bar{m} < N_\epsilon$ to guarantee that we do not leave any available power unused, and the mutual information is maximized. The precoders lie in a subspace spanned by the dominant eigenfunctions of $(h^*h)(\tau', \tau)$.

### 4.2.4 Extension to the SIMO Multi-Channel System

We can directly extend the above design process to the SIMO multi-channel system. Assume $A = hG$ where $G$ is determined by the precoders $g_k(t)$ as in (77), but $h$ is now by an $N_R \times 1$ time-varying channel impulse response vector $h = [h_1(t,\tau), \cdots, h_{N_R}(t,\tau)]'$, modelling channel effects from one transmitter to $N_R$ receivers.

The derivation and the final design process are the same except that

$$(h^*h)(\tau', \tau) = \sum_{i=1}^{N_R} \int \bar{h}_i(t, t-\tau')h_i(t, t-\tau)dt,$$

which is a summation of the correlation functions in the local time variables $(\tau', \tau)$, determined by the time-varying channel impulse responses in each channel.

So, in the design rules, the space $\mathcal{S}$ and the corresponding basis $\{\psi_k\}_{k=1}^n$ should be chosen to match the channel vector $h = [h_1(t,\tau), \cdots, h_{N_R}(t,\tau)]'$ under transmitter constraints. All other steps are the same. As long as

$$(h^*h)(\tau', \tau) = \sum_{i=1}^{N_R} \int \bar{h}_i(t, t-\tau')h_i(t, t-\tau)dt$$

is continuous, nonzero, and square integrable on $I \times I$ (I is an interval on $\mathcal{R}$), we can find such a space $\mathcal{S}$.

### 4.2.5 Canonical Coordinates and Geometry

Canonical correlations measure cosines of principal angles between random vectors in a Hilbert space. These cosines of principal angles (also called canonical correlations between canonical coordinates) between the message and measurement spaces determine error covariance, information rate, and capacity. For reduced-rank and/or quantized estimation of one random vector from another, canonical coordinate designs are known for estimation at minimum MSE and maximum information rate [6] - [7]. These designs show that reduced rank estimation must be done in a system of canonical coordinates.

In this section, we investigate the canonical coordinates between transmit vector $u$ and receive vector $v$, after the design of the optimal precoder and equalizer.

From [8], we know that the mutual information between source vector $u$ and measurement vector $v$ is $I(u,v) = \log\det(I + S)$, where $S$ is the signal-to-noise ratio matrix on the receiver side[2]. Moreover the connection between $S$ and the so-called coherence matrix $C$ is $(I - CC^H) = (I + S)^{-1}$, where $C$ is the cross correlation between whitened versions of the message $u$ and measurement $v$: $C := R_{uu}^{-1/2} R_{uv} R_{vv}^{-H/2}$ and $CC^H = R_{uu}^{-1/2} R_{uv} R_{vv}^{-1} R_{vu} R_{vv}^{-T/2}$.

From the design results in subsection 4.2.1, we find that $I(u;v) = \log\det(I + S) = \log\det(I + \sigma^{-2}\Phi^H \Lambda \Phi)$ with $\Phi^H \Lambda \Phi$ diagonal, which means the signal-to-noise ratio matrix $S$ is diagonal and furthermore, $CC^H$ is diagonal.

Thus, the optimal precoder design makes squared coherence matrix $CC^H$ diagonal. Directly from the design, we do not know if $C$ itself is diagonal or not. But, without loss of generality, $C$ can be diagonal. This means the combined effects of precoder and equalizer have been to transform the whitened source $\tilde{u} = R_{uu}^{-1/2} u$, and equalized and whitened measurement $\tilde{v} = R_{vv}^{-1/2} v$, into a system

[2]Suppose the equivalent channel model is $v = Hu + n$, where $u$ and $v$ are respectively the message and measurement vectors, $H$ is the equivalent channel transformation and $n$ is the noise vector with covariance matrix $R_{nn}$. Then $S = R_{nn}^{-1} H R_{uu} H^H$.

of canonical coordinate where variables are pairwise correlated with correlations $|E\tilde{u}_i\bar{\tilde{v}}_i|^2 = (CC^H)_{ii} = \frac{|\phi_i|^2}{|\phi_i|^2+\frac{1}{\lambda_i}}$. Because $\tilde{u}_i$ ($i$th element of $\tilde{u}$) and $\tilde{v}_i$ ($i$th element of $\tilde{v}$) have unit variance, we may call $(CC^H)_{ii}$ the cosine-squared of the angle between $\tilde{u}_i$ and $\tilde{v}_i$. It is as if the whitened source $\tilde{u}$ is communicating over $\bar{m}$ uncorrelated channels, where $\bar{m}$ is the number of channels with $|\phi_i| > 0$. In each such channel, the SNR and squared canonical correlations are $\text{SNR}_i = \frac{|\phi_i|^2}{\frac{1}{\lambda_i}}$ and $(CC^H)_i = \frac{|\phi_i|^2}{|\phi_i|^2+\frac{1}{\lambda_i}}$. The geometry of each channel is illustrated in Figure 18. In this Pythagorean decomposition of a subchannel, $|\phi_i|^2$ is the message power, $\frac{1}{\lambda_i}$ the noise power, and $\frac{|\phi_i|^2}{|\phi_i|^2+\frac{1}{\lambda_i}}$ is the cosine-squared of the angle between the whitened message and the equalized and whitened measurement. This interpretation is further illuminated in Figure 19, where the whitened variables $\tilde{u} = R_{uu}^{-1/2}u$ and $\tilde{v} = R_{vv}^{-1/2}v$ are illustrated. *It is the geometry of these whitened variables that is illuminating.* The precoder $GR_{uu}^{1/2}$ and equalizer $R_{vv}^{-1/2}F$ transform these white variables into a canonical coordinate system. Here $F$ represents the equalizer which is the matched filter under the maximizing mutual information criterion. The coloring steps $R_{uu}^{1/2}$ and $R_{vv}^{1/2}$ that take the white variables to the original variables only obscure this geometry.

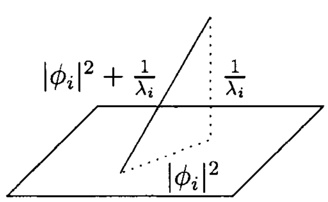



Figure 18. Pythagorean decomposition of a subchannel

## 4.3 Signaling Waveforms for Minimum Mean Square Error

We derive the precoder and equalizer that minimize the mean square error between $u$ and $Fv$. It is well known that the Wiener filter

$$F = R_{uv}R_{vv}^{-1} = R_{uu}(A^*A)^{1/2}\left[(A^*A)^{1/2}R_{uu}(A^*A)^{1/2} + \sigma^2 I\right]^{-1}$$

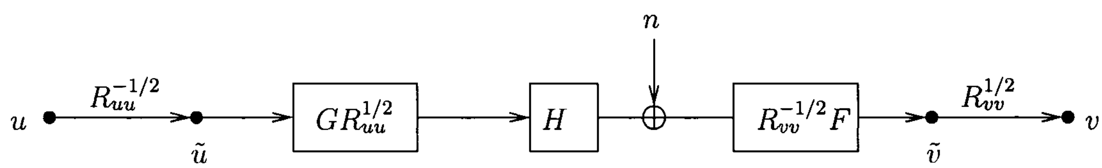



Figure 19. An equivalent system diagram with precoder and equalizer, where the precoder and the equalizer transform the whitened variables $\hat{u}$ and $\hat{v}$ into canonical coordinates.

minimizes the mean square error. After applying $F$, the mean square error between $u$ and $Fv$ is

$$\operatorname{tr}\left(R_{uu}^{-1} + \sigma^{-2}(A^*A)\right)^{-1}. \tag{84}$$

### 4.3.1 Precoding

We introduce precoder $G$ and channel filter $h$ as in (77), and assume $A = hG$. Then (84) becomes

$$\begin{aligned}
& \operatorname{tr}\left(R_{uu}^{-1} + \sigma^{-2}(A^*A)\right)^{-1} \\
= \; & \operatorname{tr}\left(R_{uu}^{-1} + \sigma^{-2}(G^*h^*hG)\right)^{-1} = \operatorname{tr}\left(R_{uu}^{-1} + \sigma^{-2}\Gamma^H Q \Gamma\right)^{-1} \\
= \; & \operatorname{tr}\left(R_{uu}^{-1} + \sigma^{-2} R_{uu}^{-1/2}\Phi^H \Lambda \Phi R_{uu}^{-1/2}\right)^{-1} = \operatorname{tr}\left(\Delta^{-1} + \sigma^{-2}\Delta^{-1/2}U^H\Phi^H\Lambda\Phi U \Delta^{-1/2}\right)^{-1} \\
= \; & \operatorname{tr}\left(\Delta(I + \sigma^{-2}\hat{\Phi}^H \Lambda \hat{\Phi})^{-1}\right),
\end{aligned}$$

where $G^*h^*hG, Q, \Gamma, V, \Lambda, \Phi$ are defined as in subsection 4.2.1. In the above formulae we introduce the eigenvalue decomposition $R_{uu} = U\Delta U^H$, and define $\hat{\Phi} := \Phi U$. So, minimizing the mean square error subject to a power constraint becomes the following optimization problem:

$$\begin{aligned}
\min \quad & \operatorname{tr}\left(\Delta(I + \sigma^{-2}\hat{\Phi}^H \Lambda \hat{\Phi})^{-1}\right) \\
\text{s.t.} \quad & \operatorname{tr}(\hat{\Phi}\hat{\Phi}^H) \leq \mathcal{P}.
\end{aligned}$$

The same problem has been solved in [9]. Here we use the majorization idea of [5] to re-solve this optimization problem. By Theorem 9.H.1.h. in [10],

$$\mathrm{tr}(\Delta(I+\sigma^{-2}\hat{\Phi}^H\Lambda\hat{\Phi})^{-1}) \geq \sum_{i=1}^{m} \delta_i \mathrm{ev}_{m-i+1}((I+\sigma^{-2}\hat{\Phi}^H\Lambda\hat{\Phi})^{-1}),$$

where $\mathrm{ev}_{m-i+1}(A)$ is the $(m-i+1)$th eigenvalue of Hermitian matrix $A$, with eigenvalues sorted in decreasing order, and $\delta_i$ is the $i$th eigenvalue of the diagonal matrix $\Delta$ with eigenvalues sorted in decreasing order. The equality holds when $(I+\sigma^{-2}\hat{\Phi}^H\Lambda\hat{\Phi})^{-1}$ is diagonal with diagonal elements in increasing order, or when $(\hat{\Phi}^H\Lambda\hat{\Phi})$ is diagonal with diagonal elements in decreasing order. Suppose $\hat{\Phi}^H\Lambda\hat{\Phi} = \Sigma$, where $\Sigma$ is an $m\times m$ diagonal matrix with diagonal elements sorted in decreasing order. W.l.o.g. $\hat{\Phi} = \Lambda^{-1/2}P\Sigma^{1/2}$, where $P$ is an $n\times m$ matrix such that $P^HP = I$. Putting this $\hat{\Phi}$ in the power constraint, we get

$$\mathrm{tr}(\hat{\Phi}\hat{\Phi}^H) = \mathrm{tr}(\Lambda^{-1}P\Sigma P^H) \geq \sum_{i=1}^{m} \lambda_i^{-1}\xi_i,$$

where $\lambda_i$ is the $i$th diagonal element of $\Lambda$, whose eigenvalues are sorted in decreasing order, and $\xi_i$ is the $i$th element of $\Sigma$. The equality holds when $P = \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix}$. Summarizing these arguments, for all $\hat{\Phi} = \Lambda^{-1/2}P\Sigma^{1/2}$ which yield the same mean square error, choosing $P = \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix}$ minimizes the power. So, the optimal $\hat{\Phi}$ is $\hat{\Phi} = \Lambda^{-1}\begin{bmatrix} \Sigma^{1/2} \\ 0_{(n-m)\times m} \end{bmatrix}$, where the diagonal elements of $\Lambda$ and $\Sigma$ are in decreasing order. The matrix optimization problem then yields the scalar optimization problem

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} \delta_i(1+\sigma^{-2}\xi_i)^{-1} \\ \text{s.t.} \quad & \sum_{i=1}^{m} \lambda_i^{-1}\xi_i \leq \mathcal{P}. \end{aligned}$$

Solving this problem we get

$$\xi_i = \sigma^2 \left[ \frac{\sqrt{\lambda_i\delta_i}(\sigma^{-2}\mathcal{P} + \sum_{k=1}^{\tilde{m}} \lambda_k^{-1})}{\sum_{k=1}^{\tilde{m}} \sqrt{\lambda_k^{-1}\delta_k}} - 1 \right]^+.$$

This yields the diagonal elements of diagonal matrix $\hat{\Phi}^H\hat{\Phi}$

$$|\phi_i|^2 = \sigma^2 \left[ \frac{\sigma^{-2}\mathcal{P} + \sum_{k=1}^{\bar{m}} \lambda_k^{-1}}{\sum_{k=1}^{\bar{m}} \sqrt{\lambda_k^{-1}\delta_k}} \sqrt{\lambda_i^{-1}\delta_i} - \lambda_i^{-1} \right]^+ . \tag{85}$$

Here $[x]^+ = \max(x, 0)$, and $\bar{m}$ is the number of channels where $\xi_i > 0$ or $|\phi_i|^2 > 0$. Summarizing, the coefficient matrix for the optimum precoder $\Gamma$ is $\Gamma = V\hat{\Phi}U^H R_{uu}^{-1/2}$, where $\hat{\Phi}$ is diagonal with squared diagonal elements in (85).

The minimized mean square error is

$$\text{MMSE} := \frac{(\sum_{i=1}^{\bar{m}} \sqrt{\lambda_i^{-1}\delta_i})^2}{\sigma^{-2}\mathcal{P} + \sum_{i=1}^{\bar{m}} \lambda_i^{-1}}.$$

To find the best subspace which minimizes the mean square error, we also choose space $\mathcal{S}$ spanned by $\varphi_i(\tau), 1 \leq i \leq N_\epsilon$ which are the eigenfunctions of $(h^*h)(\tau', \tau)$ for an arbitrary small $\epsilon$, and solve the optimization problem to get $\bar{m}$ and the optimal precoders. From the optimization solution, we see that the precoders are in the subspace spanned by the dominating eigenfunctions of $(h^*h)(\tau', \tau)$. This conclusion is the same as the conclusion reached for maximizing the mutual information. Therefore, the precoder space design is the same. Only the coefficient matrix for designing the precoder waveforms from the precoder space is different.

There is no change in the extension to the SIMO channel.

### 4.3.2 Half Canonical Coordinates and Geometry

Half canonical correlations measure correlation between source vectors and whitened measurement vectors, after they have been resolved into a half-canonical coordinate system. In this subsection, we investigate half canonical coordinates between the transmit vector $u$ and receive vector $\hat{v}$ after optimal precoding and equalizing.

We know from the last subsection that the optimal mean square error matrix is

$$\begin{aligned} M &:= (R_{uu}^{-1} + \sigma^{-2} R_{uu}^{-1/2} \Phi^H \Lambda \Phi R_{uu}^{-1/2})^{-1} \\ &= U(\Delta^{-1} + \sigma^{-2} \Delta^{-1/2} \hat{\Phi}^H \Lambda \hat{\Phi} \Delta^{-1/2})^{-1} U^H. \end{aligned}$$

Define the matrix [6] $DD^H = R_{uu} - M$, in which case

$$DD^H = U\left[\Delta - (\Delta^{-1} + \sigma^{-2}\Delta^{-1/2}\hat{\Phi}^H \Lambda \hat{\Phi} \Delta^{-1/2})^{-1}\right] U^H.$$

Here $D$ is the half coherence matrix, which is the cross covariance matrix between the transmit vector $u$ (un-whitened) and the whitened receive vector $\tilde{v}$. Then the squared half canonical correlation matrix [6] can be derived from $DD^H$ by simply applying the normal matrix $U$ as follows

$$LL^H = U^H DD^H U = \Delta - (\Delta^{-1} + \sigma^{-2}\Delta^{-1/2}\hat{\Phi}^H \Lambda \hat{\Phi} \Delta^{-1/2})^{-1}.$$

$LL^H$ is diagonal. We do not know if $L$ itself is diagonal. But, w.l.o.g. $L$ can be made diagonal easily from $LL^H$.

W.l.o.g. this means the combined effects of the precoder and equalizer have been to transform the decorrelated (but not uniform variance) source $\tilde{u} := U^H u$, with diagonal covariance $\Delta$, and equalized and whitened measurement $\tilde{v}$ into a system of half canonical coordinates where variables are only pairwise correlated, with correlation $|E\tilde{u}_i \tilde{v}_i^*|^2 = (LL^H)_i = \frac{\delta_i |\phi_i|^2}{|\phi_i|^2 + \frac{1}{\lambda_i}}$. If we normalize the variable $(\tilde{u})_i$ by its standard deviation $\sqrt{\delta_i}$, then the cosine-squared of the angle between $(\tilde{u})_i/\sqrt{\delta_i}$ and $(\tilde{v})_i$ is $|E\frac{\tilde{u}_i}{\sqrt{\delta_i}}\tilde{v}_i^*|^2 = \frac{|\phi_i|^2}{|\phi_i|^2 + \frac{1}{\lambda_i}}$. Again we see that the channel is decomposed into parallel uncorrelated channels where the SNR in each channel is $\frac{|\phi_i|^2}{\frac{1}{\lambda_i}}$, and the cosine-squared of the angle between $(\tilde{u})_i$ and $(\tilde{v})_i$ remains $\frac{|\phi_i|^2}{|\phi_i|^2 + \frac{1}{\lambda_i}}$. Figures (18) and (19) still apply, with $R_{uu}^{-1/2}$ in Figure (19) replaced by $U^H$. Then the precoder $GU$ and the equalizer $R_{vv}^{-1/2} F$ transform $\tilde{u}$ and $\tilde{v}$ into half canonical coordinate systems. This geometry is obscured in the original coordinates $(u, v)$.

## 4.4 Signaling Waveforms for Minimum Bit Error Rate

Here, we use the model and result of [11]. The transmit symbols are assumed to be equiprobable antipodal symbols with $R_{uu} = I$. Recalling that in (75) the system function is

$$v = (A^*A)^{-1/2}A^*y = (A^*A)^{1/2}u + w,$$

where $w := (A^*A)^{-1/2}A^*n$ is an $m$-dimensional complex $N(0, \sigma^2 I)$ random vector, the equalizer $F = (A^*A)^{-1/2}$ is a zero forcing equalizer. Use $A = hG$ to write $F = (G^*h^*hG)^{-1/2}$, where the definitions of $G^*h^*hG$ are the same as before. Under the minimum bit error rate criterion and under all of the above assumptions, the optimum coefficient matrix $\Gamma$ for the precoder $G$ is $\Gamma = V\Phi D_F^H$, where $V$ is the $n \times n$ unitary matrix in $Q = V\Lambda V^H$, $\Phi$ is an $n \times m$ diagonal matrix with nonzero elements only on its main diagonal, and $D_F^H$ is an $m \times m$ unitary inverse DFT matrix [11].

The probability of error is [11]

$$P_e = \frac{1}{2m}\sum_{i=1}^{m} \operatorname{erfc}\Big(\frac{1}{\sqrt{2\sigma^2(FF^H)_{ii}}}\Big) = \frac{1}{2m}\sum_{i=1}^{m} \operatorname{erfc}\Big(\frac{1}{\sqrt{2\sigma^2(\Phi^H\Lambda\Phi)^{-1}_{ii}}}\Big).$$

This function is convex with respect to $(\Phi^H\Lambda\Phi)^{-1}_{ii}$ if and only if $(\Phi^H\Lambda\Phi)^{-1}_{ii} \leq \frac{1}{3\sigma^2}$ [11]. When it is convex, using Jensen's Inequality,

$$P_e = \frac{1}{2m}\sum_{i=1}^{m} \operatorname{erfc}\Big(\frac{1}{\sqrt{2\sigma^2(\Phi^H\Lambda\Phi)^{-1}_{ii}}}\Big) \geq \frac{1}{2}\operatorname{erfc}\Big(\sqrt{\frac{m}{2\sigma^2 \operatorname{tr}(\Phi^H\Lambda\Phi)^{-1}}}\Big) = P_{e,LB}.$$

The equality holds when $(\Phi^H\Lambda\Phi)^{-1}$ has equal diagonal elements. So the optimization problem becomes

$$\min \operatorname{tr}(\Phi^H\Lambda\Phi)^{-1}. \tag{86}$$

The constraints are both the power constraint and $(\Phi^H\Lambda\Phi)^{-1}_{ii} \leq \frac{1}{3\sigma^2}$. It has been proved in [11] that if and only if $\operatorname{tr}(\Phi^H\Lambda\Phi)^{-1} \leq \frac{m}{3\sigma^2}$, there exists a unitary matrix $B$ such that $(B(\Phi^H\Lambda\Phi)^{-1}B^H)_{ii} \leq \frac{1}{3\sigma^2}$. So, the problem is simplified to minimizing

(86) subject to the power constraint. And the solution is feasible if and only if $\text{tr}(\Phi^H\Lambda\Phi)^{-1} \le \frac{m}{3\sigma^2}$. We introduce the singular value decomposition $\Phi = PTW^H$, where $P$ is a $n \times n$ unitary matrix, $T$ is a $n \times m$ diagonal matrix with nonzero elements only on its main diagonal in increasing order, and $W$ is an $m \times m$ unitary matrix. Since we have assumed that $n \ge m$, $T = \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix} \hat{T}$. Here $\hat{T}$ is a $m \times m$ diagonal matrix with diagonal elements in increasing order. So (86) becomes

$$\begin{aligned}\text{tr}(\Phi^H\Lambda\Phi)^{-1} &= \text{tr}(\hat{T}^{-H}(\begin{bmatrix} I_{m\times m} & 0_{m\times(n-m)} \end{bmatrix} P^H\Lambda P \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix})^{-1}\hat{T}^{-1}) \\ &= \text{tr}((\hat{T}^H\hat{T})^{-1}(\begin{bmatrix} I_{m\times m} & 0_{m\times(n-m)} \end{bmatrix} P^H\Lambda P \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix})^{-1}).\end{aligned}$$

The power constraint becomes $\text{tr}(\hat{T}\hat{T}^H) \le \mathcal{P}$. Define $\Sigma = (\hat{T}^H\hat{T})^{-1}$. So, $\Sigma$ is diagonal with diagonal elements in decreasing order. Define $Z = (\begin{bmatrix} I_{m\times m} & 0_{m\times(n-m)} \end{bmatrix} P^H\Lambda P \begin{bmatrix} I_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix})^{-1}$. According to Theorem 9.H.1.h. in [10], $\text{tr}(\Sigma Z) \ge \sum_{i=1}^m \xi_i \text{ev}_i(Z)$, where $\xi_i$ is the $i$th diagonal element of $\Sigma$, and $\text{ev}_i(Z)$ is the $i$th eigenvalue of $Z$, with eigenvalues in increasing order. The equality holds when $Z$ is diagonal with diagonal elements in increasing order. Since $\Lambda$ is diagonal with diagonal elements in decreasing order, the optimal $P$ is $P = I$. The problem is simplified to a scalar optimization problem, and we can easily get the solution $|t_i|^2 = \frac{\mathcal{P}\sqrt{\lambda_i^{-1}}}{\sum_{k=1}^m \sqrt{\lambda_k^{-1}}}$. If and only if $\frac{(\sum_{i=1}^m \sqrt{\lambda_i^{-1}})^2}{\mathcal{P}} \le \frac{m}{3\sigma^2}$, the solution is feasible. In case the solution is feasible, we should find a unitary matrix $B$ such that $(B(\Phi^H\Lambda\Phi)^{-1}B^H)_{ii} = (BW(T^H\Lambda T)^{-1}W^HB^H)_{ii} \le \frac{1}{3\sigma^2}$. From [11], $B = D_FW^H$ is an optimal choice since it makes the diagonal elements equal. Here $D_F$ is a unitary DFT matrix. So the optimal $\Phi$ is $\Phi = TD_F^H$. The optimal $\Gamma$ is $\Gamma = VTD_F^H$.

From above, we see that the equivalent optimization problem in (86) becomes $\frac{(\sum_{i=1}^m \sqrt{\lambda_i^{-1}})^2}{\mathcal{P}}$ after applying the optimal solution. The bigger the $\lambda_i$, the smaller the objective function. So, we still want the $\lambda_i$ as large as possible, as in the previous two designs.

The extension to the SIMO channel remains unchanged.

### 4.4.1 Canonical Coordinates and Geometry

With this optimum coefficient matrix $\Gamma$ for the precoder $G$ the squared coherence matrix is

$$CC^H = D_F \Phi^H \Lambda^{1/2} (\Lambda^{1/2} \Phi \Phi^H \Lambda^{1/2} + I)^{-1} \Lambda^{1/2} \Phi D_F^H,$$

from which we easily get the diagonal squared canonical correlation matrix by applying the DFT matrix

$$KK^H = D_F^H CC^H D_F = \Phi^H \Lambda^{1/2} (\Lambda^{1/2} \Phi \Phi^H \Lambda^{1/2} + I)^{-1} \Lambda^{1/2} \Phi^H.$$

This means the combined effects of precoder and equalizer have been to transform the white source $\hat{u} = D_F^H u$ and equalized and whitened measurement $\tilde{v}$ into a system of canonical coordinates where variables are only pairwise correlated:

$$|E\hat{u}_i \tilde{v}_i^*|^2 = (KK^H)_i = \frac{|\phi_i|^2}{|\phi_i|^2 + \frac{1}{\lambda_i}}.$$

It is as if the whitened source $D_F^H u$ is communicating over $m$ uncorrelated channels. In each, the cosine-squared is

$$(KK^H)_i = \frac{|\phi_i|^2}{|\phi_i|^2 + \frac{1}{\lambda_i}}.$$

Again, the geometry of each channel is that of previous designs. Figures (18) and (19) still apply, but now $R_{uu}^{-1/2}$ is replaced by $D_F^H$. Then the precoder $GD_F$ and equalizer $R_{vv}^{-1/2} F$ transform $\tilde{u} = D_F^H u$ and $\tilde{v} = R_{vv}^{-1/2} v$ into canonical coordinates.

## 4.5 Examples

### 4.5.1 An Example for the Time-invariant Frequency-selective Channel

In wireless communications, when the spread factor of the channel satisfies the condition $T_m B_d << 1$ [12], where $T_m$ is the time spread and $B_d$ is the doppler spread, it is possible to select signals having a signal duration $T < \Delta t_c$, and a bandwidth $W > \Delta f_c$, where $\Delta t_c \approx \frac{1}{B_d}$ is the coherence time, and $\Delta f_c \approx \frac{1}{T_m}$ is the coherence band-width. Then the signal sees a time-invariant frequency-selective channel with the tapped-delay-line channel model [12] and time-invariant channel taps

$$h(\tau) = \sum_{l=0}^{L-1} h_l \delta(\tau - \frac{l}{W}).$$

The second order channel correlation $(h^*h)(\tau', \tau)$ will be

$$(h^*h)(\tau', \tau) = \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} h_l^* h_k \delta((\tau' + \frac{l}{W}) - (\tau + \frac{k}{W})).$$

It is easy to get that the $n$th eigenvector of $(h^*h)(\tau', \tau)$ is $e^{\frac{-j2\pi n\tau}{T}}$ corresponding to the $n$th eigenvalue $|H_n|^2$ which is defined to be the squared magnitude of the complex frequency response of the channel, sampled at frequency $\frac{2\pi n}{TW}$:

$$H_n = \sum_{l=0}^{L-1} h_l e^{\frac{-j2\pi nl}{TW}}.$$

Therefore, the subspace spanned by the $N_\epsilon$ Fourier basis functions $\{e^{\frac{-j2\pi n\tau}{T}}\}_1^{N_\epsilon}$ is the best signal space for the time-invariant frequency-selective channel model. The well known OFDM technique is used to diagonalize the channel.

### 4.5.2 An Example for the Time-varying Frequency-selective Channel

If we select signals having a signal duration $T > \Delta t_c$, and a bandwidth $W > \Delta f_c$, then the signal sees a time-varying frequency-selective channel. For this time-varying frequency-selective channel, we still assume a time-spread $T_m$ and Doppler-spread $B_d$, with product $T_m B_d << 1$, which means the channel is underspread. The coherence time is approximately the inverse of the Doppler-spread $\Delta t_c \approx \frac{1}{B_d}$ and the coherence bandwidth is approximately the inverse of the time-spread $\Delta f_c \approx \frac{1}{T_m}$. In Figure 20 the rectangle associated with each of four channel descriptions is intended to convey qualitative features of the channel. For example, the scattering function $S(\nu, \tau)$ has finite support defined by $B_d$ and $T_m$.

The input-output description of the channel is

$$y(t) = \int X(f)H(t,f)e^{j2\pi ft}df.$$

We follow the arguments in [13] to constrain the bandwidth of the input signal to $W$ and constrain the output observation time to $T$. Then we get the equivalent time-varying frequency response as follows

$$\hat{H}(t,f) = \chi_T(t)H(t,f)\chi_W(f),$$

where

$$\chi_T(t) = \begin{cases} 1, & -T/2 \leq t \leq T/2 \\ 0, & \text{otherwise} \end{cases}.$$

By Fourier transforming $\hat{H}(t,f)$, we can get the scattering or ambiguity function

$$S(\nu,\tau) = \sum_{m}^{M-1}\sum_{n}^{N-1} S[m,n]\text{sinc}(T(\nu - \frac{m}{T}))\text{sinc}(W(\tau - \frac{n}{W})),$$

where $S[m,n] = \frac{1}{TW}S(\frac{m}{T},\frac{n}{W})$. $m$ and $n$ are constrained to $0 \leq m \leq M-1$ and $0 \leq n \leq N-1$ since we've assumed a time-spread $T_m$ and Doppler-spread $B_d$. So, $\frac{M}{T} = B_d$ or $M = TB_d \approx \frac{T}{\Delta t_c}$ and $\frac{N}{W} = T_m$ or $N = WT_m \approx \frac{W}{\Delta f_c}$. From $S(\nu,\tau)$, it is easy to get the time-varying impulse response as

$$h(t,\tau) = \sum_{m}^{M-1}\sum_{n}^{N-1} S[n,m]\text{sinc}(W(\tau - \frac{n}{W}))e^{-j2\pi\frac{m}{T}t}.$$

A channel identification algorithm would estimate the ambiguity coefficients $S[n,m]$ by transmitting training pulses with resolution $\frac{1}{W}$ in time and $\frac{1}{T}$ in doppler.

Note that $\text{sinc}(W(\tau - \frac{n}{W}))e^{-j2\pi t\frac{m}{T}}$, forall $n,m$ is a basis of $L^2(\mathcal{R})$, called the Gabor basis. Then $\text{sinc}(W(\tau - \frac{n}{W}))e^{-j2\pi t\frac{m}{T}}$, forall $0 \leq n \leq N-1, 0 \leq m \leq M-1$ is a basis for a subspace of $L^2(\mathcal{R})$. Although this basis does not diagonalize $(h^*h)(\tau',\tau)$, we can use it as an approximating device to construct the space $\mathcal{S}$, and then follow the design rules to find the precoder waveforms $g_j(t)$.

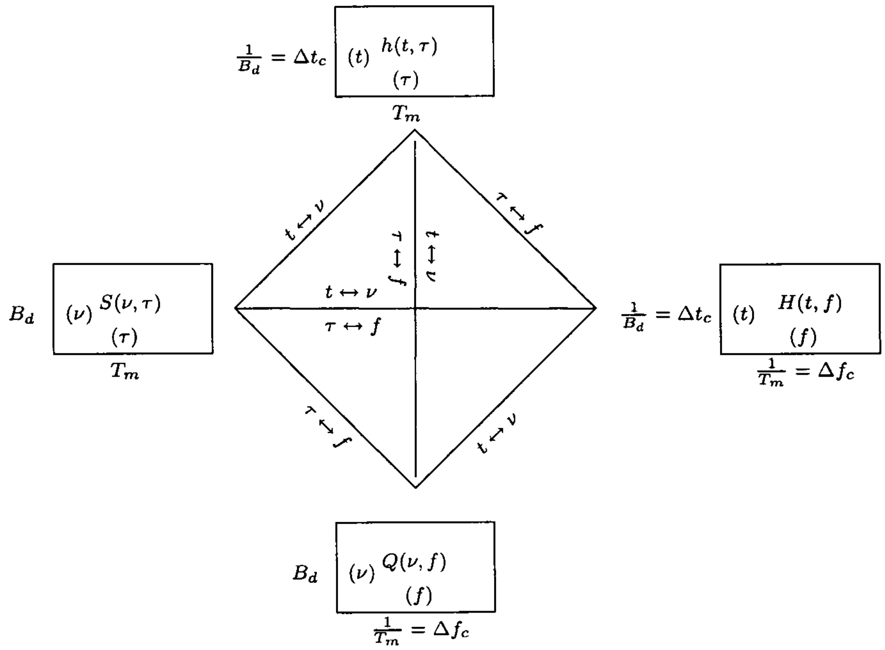



Figure 20. The four corners diagram of the four Fourier transforms of time-variant and frequency-selective channel $h(t, \tau)$

There are many other ways to find the near-optimal basis for this linear time-variant channel. Prolate spheroidal functions are an obvious, but impractical choice. In [14], the eigenfunctions are approximated, within an error bounded by the moments of the channel spread function, by multicomponent signals. In [15], the author discusses many near-optimal basis methods, and presents a novel channel partitioning and modulation technique using adaptive bases of localized complex exponentials for linear time-varying channels.

## 4.6 Practical Considerations and Conclusions

The first step in the design rule is to choose a precoder space $\mathcal{S}$ and a corresponding basis to match the channel $h(t, \tau)$, possibly under constraints on the transmitter. Under all optimization criteria we have considered, the best space to be the one spanned by a subset of the eigenvectors of $(h^*h)(\tau', \tau)$. From the examples in the last section, we find that practically, it is difficult to get exactly the eigenvalues and eigenvectors of $(h^*h)(\tau', \tau)$ for a given channel model, so we may have to choose a space that only approximates the optimal one. In this case, the choice of a subspace basis is suboptimal. In the space we choose, however, the solution is optimal. In the last example, we present many methods to find the near-optimal eigenfunctions for the linear time-variant channel.

In these analog precoder and equalizer designs we assume that the transmitter and receiver both know the channel, which requires that the channel changes slowly enough for the receiver side to estimate the channel and feed back the channel information to the transmitter. It is impractical for the receiver side to feedback the basis and coefficient matrix to the transmitter side when the channel changes fast. One way to proceed is to assume that only channel coefficients change within a fixed basis, at a slow rate. This makes it possible to only feed back the coefficient matrix.

In conclusion, the results show that analog precoder design may be carried out in two steps: first, the precoder space is matched to the analog channel, and second, the precoder functions are selected from this subspace according to standard rules [11], [5], [9], [3], [4]. The resulting designs have an illuminating canonical coordinate geometry.

## List of References

[1] J. Gubner, "Maximum mutual information signaling waveforms," Sept. 2003, unpublished notes.

[2] E. A.Lee and D. G.Messerschmitt, *Digital Communication, 2nd Edition.* Kluwer Academic Publishers: Boston/Dordrecht/London, 1994.

[3] A. Scaglione, S. Barbarossa, and G. B.Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[4] A. Scaglione, P. Stoica, S. Barbarossa, G.B.Giannakis, and H.Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[5] D. P.Palomar, J. M.Cioffi, and M. A.Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: a unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[6] L. L.Scharf, *Statistical Signal Processing.* Addison-Wesley, 1990.

[7] P. J.Schreier and L. L.Scharf, "Canonical coordinates for reduced-rank estimation of improper complex random vectors," in *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, vol. 2, 2002, pp. 1153 –1156.

[8] L. L.Scharf and C. T.Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 824–831, Mar. 2000.

[9] A. Scaglione, G. B.Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part 1: unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[10] A. W.Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* New York, NY, USA: Academic Press, 1979.

[11] Y. Ding, T. N.Davidson, Z. Luo, and K. M.Wong, "Minimum ber block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2410–2423, Sept. 2003.

[12] J. G.Proakis, *Digital Communications, 4th Edition.* McGrawHill, 2000.

[13] P. A.Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[14] S. Barbarossa and A. Scaglione, "Optimal precoding for transmissions over linear time-varying channels," in *Proc. Global Communications Conf.*, Dec. 1999, pp. 2545–2549.

[15] R. Narasimhan, "Adaptive channel partitioning and modulation for linear time-varying channels," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1313–1324, Aug. 2003.

CHAPTER 5

# Precoder and Equalizer Design in the CDMA system

In this chapter we aim to design transmitter-receiver pairs that simplify receiver design. We choose a CDMA system to demonstrate the problem and its solution.

## 5.1 Down-link Synchronous CDMA System for Flat Fading Channel

We consider the synchronous CDMA downlink system. The transmit signal of the $k$th user is

$$x_k(t) = A_k b_k s_k(t),$$

where $A_k$, $b_k$ and $s_k(t)$ are respectively the amplitude, the transmit symbol $+1$ or $-1$, and the signature waveform of the $k$th user. The received signal in the base station is the sum of $K$ such signals, from $K$ users,

$$y(t) = \sum_{k=1}^{K} A_k b_k s_k(t) + n(t).$$

On the receiver side we apply the matched filter and get the baseband equivalent vector model

$$\boldsymbol{r} = \boldsymbol{RAb} + \boldsymbol{n},$$

where $\boldsymbol{R}$ is the $K \times K$ autocorrelation matrix whose $(i,j)$th element is $\rho_{ij} = \int \bar{s}_i(t) s_j(t) dt$, $\boldsymbol{A}$ is a diagonal matrix with $k$th diagonal element $A_k$; $\boldsymbol{b}$ and $\boldsymbol{y}$ are $K$-dimensional transmit and receive vectors, and $\boldsymbol{n}$ is a $K$-dimensional noise vector with covariance matrix $\sigma^2 \boldsymbol{R}$. If we apply a $K \times K$ precoder $\boldsymbol{G}$ on the transmitter side, then the received signal is

$$\boldsymbol{r} = \boldsymbol{RGAb} + \boldsymbol{n}.$$

Then the problem of minimizing mean square error for a linear estimator of $\boldsymbol{b}$, subject to a transmit power constraint, can be formulated as the following optimization problem

$$\begin{aligned} \min \quad & \operatorname{tr}(\boldsymbol{I} + \sigma^{-2}\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G}\boldsymbol{A})^{-1} \\ \text{st.} \quad & \operatorname{tr}(\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G}\boldsymbol{A}) \leq \mathcal{P}. \end{aligned} \tag{87}$$

It is easy to get the solution

$$\boldsymbol{G} = \sqrt{\frac{\mathcal{P}}{r}}\boldsymbol{U}\boldsymbol{\Lambda}^{\dagger 1/2}\boldsymbol{V}^H\boldsymbol{A}^{-1},$$

where we decompose $\boldsymbol{R}$ into $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ and $\boldsymbol{V}$ is an arbitrary $K \times K$ unitary matrix. $\boldsymbol{\Lambda}^\dagger$ is the pseudo-inverse of $\boldsymbol{\Lambda}$.

With this precoder, the MMSE matrix is

$$\begin{aligned} \text{MMSE} &:= (\boldsymbol{I} + \sigma^{-2}\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G}\boldsymbol{A})^{-1} \\ &= (\boldsymbol{I} + \frac{\mathcal{P}}{r\sigma^2}\hat{\boldsymbol{I}})^{-1}. \end{aligned} \tag{88}$$

where $\hat{\boldsymbol{I}}$ is a diagonal matrix consisting of a number of 1s equal to the rank of $\boldsymbol{R}$ and remaining 0s.

Since the covariance matrix of the received vector $r$ is

$$\begin{aligned} \boldsymbol{R}_{rr} &= \boldsymbol{R}\boldsymbol{G}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{R} + \sigma^2\boldsymbol{R} \\ &= (\boldsymbol{R}\boldsymbol{G}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{G}^H + \sigma^2\boldsymbol{I})\boldsymbol{R}, \end{aligned} \tag{89}$$

the Wiener filter on the receiver side is

$$\begin{aligned} \boldsymbol{R}_{br}\boldsymbol{R}_{rr}^{-1} &= \boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{R}^H\boldsymbol{R}^{-1}(\boldsymbol{R}\boldsymbol{G}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{G}^H + \sigma^2\boldsymbol{I})^{-1} \\ &= \boldsymbol{A}^H\boldsymbol{G}^H(\boldsymbol{R}\boldsymbol{G}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{G}^H + \sigma^2\boldsymbol{I})^{-1}. \end{aligned} \tag{90}$$

On the receiver side, the matrix inverse is the most computationally intensive part in the Wiener filter. Practically, we may use the steepest descent or conjugate gradient algorithms to iteratively solve the quadratic problem to approximate the matrix inverse. With the optimal precoder $\boldsymbol{G}$, the matrix inverse part of the Wiener filter becomes

$$\boldsymbol{RGAA}^H\boldsymbol{G}^H + \sigma^2\boldsymbol{I} = \boldsymbol{U}(\frac{\mathcal{P}}{r}\hat{\boldsymbol{I}} + \sigma^2\boldsymbol{I})\boldsymbol{U}^H. \tag{91}$$

This matrix has one distinct eigenvalue when the rank of $\boldsymbol{R}$ is $K$, or when all the $K$ signature waveforms are linearly independent of each other. When the rank of $\boldsymbol{R}$ is less than $K$, the matrix has two distinct eigenvalues. The conjugate gradient algorithm will converge in $n$ steps, where $n$ is the number of the distinct eigenvalues of the matrix (91). So, with the precoder which minimizes the mean square error, the conjugate gradient algorithm converges in one or two steps.

### 5.1.1 Performance Analysis

We investigate the channel model again. Suppose we use the DS-CDMA waveform

$$s_k(t) = \sum_{i=1}^{N} c_{ki}\psi(t - iT_c).$$

On the receiver side the received vector can be expressed as

$$\boldsymbol{y} = \boldsymbol{CGAb} + \boldsymbol{n},$$

where $\boldsymbol{C}$ is an $N \times K$ signature matrix and $\boldsymbol{y}$ is the received vector sampled at the chip rate. When $< \psi(t - iT_c), \psi(t - jT_c) >= \delta_{ij}$, then $\boldsymbol{R} = \boldsymbol{C}^H\boldsymbol{C}$. So we can decompose $\boldsymbol{C}$ as $\boldsymbol{C} = \boldsymbol{W}\Lambda^{1/2}\boldsymbol{U}^H$ where $\boldsymbol{W}$ is an $N \times K$ matrix such that $\boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{I}$. Putting the optimal $\boldsymbol{G}$ into the formula, we get that

$$\boldsymbol{y} = \sqrt{\frac{\mathcal{P}}{r}}\boldsymbol{W}\boldsymbol{V}^H\boldsymbol{b} + \boldsymbol{n}.$$

We see that the equivalent signatures after the precoder are orthonormal to each other if $K \leq N$, or with the appropriate choice of $\boldsymbol{V}$ are sequences that meet the so-called Welch-bound equality (WBE) when $K \geq N$, and the maximum power is allocated to each user.

The optimal MMSE for the $i$th user is

$$\text{MMSE}_i = ((\boldsymbol{I} + \frac{\mathcal{P}}{r\sigma^2}\boldsymbol{V}\hat{\boldsymbol{I}}\boldsymbol{V}^H)^{-1})_{ii},$$

where $\text{MMSE}_i$ is the $i$th diagonal element of the MMSE matrix defined in 88. If the rank of $\boldsymbol{R}$ is $K$, the MMSE for all the users are equal. Since $\text{MMSE}_i = \frac{1}{1+\text{SIR}_i}$ [1], where $\text{SIR}_i$ is the signal to interference ratio for user $i$, the SIR of each user is $\frac{\mathcal{P}}{r\sigma^2}$.

If the rank of $\boldsymbol{R}$ is less than $K$, the MMSE of each user and therefore the SIR of each user depends on the unitary matrix $\boldsymbol{V}$. We can design $\boldsymbol{V}$ to change each user's SIR. The eigenvalue vector of the MMSE matrix $[1, \cdots, 1, \frac{1}{1+\frac{\mathcal{P}}{r\sigma^2}}, \cdots, \frac{1}{1+\frac{\mathcal{P}}{r\sigma^2}}]^T$ majorizes the diagonal vector $[\text{MMSE}_1, \cdots, \text{MMSE}_K]^T$ of the MMSE matrix. Since $\frac{1}{1+\text{SIR}_i} := \text{MMSE}_i$, we can design $\boldsymbol{V}$ to make $\text{SIR}_i$ satisfy some SIR requirement for each user subject to the constraint that $\frac{1}{1+\text{SIR}_i}$ is majorized by the eigenvalue vector.

## 5.2 Uplink Multipath CDMA System

We consider the uplink of the multipath synchronous CDMA system. The transmit signal of the $k$th user is

$$x_k(t) = \sum_i A_k[i] b_k[i] s_k(t - iT),$$

where $A_k[i]$, $b_k[i]$ and $s_k(t)$ are respectively the amplitude of the $i$th symbol, the $i$th transmit symbol $+1$ or $-1$, and the signature waveform of the $k$th user. $s_k(t) = \sum_{i=1}^{N} c_{ki}\psi(t - iT_c)$, and $T$ is the symbol period. The multipath channel for user $k$ is

$$h_k(t) = \sum_{l=0}^{L_k-1} h_{kl}\delta(t - \tau_{kl}).$$

Then the received signal in the base station is

$$y(t) = \sum_{k=1}^{K} \sum_{i} \sum_{l=0}^{L_k-1} A_k[i] b_k[i] h_{kl} s_k(t - \tau_{kl} - iT).$$

Suppose the chip period is $T_c$, and the processing gain is $N$ such that $T = NT_c$. Assume $\tau_{kl}$ is an integer multiple of $T_c$. Define $L := \max\{L_1, \cdots, L_K\}$. Then the received signal becomes

$$y(t) = \sum_{k=1}^{K} \sum_{i} \sum_{l=0}^{L-1} A_k[i] b_k[i] h_{kl} s_k(t - lT_c - iNT_c).$$

Notice that some $h_{kl}$ may be zero.

Assume the system is synchronous, or $\tau_{k0} = 0, \forall k$. Assume the delay spread of the channel is far less than the symbol period, so that the inter-symbol interference can be ignored. Or, $L << N$, so we can insert a guard interval to eliminate the inter-symbol interference.

On the receiver side we use the chip matched filter and sample at the chip rate. Then we get

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{c}_k A_k b_k$$

during one symbol period. Here $\boldsymbol{y}_k$ is an $(N + L - 1)$ dimensional received vector from user $k$. Theoretically, $\boldsymbol{c}_k$ is an N dimensional signature vector $\boldsymbol{c}_k = [c_{k1}, \cdots, c_{kN}]^T$, and $\boldsymbol{H}_k$ is an $(N + L - 1) \times N$ dimensional Toeplitz matrix with the first column $[h_{k,0}, \cdots, h_{k,L-1}, 0, \cdots, 0]^T$.

Define $(N + L - 1) \times K$ dimensional matrix $\boldsymbol{C}_H = [[\boldsymbol{H}_1 \boldsymbol{c}_1], \cdots, [\boldsymbol{H}_K \boldsymbol{c}_K]]$. Then the received $N + L - 1$ vector from all the users is

$$\boldsymbol{y} = \boldsymbol{C}_H \boldsymbol{A} \boldsymbol{b} + \boldsymbol{n},$$

where $\boldsymbol{A}$ is a $K \times K$ diagonal matrix whose $k$th diagonal element is $A_k$, $\boldsymbol{b}$ is a $K$ dimensional vector with $k$th element $b_k$, and $\boldsymbol{n}$ is a $(N + L - 1)$ dimensional vector with covariance matrix $\sigma^2 \boldsymbol{I}$.

On the receiver side, we use the Wiener filter to minimize the mean square error

$$\boldsymbol{F} = \boldsymbol{R}_{by}\boldsymbol{R}_{yy}^{-1}.$$

Applying the precoder, we get the following optimization problem:

$$\begin{aligned} \min \quad & \operatorname{tr}(\sigma^2\boldsymbol{I} + \boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{C}_H^H\boldsymbol{C}_H\boldsymbol{G}\boldsymbol{A})^{-1} \\ \text{s.t.} \quad & \text{ev}_{\max}(\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{C}^H\boldsymbol{C}\boldsymbol{G}\boldsymbol{A}) \leq \mathcal{P}_0. \end{aligned} \tag{92}$$

In the uplink synchronous CDMA system, we assume we can not distribute power between users. So, we constrain the peak power of each user by constraining the maximum eigenvalue, which dominates all diagonal elements of $\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{C}^H\boldsymbol{C}\boldsymbol{G}\boldsymbol{A}$.

Suppose $\boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$, Decompose $\boldsymbol{G}$ into $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{\Lambda}^{\dagger 1/2}\hat{\boldsymbol{\Phi}}\boldsymbol{A}^{-1}$ and define

$$\boldsymbol{\Gamma} := \boldsymbol{\Lambda}^{\dagger 1/2}\boldsymbol{U}^H\boldsymbol{C}_H^H\boldsymbol{C}_H\boldsymbol{U}\boldsymbol{\Lambda}^{\dagger 1/2} = \boldsymbol{V}\boldsymbol{T}\boldsymbol{V}^H.$$

Let $\boldsymbol{\Phi} = \boldsymbol{V}\hat{\boldsymbol{\Phi}}$. Then the optimization problem becomes

$$\begin{aligned} \min \quad & \operatorname{tr}(\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^H\boldsymbol{T}\boldsymbol{\Phi})^{-1} \\ \text{s.t.} \quad & \text{ev}_{\max}(\boldsymbol{\Phi}^H\boldsymbol{\Phi}) \leq \mathcal{P}_0, \end{aligned} \tag{93}$$

where $\boldsymbol{\Phi}$ is an arbitrary $K \times K$ matrix. Note that the diagonal elements of the diagonal matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{T}$ are in decreasing order.

Solving this problem, we get the optimal $\boldsymbol{\Phi} = \boldsymbol{D}\boldsymbol{W}^H$ where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements

$$d_i = \begin{cases} \sqrt{\mathcal{P}_0} & \forall i \text{ s.t. } t_i \neq 0 \\ 0 & \forall i \text{ s.t. } t_i = 0 \end{cases}, \tag{94}$$

and $\boldsymbol{W}$ is an arbitrary $K \times K$ unitary matrix. Note that diagonal elements of $\boldsymbol{T}$ are the eigenvalues of $\boldsymbol{\Lambda}^{\dagger 1/2}\boldsymbol{U}^H\boldsymbol{C}_H^H\boldsymbol{C}_H\boldsymbol{U}\boldsymbol{\Lambda}^{\dagger 1/2}$ and depend on both the signature and the channel. The measurement covariance is

$$\boldsymbol{R}_{yy} = \boldsymbol{C}_H\boldsymbol{G}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{C}_H^H + \sigma^2\boldsymbol{I}.$$

Then the eigenvalues of $\boldsymbol{R}_{yy}$ are equal to the eigenvalues of

$$\boldsymbol{A}^H\boldsymbol{G}^H\boldsymbol{C}_H^H\boldsymbol{C}_H\boldsymbol{G}\boldsymbol{A} + \sigma^2\boldsymbol{I} = \boldsymbol{\Phi}^H\boldsymbol{T}\boldsymbol{\Phi} + \sigma^2\boldsymbol{I},$$

which are $1 + \mathcal{P}_0 t_i$ using the precoder that minimizes the mean square error.

The precoder that minimizes the mean square error does not give a simple MMSE receiver unless the matrix $\boldsymbol{T}$ has a small number of distinct eigenvalues. But this depends on both the signature and the channel, which we can not control. To get a simple MMSE receiver, where conjugate gradient recursions converge in a small number of steps, we need to design $\boldsymbol{\Phi}$ to make the number of distinct $1 + |d_i|^2 t_i, \forall i$ smaller. Suppose the purpose is to get $M$ distinct eigenvalues. We need to divide the eigenvalues $1 + |d_i|^2 t_i$ into $M$ groups and make the eigenvalues in each group equal. Without loss of generality, we do not consider the eigenvalues with $t_i = 0$. Suppose all $t_i \neq 0$. From the optimization problem above, we know that $|d_i|^2 \leq \mathcal{P}_0, \forall i$ and the larger the $|d_i|^2$, the smaller the minimum mean square error. If for $i > j$, $|d_i|^2 t_i = |d_j|^2 t_j$, since $t_i \geq t_j \neq 0$, $|d_i|^2 \leq |d_j|^2$. So, in the same group where the eigenvalues are equal, suppose $\hat{t}_i$ is the smallest $t_i$ in this group, then the corresponding $|\hat{d}_i|^2 = \mathcal{P}_0$. The eigenvalues are equal to $1 + \mathcal{P}_0 \hat{t}_i$ for this whole group. We order all $1 + |d_i|^2 t_i$ according to the decreasing order of $t_i$. To make the minimum mean square error as small as possible, we have to put the consecutive elements into one group. If we put the last $K - (M - 1)$ eigenvalues into one group, we get a solution similar to the reduced rank MMSE solution, but better because we do not set the eigenvalues equal to zero.

Suppose we divide the $K$ eigenvalues into $M$ groups where the $p$th group contains $k_p$ elements, and $\sum_{p=1}^{M} k_p = K$. Suppose $\hat{t}_p$ is the smallest $t_i$ in the $p$th group. So, $\hat{t}_p = t_{\sum_{i=1}^{p} k_i}$. Set all the eigenvalues in the $p$th group equal to $1 + \mathcal{P}_0 \hat{t}_p$. The optimal precoder that minimizes the mean square error has eigenvalues $1 + \mathcal{P}_0 t_i, \forall i$. Then the method looks like quantizing the $K$ eigenvalues into $M$ sets of approximately equal eigenvalues. The loss of minimum mean square error is

$$\sum_{p=1}^{M} \sum_{i=1+\sum_{j=1}^{p-1} k_j}^{\sum_{j=1}^{p} k_j} \left( \frac{1}{1+\mathcal{P}_0 \hat{t}_p} - \frac{1}{1+\mathcal{P}_0 t_i} \right). \tag{95}$$

The problem is how to pick the $M$ eigenvalues to make the MMSE loss smallest. The optimal solution can be got by exhaustive search for all the possibilities. But the computation will take $\frac{K(K-1)\cdot(K-M+1)}{1\cdot M}$ steps and is time consuming. This is not what we want. Next, we'll give a simple algorithm to get a suboptimal solution.

Predivide the $K$ eigenvalues into $M$ groups by guess. Define $\text{ev}_i = \frac{1}{1+\mathcal{P}_0 t_i}$. From (95), we notice that the MMSE loss depends on the difference between adjacent $\text{ev}_i - \text{ev}_{i-1}$ and the number of eigenvalues in the same group. So, during the original guess, we try not to put the eigenvalues with comparatively big differences $\text{ev}_i - \text{ev}_{i-1}$ into one group. And we try not to put too many eigenvalues into one group. Or, in this quantization procedure, we try to put those eigenvalues with small differences into one group.

After the pre-division, we optimize the group division by searching between the adjacent groups. In this way, we get a suboptimal solution with computational complexity not more than $2K$.

Till now, we have the optimal or suboptimal precoder which minimizes the mean square error when a Wiener filter is used on the receiver side. But how is this precoder applied to each mobile user? Suppose the base station can estimate the amplitude matrix $\boldsymbol{A}$ for each user and get the optimal precoder. The signature vector $[s_1(t), s_2(t), \cdots, s_K(t)]$ can be updated as $[s_1(t), s_2(t), \cdots, s_K(t)]\boldsymbol{G}$, and the updated signatures are sent back to each mobile user from the base station. The resulting transmit signals, after matched filtering on the receiver side are $\boldsymbol{P}\boldsymbol{T}^{1/2}\sqrt{\mathcal{P}_0}\boldsymbol{W}^H$, where $\boldsymbol{P}$ is a unitary matrix from the decomposition $\boldsymbol{C}_H \boldsymbol{U} \boldsymbol{\Lambda}^{\dagger 1/2} = \boldsymbol{P}\boldsymbol{T}^{1/2}\boldsymbol{V}$.

## List of References

[1] S. Verdu, *Multiuser Detection.* Cambridge University Press, 1998.