UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Performance of Non-Gaussian Distribution Based Communication and Compressed Sensing Systems**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Hwan Joon Kwon

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor Sanjoy Dasgupta
Professor William S. Hodgkiss
Professor Young-Han Kim
Professor Laurence B. Milstein

2013

The dissertation of Hwan Joon Kwon is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2013

DEDICATION

To my family.

# EPIGRAPH

*Chance favors the prepared mind.*

—Louis Pasteur

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It has been a great privilege to spend several years in the Department of Electrical and Computer Engineering at University of California, San Diego, to pursue my Ph.D. study. Certainly, completing Ph.D. degree is one of the most challenging and the most memorable activities in my life. Fortunately, I have met many people with whom I have shared the best and worst moments of my Ph.D. journey and they have made this dissertation a reality. I owe my gratitude to all those people and they will always remain dear to me.

My first debt of gratitude must go to my advisor, Professor Bhaskar D. Rao. He has patiently guided me through the doctorial program so that I have been able to proceed. He has always given me great freedom to pursue independent work, and at the same time supportive and productive advice to recover when my steps faltered. His unflagging encouragement helped me overcome many crisis situations and complete this dissertation. It has been a great pleasure working under his supervision.

I am also deeply indebted to my co-advisor, Professor Young-Han Kim for long discussions, insightful comments, and constructive criticisms that helped me sort out the technical details of my work. He has enforced strict validations for each research result and has heightened the completeness of the work.

I would also like to express my heartfelt appreciation to other members on my Ph.D. committee, Professor Laurence B. Milstein, Professor William S. Hodgkiss, and Professor Sanjoy Dasgupta. I have been fortunate to have these wonderful teachers and researchers in my committee, and I enjoyed your colorful and insightful lectures.

Members of Digital Signal Processing Lab. also deserve my sincerest thanks for their friendship and assistance. I could obtain valuable insights and many ideas from many senior lab-mate, Yogananda Isukapalli, Matthew Pugh, Seong-Ho Hur, and Liwen Yu. Special thanks to my small brother, Yichao Huang for fruitful and productive discussions and tremendous help. I am also thankful to my lab fellows in the research group, Sagnik Ghosh, Yuzhe Jin, Nandan Das, Sheu-Sheu Tan, Alireza Masnadi-Shirazi, Zhilin Zhang, Furkan Kavasoglu, PhuongBang Nguyen,

Xiaofei Chen and Oleg Tanchuk for many constructive discussions in our group meeting.

I also thankful to Yu Xiang and Lele Wang for useful discussion especially on information theory. I am lucky to know many friends during my stay at UCSD and it is my pleasure to meet Suk-Ryool Kang and Suk-Hun Cho. I am especially thankful to Zuchul Lee and Jihun Lee for being incredible friends and having frank conversations on our life and future.

Most importantly, I would like to express my heart-felt gratitude to my family. None of my work would have been possible without your belief, encouragement, patience, and sacrifice. In particular, I would like to express my heart-felt gratitude to my beautiful wife, Su-Kyoung Baek and my lovely and smart son, Seewoo Kwon who mean more to me than anything in the world.

Finally, I appreciate the financial support from NSF that funded parts of the research discussed in this dissertation.

This dissertation is a collection of papers, currently being prepared for submission for publication. The text of Chapter 2, in part, is a reprint of the paper, H. Kwon, B. Rao, and Y. Kim, "Uniform Power Allocation with Thresholding for Rayleigh Fading and QAM Inputs", in preparation for submission. The text of Chapter 3, in part, is a reprint of the paper, H. Kwon, B. Rao, and Y. Kim, "Limits on Support Recovery of Sparse signals: Measurement Matrix and Block-Sparsity", in preparation for submission. The dissertation author is the primary researcher and author, and the co-authors listed above contributed to or supervised the research which forms the basis of this dissertation.

VITA

| | |
|---|---|
| 1998 | Bachelor of Science<br>Electronic Communication Engineering<br>Hanyang University, Seoul, Korea |
| 2000 | Master of Science<br>Electronic Communication Engineering<br>Hanyang University, Seoul, Korea |
| 2000-2008 | Research and Standards Engineer<br>Samsung Electronics Co., Ltd., Suwon, Korea |
| 2009-2010 | Teaching Assistant<br>Department of Electrical and Computer Engineering<br>University of California, San Diego, La Jolla, CA |
| 2011-2013 | Research Assistant<br>Department of Electrical and Computer Engineering<br>University of California, San Diego, La Jolla, CA |
| 2013 | Doctor of Philosophy<br>Electrical Engineering (Communication Theory and Systems)<br>University of California, San Diego, La Jolla, CA |

PUBLICATIONS

**Journal Publications**

H. Kwon, Y. Kim, and B. Rao, "Limits on Support Recovery of Sparse signals: Measurement Matrix and Block-Sparsity", in preparation for submission.

H. Kwon, Y. Kim, and B. Rao, "Uniform Power Allocation with Thresholding for Rayleigh Fading and QAM Inputs", in preparation for submission.

H. Kwon, Y. Kim, J. Han, D. Kim, H. Lee, and Y. Kim, "Performance Evaluation of High Speed Packet Enhancement on cdma2000 1xEV-DV", *IEEE Communications Magazine*, vol. 43, no. 4, Apr. 2005.

**Conference Proceedings**

H. Kwon, Y. Kim, and B. Rao, "Uniform Power Allocation with Thresholding over Rayleigh Slow Fading Channels with QAM Inputs", *IEEE Vehicular Technology Conference (VTC)*, Las Vegas, NV, Sep. 2013.

H. Kwon and B. Rao, "On the Benefits of the Block-Sparsity Structure in Sparse Signal", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

H. Kwon and B. Rao, "Uniform Bit and Power Allocation with Subcarrier Selection for Coded OFDM Systems", *IEEE Vehicular Technology Conference (VTC)*, Budapest, Hungary, May 2011.

Z. Pi, H. Kwon, J. Han, D. Kim, and D. Kim, "Statistics of downlink interference in multi-cell OFDMA systems", *Wireless World Research Forum 20th meeting*, Apr. 2008.

H. Kwon, J. Yu, Y. Lim, D. Kim, H. LEE, "An efficient resource allocation and multiplexing scheme in OFDMA systems", *Wireless World Research Forum 20th meeting*, Nov. 2007.

Z. Pi, J. Han, D. Kim, H. Kwon, and D. Kim, "An analysis of forward link interference for OFDMA systems", *Samsung Tech Conference*, Nov. 2006.

H. Kwon, Y. Kim, J. Han, and D. Kim, "Impact of High-Speed Packet Transmissions on Voice User in the Reverse Link of CDMA Systems", *IEEE Vehicular Technology Conference (VTC)*, Sep. 2004.

Y. Kim, H. Kwon, J. Han, and D. Kim, "Performance of HARQ and the Effect of Imperfect Power Control in CDMA Reverse Link", *IEEE Global Communications Conferences (Globecom)*, Nov. 2004.

Y. Kim, J. Han, H. Kwon, and D. Kim, "Effect of Imperfect Power Control on the Performance of HARQ in cdma2000 Reverse Link", *IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2004.

H. Kwon, Y. Kim, J. Han, and D. Kim, "An Efficient Radio Resource Management Technique for the Reverse Link in cdma2000 1xEV-DV", *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2003.

H. Kwon, "Mini-slot Assignment Algorithm and Back-off Window Setting for DOCSIS in HFC CATV Network", *Korean Communication Conference*, Oct. 1999.

ABSTRACT OF THE DISSERTATION

**Performance of Non-Gaussian Distribution Based Communication and Compressed Sensing Systems**

by

Hwan Joon Kwon

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California, San Diego, 2013

Professor Bhaskar D. Rao, Chair

Gaussian distribution is often assumed for the signals in the analysis of and in the design of communication systems and signal processing systems, although Gaussian signals can never be realized in practice. Indeed, Gaussian distribution has proven optimal in many problems of communication and signal processing, e.g., the channel input with a Gaussian distribution achieves the channel capacity of a communication channel. Moreover, many problems of communication and signal processing are mathematically tractable when the Gaussian signal distribution is assumed. This dissertation is concerned with the performance of the systems (or algorithms) in which the actual signal distribution is not Gaussian. In particular,

we study the performance loss of an optimal system with non-Gaussian signals in comparison with the system performance with the optimal Gaussian signals. In addition, when the actual signal distribution is non-Gaussian, we investigate the performance of the practical algorithm that has been derived under the assumption that the signal distribution is Gaussian. Two well-known problems in communication and signal processing are investigated.

First, we study a communication problem, in particular, the power allocation problem that minimizes the outage probability over a slow Rayleigh fading channel or maximizes the mutual information over a fast Rayleigh fading channel, where the channel input is equiprobable QAM signal constellations. The mercury/water-filling (MWF) power allocation is optimal for this problem, while the water-filling (WF) power allocation is optimal if the channel input is Gaussian rather than QAM signals. We show that WF performs close to MWF as long as the constellation size is appropriately chosen, more specifically, the MWF performance itself is not limited by having too small a constellation size. In addition, we study a simple practical power allocation policy, uniform power allocation with thresholding (UPAT) that assigns nonzero constant power only to a subset of the fading blocks. The UPAT can significantly alleviate the feedback overhead and the complexity compared to MWF and WF. We show that the optimal UPAT, namely, the UPAT with the optimal threshold, performs close to MWF as long as the constellation size is large enough.

Next, we study a signal processing problem, in particular, the asymptotic performance limits of reliably recovering the support of block-sparse signals (including scalar-sparse signals as a special class) through an arbitrarily distributed random measurement matrix (including Gaussian) in a noisy setting. Sharp sufficient and necessary conditions for asymptotically reliable support recovery are derived in terms of the signal dimension, the number of nonzero blocks, the block size, the number of measurements, the distribution of the random measurement matrix, and signal-to-noise ratio (SNR) of each nonzero block. The results reveal the effect of the distribution of the random measurement matrix on the number of measurements required for asymptotically reliable support recovery. They also

unveil how much we can potentially reduce the number of measurements required for asymptotically reliable support recovery, when a signal is block-sparse and its structure is known, by making use of the block-sparsity structure compared to treating the signal as being scalar-sparse.

# Chapter 1

# Introduction

Consider a simple communication channel $Y(i) = X(i) + Z(i)$, referred to as additive white Gaussian noise (AWGN) channel, where $i = 1, 2, ...$ is the time index, $Y(i) \in \mathbb{R}$ is the received signal at time $i$, $X(i) \in \mathbb{R}$ is the transmitted signal at time $i$, and $Z(i) \in \mathbb{R}$ is the AWGN at time $i$. It is assumed that $Z(i)$ is independently and identically distributed (i.i.d.) according to Gaussian (normal) distribution $\mathcal{N}(0, \sigma_z^2)$. One wishes to send information over the AWGN channel under the average transmit power constraint

$$\frac{1}{N}\sum_{i=1}^{N} X(i)^2 \leq \sigma_x^2 \tag{1.1}$$

where $N$ is the number of channel uses. The goal is to successfully send as much information as possible. The maximum rate of information that can be reliably transmitted over a given communication channel is referred to as the channel capacity [91]. Then, the capacity of the AWGN channel is given by

$$C = \frac{1}{2}\log_2(1 + \frac{\sigma_x^2}{\sigma_z^2}). \tag{1.2}$$

The capacity of the AWGN channel $C$ is achieved when $X(i)$ is i.i.d. according to Gaussian distribution $\mathcal{N}(0, \sigma_x^2)$.

In addition to the simple AWGN channel, for many other communication channels such as parallel Gaussian channels, multiple access channel (MAC), broadcast channel, and multiple input and multiple output (MIMO) channel, if

the noise is AWGN, the channel capacity is achieved when the channel input is Gaussian-distributed [91].

Now, let us consider a signal processing problem. Suppose the signal of interest is $\mathbf{X} \in \mathbb{R}^m$, and $\mathbf{X}$ is said to be sparse when the signal dimension $m$ is large and only a few elements of $\mathbf{X}$ are nonzero whereas the rest of the elements are zero. One wishes to estimate $\mathbf{X}$ via the linear measurements $\mathbf{Y} = A\mathbf{X} + \mathbf{Z}$, where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix and $\mathbf{Z} \in \mathbb{R}^n$ is the measurement noise. The goal is to estimate $\mathbf{X}$ from as few measurements as possible. This problem has received much attention in many disciplines motivated by a broad array of applications such as compressed sensing [32, 33], biomagnetic inverse problems [34, 35], and image processing [36, 37] (see Chapter 3.1 for more details).

In many applications, finding the positions of the nonzero elements in $\mathbf{X}$, known as *support recovery*, is important. Such applications include EEG/MEG of medical imaging [62, 63] and spectrum sensing in cognitive radio systems [75] (see Chapter 3.1 for more details). It was shown in [81] that the number of measurements required for asymptotically reliable support recovery of the sparse signal $\mathbf{X}$ can be minimized when the elements of the measurement matrix $A$ are Gaussian-distributed.

In addition to the problems introduced above, the Gaussian signal distribution provides the optimal performance in many other problems of communication and signal processing under the AWGN assumption. Moreover, the Gaussian distribution is analytically convenient, since many problems of communication and signal processing have closed-form expressions and are mathematically tractable when the signal is assumed to be Gaussian-distributed. In contrast, a non-Gaussian distribution is mathematically more complicated and is often not mathematically tractable even though the non-Gaussian distribution is a simple distribution, e.g., Bernoulli symmetric distribution. For instance, consider the AWGN channel $Y(i) = X(i) + Z(i)$. Since $Z(i)$ is Gaussian, $Y(i)$ is also Gaussian if $X(i)$ is Gaussian, which makes the problem easier. However, if $X(i)$ is Bernoulli symmetric, e.g., $P(X(i) = +1) = 0.5$ and $P(X(i) = -1) = 0.5$, where $P(B)$ indicates the probability of event $B$. Then, the distribution of $Y(i)$ is the mix-

ture of Bernoulli symmetric and Gaussian, which is not simple for mathematical manipulation.

Due to its optimality and analytical convenience discussed above, the Gaussian distribution is often assumed for the signals in the design of and in the analysis of the systems of communication and signal processing. However, Gaussian-distributed signals are not employed for practical reasons, e.g., the Gaussian distribution requires infinite granularity as well as infinite peak power. Rather, practical signals are usually drawn from discrete constellations, e.g., $\pm 1$, which may significantly depart from the Gaussian idealization. Thus, it is important to understand the effect of non-Gaussian distribution on the system performance.

In this dissertation, we are particularly interested in the following questions.

- How much is the performance loss from a non-Gaussian signal distribution compared to the optimal Gaussian distribution, especially for the signal distributions commonly used in practice?
- What are the conditions under which a non-Gaussian distribution performs close to the Gaussian distribution?
- When the actual signal distribution is non-Gaussian, how much is the performance loss from the algorithm that is optimal in case the signal distribution is Gaussian, compared to the solution optimal for the given distribution?

To answer these questions, we study two well-known problems in communication and signal processing with AWGN under the assumption of non-Gaussian distribution: (1) power allocation over fading channels with QAM inputs and (2) limits on support recovery of sparse signals.

## 1.1 Contributions of the Dissertation

In the sequel, we provides a brief introduction and summarize the main contributions for each of the research topics.

### 1.1.1 Power Allocation over Fading Channels with QAM Inputs

The performance of a communication system over a fading channel can be substantially improved by adapting transmit power according to the channel gains [2–11, 91]. We study the problem of power allocation over Rayleigh block-fading channels [2] with QAM inputs. The performance is measured by the outage probability for nonergodic fading channels and by the ergodic mutual information (MI) for ergodic fading channels. We investigate three power allocation schemes: (1) waterfilling [91] (WF) that is optimal when the channel inputs are Gaussian-distributed, (2) mercury/water-filling (MWF) that is optimal in our setting [4], and (3) uniform power allocation with Thresholding (UPAT) that is a simple and practical solution with reduced feedback requirement. It should be noted that it is mathematically challenging to derive the exact performance results, since the performance metrics (outage probability and ergodic MI) involve the MI of an AWGN channel with QAM input, which does not have a closed-form expression. In our work, we rely on asymptotic analysis, engineering intuition, and comprehensive numerical work. The main results of the paper are summarized as follows.

(1) **Near-optimality of waterfilling and UPAT**: Through mathematical asymptotic analysis and numerical simulations, we show that WF and the UPAT with the optimal threshold, namely, the optimal UPAT perform close to MWF as long as the constellation size is chosen appropriately not to limit the performance.

(2) **Constellation size selection rule**: Taking into account the performance and complexity tradeoff induced by the constellation size, we study the constellation size selection problem. The goal is keeping the constellation size as small as possible while not limiting the performance. We propose the following rule: *minimize the constellation size while achieving the maximum diversity.*

(3) **Gain of the optimal UPAT**: We quantify the gain of the optimal UPAT over the uniform power allocation (UPA) that evenly assigns the total power

across the blocks. We show that the gain in average transmit power increases without bound as the number of independent fading blocks, $B$, increases, but rapidly increases in dB scale only in the low $B$ regime.

(4) **A simple UPAT scheme**: We present a simple method to determine the threshold value for UPAT. Compared to the optimal UPAT, this method significantly reduces the computational complexity with minimal performance loss.

### 1.1.2 Limits on Support Recovery of Sparse Signals: Arbitrarily Distributed Random Measurement Matrices and Block-Sparse Signals

We study the asymptotic performance limits of reliably recovering the support of block-sparse signals (including scalar-sparse signals as a special class) through an arbitrarily distributed random measurement matrix in the Gaussian noise scenario. The main contributions are summarized as follows.

(1) **Information-theoretic analytical framework**: We interpret the problem of recovering of signal support of block-sparse signals as a problem of communication over multiple input single output (MISO) multiple access channel (MAC). Based on this connection, we establish an information-theoretic analytical framework to unearth the performance limits in the support recovery of block-sparse signals. The new perspective also leads to the opportunity of leveraging the rich results and insights available in information theory to help understand the performance limits of block-sparse signal recovery.

(2) **Sufficient and necessary conditions for exact support recovery**: We derive sharp sufficient and necessary conditions for asymptotically reliable support recovery in terms of the signal dimension, the number of nonzero blocks, the block size, the number of measurements, the distribution of the random measurement matrix, and signal-to-noise ratio (SNR) of each nonzero block.

(3) **Suboptimality of a non-Gaussian measurement matrix**: We show that the loss from a non-Gaussian measurement matrix can be significant when both the constellation size and the number of nonzero elements are small and SNR is large.

(4) **Benefit of the block-sparsity**: We identify and discuss three factors of block-sparse signals that can reduce the number of measurement required for reliable support recovery of block-sparse signals, increased SNR, Reduced effective number of nonzero elements, and diversity.

## 1.2 Dissertation Outline

The remainder of the dissertation is organized as follows.

In Chapter 2, we study the problem over fading channels with QAM inputs. Three power allocation schemes, MWF, WF, and UPAT are compared in terms of outage probability and ergodic MI. We derive the conditions under which WF and UPAT perform close to the optimal MWF. We also develop a constellation size selection rule that provides a good compromise between performance and complexity. In addition, we present a simple method to determine the threshold value for UPAT.

In Chapter 3, we study the problem of support recovery of block-sparse signals through an arbitrarily distributed random measurement matrix. The support recovery problem is connected to the problem of communication over MISO MAC. Based on this connection, we develop an information-theoretic analytical framework to derive sufficient and necessary conditions for exact support recovery. Then, we discuss the effect of non-Gaussian distribution on the measurement matrix and the benefit of the block-sparsity.

Finally, Chapter 4 concludes the dissertation.

# Chapter 2

# Power Allocation over Fading Channels with QAM Inputs

## 2.1 Introduction

The performance of a communication system over a fading channel can be substantially improved by adapting transmit power according to the channel gains [2–11, 91]. The waterfilling (WF) power allocation [91] in conjunction with a Gaussian input distribution minimizes the outage probability over a nonergodic block-fading channel (also referred to as slow fading channel) [2] as well as maximizes the ergodic mutual information (MI) over an ergodic fading channel (also referred to as fast fading channel) [3]. In practice, however, the channel inputs must be drawn from discrete constellations such as phase shift keying (PSK), pulse amplitude modulation (PAM), and quadrature amplitude modulation (QAM). For these non-Gaussian inputs, the *mercury/water-filling* (MWF) power allocation is optimal for outage probability and ergodic MI [4].

The amount of overhead to enable transmit power adaptation is often significant in practice [5], [6], especially in a frequency-division duplex (FDD) system in which the receiver has to feed back channel gain information to the transmitter. Once the power levels of the fading blocks are determined, they should be signaled to the receiver so that the received signals can be properly decoded [96], [16].

In addition, the optimal MWF power allocation involves the inverse minimum mean-square error (MMSE) functions [4], an exact implementation of which can be excessively complex in practice [7].

To overcome these difficulties, a simple suboptimal power allocation policy, *uniform power allocation with thresholding* (UPAT), that assigns nonzero constant power only to a subset of the fading blocks has received much attention [5,6,8–11], due to the remarkably relaxed overhead requirements [5], [6] as well as the simple transceiver structure [11]. Since the power levels of the selected blocks are uniform, complete information of the channel gains and the transmit power levels does not have to be exchanged [5].

Much effort has been devoted to developing simple methods to determine the threshold value (or equivalently, the subset of the blocks to which nonzero constant power is assigned) for UPAT and to analyzing the resulting performance. Dardari [5] proposed a simple UPAT scheme, referred to as the ordered subcarrier selection algorithm (OSSA) that always selects half of the blocks with higher channel gains, and analyzed its performance in terms of uncoded bit-error rates (BERs) and packet-error rates (PERs) with Reed–Solomon coding. Kwon and Rao [6] proposed a UPAT scheme that maximizes an equivalent signal-to-noise ratio (SNR) and evaluated the corresponding PER performance of turbo codes. The results in [5] and [6] show that a UPAT with an appropriate threshold can significantly improve the performance compared to no power adaptation. Then, a fundamental question in this context is: *How far from optimal is the UPAT policy?* This question has been studied in [8–11] under the assumption of Gaussian inputs over Rayleigh fast fading channels, where several bounds on the ergodic MI loss from various UPAT schemes compared to the optimal power allocation (waterfilling in this setup) have been numerically derived. However, these results do not naturally extend to non-Gaussian practical constellations. More importantly, the insights from the results of Gaussian inputs into design of practical systems with non-Gaussian inputs are limited, since the optimal power allocation for Gaussian inputs is quite different from that for non-Gaussian inputs and the results for Gaussian inputs shed no light on the constellation size which is one of the most

crucial parameters in practical systems.

In this paper, we study three power allocation schemes, WF, MWF, and UPAT for Rayleigh fading and equiprobable QAM inputs. We are particularly interested in the UPAT policy. We focus primarily on nonergodic block-fading channels and study the outage probability performance. We also consider ergodic fading channels and study the ergodic MI performance in a parallel manner. Note that it is mathematically challenging to derive the exact performance results, since the performance metrics (outage probability and ergodic MI) involve the MI of an AWGN channel with QAM input, which does not have a closed-form expression. In our work, we rely on asymptotic analysis, engineering intuition, and comprehensive numerical work. To give a perspective on the effect of the constellation size on the performance, we recall that the performance of any power allocation scheme improves with the growing constellation size at the expense of additional complexity (see Section 2.4 and Section 2.5 for details). The main results of the paper are summarized as follows.

1) **Near-optimality of waterfilling and UPAT**: Through mathematical asymptotic analysis and numerical simulations, we show that WF and the optimal UPAT performs close to MWF if the constellation size $M$ is large enough that $\log M \gg R$, where $R$ is the fixed target transmission rate. This condition $\log M \gg R$ turns out to define a natural system operating point. As we show through numerical results, if $\log M \approx R$, WF, MWF, and the optimal UPAT perform poorly due to having too small $M$ and their performance can be significantly improved by using a larger $M$. From these results, we can conclude that for a given target transmission rate, WF and the optimal UPAT performs close to MWF as long as the constellation size is chosen appropriately not to limit the performance.

2) **Constellation size selection rule**: Taking into account the performance and complexity tradeoff induced by the constellation size, we study the constellation size selection problem. The goal is keeping the constellation size as small as possible while not limiting the performance. We propose the following rule: *minimize the constellation size while achieving the maximum*

*diversity*. Numerical results show that (i) the proposed rule meets the above goal and (ii) with the proposed rule, the optimal UPAT performs close to MWF.

3) **Gain of the optimal UPAT**: We quantify the gain of the optimal UPAT over the uniform power allocation (UPA) that evenly assigns the total power across the blocks. The focus is on the low $R$ regime since the gain fades away as $R$ increases (see Section 2.6). We show that the gain in average transmit power increases without bound as the number of independent fading blocks, $B$, increases, but rapidly increases in dB scale only in the low $B$ regime.

4) **A simple UPAT scheme**: We present a simple method to determine the threshold value for UPAT. Compared to the optimal UPAT, this method significantly reduces the computational complexity with minimal performance loss.

5) **Extension to ergodic fading**: We extend our work to ergodic fading channels and study the ergodic MI performance. We show through numerical simulations that the same conclusion holds, that is, the optimal UPAT performs close to MWF as long as the constellation size is chosen appropriately not to limit the performance.

The rest of the paper is organized as follows. Section II presents the system model and the performance metric. Section III formulates the outage probability with MWF and the optimal UPAT. Section IV discusses the outage performance results. Section V studies the constellation size selection problem. Section VI quantifies the gain of the optimal UPAT over the UPA. Section VII presents a simple UPAT scheme. Section VIII considers ergodic fading channels. Finally, Section IX concludes the paper with a few remarks.

Throughout, $\mathbb{R}^m$ and $\mathbb{C}^m$ denote the $m$-dimensional real and complex Euclidean spaces, respectively. The circular symmetric complex normal distribution is denoted by $\mathcal{N}_{\mathbb{C}}$. For a vector $\mathbf{a} = (a_1, ..., a_n)$, $\langle \mathbf{a} \rangle \triangleq \frac{1}{n} \sum_{i=1}^{n} a_i$. Component-wise inequalities are denoted by $\preceq$ and $\succeq$. The probability of an event $A$ is denoted by

$\mathsf{P}(A)$. Finally, $\log(\cdot)$ and $\ln(\cdot)$ denote the logarithm to base two and the natural logarithm, respectively.

## 2.2 System Model and Performance Metric

### 2.2.1 System Model

Consider transmission over a nonergodic block-fading channel [7], [13], consisting of $B$ blocks of $L$ channel uses, where block $i = 1, 2, ..., B$ undergoes a random channel gain $H_i$ that is constant during the block and is independently and identically distributed (i.i.d.) across the blocks. Assume that the channel inputs to the blocks are independently and uniformly distributed over the standard $M$-QAM [17] constellation set $\mathcal{S}_M$, where $M \in \{2^{2j} : j = 1, 2, ...\}$ and $\frac{1}{M}\sum_{s\in\mathcal{S}_M} |s|^2 = 1$. Suppose that $\{|H_i|\}_{i=1}^B$ is known to the transmitter so that transmit power for each block can be adapted to the channel strength, subject to an average power constraint $P$. To describe a power allocation scheme, it is convenient to define $\gamma_i \triangleq P|H_i|^2$ which indicates the SNR in block $i$ with the uniform power allocation (UPA). Let $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^B$. Then, a power allocation scheme is described as $\boldsymbol{p}(\boldsymbol{\gamma}; M) = \{p_i(\boldsymbol{\gamma}; M)\}_{i=1}^B$, where $p_i(\boldsymbol{\gamma}; M) \geq 0$ indicates the normalized transmit power of block $i$, i.e., $\langle \boldsymbol{p}(\gamma; M)\rangle \leq 1$. Note that $\boldsymbol{p}(\boldsymbol{\gamma}; M)$ depends on $\{|H_i|^2\}_{i=1}^B$, $P$, and $B$ through $\boldsymbol{\gamma}$. The channel output vector $\mathbf{Y}_i \in \mathbb{C}^L$ in block $i$, is given by

$$\mathbf{Y}_i = H_i\sqrt{p_i(\boldsymbol{\gamma}; M)P}\ \mathbf{S}_i + \mathbf{Z}_i, \quad i = 1, 2, ..., B, \tag{2.1}$$

where $\mathbf{S}_i \in \mathcal{S}_M^L$ is the $M$-QAM channel input vector and $\mathbf{Z}_i \sim \mathcal{N}_{\mathbb{C}}(0, \mathrm{I})$ is the Gaussian channel noise vector. Note that $p_i(\boldsymbol{\gamma}; M)\gamma_i$ corresponds to the instantaneous SNR in block $i$. Throughout, we assume the Rayleigh fading $H_i \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and therefore the probability density function (pdf) of $\gamma_i$ is given by $f_{\gamma_i}(\xi) = \frac{1}{P}e^{-\frac{1}{P}\xi}, \xi \geq 0$. We assume that $\{H_i\}_{i=1}^B$, $M$, $P$, and $\boldsymbol{p}(\boldsymbol{\gamma}; M)$ are known to the receiver so that the received signals can be properly decoded. We also assume that $L$ is sufficiently large so that the input–output MI of the channel is meaningful. For the sake of convenience, blocks with high and low $\gamma_i$ are referred to as strong

and weak blocks, respectively.

### 2.2.2 Performance Metric: Outage Probability

When $B$ is finite, the system in (2.1) models a slow fading channel. In particular, when $B$ is small, e.g., $B = 4$ or $B = 8$, the system in (2.1) can model a transmission over a small number of independent fading blocks, for instance, a multiple-input and multiple-output (MIMO) transmission using singular value decomposition (SVD) [14] with rank $B$ over $L$ symbols in time where the channel is independent across the $B$ decomposed single-input and single-output (SISO) channels but is constant across the $L$ symbols in time. The system in (2.1) with a small $B$ can also model an orthogonal frequency division multiplexing (OFDM) transmission with a large number of subcarriers whose channel gains are highly correlated. We assume $L$ is arbitrarily large which makes the mutual information operationally significant. For finite $B$, we consider the outage probability as the performance metric that can provide a lower bound on the codeword error probability of arbitrary coding schemes.

Recall that the MI of the AWGN channel $Y = \sqrt{\rho}S + Z$ under the uniform distribution of the input $S$ over $\mathcal{S}_M$ is given [18] by

$$I_M^{\mathrm{AW}}(\rho) = \log M - \frac{1}{M} \sum_{s \in \mathcal{S}_M} \mathsf{E}\left[ \log \left( \sum_{s' \in \mathcal{S}_M} e^{-|\sqrt{\rho}(s-s')+Z|^2+|Z|^2} \right) \right] \tag{2.2}$$

where the expectation $\mathsf{E}[\cdot]$ is with respect to the random noise $Z$. Plots of (2.2) for different values of $M$ are shown in Fig. 2.1, where the MI of the AWGN channel with the Gaussian input (i.e., the AWGN channel capacity $\log(1 + \rho)$) is also presented for comparison.

Then, the instantaneous MI is defined [2] by

$$I_M(\boldsymbol{\gamma}, \boldsymbol{p}(\boldsymbol{\gamma}; M)) \triangleq \frac{1}{B} \sum_{i=1}^{B} I_M^{\mathrm{AW}}(p_i(\boldsymbol{\gamma}; M)\gamma_i). \tag{2.3}$$

For a fixed target transmission rate $R$, the *outage probability* is defined [2]

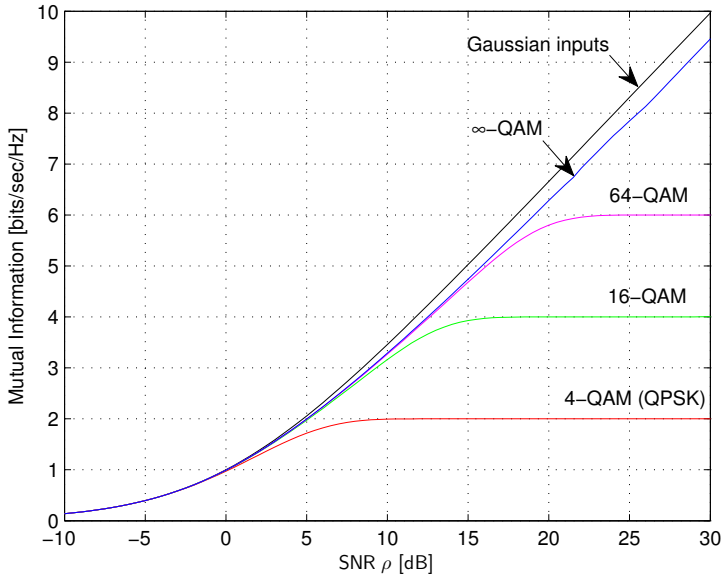


**Figure 2.1**: Mutual information of the AWGN channel with equiprobable $M$-QAM input constellations.

by

$$P_M^{\text{out}}(\boldsymbol{p}(\boldsymbol{\gamma}; M), P, R) \triangleq \mathsf{P}\{I_M(\boldsymbol{\gamma}, \boldsymbol{p}(\boldsymbol{\gamma}; M)) < R\} \tag{2.4}$$

where the probability is with respective to the random $\boldsymbol{\gamma}$.

## 2.3 Power Allocation Schemes

### 2.3.1 Optimal Power Allocation: Mercury/water-filling

The outage probability minimization problem is formulated as

$$\begin{aligned} \text{minimize} \quad & P_M^{\text{out}}(\boldsymbol{p}(\boldsymbol{\gamma}; M), P, R) \\ \text{subject to} \quad & \langle \boldsymbol{p}(\boldsymbol{\gamma}; M) \rangle \leq 1 \\ & \boldsymbol{p}(\boldsymbol{\gamma}; M) \succeq 0. \end{aligned} \tag{2.5}$$

We refer to the solution to this problem as the *optimal power allocation*, denoted by $\boldsymbol{p}^{\text{opt}}(\boldsymbol{\gamma}; M)$. Since the outage probability is minimized when the instantaneous

MI (2.3) is maximized for each realization of $\boldsymbol{\gamma}$, we have [2]

$$\boldsymbol{p}^{\mathrm{opt}}(\boldsymbol{\gamma}; M) = \arg \max_{\substack{\langle \boldsymbol{p}(\gamma;M)\rangle \leq 1 \\ \boldsymbol{p}(\gamma;M) \succeq 0}} I_M(\boldsymbol{\gamma}, \boldsymbol{p}(\boldsymbol{\gamma}; M)). \tag{2.6}$$

Let $\mathsf{MMSE}_M(\rho)$ denote the MMSE incurred in the estimation of an equiprobable $M$-QAM symbol over the AWGN channel with SNR $\rho$. By the Lagrangian duality, the Karush–Kuhn–Tucker (KKT) conditions [19], and the relationship $\frac{d}{d\rho} I_M^{\mathrm{AW}}(\rho) = \frac{1}{\ln 2}\mathsf{MMSE}_{\mathsf{M}}(\rho)$ between MI and MMSE [20], the solution to (2.6) is given [4] by

$$p_i^{\mathrm{opt}}(\boldsymbol{\gamma}; M) = \begin{cases} \frac{1}{\gamma_i}\mathsf{MMSE}_M^{-1}\left(\frac{\lambda_{\mathrm{MWF}}}{\gamma_i}\right), & \gamma_i \geq \lambda_{\mathrm{MWF}} \\ 0, & \gamma_i < \lambda_{\mathrm{MWF}} \end{cases} \tag{2.7}$$

where the SNR threshold $\lambda_{\mathrm{MWF}}$ is chosen so that the average power constraint is satisfied with equality. The solution (2.7) is often referred to as *mercury/water-filling* [4]. Substituting (2.7) into (2.4) yields the outage probability with the optimal power allocation

$$\begin{aligned} &P_M^{\mathrm{out}}(\boldsymbol{p}^{\mathrm{opt}}(\boldsymbol{\gamma}; M), P, R) = \\ &\mathsf{P}\left\{\frac{1}{B}\sum_{i:\gamma_i \geq \lambda_{\mathrm{MWF}}} I_M^{\mathrm{AW}}\left(\mathsf{MMSE}_M^{-1}\left(\frac{\lambda_{\mathrm{MWF}}}{\gamma_i}\right)\right) < R\right\}. \end{aligned} \tag{2.8}$$

### 2.3.2 Waterfilling Power Allocation

The power allocation with the waterfilling policy $\boldsymbol{p}^{\mathrm{wf}}(\boldsymbol{\gamma})$ can formulated as

$$\boldsymbol{p}^{\mathrm{wf}}(\boldsymbol{\gamma}) = \arg \max_{\boldsymbol{p}(\gamma)} \sum_{i=1}^{B} \log_2(1 + p_i(\boldsymbol{\gamma})\gamma_i) \tag{2.9}$$

$$\text{subject to } \langle \boldsymbol{p}(\gamma)\rangle \leq 1 \text{ and } \boldsymbol{p}(\gamma) \succeq 0. \tag{2.10}$$

Note that the waterfilling solution is independent of $M$.

The solution to the above problem is given by [91]

$$p_i^{\mathrm{wf}}(\boldsymbol{\gamma}) = \begin{cases} \frac{1}{\lambda_{\mathrm{wf}}} - \frac{1}{\gamma_i}, & \gamma_i \geq \lambda_{\mathrm{wf}} \\ 0, & \gamma_i < \lambda_{\mathrm{wf}} \end{cases} \tag{2.11}$$

where SNR threshold $\lambda_{\mathrm{wf}}$ is chosen such that the average power constraint is met with equality. As seen in (2.11), the waterfilling assigns more power to stronger blocks regardless of $M$ if $\gamma_i$ is larger than or equal to the threshold.

### 2.3.3 Uniform Power Allocation with Thresholding

In a UPAT scheme, nonzero constant power is assigned to a set of selected blocks while zeropower is assigned to the other blocks. Let $\gamma_1^o \le \gamma_2^o \le \cdots \le \gamma_B^o$ be the ordered $\boldsymbol{\gamma}$ sequence. Then, a UPAT scheme can be described as

$$p_i^{\mathrm{UPAT}}(\boldsymbol{\gamma}; M) = \begin{cases} \frac{B}{B-N_{\mathrm{UPAT}}}, & \gamma_i \ge \lambda_{\mathrm{UPAT}} \\ 0, & \gamma_i < \lambda_{\mathrm{UPAT}} \end{cases} \tag{2.12}$$

where $\lambda_{\mathrm{UPAT}}$ is the SNR threshold, $0 \le N_{\mathrm{UPAT}} < B$ is the number of zero-power blocks, and $\lambda_{\mathrm{UPAT}} = \gamma^o_{N_{\mathrm{UPAT}}+1}$. Therefore, a UPAT scheme is completely defined by how to determine the value of $N_{\mathrm{UPAT}}$ (or the value of $\lambda_{\mathrm{UPAT}}$) depending on $\boldsymbol{\gamma}$ and $M$. The optimal value of $N_{\mathrm{UPAT}}$ that maximizes the instantaneous MI is

$$N^*_{\mathrm{UPAT}} = \arg \max_{0 \le n < B} \sum_{i=n+1}^{B} I_M^{\mathrm{AW}}\left(\frac{B}{B-n}\gamma_i^o\right). \tag{2.13}$$

The corresponding SNR threshold is denoted by $\lambda^*_{\mathrm{UPAT}} (= \gamma^o_{N^*_{\mathrm{UPAT}}+1})$. The UPAT with $N^*_{\mathrm{UPAT}}$ is referred to as the *optimal UPAT*, denoted by $\boldsymbol{p}^{\mathrm{UPAT}^*}(\boldsymbol{\gamma}; M)$. Substituting (2.12) and (2.13) into (2.4) yields the outage probability with the optimal UPAT

$$\begin{aligned} &P_M^{\mathrm{out}}(\boldsymbol{p}^{\mathrm{UPAT}^*}, P, R) \\ &= \mathsf{P}\left\{\frac{1}{B}\sum_{i=N^*_{\mathrm{UPAT}}+1}^{B} I_M^{\mathrm{AW}}\left(\frac{B}{B-N^*_{\mathrm{UPAT}}}\gamma_i^o\right) < R\right\}. \end{aligned} \tag{2.14}$$

### 2.3.4 Examples

We now discuss a set of examples that show power allocations of MWF and the optimal UPAT, corresponding instantaneous MIs, and the power loss $\Delta P$ that indicates the additional $P$ required for the optimal UPAT to achieve the same instantaneous MI as MWF. The examples are with respect to two specific realizations of $\boldsymbol{\gamma}$ but provide some insight into the outage probability results that come from a random $\boldsymbol{\gamma}$, discussed in subsequent sections.

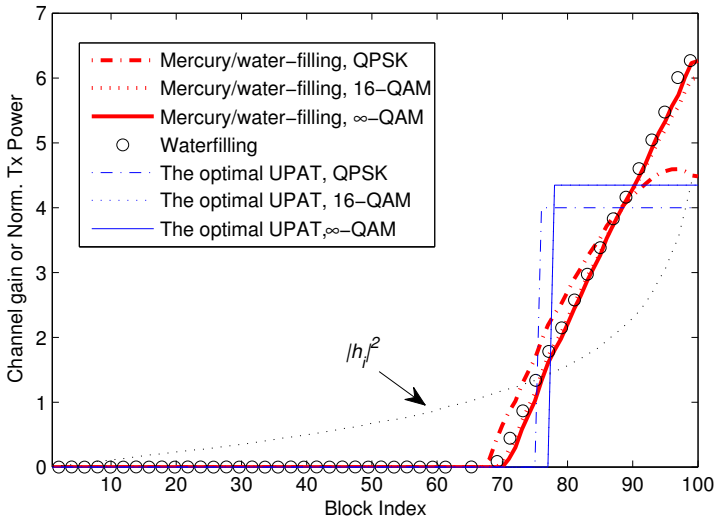


**Figure 2.2**: Power allocation results of mercury/water-filling, waterfilling, and the optimal UPAT, when $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1$ ($\frac{1}{B}\sum_{i=1}^{B}\gamma_i = -10$ dB).

We consider a specific channel gain vector $\{h_i\}_{i=1}^{B}$ with $B = 100$, where its elements were i.i.d. drawn according to $\mathcal{N}_{\mathbb{C}}(0,1)$, normalized and ordered[1] such that $\frac{1}{B}\sum_{i=1}^{B}|h_i|^2 = 1$ and $|h_1| \leq \cdots \leq |h_B|$ (see Fig. 2.2 and Fig. 2.3). Two different values of $P$, $P_1 = -10$ dB and $P_2 = 13$ dB are examined. Three constellation sizes, $M = 4, 16, \infty$ are considered. Let $\boldsymbol{\gamma}_1 = \{P_1|h_i|^2\}_{i=1}^{B}$ and $\boldsymbol{\gamma}_2 = \{P_2|h_i|^2\}_{i=1}^{B}$. Then, $\langle\boldsymbol{\gamma}_1\rangle = -10$ dB and $\langle\boldsymbol{\gamma}_2\rangle = 13$ dB. Since MWF and the optimal UPAT depend on $(\boldsymbol{\gamma}, M)$, there are 6 different cases, i.e., $M = 4, 16, \infty$ for each $\boldsymbol{\gamma}$.

The power allocations of MWF and the optimal UPAT are shown in Fig. 2.2 and Fig. 2.3 for $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_2$, respectively. For each $\boldsymbol{\gamma}$, we also present the power allocations of waterfilling that does not depend on $M$. Table 2.1 shows the instantaneous MI of MWF, $I_M(\boldsymbol{p}^{\mathrm{opt}})$, the instantaneous MI of the optimal UPAT, $I_M(\boldsymbol{p}^{\mathrm{UPAT}^*})$, the loss $\Delta I_M$ of the optimal UPAT in instantaneous MI compared to MWF, and the power loss $\Delta P$.

The results in Fig. 2.2, Fig. 2.3, and Table 2.1 can be explained by the following properties of $I_M^{\mathrm{AW}}(\rho)$ (see Fig. 2.1). Note that the sum of $I_M^{\mathrm{AW}}(p_i\gamma_i)$ is the objective function in the relevant optimization problems (2.6) and (2.13).

[1]The ordering is merely intended to clearly show the dependency of power allocation on the channel gains

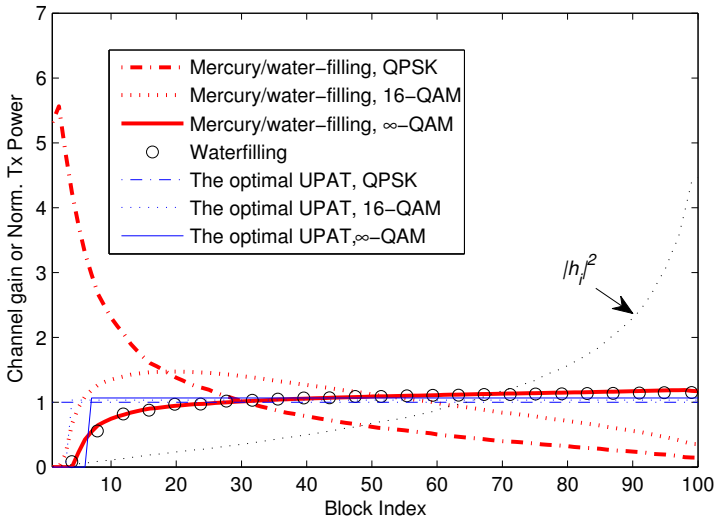


**Figure 2.3**: Power allocation results of mercury/water-filling, waterfilling, and the optimal UPAT, when $\boldsymbol{\gamma} = \boldsymbol{\gamma}_2$ ($\frac{1}{B}\sum_{i=1}^{B}\gamma_i$ =13 dB).

**Table 2.1**: Instantaneous MI and additional $P$ required by $\boldsymbol{p}^{\mathrm{UPAT}*}$

| $(\boldsymbol{\gamma}, M)$ | $I_M(\boldsymbol{p}^{\mathrm{opt}})$ | $I_M(\boldsymbol{p}^{\mathrm{UPAT}*})$ | $\Delta I_M$ | $\Delta P$ |
|---|---|---|---|---|
| $(\boldsymbol{\gamma}_1, 4)$ | 0.2314 | 0.2288 | 1.2 % | 0.07 dB |
| $(\boldsymbol{\gamma}_1, 16)$ | 0.2385 | 0.2341 | 1.8 % | 0.12 dB |
| $(\boldsymbol{\gamma}_1, \infty)$ | 0.2394 | 0.2349 | 1.9 % | 0.12 dB |
| $(\boldsymbol{\gamma}_2, 4)$ | 1.9499 | 1.8630 | 4.5 % | 4.4 dB |
| $(\boldsymbol{\gamma}_2, 16)$ | 3.1917 | 3.1199 | 2.3 % | 0.5 dB |
| $(\boldsymbol{\gamma}_2, \infty)$ | 3.5950 | 3.5871 | 0.2 % | 0.03 dB |

(i) Given $M$, if $\rho$ is so low that $I_M^{\mathrm{AW}}(\rho) \ll \log M$, $I_M^{\mathrm{AW}}(\rho)$ is not much different from $\log(1+\rho)$. In particular, $I_M^{\mathrm{AW}}(\rho) \approx \log(1+\rho)$ in the low $\rho$ regime [4].

(ii) Given $M$, if $\rho$ is so high that $I_M^{\mathrm{AW}}(\rho) \approx \log M$, $I_M^{\mathrm{AW}}(\rho)$ slowly approaches its maximum $\log M$ as $\rho$ increases. Therefore, in this regime, $I_M^{\mathrm{AW}}(\rho)$ does not vary much as $\rho$ changes.

Due to property (i), if $(\boldsymbol{\gamma}, M)$ is such that $I_M^{\mathrm{AW}}(\gamma_i) \ll \log M$ for most of the blocks, e.g., $(\boldsymbol{\gamma} = \boldsymbol{\gamma}_1, M = 16)$, $(\boldsymbol{\gamma} = \boldsymbol{\gamma}_1, M = \infty)$, and $(\boldsymbol{\gamma} = \boldsymbol{\gamma}_2, M = \infty)$, MWF is not much different from waterfilling, i.e., cutting off weakest blocks and then assigning more power to stronger blocks. This is because the solution to the

problem (2.6) is waterfilling if $I_M^{\mathrm{AW}}(p_i\gamma_i)$ is replaced by $\log(1+p_i\gamma_i)$. In this case, as shown in [8–11] for Gaussian inputs, $\Delta I_M$ and $\Delta P$ are not significant.

Due to property (ii), on the contrary, if $(\boldsymbol{\gamma}, M)$ is such that $I_M^{\mathrm{AW}}(\gamma_i) \approx \log M$ for most of the blocks, e.g., $(\boldsymbol{\gamma} = \boldsymbol{\gamma}_2, M = 4)$, MWF tends to assign power inversely proportionally to $\gamma_i$ [4]. Intuitively, this is because with relatively small power, stronger blocks can still provide per-block MI $I_M^{\mathrm{AW}}(p_i\gamma_i)$ close to $\log M$. By assigning more power to weaker blocks, MWF increases per-block MI of weaker blocks, while minimally decreasing per-block MI of stronger blocks. This is how MWF maximizes the average of per-block MI, i.e., the instantaneous MI. In contrast, the optimal UPAT just activates most of the blocks due to the uniform power constraint and therefore the optimal UPAT tends to be the UPA. Interestingly, although the power allocations of MWF and the optimal UPAT are quite different from each other, $\Delta I_M$ is not so significant since the instantaneous MI of the optimal UPAT is close to $\log M$ and there is not much room to improve. However, even though $\Delta I_M$ is small, $\Delta P$ can be significant. This is because the optimal UPAT is similar to the UPA and therefore if $P$ increases, a large portion of the increased power is assigned to the blocks that could provide per-block MI close to $\log M$ without the increased power, which makes increasing $P$ very inefficient in terms of increasing the instantaneous MI. Therefore, a large $\Delta P$ is needed to increase the instantaneous MI by a small amount.

## 2.4 Near-optimality of UPAT

In this section, we discuss the outage performance loss from the optimal UPAT compared to MWF. We first show that the optimal UPAT performs close to MWF as long as the constellation size is sufficiently large. We then show that the constellation size should be sufficiently large for both MWF and the optimal UPAT in order not to suffer from a huge performance degradation.

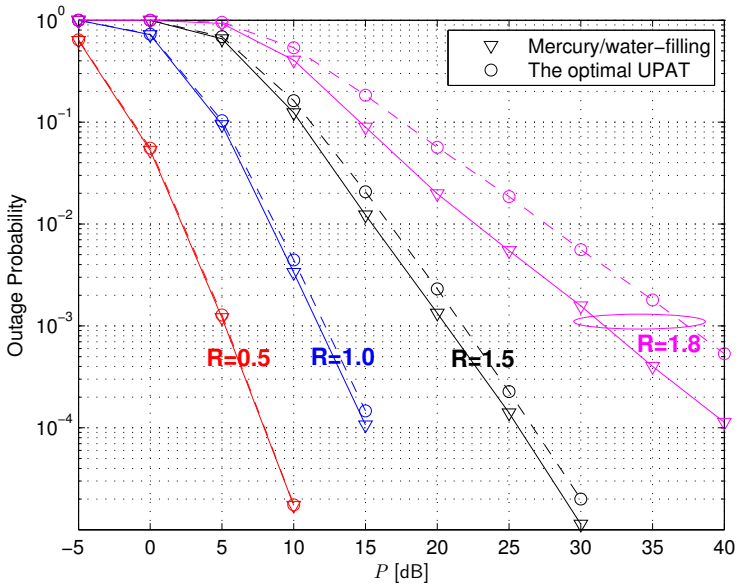


**Figure 2.4**: Outage performance of the optimal power allocation (mercury/water-filling) and the optimal UPAT for different target transmission rates ($R = 0.5$, 1.0, 1.5, 1.8) when the number of blocks is 4 ($B = 4$) and $M = 4$ (QPSK).

### 2.4.1 The Optimal UPAT is Near-Optimal If $\log M \gg R$

Fig. 2.4 shows the outage probability of MWF and the optimal UPAT when $M = 4$ (QPSK), $R = 0.5, 1.0, 1.5, 1.8$, and $B = 4$. We observe that when $R = 0.5$, the optimal UPAT performs near MWF. But, as $R$ increases, the performance loss from the optimal UPAT increases. In particular, the loss is significant when $R \approx \log M$, e.g., at $10^{-3}$ outage rate, the loss is about 1 dB and 6 dB for $R = 1.5$ and $R = 1.8$, respectively.

The performance loss of the optimal UPAT for various values of $M$, $R$, and $B$ is summarized in Fig. 2.5. Each curve in the figure indicates the additional $P$ required for the optimal UPAT to achieve the same outage probability $10^{-3}$ as MWF. The results clearly show that the loss due to the optimal UPAT is marginal when $R \ll \log M$, increases with $R$, and becomes significant when $R \approx \log M$.

Intuitively, the above results can be explained as follows. By the definition of the outage probability (2.4), when $R \ll \log M$, the outage events occur when $I_M^{\mathrm{AW}}(\gamma_i) \ll \log M$ for most of the blocks. As discussed in the previous section,

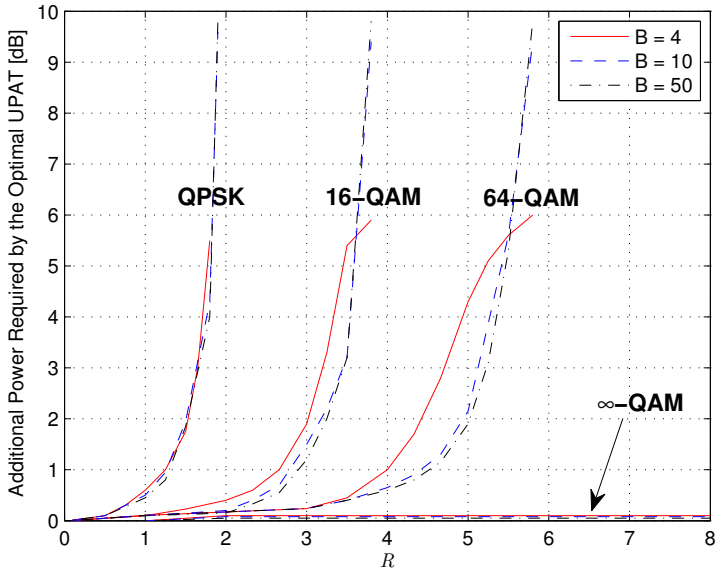


**Figure 2.5**: UPAT suboptimality for various $(M, R, B)$: each curve indicates the additional $P$ required for the optimal UPAT at outage rate $10^{-3}$ to achieve the same outage probability as mercury/water-filling.

for $\boldsymbol{\gamma}$ realizations such that $I_M^{\mathrm{AW}}(\gamma_i) \ll \log M$ for most of the blocks, MWF is not much different from waterfilling and the additional power required for the optimal UPAT to achieve the same instantaneous MI as MWF is not significant.

When $R \approx \log M$, in contrast, the outage events occur unless $I_M^{\mathrm{AW}}(\gamma_i) \approx \log M$ for most of the blocks. For $\boldsymbol{\gamma}$ realizations such that $I_M^{\mathrm{AW}}(\gamma_i) \approx \log M$ for most of the blocks, the optimal power allocation tends to be inversely proportional to $\gamma_i$, while the optimal UPAT tends to be the UPA. In this case, the additional power required for the optimal UPAT is significant, as discussed in the previous section.

An asymptotic behavior of these observations is proved by the following proposition.

**Proposition 1**

$$\lim_{R\to 0} \frac{P_M^{out}(\boldsymbol{p}^{UPAT*}(\boldsymbol{\gamma}; M), P, R)}{P_M^{out}(\boldsymbol{p}^{opt}(\boldsymbol{\gamma}; M), P, R)} = 1.$$

*Proof:* See Appendix 2.10.1. ■

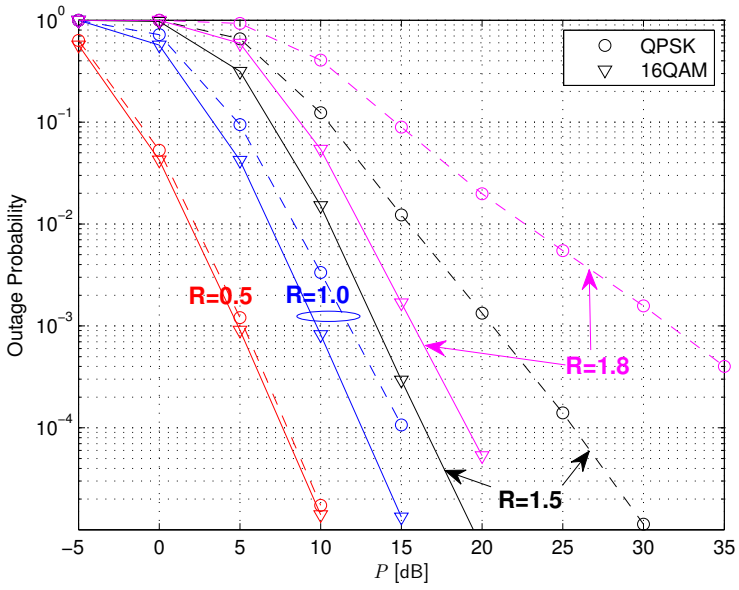


**Figure 2.6**: Outage performance gain of 16-QAM over QPSK with the optimal power allocation (mercury/water-filling) when $B = 4$.

### 2.4.2 A System Should Operate in the Regime in Which $\log M \gg R$

The following property of the outage probability is important for subsequent discussion.

**Lemma 1** *$P_M^{out}(\boldsymbol{p}(\boldsymbol{\gamma}; M), P, R)$ is a decreasing function of the constellation size $M$.*

*Proof:* Since $I_M^{\mathrm{AW}}(\rho)$ is an increasing function of $M$ [21], [22], $I_M(\boldsymbol{\gamma}, \boldsymbol{p}(\boldsymbol{\gamma}; M))$ is an increasing function of $M$. The proof follows by (2.4). ■

Lemma 1 implies that the outage probability with $\infty$-QAM inputs not only provides an analytical insight into the asymptotic behavior for large $M$ but also serves as a lower bound for the outage probability of any finite $M$.

Now, we show through numerical results that the condition $\log M \gg R$ is a necessary prerequisite for a system to perform well. Fig. 2.6 shows the performance gain of 16-QAM over QPSK (4-QAM) with MWF for $B = 4$ fading blocks. We observe that when $R = 0.5$, QPSK and 16-QAM exhibit a similar performance.

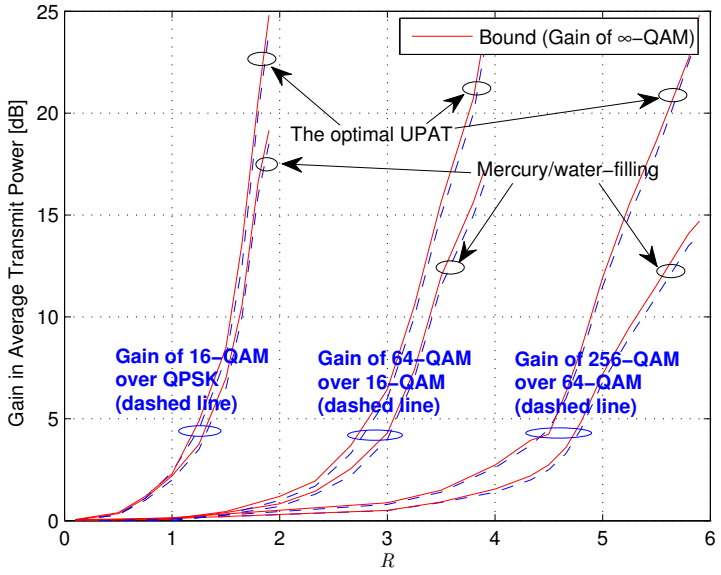


**Figure 2.7**: Gain from using a larger constellation size in average power $P$ at $10^{-3}$ outage probability when $B = 4$.

But, as $R$ increases, the gain of 16-QAM increases. In particular, when $R$ is close to the maximum achievable rate of QPSK (i.e., $\log M = 2$), the gain is significant, e.g., 16 dB gain at $10^{-3}$ outage probability when $R = 1.8$.

Fig. 2.7 summarizes the gain from using a larger $M$ for various values of $M$ and $R$ when $B = 4$. The dashed line indicates the gain in $P$ from using one-step larger $M$ at $10^{-3}$ outage probability. The solid line indicates the gain of $\infty$-QAM, which bounds the gain from using any larger (finite) $M$. We note the following observations. If $\log M \gg R$, the gain from using a larger $M$ is marginal for both MWF and the optimal UPAT. In contrast, if $\log M \approx R$, the performance of them can be significantly improved by even one-step larger $M$. In other words, both MWF and the optimal UPAT perform poorly in this regime, due to having too small $M$.

Intuitively, these results can be explained as follows. If $R \ll \log M$, the outage events occur when $I_M^{\mathrm{AW}}(\gamma_i) \ll \log M$ for most of the blocks. For $\boldsymbol{\gamma}$ realizations such that $I_M^{\mathrm{AW}}(\gamma_i) \ll \log M$ for most of the blocks, increasing $M$ does not significantly increase the instantaneous MI, since $I_M^{\mathrm{AW}}(\gamma_i) \approx I_{M'}^{\mathrm{AW}}(\gamma_i)$, $M < M'$,

for most of the blocks (see Fig. 2.1). Therefore, increasing $M$ does not significantly decrease the outage probability. In contrast, if $R \approx \log M$, the outage events occur unless $I_M^{\mathrm{AW}}(\gamma_i) \approx \log M$ for most of the blocks. For $\boldsymbol{\gamma}$ realizations such that $I_M^{\mathrm{AW}}(\gamma_i) \approx \log M$ for most of the blocks, increasing $M$ can significantly increase the instantaneous MI, since $I_M^{\mathrm{AW}}(\gamma_i) \ll I_{M'}^{\mathrm{AW}}(\gamma_i)$, $M < M'$, for most of the blocks (see Fig. 2.1). Therefore, increasing $M$ can substantially decrease the outage probability.

### 2.4.3 Summary

If $\log M \gg R$ the optimal UPAT performs close to MWF; otherwise, the suboptimality of the optimal UPAT can be significant. However, if $\log M \approx R$, both MWF and the optimal UPAT perform poorly and therefore a system should avoid operating in this regime. In conclusion, for a given target transmission rate $R$, the optimal UPAT is near-optimal as long as the constellation size $M$ is chosen appropriately not to limit the performance.

## 2.5 Constellation Size Selection

Although a larger constellation size provides a better outage performance (Lemma 1), it is more complex to implement in practice [23]. In this section, we study the constellation size selection problem under the assumption that the values of $B$ and $R$ are given. The goal is keeping the constellation size as small as possible while not limiting the performance (i.e., while minimizing the performance loss compared to $\infty$-QAM).

An interesting observation from Fig. 2.6 is that the outage performance gain of 16-QAM over QPSK is mainly linked to the slope of the outage probability curve. When $R = 0.5$, the slope with 16-QAM seems essentially identical to that with QPSK and the difference in required $P$ to achieve the same outage probability is not significant. However, when $R = 1.5$ and $R = 1.8$, the slopes with 16-QAM are much steeper than those with QPSK and the difference in slope makes a significant difference in required $P$ at a low outage probability.

Based on this observation, we propose the following constellation size selection rule: *minimize the constellation size while maximizing the slope*. The asymptotic slope in log-log scale is referred to as the *outage diversity* [24], which is defined by

$$d \triangleq \lim_{P\to\infty} -\frac{\log P_M^{\text{out}}(\boldsymbol{p}(\boldsymbol{\gamma}; M), P, R)}{\log P}. \tag{2.15}$$

For both the UPA [24] and MWF [25], the outage diversity of the Rayleigh block-fading channel in (2.1) is given by

$$d = 1 + \left\lfloor B\left(1 - \frac{R}{\log M}\right)\right\rfloor. \tag{2.16}$$

We can readily show that the outage probability for the optimal UPAT is also given by (2.16) using the fact that $P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}) \geq P_M^{\text{out}}(\boldsymbol{p}^{\text{UPAT}^*}) \geq P_M^{\text{out}}(\boldsymbol{p}^{\text{opt}})$ where $\boldsymbol{p}^{\text{UPA}} = (1, ..., 1)$.

Therefore, for both MWF and the optimal UPAT, the constellation size selection rule that minimizes the constellation size while maximizing the outage diversity can be formulated from (2.16) as

$$\min\big\{M : \log M \geq RB,\ M = 2^{2i},\ i = 1, 2, ...\big\}. \tag{2.17}$$

Note that for any given $B$ and $R$, the constellation size according to this rule provides the maximum diversity $B$. For example, when $B = 4$, the rule results in: QPSK for $0 < R \leq 0.5$, 16-QAM for $0.5 < R \leq 1.0$, 64-QAM for $1.0 < R \leq 1.5$, and so forth.

Fig. 2.8 shows the outage probability with the proposed constellation selection rule, where $B = 4$, $R = 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4$, and the $M$ value for each $(B, R)$ follows (2.17). The solid lines indicate the performance bounds obtained with MWF and $\infty$-QAM inputs. First, we observe that the slopes of all the curves are the same in the low outage probability regime, which confirms that the proposed rule leads to the maximum outage diversity for both MWF and the optimal UPAT. Second, we observe that the gaps in $P$ between the performance bounds and the results with the proposed rule are marginal for all the cases, which indicates that the performance gain from using an $M$ larger than the proposed

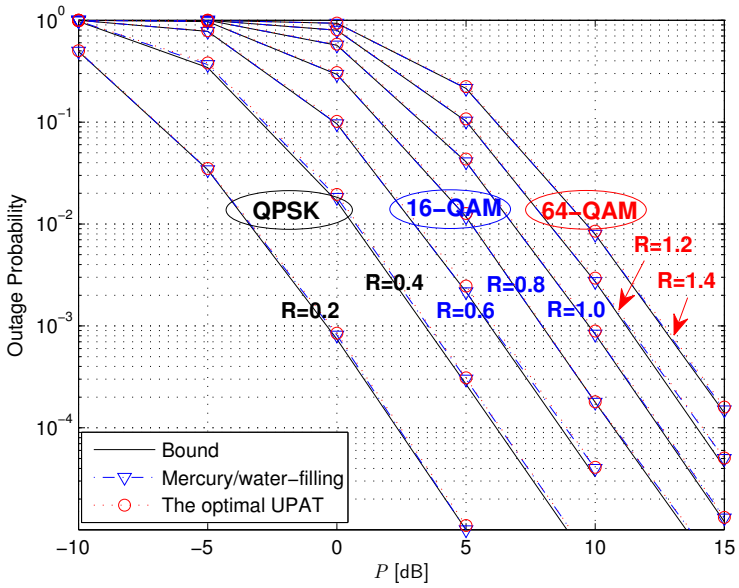


**Figure 2.8**: Outage probability when $B = 4$ and the constellation size $M$ is chosen according to (2.17) depending on $(B, R)$. The solid lines denote the performance bounds obtained with mercury/waterfilling and $\infty$-QAM inputs.

rule is minimal. Since the performance loss from having an $M$ smaller than the proposed rule can be significant due to less diversity, especially in case $B$ is small and/or a low outage probability is concerned, we can conclude that the proposed rule meets our goal, that is, to choose as small constellation size as possible while not limiting the performance. Finally, we observe that with the proposed rule, the optimal UPAT performs close to MWF.

## 2.6 Gain of the Optimal UPAT over the UPA

In this section, we analyze the gain of the optimal UPAT over the UPA (i.e., no power adaptation). To support a high $R$, $\gamma_i$ should be large for many blocks. For such $\boldsymbol{\gamma}$ realizations, the optimal UPAT tends to be the UPA as discussed in Section 2.3.4 and therefore the gain from the optimal UPAT fades away as $R$ increases. Thus, we focus on the low $R$ regime. The amount of the gain is evaluated from two different perspectives: (i) gain in outage probability for a given $P$ and (ii) gain

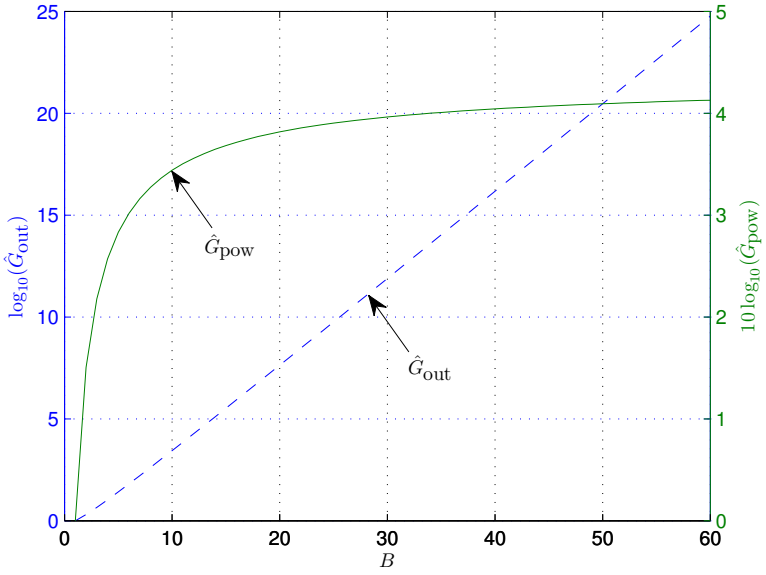


**Figure 2.9**: Approximate gains of the optimal UPAT over the UPA in terms of outage probability ($\hat{G}_{\text{out}}$ in (2.20)) and required $P$ to achieve the same outage probability ($\hat{G}_{\text{pow}}$ in (2.23)) in the low $R$ and high $P$ regime.

in required $P$ to have the same outage probability.

### 2.6.1 Outage Probability Gain

We define the outage probability gain of the optimal UPAT over the UPA as

$$G_{\text{out}} \triangleq \frac{P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}, P, R)}{P_M^{\text{out}}(\boldsymbol{p}^{\text{UPAT}^*}(\boldsymbol{\gamma}; M), P, R)} \tag{2.18}$$

where $\boldsymbol{p}^{\text{UPA}} = (1, ..., 1)$. The next proposition characterizes $G_{\text{out}}$ in the low $R$ regime.

**Proposition 2** *In the low $R$ regime, the outage probability gain of the optimal UPAT over the UPA is approximated by*

$$\tilde{G}_{out} = \frac{\Gamma_l(B, \frac{BR\ln 2}{P})}{(B-1)!(1-e^{-\frac{R\ln 2}{P}})^B} \tag{2.19}$$

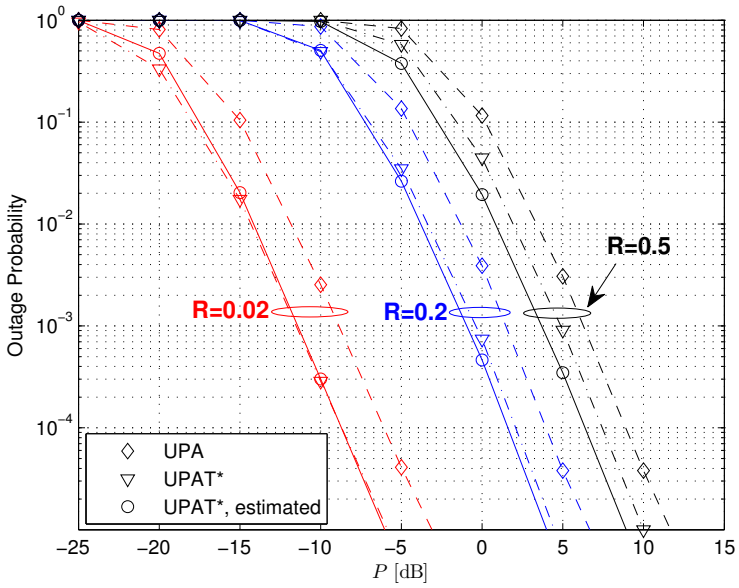


**Figure 2.10**: Outage probability when $M = 16$, $B = 4$, and $R = 0.02, 0.2, 0.5$. 'UPA' and 'UPAT*' indicate the uniform power allocation and the optimal UPAT, respectively. 'UPAT*, estimated' corresponds to $\frac{P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}},P,R)}{\tilde{G}_{\text{out}}}$ where $\tilde{G}_{\text{out}}$ is given by (2.19).

*where* $\Gamma_l(s,x) = \int_0^x t^{s-1}e^{-t}dt$ *is the lower incomplete gamma function [27]. Furthermore, in the high* $P$ *regime,* $\tilde{G}_{out}$ *is approximated by*

$$\hat{G}_{out} = \frac{B^{B-1}}{(B-1)!}. \tag{2.20}$$

*Proof:* See Appendix 2.10.2. ■

Note that gain in the high $P$ regime reflects gain in the low outage probability regime, e.g., $P_{\text{out}} = 10^{-3}$ or lower. Proposition 2 indicates that when $R$ is low, the outage probability gain in the high power regime depends only on $B$ and exponentially increases without bound as $B$ increases, as shown in Fig. 2.9. Therefore, for a given $B$, $G_{\text{out}}$ behaves like a constant in the low $R$ and high $P$ regime, indicating that the outage probability curves for the optimal UPAT and the UPA are parallel to each other, which agrees with the results in the previous section that both the schemes have the same outage diversity.

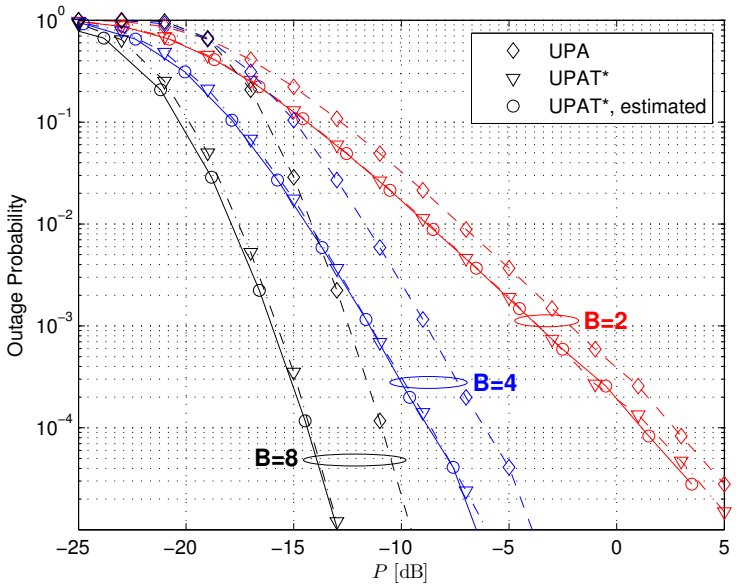


**Figure 2.11**: Outage probability when $M = 16$, $B = 2, 4, 8$, and $R = 0.02$. 'UPA' and 'UPAT*' indicate the uniform power allocation and the optimal UPAT, respectively. 'UPAT*, estimated' corresponds to the left-shifted $P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}, P, R)$ by $\tilde{G}_{\text{pow}}$ in (2.22).

Fig. 2.10 compares the outage probability between the optimal UPAT and the UPA when $M = 16$, $B = 4$, and $R = 0.02, 0.2, 0.5$. The solid lines correspond to $\frac{P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}, P, R)}{\tilde{G}_{\text{out}}}$, showing how accurately $\tilde{G}_{\text{out}}$ in (2.19) estimates the actual outage probability gain of the optimal UPAT. As expected, $\tilde{G}_{\text{out}}$ precisely estimates the actual gain when $R$ is low, while it tends to overestimate the gain as $R$ increases.

### 2.6.2 Average Transmit Power Gain

Next, we analyze the gain of the optimal UPAT over the UPA in terms of the required $P$ to achieve the same outage probability. We first define $P_{\text{UPAT}^*}(P)$ as the average transmit power required for the optimal UPAT to achieve the same outage probability as the UPA with transmit power $P$, i.e., $P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}, P, R) = P_M^{\text{out}}(\boldsymbol{p}^{\text{UPAT}^*}(\boldsymbol{\gamma}; M), P_{\text{UPAT}^*}(P), R)$. Then, we define the power gain of the optimal

UPAT as

$$G_{\text{pow}} \triangleq \frac{P}{P_{\text{UPAT}^*}(P)}. \tag{2.21}$$

The next proposition characterizes $G_{\text{pow}}$ in the low $R$ regime.

**Proposition 3** *In the low R regime, the average transmit power gain of the optimal UPAT over the UPA is approximated by*

$$\tilde{G}_{pow} = -\frac{P}{R\ln 2}\log\left(1-\left(\frac{\Gamma_l(B,\frac{BR\ln 2}{P})}{(B-1)!}\right)^{\frac{1}{B}}\right) \tag{2.22}$$

*Furthermore, in the high P regime, $\tilde{G}_{pow}$ is approximated by*

$$\hat{G}_{pow} = B(B!)^{-\frac{1}{B}}. \tag{2.23}$$

*Proof:* See Appendix 2.10.3. ■

Proposition 3 indicates that when $R$ is low and $P$ is high, the power gain of the optimal UPAT depends only on $B$ and increases without bound as $B$ increases. As shown in Fig. 2.9, the power gain in dB rapidly increases as $B$ increases, only in the low $B$ regime.

Fig. 2.11 compares the outage probability between the optimal UPAT and the UPA when $M = 16$, $B = 2, 4, 8$, and $R = 0.02$. The solid lines correspond to the left-shifted $P_M^{\text{out}}(\boldsymbol{p}^{\text{UPA}}, P, R)$ by $\tilde{G}_{\text{pow}}$, showing how accurately $\tilde{G}_{\text{pow}}$ in (2.22) estimates the actual power gain of the optimal UPAT. We observe that (i) $\tilde{G}_{\text{pow}}$ precisely estimates the actual power gain of the optimal UPAT when $R$ is low and (ii) the power gain of the optimal UPAT increases as $B$ increases, which agrees with the analytical result in (2.23).

## 2.7 A Simple Algorithm for UPAT

The optimal UPAT requires finding the optimal number of zero-power blocks ($N^*_{\text{UPAT}}$) or equivalently finding the optimal SNR threshold value ($\lambda^*_{\text{UPAT}}$)

for a given $(\boldsymbol{\gamma}, M)$. Since the objective function in (2.13) is non-convex with respect to $n$, it is not easy to develop an efficient algorithm for finding the optimal solution. In case $I_M^{\mathrm{AW}}(\rho)$ is implemented by a lookup table, a naive algorithm for solving (2.13) would include $\frac{1}{2}B(B+1)$ series of multiplications, additions, divisions, and interpolations, which can be burdensome if $B$ is large. To tackle this problem, we derive a simple method to determine the value of $N_{\mathrm{UPAT}}$, leading to marginal performance loss compared to the optimal thresholds.

Consider the fact that $I_M^{\mathrm{AW}}(\rho) \approx \frac{1}{\ln 2}\rho$ in the low $\rho$ regime [4]. Let $\eta$ be the maximum $\alpha \in \mathbb{R}$ satisfying that $I_M^{\mathrm{AW}}(\rho) \approx \frac{1}{\ln 2}\rho$ for $0 \le \rho \le \alpha$. Let $\boldsymbol{p}^{\max}(\boldsymbol{\gamma})$ denote the power allocation scheme that assigns the entire power to the block with the highest $\gamma_i$. Then, we can readily show that

$$\boldsymbol{p}^{\max}(\boldsymbol{\gamma}) = \arg \max_{\substack{\langle \boldsymbol{p}(\boldsymbol{\gamma};M)\rangle \le 1 \\ \boldsymbol{p}(\boldsymbol{\gamma};M) \succeq 0}} \frac{1}{B}\sum_{i=1}^{B} \frac{1}{\ln 2} p_i(\boldsymbol{\gamma};M)\gamma_i. \tag{2.24}$$

Therefore, if $\boldsymbol{p}(\boldsymbol{\gamma};M)$ is such that $p_i \neq 0$ and $p_i\gamma_i < \eta$ for some $i$, i.e., if the instantaneous SNRs for some blocks are within the near-linear regime of the MI curve of the AWGN channel, then $\boldsymbol{p}(\boldsymbol{\gamma};M)$ is inefficient in the approximate sense of maximizing the instantaneous MI.

Based on this argument, we propose to set $N_{\mathrm{UPAT}}$ such that the resulting minimum nonzero instantaneous SNR is equal to $\eta$, i.e., $\frac{B}{B-N_{\mathrm{UPAT}}}\gamma^o_{N_{\mathrm{UPAT}}+1} = \eta$. As a result, the proposed method activates as many blocks as possible while guaranteeing that there is no block whose instantaneous SNR is less than $\eta$. The proposed method can be formulated as follows.

(i) If $B\gamma^o_B < \eta$, $N_{\mathrm{UPAT}} = B-1$.

(ii) Otherwise,

$$N_{\mathrm{UPAT}} = \min\big\{k : \frac{B}{B-k}\gamma^o_{k+1} \ge \eta, \quad 0 \le k < B\big\}. \tag{2.25}$$

The case when $B\gamma^o_B < \eta$ in (i) corresponds to the case when all the blocks are so weak that even if the total power is assigned to the strongest block, its instantaneous SNR is lower than $\eta$. In this case, $N_{\mathrm{UPAT}} = B-1$ is near-optimal due to (2.24).

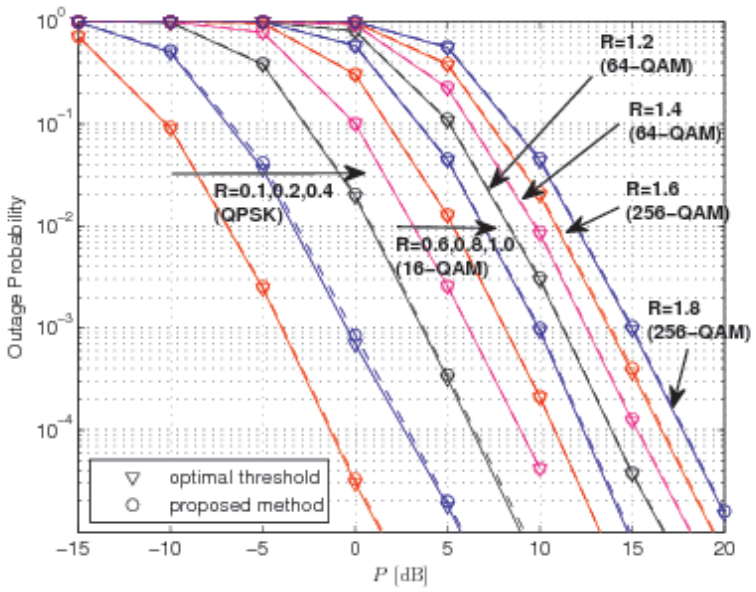


**Figure 2.12**: Comparison of outage probability between the UPAT schemes with the proposed thresholding method and the optimal thresholds when $B = 4$. For each $(B, R)$, the value of $M$ is chosen according to (2.17). In the proposed method, $\eta = -3$ dB for all the cases.

In contrast, if all the blocks are so strong that with the UPA, even the weakest block has the instantaneous SNR equal to or higher than $\eta$, then the UPAT with (2.25) turns into the UPA.

Note that the proposed method can significantly reduce the computational complexity compared to directly solving (2.13) because it does not actually compute the instantaneous MI for each possible value of $N_{\text{UPAT}}$ and the monotonicity of $\frac{B}{B-k}\gamma^o_{k+1}$ with respect to $k$ can be used to develop an efficient algorithm for solving (2.25), e.g., a binary search [26] can be applied.

Fig. 2.12 compares the outage probability between the optimal UPAT and the UPAT with the proposed thresholding method, when $B = 4$. Various values of $R$ from 0.1 to 1.8 are examined. For each $(B, R)$, the constellation size $M$ is chosen according to the proposed $M$ selection rule (2.17). For all the cases, $\eta$ is set to $-3$ dB. We observe that the proposed method performs close to the optimal thresholds for all the cases.

## 2.8 Extension to Ergodic Fading Channels

### 2.8.1 System Model and Performance Metric

The system model with an ergodic fading channel is obtained from (2.1) by letting $B \to \infty$. We assume that the channel gain process $\{H_i\}_{i=1}^{\infty}$ is stationary and ergodic. Then, it suffices [4] to consider power allocation schemes of the form $p_i(\boldsymbol{\gamma}; M) = p(\gamma_i; M, P)$ which depends on the pdf of $\gamma_i$, $f_{\gamma_i}(\xi) = \frac{1}{P}e^{-\frac{1}{P}\xi}, \xi \geq 0$, through $P$.

The ergodic MI is defined [4] by

$$\begin{aligned} I_M^{\infty}(P, p(\gamma; M, P)) &\triangleq \lim_{B\to\infty} \frac{1}{B}\sum_{i=1}^{B} I_M^{\mathrm{AW}}(p(\gamma_i; M, P)\gamma_i) \\ &= \int_0^{\infty} I_M^{\mathrm{AW}}(p(\gamma; M, P)\gamma) f_{\gamma_i}(\gamma) d\gamma. \end{aligned} \tag{2.26}$$

Since $I_M^{\mathrm{AW}}(p(\gamma_i; M, P)\gamma_i)$ is an increasing function of $M$, $I_M^{\infty}(P, p(\gamma; M, P))$ is an increasing function of $M$. Therefore, the ergodic MI with $\infty$-QAM serves an upper bound on the ergodic MI for any finite $M$.

### 2.8.2 Power Allocation Schemes

#### Optimal Power Allocation

The ergodic MI maximization problem is formulated as

$$\begin{aligned} \text{maximize} \quad & I_M^{\infty}(P, p(\gamma; M, P)) \\ \text{subject to} \quad & \int_0^{\infty} p(\gamma; M, P)\frac{1}{P}e^{-\frac{1}{P}\gamma} d\gamma \leq 1 \\ & p(\gamma; M, P) \geq 0. \end{aligned} \tag{2.27}$$

The solution to the above problem, the *optimal power allocation* is given [4] by

$$p^{\mathrm{opt}}(\gamma; M, P) = \begin{cases} \frac{1}{\gamma}\mathsf{MMSE}_M^{-1}(\frac{\lambda_{\mathrm{MWF}}}{\gamma}), & \gamma \geq \lambda_{\mathrm{MWF}} \\ 0, & \gamma < \lambda_{\mathrm{MWF}} \end{cases} \tag{2.28}$$

where the SNR threshold $\lambda_{\mathrm{MWF}}$ is chosen so that the average power constraint is satisfied with equality. Substituting (2.28) into (2.26) yields the ergodic MI with

the optimal power allocation

$$I_M^{\infty}(P, p^{\text{opt}}(\gamma; M, P)) = \int_{\lambda_{\text{MWF}}}^{\infty} I_M^{\text{AW}}\left(\mathsf{MMSE}_M^{-1}\left(\frac{\lambda_{\text{MWF}}}{\gamma}\right)\right)\frac{1}{P}e^{-\frac{1}{P}\gamma}d\gamma. \tag{2.29}$$

**Uniform Power Allocation with Thresholding**

A UPAT scheme is given by

$$p^{\text{UPAT}}(\gamma; M, P) = \begin{cases} p_0, & \gamma \geq \lambda_{\text{UPAT}} \\ 0, & \gamma < \lambda_{\text{UPAT}} \end{cases} \tag{2.30}$$

where $\lambda_{\text{UPAT}}$ is the SNR threshold and a constant power level $p_0 > 0$ depends on $\lambda_{\text{UPAT}}$. The optimal SNR threshold that maximizes the ergodic MI is given by

$$\lambda_{\text{UPAT}}^* = \arg\max_{\lambda \geq 0} \int_{\lambda}^{\infty} I_M^{\text{AW}}(p_0\gamma)\frac{1}{P}e^{-\frac{1}{P}\gamma}d\gamma \tag{2.31}$$

$$\text{subject to} \int_{\lambda}^{\infty} p_0\frac{1}{P}e^{-\frac{1}{P}\gamma}d\gamma \leq 1. \tag{2.32}$$

From the average power constraint (2.32), we have

$$p_0 = e^{\frac{\lambda}{P}}. \tag{2.33}$$

Thus, (2.31) and (2.32) reduce to

$$\lambda_{\text{UPAT}}^* = \arg\max_{\lambda \geq 0} \int_{\lambda}^{\infty} I_M^{\text{AW}}(e^{\frac{\lambda}{P}}\gamma)\frac{1}{P}e^{-\frac{1}{P}\gamma}d\gamma. \tag{2.34}$$

The UPAT with the optimal threshold $\lambda_{\text{UPAT}}^*$ is referred to as the *optimal UPAT* and denoted by $p^{\text{UPAT}*}(\gamma; M, P)$. Substituting (2.33) and (2.34) into (2.26) yields the ergodic MI with the optimal UPAT

$$I_M^{\infty}(P, p^{\text{UPAT}*}(\gamma; M, P)) = \int_{\lambda_{\text{UPAT}}^*}^{\infty} I_M^{\text{AW}}\big(e^{\frac{\lambda_{\text{UPAT}}^*}{P}}\gamma\big)\frac{1}{P}e^{-\frac{1}{P}\gamma}d\gamma. \tag{2.35}$$

### 2.8.3 Ergodic Mutual Information and Constellation Size

Fig. 2.13 shows the ergodic MI for MWF and the optimal UPAT when $M = 4, 16, 64, 256$. We observe that the difference along the vertical axis between

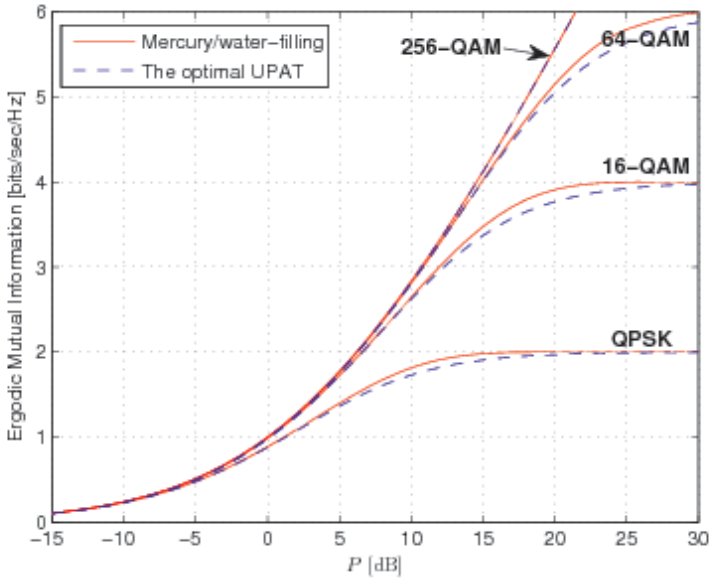


**Figure 2.13**: Ergodic mutual information of the Rayleigh fading channel with the mercury/waterfilling (optimal) and the optimal UPAT.

the power allocation schemes, i.e., the loss in ergodic MI due to the optimal UPAT compared to the optimal power allocation is insignificant for all the cases. However, the difference along the horizontal axis, i.e., the additional $P$ required for the optimal UPAT to achieve the same ergoridc MI as the optimal power allocation is significant when the ergodic MI is close to $\log M$. Note that in order to support the ergodic MI close to $\log M$, the required $P$ has to be so high that per-block MI is close to $\log M$ with high probability. For such a $(M, P)$, as discussed in Section 2.3.4, the optimal power allocation tends to be inversely proportional to $\gamma$, while the optimal UPAT tends to be the UPA. In this case, the additional $P$ required for the optimal UPAT is significant. However, in this regime, the required $P$ for the both MWF and the optimal UPAT could be significantly reduced by a larger $M$. With a larger $M$, the performance loss of the optimal UPAT compared to MWF becomes marginal. Therefore, we reach the same conclusion as that in the nonergodic fading case, that is, *the optimal UPAT performs near-optimal as long as the constellation size $M$ is appropriately chosen not to limit the performance.*

With the goal of keeping the constellation size as small as possible while

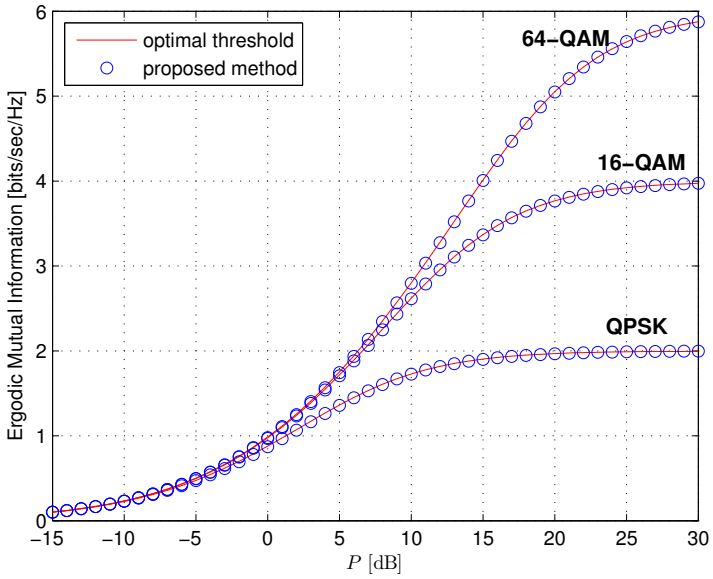


**Figure 2.14**: Comparison of the ergodic mutual information between the UPAT schemes with the optimal thresholds and with the thresholds according to the proposed rule. In the proposed rule, the $\eta$ value is set to $-3$ dB regardless of $M$ and $P$.

not limiting the performance, based on the results in Figure 2.13, a reasonable $M$ selection rule is: QPSK for $0 < I_M^\infty \leq 0.5$, 16QAM for $0.5 < I_M^\infty \leq 1.8$, 64QAM for $1.8 < I_M^\infty \leq 3.8$, and so forth. With this rule, the optimal UPAT performs close to MWF.

### 2.8.4 A Simple UPAT Scheme

We propose a simple thresholding method that alleviates the computational complexity of the optimal SNR threshold given by (2.34). We apply the same idea as in the nonergodic fading case, that is, setting $\lambda_{\text{UPAT}}$ such that the resulting minimum nonzero instantaneous SNR equals $\eta$, i.e., $p_0\lambda_{\text{UPAT}} = \eta$. By (2.33), the proposed SNR threshold is the solution to

$$\lambda_{\text{UPAT}} e^{\frac{\lambda_{\text{UPAT}}}{P}} = \eta. \tag{2.36}$$

The performance of the above method for $M = 4, 16, 64$ is presented in Fig.

2.14, where the $\eta$ value was set to $-3$ dB regardless of $M$. We observe that the proposed method performs close to the optimal thresholds for all the constellation sizes considered and over the entire range of $P$.

## 2.9 Concluding Remarks

Compared to the optimal mercury/water-filling (MWF) power allocation scheme, a uniform power allocation with thresholding (UPAT) scheme can significantly reduce the overhead requirements with a simple transceiver structure. In this paper, we have shown that for Rayleigh fading channels with QAM inputs, the UPAT policy achieves a near-optimal performance, as long as the constellation size and the threshold value are appropriately chosen. We propose a constellation size selection rule and a simple, but yet close to optimal, thresholding method. Our thresholding method is applicable to practical systems with a large number of fading blocks such as orthogonal frequency-division multiplexing (OFDM) with a larger number of subcarriers and has potential applications to singular value decomposition (SVD) multiple-input and multiple-output (MIMO) with a large number of antennas (e.g., massive MIMO [28], [29]), for which determining the number of active subcarriers and the number of spatial streams (sometimes referred to as rank) may have prohibitively high complexity. We leave it as future work to extend our result to fading distributions other than Rayleigh and to multiple-user scheduling scenarios as in orthogonal frequency division multiple access (OFDMA) systems.

The text of this chapter, in part, is a reprint of the paper, H. Kwon, B. Rao, and Y. Kim, "Uniform Power Allocation with Thresholding for Rayleigh Fading and QAM Inputs", in preparation for submission. The dissertation author is the primary researcher and author, and the co-authors contributed to or supervised the research which forms the basis of this chapter.

## 2.10 Appendices

### 2.10.1 Proof of Proposition 1

Let $\boldsymbol{p}^{\max}(\boldsymbol{\gamma})$ denote the power allocation scheme that assigns the entire power to the block with the highest $\gamma_i$. Let $\gamma_{\max} = \max\{\gamma_1, \gamma_2, \ldots, \gamma_B\}$. Then, for any $R > 0$,

$$\mathsf{P}\big(\frac{1}{B} I_M^{\mathrm{AW}}(\gamma_{\max}) < R\big) \tag{2.37}$$

$$= \mathsf{P}\big(I_M(\boldsymbol{\gamma}, \boldsymbol{p}^{\max}(\boldsymbol{\gamma})) < R\big) \tag{2.38}$$

$$\geq \mathsf{P}\big(I_M(\boldsymbol{\gamma}, \boldsymbol{p}^{\mathrm{UPAT}^*}(\boldsymbol{\gamma}; M)) < R\big) \tag{2.39}$$

$$\geq \mathsf{P}\big(I_M(\boldsymbol{\gamma}, \boldsymbol{p}^{\mathrm{opt}}(\boldsymbol{\gamma}; M)) < R\big) \tag{2.40}$$

$$= \mathsf{P}\left(\frac{1}{B}\sum_{i=1}^{B} I_M^{\mathrm{AW}}(p_i^{\mathrm{opt}}(\boldsymbol{\gamma}; M)\gamma_i) < R\right) \tag{2.41}$$

$$\geq \mathsf{P}\left(\frac{1}{B}\sum_{i=1}^{B} \frac{1}{\ln 2} p_i^{\mathrm{opt}}(\boldsymbol{\gamma}; M)\gamma_i < R\right) \tag{2.42}$$

$$\geq \mathsf{P}\left(\frac{1}{B}\sum_{i=1}^{B} \frac{1}{\ln 2} p_i^{\max}(\boldsymbol{\gamma})\gamma_i < R\right) \tag{2.43}$$

$$= \mathsf{P}\left(\frac{1}{B\ln 2}\gamma_{\max} < R\right), \tag{2.44}$$

where (2.39) is due to the definition of $\boldsymbol{p}^{\mathrm{UPAT}^*}(\boldsymbol{\gamma}; M)$, (2.42) is due to $I_M^{\mathrm{AW}}(\rho) \leq (1/\ln 2)\rho$ for all $\rho$ [4], and (2.43) is due to (2.24). Therefore, Proposition 1 can be proved by showing that

$$\lim_{R\to 0} \frac{\mathsf{P}\left(I_M^{\mathrm{AW}}(\gamma_{\max}) < BR\right)}{\mathsf{P}\left(\gamma_{\max} < (\ln 2)BR\right)} = 1. \tag{2.45}$$

Let $f(x)$ denote the pdf of $\gamma_{\max}$, i.e., $f(x) = \frac{d}{dx}\mathsf{P}\left(\gamma_{\max} < x\right) = \frac{d}{dx}\big(1 - e^{-\frac{x}{P}}\big)^B$. Then, $f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x)$ is well-defined for every $n \geq 1$, where $f^{(0)}(x) = f(x)$.

Let $g(\rho) = I_M^{\mathrm{AW}}(\rho)$. Then,

$$\begin{aligned}
\lim_{R\to 0} \frac{\mathsf{P}\Big(I_M^{\mathrm{AW}}(\gamma_{\max}) < BR\Big)}{\mathsf{P}\Big(\gamma_{\max} < (\ln 2)BR\Big)} &= \lim_{x\to 0} \frac{\mathsf{P}\Big(I_M^{\mathrm{AW}}(\gamma_{\max}) < x\Big)}{\mathsf{P}\Big(\gamma_{\max} < (\ln 2)x\Big)} \\
&\overset{(a)}{=} \lim_{x\to 0} \frac{f(g^{-1}(x))\frac{d}{dx}g^{-1}(x)}{f((\ln 2)x)\cdot \ln 2} = \lim_{x\to 0} \frac{f^{(0)}(g^{-1}(x))}{f^{(0)}((\ln 2)x)} \\
&\overset{(b)}{=} \cdots = \lim_{x\to 0} \frac{f^{(k)}(g^{-1}(x))}{f^{(k)}((\ln 2)x)} = \lim_{x\to 0} \frac{f^{(k)}(0)}{f^{(k)}(0)} = 1,
\end{aligned}$$

where $(a)$ follows [4] by $\lim_{\rho\to 0} \frac{\rho}{I_M^{\mathrm{AW}}(\rho)} = \ln 2$, $(b)$ follows by the L'Hôpital's rule, and $k$ is the smallest integer such that $f^{(k)}(0) > 0$ (there must exist one since $f(x)$ is a pdf).

### 2.10.2 Proof of Proposition 2

When $R$ is close to zero, the outage events occur when $I_M^{\mathrm{AW}}(p_i(\boldsymbol{\gamma}; M)\gamma_i)$ is zero or close to zero for all $i$. Let $\boldsymbol{p}^{\mathrm{UPA}} = (1, ..., 1)$ denote the UPA. Then, in the low $R$ regime,

$$\begin{aligned}
P_M^{\mathrm{out}}(\boldsymbol{p}^{\mathrm{UPA}}, P, R) &\overset{(c)}{\approx} \mathsf{P}\left(\frac{1}{B\ln 2}\sum_{i=1}^{B}\gamma_i < R\right) \\
&= \mathsf{P}\left(\sum_{i=1}^{B}|H_i|^2 < \frac{BR\ln 2}{P}\right) = \frac{\Gamma_{\mathrm{l}}(B, \frac{BR\ln 2}{P})}{(B-1)!}
\end{aligned}$$

where $(c)$ is due to the fact that $I_M^{\mathrm{AW}}(\rho) \approx (1/\ln 2)\rho$ in the low $\rho$ regime [4].

For the optimal UPAT, in the low $R$ regime,

$$\begin{aligned}
P_M^{\mathrm{out}}(\boldsymbol{p}^{\mathrm{UPAT}^*}, P, R) &\overset{(d)}{\approx} P_M^{\mathrm{out}}(\boldsymbol{p}^{\max}, P, R) = \mathsf{P}\left(\frac{1}{\ln 2}\gamma_{\max} < R\right) \\
&= \mathsf{P}\left(|H_{\max}|^2 < \frac{R\ln 2}{P}\right) = \left(1 - e^{-\frac{R\ln 2}{P}}\right)^B
\end{aligned}$$

where $(d)$ is due to (2.45), $\gamma_{\max} = \max\{\gamma_1, ..., \gamma_B\}$, and $|H_{\max}|^2 = \frac{\gamma_{\max}}{P}$. Therefore,

$$G_{\mathrm{out}} \approx \frac{\Gamma_{\mathrm{l}}(B, \frac{BR\ln 2}{P})}{(B-1)!(1-e^{-\frac{R\ln 2}{P}})^B} \triangleq \tilde{G}_{\mathrm{out}}.$$

The lower incomplete gamma function $\Gamma_{\rm l}(s,x)$ can be expanded as

$$\Gamma_{\rm l}(s,x) = \frac{x^s}{s} e^{-x} K(1, s+1, x) \tag{2.46}$$

where $K$ is Kummer's confluent hypergeometric function given by

$$\begin{aligned} K(1,s+1,x) = 1 + \frac{x}{s+1} + \frac{x^2}{(s+1)(s+2)} \\ + \frac{x^3}{(s+1)(s+2)(s+3)} + \cdots . \end{aligned} \tag{2.47}$$

Therefore,

$$\begin{aligned} \lim_{P\to\infty} \tilde{G}_{\rm out} &= \lim_{P\to\infty} \frac{\frac{(\frac{BR\ln 2}{P})^B}{B} e^{-(\frac{BR\ln 2}{P})} K(1, B+1, (\frac{BR\ln 2}{P}))}{(B-1)!(1-e^{-\frac{R\ln 2}{P}})^B} \\ &= \frac{B^{B-1}}{(B-1)!}. \end{aligned}$$

### 2.10.3 Proof of Proposition 3

As shown in Appendix 2.10.2, in the low $R$ regime,

$$P_M^{\rm out}(\boldsymbol{p}^{\rm UPA}, P, R) \approx \frac{\Gamma_{\rm l}(B, \frac{BR\ln 2}{P})}{(B-1)!}.$$

and

$$P_M^{\rm out}(\boldsymbol{p}^{\rm UPAT^*}(\boldsymbol{\gamma}; M), P, R) \approx \left(1 - e^{-\frac{R\ln 2}{P}}\right)^B.$$

By the definition of $P_{\rm UPAT^*}$ in Subsection 2.6.2, in the low $R$ regime,

$$\frac{\Gamma_{\rm l}(B, \frac{BR\ln 2}{P})}{(B-1)!} \approx \left(1 - e^{-\frac{R\ln 2}{P_{\rm UPAT^*}}}\right)^B.$$

Therefore, in the low $R$ regime,

$$G_{\rm pow} \approx -\frac{P\ln 2}{R} \log\left(1 - \left(\frac{\Gamma_{\rm l}(B, \frac{BR\ln 2}{P})}{(B-1)!}\right)^{\frac{1}{B}}\right) \triangleq \tilde{G}_{\rm pow}.$$

It follows that

$$\begin{aligned} \lim_{P\to\infty} \tilde{G}_{\rm pow} &= \lim_{P\to\infty} -\frac{P}{R\ln 2} \log\left(1 - \left(\frac{\Gamma_{\rm l}(B, \frac{BR\ln 2}{P})}{(B-1)!}\right)^{\frac{1}{B}}\right) \\ &\overset{(e)}{=} \lim_{P\to\infty} -\frac{P}{R\ln 2} \log\left(1 - \frac{RB\ln 2}{P}(B(B-1)!)^{-\frac{1}{B}}\right) \\ &= B(B!)^{-\frac{1}{B}}. \end{aligned}$$

where $(e)$ follows by (2.46) and (2.47).

# Chapter 3

# Limits on Support Recovery of Sparse Signals: Arbitrarily Distributed Random Measurement Matrices and Block-Sparse Signals

## 3.1 Introduction

Recent years have witnessed a great deal of work on the problem of recovering a sparse signal $\mathbf{X} \in \mathbb{R}^m$ in high dimension with a small number of nonzero elements. This problem involves the estimation of $\mathbf{X}$ via linear measurements $\mathbf{Y} = A\mathbf{X} + \mathbf{Z}$, where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix, and $\mathbf{Z}$ is the measurement noise. The goal is to estimate the signal $\mathbf{X}$ from as few measurements as possible. This problem has been motivated by a broad spectrum of applications, including compressed sensing [32, 33], biomagnetic inverse problems [34, 35], image processing [36, 37], bandlimited extrapolation and spectral estimation [38], robust regression and outlier detection [39], speech processing [40], channel estimation [41, 42], echo cancellation [43, 44], and wireless communication [45]. Com-

putationally efficient algorithms for solving the problem have been proposed in different settings, e.g., matching pursuit (MP) [46], orthogonal matching pursuit (OMP) [47], LASSO [48], basis pursuit (BP) [49], FOCUSS [34], and sparse Bayesian learning [50]. In addition, analysis (e.g., [32, 51–58]) has been developed to establish conditions under which a given algorithm succeeds with high probability.

In various applications, the nonzero elements of $\mathbf{X}$ often take place in clusters, referred to as *block-sparse* [59–61]. For instance, in magnetoencephalography (MEG) and electroencephalography (EEG) of medical imaging [62–64], the brain activities are in localized regions rather than at a single point, which can be modeled as a block-sparse signal. Another example is communication channel modeling where an ideal sparse channel consisting of a few specular multi-path components has a discrete time, bandlimited, baseband representation which exhibits a block-sparse structure with the block centers determined by the arbitrary arrival times of the multi-path components. Other examples are multi-band signals [65–68], measurements of gene expression levels [69], Neuromagnetic source imaging [35], and DNA microarrays [69]. Since there is structure in the sparse signal, if the additional information is judiciously utilized in a sparse signal recovery algorithm, the performance can be improved. Such algorithms have been proposed together with mathematical tools to analyze the performance of the proposed algorithms [59, 60, 70–74].

Of particular importance is to find the positions of the nonzero elements in $\mathbf{X}$, known as *support recovery*, in a wide variety of applications such as EEG/MEG of medical imaging [62, 63], spectrum sensing in cognitive radio systems [75], multiuser detection in communication systems [45], graphical model selection [76], and signal denoising [49]. Information-theoretic tools have proven successful to provide sufficient and necessary conditions to characterize the asymptotic performance limits of optimal algorithms for support recovery with additive white Gaussian noise (AWGN), regardless of computational complexity [77–82].

One of the most fundamental questions in the problem of estimating $\mathbf{X}$ or its support is what are the conditions on the measurement matrix $A$, under

which reliable estimation is possible. A well-known tool for characterizing the recovery ability of a measurement matrix is the restricted isometry property (RIP) [83, 84]. It has been shown that if a measurement matrix satisfies the RIP with an appropriate restricted isometry constant, BP can exactly recover $\mathbf{X}$ in the noiseless setting, i.e., $Z_i = 0$ for all $i$. Another way to characterize the recovery ability is the coherence measure [57, 85–88]. It has been proved that under proper conditions on the coherence measure, $\mathbf{X}$ can be accurately recovered by BP, MP, and OMP in the noiseless setting. For block-sparse signals, the RIP and the coherence measure have been generalized to the block-RIP and the block-coherence measure, respectively [70]. Then, an interesting question that arises in this context is how to construct a *good* measurement matrix. In fact, a measurement matrix whose elements are independently and identically distributed (i.i.d.) realizations of certain zero-mean distributions (e.g., Gaussian and symmetric Bernoulli), referred to as *random measurement matrix*, has been shown to satisfy the RIP (or block-RIP) and the required conditions on the coherence measure with high probability [32, 52, 89]. In addition, a collection of deterministic matrices that satisfy the RIP was established in [90], formed by deterministic selection of rows of Fourier matrices. In analysis of asymptotic performance bounds (i.e., $m \to \infty$) in the AWGN setting, the zero-mean Gaussian distribution has been shown to achieve the optimal performance of support recovery. However, non-Gaussian distributions have not been well studied, and only limited results are available, e.g., a sharp necessary condition for asymptotically reliable support recovery was presented in [78], which is commonly applicable to any distribution with zero mean and unit variance including the Gaussian distribution.

In this paper, we study two interesting questions on the support recovery of sparse signals through measurements with AWGN: (1) what is the effect of the distribution of the random measurement matrix on the asymptotic performance limits for exact support recovery?, and (2) especially in case a signal is block-sparse how much can we reduce the number of measurements for exact support recovery by exploiting the block-sparsity structure? We focus on the scenario where the number of nonzero elements $k$ in the sparse signal is fixed, which has been

observed in practical applications, e.g., medical applications such as EEG/MEG and the GPS signal acquisition (refer to [82] for details). Our contributions in regard to each of the above two questions are summarized below.

**(1) Effect of the distribution of the random measurement matrix**: For simplicity and clarity of presentation, we first restrict our attention to conventional sparse signals [32, 33], referred to as scalar-sparse signals. For a random measurement matrix $A$ whose distribution $\mathcal{P}$ is arbitrarily given, we derive the asymptotic performance limits for reliable support recovery of scalar-sparse signals in the AWGN setting. The performance limits involves the signal dimension $m$, the number of nonzero elements $k$, the number of measurements $n$, the nonzero value vector $\mathbf{w}$, the distribution $\mathcal{P}$, and the signal-to-noise ratio (SNR) of each nonzero element $\gamma_i$ (defined in (3.3)). In particular, we show that, when $k$ is fixed, $n = (\log m)/C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ is sufficient and necessary for asymptotically successful support recovery. We provide a complete characterization of $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ as a function of the maximum achievable sum-rates of $(2^k - 1)$ Gaussian-noise multiple-access channels (MACs) [91], where the maximum achievable sum-rate of each Gaussian MAC depends on the distribution $\mathcal{P}$ and the SNRs of each nonzero element. Together with interpretations of our main result, we demonstrate the conditions under which a non-Gaussian random measurement matrix performs close to the Gaussian random measurement matrix.

The main result is inspired by the connection between the problem of support recovery of scalar-sparse signals and the problem of communication over Gaussian MAC. This connection was introduced in [81] where only Gaussian random measurement matrices are considered. In this paper, we make this connection more specific and tighter as well as generalize it to the support recovery problem with arbitrarily distributed random measurement matrices. Especially, we show that $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ is the same as the symmetric maximum achievable rate (SMAR) of a conventional Gaussian MAC with a given channel input distribution, where the SMAR is defined to be the highest rate at which all the messages from each sender can be reliably decoded at the receiver. The SMAR of a conventional Gaussian MAC can be simply obtained from the known result of optimal achievable rate

region for the Gaussian MAC. This clear relationship leads to the opportunity of leveraging the rich results and insights available in network information theory to help understand the performance limits of sparse signal recovery. In addition, the approach of interpreting our original support recovery problem as a Gaussian MAC communication problem has motivated us to develop the proof of the main result. It should be noted that the proof techniques in [81] do not naturally extend to non-Gaussian random measurement matrices and we develop new techniques to prove our main result, which are applicable to any distribution (including Gaussian) of the random measurement matrix.

**(2) Benefit of the block-sparsity structure**: We derive the asymptotic performance limits for reliable support recovery of block-sparse signals in the AWGN setting. The performance limits involves the signal dimension $m$, the number of nonzero blocks $k_b$, the block size $b$, the number of measurements $n$, the nonzero value vector $\mathbf{w}$, the distribution $\mathcal{P}$, and the SNR of each nonzero block $\gamma_i^{(b)}$ (defined in (3.33)). In particular, we show that, when the number of nonzero element $k$ is fixed, $n = (\log m)/C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w})$ is sufficient and necessary for asymptotically successful support recovery. We give a complete characterization of $C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w})$ as a function of the maximum achievable sum-rates of $(2^{k_b} - 1)$ Gaussian-noise multi-input and single-output (MISO) MACs [91], where the maximum achievable sum-rate of each Gaussian MISO MAC depends on the distribution $\mathcal{P}$ and the SNRs of each nonzero block.

Based on the asymptotic sufficient and necessary condition, we discuss how much we can reduce the number of measurements for asymptotically accurate support recovery by exploiting the block-sparsity structure. We identify three factors by which the number of measurements can be potentially reduced: increased SNR, reduced effective number of nonzero elements (i.e., not individual elements but blocks), and diversity. In addition, we discuss in which environment and how much the above factors take effect.

The main results are motivated by an interesting interpretation of the problem of support recovery for block-sparse signals, where the original problem is viewed as a communication problem over Gaussian MISO MAC with each sender

equipped with multiple transmit antennas whose number corresponds to the block size of the block-sparse signal. We show that $C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w})$ is identical to the SMAR of a conventional Gaussian MISO MAC, which can be readily derived from the known result of the optimal achievable rate region of the conventional Gaussian MISO MAC. As in the case of scalar-sparse signal recovery, this explicit relationship makes a number of results and insights on MISO MAC communication more accessible to help understand the performance limits of block-sparse signal recovery.

The rest of this paper is organized as follows[1]. In section II, we consider the problem of recovering the support of scalar-sparse signals where the distribution of the random measurement matrix is arbitrarily given. First, we present the signal model and formulate the problem. We then discuss an interpretation of the problem via multiple-access communication, to motivate the main results and their proof techniques. Next, we present the asymptotic sufficient and necessary condition for exact support recovery, along with discussion on the effect of distribution of the measurement matrix on the asymptotic performance limits. In section III, we study the problem of recovering the support of block-sparse signals in a parallel manner. After presenting signal model and problem formulation, we introduce an important interpretation of the problem by relating it to a MISO MAC communication problem. We then present the asymptotic sufficient and necessary condition for exact support recovery, in conjunction with a discussion on the benefit of the block-sparsity structure. Section IV concludes the paper with further discussions.

Throughout this paper, a set is a collection of unique objects. Let $\mathbb{R}^m$ denote the $m$-dimensional real Euclidean space. Let $\mathbb{N} = \{1, 2, 3, \dots\}$ denote the set of natural numbers. Let $[k]$ denote the set $\{1, 2, \dots, k\}$. The notation $|S|$ denotes the cardinality of set $S$, $\|\mathbf{x}\|$ denotes the $\ell_2$-norm of a vector $\mathbf{x}$, and $\|A\|_F$ denotes the Frobenius norm of a matrix $A$.

[1]The two main subjects of the paper, the effect of the distribution of the random measurement matrix and the benefit of the block-sparsity structure, could be treated at the same time by considering the problem of support recovery for block-sparse signals with a given probability distribution of the random measurement matrix. However, for ease of exposition, we first focus on the first subject with scalar-sparse signals and then study the second subject with block-sparse signals.

## 3.2 Limits on Support Recovery of Scalar-Sparse Signals: Effect of the Measurement Matrix

In this section, we restrict our attention to the scalar-sparse signals. Of special interest is the effect of the distribution of the random measurement matrix on the asymptotic performance limits of support recovery.

### 3.2.1 Signal Model and Problem Formulation

Let $\mathbf{w} = (w_1, \ldots, w_k)^\intercal \in \mathbb{R}^k$, where $w_i \neq 0$ for all $i$. Let $\mathbf{S} = (S_1, \ldots, S_k)^\intercal \in [m]^k$ be such that $S_1$,...,$S_k$ are chosen uniformly at random from $[m]$ without replacement. Then, the signal of interest $\mathbf{X} = \mathbf{X}(\mathbf{w}, \mathbf{S})$ is generated as

$$X_s = \begin{cases} w_j & \text{if } s = S_j, \\ 0 & \text{if } s \notin \{S_1, \ldots, S_k\}. \end{cases} \tag{3.1}$$

Thus, the support of $\mathbf{X}$ is $\text{supp}(\mathbf{X}) = \{S_1, \ldots, S_k\}$ and $|\text{supp}(\mathbf{X})| = k$. We measure $\mathbf{X}$ through the linear operation

$$\mathbf{Y} = A\mathbf{X} + \mathbf{Z} \tag{3.2}$$

where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix, $\mathbf{Z} \in \mathbb{R}^n$ is the measurement noise, and $\mathbf{Y} \in \mathbb{R}^n$ is the noisy measurement. We further assume that the elements of the measurement matrix $A$ are i.i.d. realizations of the probability distribution $\mathcal{P}$, either a probability mass function (pmf) or a probability density function (pdf) with zero mean and variance $\sigma_a^2$, and the noise $Z_i$ is i.i.d. according to the Gaussian distribution $\mathcal{N}(0, \sigma_z^2)$. Let us define the SNR of the $i$-th nonzero element as

$$\gamma_i \triangleq w_i^2 \frac{\sigma_a^2}{\sigma_z^2}. \tag{3.3}$$

We assume that $k$, $A$, and $\mathcal{N}(0, \sigma_z^2)$ are known, but $\mathbf{w}$ is unknown.

Upon observing the noisy measurement $\mathbf{Y}$, the goal is to recover the support of the scalar-sparse signal $\mathbf{X}$. A support recovery map is defined as

$$d : \mathbb{R}^n \longmapsto 2^{[m]}. \tag{3.4}$$

Given the signal model (3.1), the measurement model (3.2), and the support recovery map (3.4), the performance metric is defined to be the average probability of error in support recovery, i.e.,

$$\mathsf{P}\{d(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}))\}$$

for each (unknown) signal value vector $\mathbf{w} \in \mathbb{R}^k$. Note that the probability here is taken over the random signal support vector $\mathbf{S}$, the random measurement matrix $A$, and the random noise $\mathbf{Z}$.

### 3.2.2 Interpretation of Support Recovery via Multi-User Communication

An analogy was drawn in [81] between the problem of recovering the support of a scalar-sparse signal through a Gaussian random measurement matrix and a communication problem over a MAC with Gaussian inputs (Gaussian random codebooks). We make this connection more specific and generalize it to the problem of support recovery via arbitrarily distributed random measurement matrix.

We first define a conventional Gaussian MAC problem in which encoding is restricted to random code ensembles with a given probability distribution, referred to as the *conventional MAC* (*CMAC*) problem. We then define the problem of communication over a Gaussian MAC that is equivalent to the support recovery problem, referred to as the *equivalent MAC* (*EMAC*) problem. Finally, we relate the performance limit of EMAC to known results on CMAC, by discussing the differences between the two problems and their expected effects on the performance limits, which will be proved in the next subsection.

This approach not only motivates the intuition behind our results on the support recovery problem but also facilities the development of the proof techniques. We will extend this approach to the problem of support recovery for block-sparse signals in Section 3.3.

**Conventional MAC (CMAC)**

Consider the $k$-sender Gaussian MAC [93]. Sender $i$ has access to a randomly generated codebook $A^{(i)} = \{\mathbf{A}_1^{(i)}, \cdots, \mathbf{A}_{m^{(i)}}^{(i)}\}$, where $\mathbf{A}_j^{(i)} \in \mathbb{R}^n$ is a codeword whose elements are i.i.d. realizations of a probability distribution $\mathcal{P}$ with zero mean and variance $\sigma_a^2$, and $m^{(i)}$ is the number of codewords in $A^{(i)}$. Codebooks of different senders are independent of each other. The rate of sender $i$, $R^{(i)} = (\log m^{(i)})/n$. To transmit a message, each sender chooses a codeword from its codebook. Let $S_i$, uniformly distributed over $[m^{(i)}]$, denote the codeword index chosen by sender $i$. Then, the received signal $\mathbf{Y} \in \mathbb{R}^n$ at the receiver is

$$\mathbf{Y} = w_1 \mathbf{A}_{S_1}^{(1)} + \cdots + w_k \mathbf{A}_{S_k}^{(k)} + \mathbf{Z} \tag{3.5}$$

where $w_i \in \mathbb{R}$ is the channel gain associated with sender $i$ and $\mathbf{Z} \in \mathbb{R}^n$ is the noise with components $Z_j$ i.i.d. according to $\mathcal{N}(0, \sigma_z^2)$. Upon receiving $\mathbf{Y}$, the receiver determines the codeword indices transmitted by each sender, $(\hat{S}_1, \ldots, \hat{S}_k)$. The channel gain vector $\mathbf{w} = (w_1, \ldots, w_k)$ is known to the receiver.

Since the senders interfere with each other, there is an inherent tradeoff among their operating rates. The notion of achievable rate region is introduced to capture this tradeoff by characterizing all possible rate tuples $\mathcal{R} \triangleq (R^{(1)}, R^{(2)}, \cdots, R^{(k)})$ at which the decoding error probability, $\mathsf{P}\{(\hat{S}_1, \ldots, \hat{S}_k) \neq (S_1, \ldots, S_k)\}$ when averaged over the random code ensemble, diminishes as $m \to \infty$. The optimal achievable rate region can be characterized

$$\mathcal{R} = \left\{ (R^{(1)}, \ldots, R^{(k)}) : \sum_{i \in \mathcal{T}} R^{(i)} \leq I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}), \forall \mathcal{T} \subseteq [k] \right\} \tag{3.6}$$

where

$$I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) \triangleq I((V_j : j \in \mathcal{T}); Y_{\mathcal{T}}) \tag{3.7}$$

is the maximum achievable sum-rate of the $|\mathcal{T}|$-sender Gaussian MAC

$$Y_{\mathcal{T}} = \sum_{j \in \mathcal{T}} w_j V_j + Z. \tag{3.8}$$

Here the inputs $(V_j : j \in \mathcal{T})$ are i.i.d. according to $\mathcal{P}$, $w_j$ is a constant channel gain associated with the $j$-th sender, and the noise $Z$ is distributed according to

$\mathcal{N}(0, \sigma_z^2)$. For example, when $k = 2$, $\mathcal{T} = \{1, 2\}$, and $\mathcal{P} = \mathcal{N}(0, \sigma_a^2)$,

$$I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) = \frac{1}{2} \log\left(1 + (w_1^2 + w_2^2)\frac{\sigma_a^2}{\sigma_z^2}\right).$$

Another example which will be useful for later discussion is the case in which $k = 2$, $\mathcal{T} = \{1, 2\}$, and $\mathcal{P}$ is a symmetric Bernoulli distribution with $\mathsf{P}(V_j = \pm\sigma_a) = 1/2$. Then,

$$I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) = 2 - \frac{1}{4}\sum_{v \in \mathcal{V}} \mathsf{E}_Z\left[\log\left(\sum_{v' \in \mathcal{V}} e^{\frac{-(v-v'+Z)^2+Z^2}{2\sigma_z^2}}\right)\right] \tag{3.9}$$

where $\mathcal{V} = \{(|w_1| + |w_2|)\sigma_a, (|w_1| - |w_2|)\sigma_a, (-|w_1| + |w_2|)\sigma_a, (-|w_1| - |w_2|)\sigma_a\}$. Note that when $|w_1| \neq |w_2|$, (3.9) is nothing but the input–output mutual information of the standard AWGN channel $Y = X + Z$ with equiprobable 4-PAM input constellations [94] where distances between consecutive signal points are not uniform. In particular, when $|w_1| = 2|w_2|$, (3.9) becomes the mutual information of the standard AWGN channel with the standard (i.e., equiprobable and equally spaced) 4-PAM input at SNR $5w_1^2\sigma_a^2/(2\sigma_z^2)$. Let $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ denote the highest rate at which all the messages from each sender can be reliably decoded at the receiver, referred to as the *symmetric maximum achievable rate (SMAR)* of CMAC. Then, $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ can be readily obtained from $\mathcal{R}$ as

$$C_{\text{sym}}(\mathcal{P}, \mathbf{w}) = \max\left\{R^{(1)} : R^{(1)} \in \mathcal{R} \cap \mathcal{D}\right\} \tag{3.10}$$

where $\mathcal{D} = \left\{(R^{(1)}, \ldots, R^{(k)}) : R^{(1)} = \cdots = R^{(k)}\right\}$. Thus,

$$C_{\text{sym}}(\mathcal{P}, \mathbf{w}) = \min_{\mathcal{T} \subseteq [k]}\left[\frac{1}{|\mathcal{T}|} I_{\mathcal{T}}(\mathcal{P}, \mathbf{w})\right]. \tag{3.11}$$

**Equivalent MAC (EMAC)**

Now, we define the EMAC problem that interprets our original support recovery problem as a communication problem over a Gaussian MAC. The measurement model in (3.2) can be alternatively represented by

$$\mathbf{Y} = X_{S_1}\mathbf{A}_{S_1} + \cdots + X_{S_k}\mathbf{A}_{S_k} + \mathbf{Z} = w_1\mathbf{A}_{S_1} + \cdots + w_k\mathbf{A}_{S_k} + \mathbf{Z} \tag{3.12}$$

where $\mathbf{A}_s \in \mathbb{R}^n$ corresponds to the $s$-th column of $A$ of which elements are i.i.d. realizations of a probability distribution $\mathcal{P}$. Contrasting (3.12) to (3.5), we can view the support recovery problem as the following communication problem: There are $k$ senders and a common receiver. All the senders use the common codebook $A = \{\mathbf{A}_1, \ldots, \mathbf{A}_m\}$ where $\mathbf{A}_s \in \mathbb{R}^n$ is the $s$-th codeword. The codeword index chosen by sender $s$ and the corresponding codeword are $S_s$ and $\mathbf{A}_{S_s}$, respectively. The common rate, $R = (\log m)/n$. Upon receiving $\mathbf{Y}$, the receiver determines the *set* of codeword indices transmitted by the $k$ senders, $\{\hat{S}_1, \ldots, \hat{S}_k\}$. The channel gain vector $\mathbf{w}$ is unknown to the receiver.

Let $R_{\max}$ denote the maximum $R$ at which the probability of decoding error, $\mathsf{P}\{\{\hat{S}_1, \ldots, \hat{S}_k\} \neq \{S_1, \ldots, S_k\}\}$, diminishes as $m \to \infty$. Since determining the set of codeword indices in the EMAC problem is identical to recovering the signal support in the original problem, the asymptotic sufficient and necessary condition for exact support recovery is given by

$$\frac{\log m}{n} = R_{\max}. \tag{3.13}$$

**Connecting EMAC to CMAC**

Now, let us relate $R_{\max}$ to the known result $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$. The discussion below facilitates interpretation of the main result in the next subsection that $(\log m)/n = C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ is the asymptotic sufficient and necessary conditions for exact support recovery. We first identify the differences between EMAC and CMAC, and then discuss why these differences do not affect the asymptotic performance limit (thus, $R_{\max} = C_{\text{sym}}(\mathcal{P}, \mathbf{w})$):

(i) *Recovery of the codeword index set*: The set of the codeword indices is estimated in EMAC, whereas the ordered tuple of the codeword indices are estimated in CMAC.

(ii) *Common codebook*: The codebook is shared by all senders in EMAC, whereas individual codebooks are used in CMAC.

(iii) *Unknown channel gains at the receiver*: The channel gains are unknown to the receiver in EMAC, whereas they are known to the receiver in CMAC.

For ease of explanation, let us define the following two interim problems.

(P1) The MAC problem that is the same as the CMAC problem except that the receiver recovers the *set* of codeword indices in P1. The SMAR of P1 is referred to as $C_{\mathrm{P1}}$.

(P2) The MAC problem that is the same as P1 except that all the senders share a common codebook, and no two codeword indices are the same. The SMAR of P2 is referred to as $C_{\mathrm{P2}}$.

We first show that $C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w}) = C_{\mathrm{P1}} = C_{\mathrm{P2}}$.

Let us compare CMAC and P1. Since the codeword index set is recovered in P1, the decoding error probability in P1 is lower than or equal to that in CMAC, i.e., $C_{\mathrm{P1}} \geq C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$. By adapting the standard Fano's inequality (see Appendix B), it can be easily shown that $C_{\mathrm{P1}} \leq C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$. Thus, $C_{\mathrm{P1}} = C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$, which indicates that recovering the codeword index set does not increase the SMAR compared to recovering the codeword index vector in an asymptotic sense.

Next, let us compare P1 and P2. Recall that in this paper, we only consider the case where $k$ is fixed as $m \to \infty$. Thus, we can disregard the events that codeword indices chosen by different senders are identical since the probability of the events tends to zero as $m \to \infty$. Since all codewords in both P1 and P2 are random realizations of the same probability distribution, *common codebook vs. individual codebooks* does not make any difference in error probability averaged over the codeword ensembles. Therefore, $C_{\mathrm{P1}} = C_{\mathrm{P2}}$, which implies along with the result $C_{\mathrm{P1}} = C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$ that if the channel gains were known in EMAC, then $R_{\max} = C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$.

Since the channel gains are unknown at the receiver in EMAC, $R_{\max} \leq C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$. We show in Appendix A that $C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$ is achievable in EMAC. Thus, $R_{\max} = C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$, which implies that the SMAR of CMAC does not decrease even when the channel gains are unknown.

The above discussion clearly exteriorizes the connection between the problems of support recovery of sparse signals and communication over Gaussian MAC. In addition, although the result has been obtained for revealing the performance

limits of support recovery, it may be of interest on its own in information theory, since decoding in EMAC should be performed without any explicit channel gain information and without any channel training procedure (e.g., through pilot symbols), but can still achieve the same performance as decoding with the channel gains known. More details of this aspect will be discussed in the next subsection.

### 3.2.3 Main Results

Now, we provide sufficient and necessary conditions for asymptotically reliable support recovery of scalar-sparse signals, of which problem is formulated in Section 3.2.1. Note that we consider the support recovery of a sequence of sparse signals generated with the same signal value vector $\mathbf{w}$. In particular, we assume that $k$ is fixed. In the two theorems below, the subscript in $n_m$ denotes possible dependence between the number of measurements $n$ and the signal dimension $m$. The proofs of the next two theorems are presented in Appendices 3.5 and 3.6, respectively.

**Theorem 1** *If*

$$\limsup_{m\to\infty} \frac{\log m}{n_m} < C_{sym}(\mathcal{P}, \mathbf{w}) \tag{3.14}$$

*where $C_{sym}(\mathcal{P}, \mathbf{w})$ is given by (3.11), then there exists a sequence of support recovery maps $\{d^{(m)}\}_{m=k}^{\infty}$, $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{[m]}$, such that*

$$\lim_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq supp(\mathbf{X}(\mathbf{w}, \mathbf{S}))\} = 0. \tag{3.15}$$

**Theorem 2** *If*

$$\limsup_{m\to\infty} \frac{\log m}{n_m} > C_{sym}(\mathcal{P}, \mathbf{w}), \tag{3.16}$$

*then for any sequence of support recovery maps $\{d^{(m)}\}_{m=k}^{\infty}$, $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{[m]}$,*

$$\liminf_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq supp(\mathbf{X}(\mathbf{w}, \mathbf{S}))\} > 0. \tag{3.17}$$

Theorems 1 and 2 together indicate that $\eta \triangleq n/(\log m) = 1/C_{\mathrm{sym}}(\mathcal{P}, \mathbf{w})$ is the normalized sufficient and necessary (NSN) number of measurements for exact

support recovery[2]. The constant $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ explicitly captures the role of each parameter: the number of nonzero elements $k$, the SNR of each nonzero element $\gamma_i$, and the distribution of the measurement matrix $\mathcal{P}$.

### 3.2.4 Significance of the Main Results in Information Theory

The proof of Theorem 2 for the necessary condition employs the assumption that the values of the nonzero elements are known. This indicates, from the communication perspective discussed in the previous subsection, that the SMAR of the CMAC problem is not reduced even when the channel gains are unknown to the receiver and no channel training procedure (e.g., with pilot symbols) is allowed. Note that the decoding technique described in Appendix A produces estimates of both the codewords and the channel gains only through the received signal. We can readily derive a more general result, by slightly modifying our proofs, that the optimal achievable rate region of the Gaussian MAC with an arbitrary input distribution does not decrease when the channel gains are not known to the receiver and no channel training procedure is allowed.

### 3.2.5 Effect of the Distribution of the Measurement Matrix

We further explore the effect of the distribution of the measurement matrix on the asymptotic performance limits in the support recovery of sparse signals with fixed number of nonzero elements. Theorems 1 and 2 indicate that among all distributions, the Gaussian distribution minimizes the NSN number and show how much the NSN number increases for a non-Gaussian measurement matrix, depending on the number of nonzero elements $k$ and the SNR of each nonzero element $\gamma_i$. To obtain more insight, we take a closer look into a class of uniform distributions which is often studied in the literature for the problem of support recovery for sparse signals.

[2]The NSN number of measurements for exact support recovery is simply referred to as the NSN number throughout the paper.

Consider a set $\mathcal{M}$ that corresponds to the standard (equally spaced) $M$-PAM constellations [94], i.e., $\mathcal{M} = \{\pm r, \pm 3r, \ldots, \pm(M-1)r\}$, where $M$ is an even natural number and $r \in \mathbb{R}$ satisfies that $\frac{2}{M}\{r^2 + (3r)^2 + \cdots + ((M-1)r)^2\} = \sigma_a^2$. Let $\mathcal{P}_M$ denote the uniform distribution (pmf) over $\mathcal{M}$. For example, if $M = 2$, then $\mathcal{M} = \{\pm\sigma_a\}$ and

$$\mathcal{P}_2(V = v) = \begin{cases} \frac{1}{2} & \text{for } v = \pm\sigma_a, \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

Let $\mathcal{G}$ denote the Gaussian distribution $\mathcal{N}(0, \sigma_a^2)$. We compare the NSN number between $\mathcal{P}_M$ and $\mathcal{G}$, or equivalently, compare $C_{\text{sym}}(\mathcal{P}_M, \mathbf{w})$ with $C_{\text{sym}}(\mathcal{G}, \mathbf{w})$. Let us define

$$\eta_{\text{excess}} \triangleq \frac{C_{\text{sym}}(\mathcal{G}, \mathbf{w})}{C_{\text{sym}}(\mathcal{P}_M, \mathbf{w})} \tag{3.19}$$

which measures the excess number of measurement due to using a non-Gaussian distribution.

**Uniform distribution is near-optimal when $k$ is large**

It can be readily shown that for both $\mathcal{P}_M$ and $\mathcal{G}$,

$$C_{\text{sym}}(\mathcal{P}, \mathbf{w}) = \min_{\mathcal{T} \subseteq [k]} \left[ \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) \right] = \frac{1}{k} I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) \tag{3.20}$$

when $|w_1| = \cdots = |w_k|$. For other values of $\mathbf{w}$, we can check that (3.20) still tends to hold as $k$ increases (i.e., (3.20) holds for any $\mathbf{w}$, as $k \to \infty$). Let $V_{\mathcal{P}_M} = \sum_{j=1}^k w_j V_j$ with $V_j$ i.i.d. $\sim \mathcal{P}_M$. Let $V_{\mathcal{G}} = \sum_{j=1}^k w_j V_j$ with $V_j$ i.i.d. $\sim \mathcal{G}$. Then, as $k$ increases, the distribution of $V_{\mathcal{P}_M}$ tends to that of $V_{\mathcal{G}}$ by the central limit theorem. It immediately follows that $\eta_{\text{excess}}$ tends to one as $k$ increases, i.e., the NSN number with $\mathcal{P}_m$ is approximately the same as that with $\mathcal{G}$ in the high $k$ regime.

**The excess NSN number required by a uniform distribution compared to the Gaussian distribution can be significant when both $M$ and $k$ are small and SNR is large**

By (3.10) and the definition of $I_{\mathcal{T}}(\mathcal{P}, \mathbf{w})$ in (3.37) (also, refer to (3.9)),

$$C_{\text{sym}}(\mathcal{P}_M, \mathbf{w}) < \log M \tag{3.21}$$

and

$$\lim_{\gamma_i \to \infty, \ldots, \gamma_k \to \infty} C_{\text{sym}}(\mathcal{G}, \mathbf{w}) = \infty. \tag{3.22}$$

Thus,

$$\lim_{\gamma_i \to \infty, \ldots, \gamma_k \to \infty} \eta_{\text{excess}} = \infty \tag{3.23}$$

for any finite $M$.

To deliver more insight, we discuss a simple example below. First, note that when $M_1 > M_2$,

$$C_{\text{sym}}(\mathcal{P}_{M_1}, \mathbf{w}) > C_{\text{sym}}(\mathcal{P}_{M_2}, \mathbf{w}) \tag{3.24}$$

since $I_{\mathcal{T}}(\mathcal{P}_{M_1}, \mathbf{w}) > I_{\mathcal{T}}(\mathcal{P}_{M_2}, \mathbf{w})$ for any $\mathcal{T} \subseteq [k]$ (recall that for $M$-PAM input, a larger constellation size leads to a higher input–output mutual information of the standard AWGN channel $Y = X + Z$ [96, 97]). Therefore, given $k$ and $\mathbf{w}$, the NSN number is a decreasing function of $M$. Consider the case when $k = 2$ and $M = 2$. This is a worst case for $\mathcal{P}_M$ in the sense that $M = 2$ is the minimum possible number and as discussed above, $\eta_{\text{excess}}$ tends to increase as $k$ decreases (in fact, the worst case is $k = 1$ and $M = 2$, but this case is trivial). Without loss of generality, we assume that $|w_1| < |w_2|$. Then, by (3.10),

$$C_{\text{sym}}(\mathcal{G}, (w_1, w_2)) = \min\left[\frac{1}{2}\log\left(1 + w_1^2\frac{\sigma_a^2}{\sigma_z^2}\right), \frac{1}{4}\log\left(1 + (w_1^2 + w_2^2)\frac{\sigma_a^2}{\sigma_z^2}\right)\right] \tag{3.25}$$

and

$$C_{\text{sym}}(\mathcal{P}_2, (w_1, w_2)) = \min\left[I_{\{1\}}(\mathcal{P}_2, (w_1)), \frac{1}{2}I_{\{1,2\}}(\mathcal{P}_2, (w_1, w_2))\right] \tag{3.26}$$

where

$$I_{\{1\}}(\mathcal{P}_2, (w_1)) = 1 - \frac{1}{2}\sum_{v \in \mathcal{V}_1} \mathsf{E}_Z\left[\log\left(\sum_{v' \in \mathcal{V}_1} e^{\frac{-(v-v'+Z)^2+Z^2}{2\sigma_z^2}}\right)\right] \tag{3.27}$$

with $\mathcal{V}_1 = \{\pm|w_1|\sigma_a\}$ and

$$I_{\{1,2\}}(\mathcal{P}_2, (w_1, w_2)) = 2 - \frac{1}{4}\sum_{v \in \mathcal{V}_2} \mathsf{E}_Z\left[\log\left(\sum_{v' \in \mathcal{V}_2} e^{\frac{-(v-v'+Z)^2+Z^2}{2\sigma_z^2}}\right)\right] \tag{3.28}$$

with $\mathcal{V}_2 = \{(|w_1|+|w_2|)\sigma_a, (|w_1|-|w_2|)\sigma_a, (-|w_1|+|w_2|)\sigma_a, (-|w_1|-|w_2|)\sigma_a\}$. Note that (3.27) is nothing but the mutual information of the standard AWGN channel with 2-PAM input at SNR $\gamma_1 = w_1^2\sigma_a^2/\sigma_z^2$, which is an increasing function of $\gamma_1$ and

$$\lim_{\gamma_1 \to \infty} I_{\{1\}}(\mathcal{P}_2, (w_1)) = 1. \tag{3.29}$$

The result in (3.28) corresponds to the mutual information of the standard AWGN channel with unequally spaced 4-PAM input constellations at SNR $\gamma \triangleq \frac{1}{2}(\gamma_1 + \gamma_2)$ (averaged over 4-PAM symbols), which is an increasing function of $\gamma$ and

$$\lim_{\gamma \to \infty} I_{\{1,2\}}(\mathcal{P}_2, (w_1, w_2)) = 2. \tag{3.30}$$

Thus, $C_{\text{sym}}(\mathcal{P}_2, (w_1, w_2)) < 1$ is the minimum of the mutual information of the AWGN channel with 2-PAM input and a half of the mutual information of the AWGN channel with 4-PAM input, as a function of SNRs $\gamma_1$ and $\gamma$ (refer to [96] and [97] for the mutual information curves for 2-PAM and 4-PAM). We can check that when either $\gamma_1$ or $\gamma$ is low, $\eta_{\text{excess}} \approx 1$. In contrast, $\eta_{\text{excess}}$ grows without bound, as both $\gamma_1$ and $\gamma$ increases.

## 3.3 Limits on Support Recovery of Block-Sparse Signals: Benefit of the Block-Sparsity Structure

Now, let us study the case in which the sparse signal is block-sparse, i.e, the nonzero elements of the sparse signal appear in clusters. We consider a commonly used block-sparse signal model [59, 60, 71], where clusters of nonzero elements are of the same size. We investigate the asymptotic performance limits on support recovery of block-sparse signals. Of particular interest is the benefit of making use of the block-sparsity structure, in terms of the number of measurements required for successful support recovery.

### 3.3.1 Signal Model and Problem Formulation

Let $\mathbf{w}_i = (w_{(i-1)b+1}, ..., w_{ib})^\intercal \in \mathbb{R}^b$ with $w_j \neq 0$ for all $j$. Let $\mathbf{w} = (\mathbf{w}_1^\intercal, ..., \mathbf{w}_{k_b}^\intercal)^\intercal \in \mathbb{R}^k$. Thus, $k = b \cdot k_b$. Let $\mathbf{S} = (S_1, ..., S_{k_b})^\intercal \in [m_b]^{k_b}$ be such that $S_1$, ..., $S_{k_b}$ are chosen uniformly at random from $[m_b]$ without replacement. Let $m = b \times m_b$. Let $\mathbf{X}_i = (X_{(i-1)b+1}, ..., X_{ib})^\intercal \in \mathbb{R}^b$. Then, the signal of interest $\mathbf{X} = \mathbf{X}(\mathbf{w}, \mathbf{S}, b) = (\mathbf{X}_1^\intercal, ..., \mathbf{X}_{m_b}^\intercal)^\intercal \in \mathbb{R}^m$ is generated as

$$\mathbf{X}_s = \begin{cases} \mathbf{w}_j & \text{if } s = S_j, \\ \mathbf{0} & \text{if } s \notin \{S_1, ..., S_{k_b}\} \end{cases} \tag{3.31}$$

where $\mathbf{0}$ is the all-zero column vector of size $b$. Thus, the nonzero elements of $\mathbf{X}$ appear by the block of size $b$. The set $\text{supp}_\text{b}(\mathbf{X}) = \{S_1, ..., S_{k_b}\}$ is referred to as the *block-support* of $\mathbf{X}$. Note that $\mathbf{X}$ becomes scalar-sparse when $b = 1$.

We measure $\mathbf{X}$ through the linear operation

$$\mathbf{Y} = A\mathbf{X} + \mathbf{Z} \tag{3.32}$$

where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix, $\mathbf{Z} \in \mathbb{R}^n$ is the measurement noise, and $\mathbf{Y} \in \mathbb{R}^n$ is the noisy measurement. We further assume that the elements of the measurement matrix $A$ are i.i.d. realizations of the probability distribution $\mathcal{P}$, either a pmf or a pdf with zero mean and variance $\sigma_a^2$, and the noise $Z_i$ is i.i.d. according to $\mathcal{N}(0, \sigma_z^2)$. The SNR of the $i$-th block is defined as

$$\gamma_i^{(b)} \triangleq \|\mathbf{w}_i\|^2 \frac{\sigma_a^2}{\sigma_z^2}. \tag{3.33}$$

We assume that $b$, $k_b$, $A$, and $\mathcal{N}(0, \sigma_z^2)$ are known, but $\mathbf{w}$ is unknown.

Upon observing the noisy measurement $\mathbf{Y}$, the goal is to recover the support (equivalently, block-support) of the block-sparse signal $\mathbf{X}$, $\text{supp}(\mathbf{X})$. A support recovery map is defined as

$$d : \mathbb{R}^n \longmapsto 2^{[m]}. \tag{3.34}$$

Given the signal model (3.31), the measurement model (3.32), and the support recovery map (3.34), the performance metric is defined to be the average probability of error in support recovery, i.e.,

$$\mathsf{P}\{d(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}, b))\}$$

for each (unknown) signal value vector $\mathbf{w} \in \mathbb{R}^k$. Note that the probability here is taken over the random signal block-support vector $\mathbf{S}$, the random measurement matrix $A$, and the random noise $\mathbf{Z}$.

### 3.3.2 Interpretation of Support Recovery via Multi-User Communication

We introduce an interesting interpretation of the problem of support recovery of block-sparse signals by relating it to a multiple-input single-output (MISO) MAC communication problem. This can be viewed as an extension of the connection, discussed in the previous section, between the problem of support recovery of scalar-sparse signals and a MAC communication problem.

As in the case of support recovery of scalar-sparse signals, we first define a conventional Gaussian MISO MAC problem in which encoding is restricted to random code ensembles with a given probability distribution, referred to as the *conventional MISO MAC (CMMAC)* problem. We then define a communication problem over a Gaussian MISO MAC that is equivalent to the problem of support recovery of block-sparse signals, referred to as the *equivalent MISO MAC (EMMAC)* problem. Finally, we relate the performance limit of the EMMAC to known results on the CMMAC, by discussing the differences between the two problems and their expected effects on the performance limit, which will be proved in the next subsection.

This approach not only motivates the intuition behind our main results presented in the next subsection but also facilities the development of the proof techniques.

#### Conventional MISO MAC (CMMAC)

Suppose $k_b$ senders wish to transmit messages to a common receiver. Each sender is equipped with $b$ transmit antennas and the receiver is equipped with a single receive antenna. Sender $i$ has access to a randomly generated codebook $A^{(i)} = \{A_1^{(i)}, \cdots, A_{m_b^{(i)}}^{(i)}\}$, where $A_j^{(i)} \in \mathbb{R}^{n \times b}$ is a MISO codeword of which elements

are i.i.d. realizations of the probability distribution $\mathcal{P}$ and $m_b^{(i)}$ is the number of MISO codewords in $A^{(i)}$. The rate of sender $i$, $R^{(i)} = (\log m_b^{(i)})/n$. To transmit a message, each sender chooses a MISO codeword from its codebook. Let $S_i$ denote the MISO codeword index chosen by sender $i$. Then, the received signal $\mathbf{Y} \in \mathbb{R}^n$ at the receiver is

$$\mathbf{Y} = A_{S_1}^{(1)}\mathbf{w}_1 + \cdots + A_{S_{k_b}}^{(k_b)}\mathbf{w}_{k_b} + \mathbf{Z} \tag{3.35}$$

where $\mathbf{w}_i \in \mathbb{R}^b$ is the MISO channel gain associated with sender $i$ and $\mathbf{Z} \in \mathbb{R}^n$ is the noise with components $Z_j$ i.i.d. according to $\mathcal{N}(0, \sigma_z^2)$. Upon receiving $\mathbf{Y}$, the receiver determines the MISO codeword indices transmitted by each sender, $(\hat{S}_1, \ldots, \hat{S}_{k_b})$. The channel gain vector $\mathbf{w}$ is known to the receiver.

The optimal achievable rate region of CMMAC can be characterized

$$\mathcal{R} = \left\{ (R^{(1)}, \ldots, R^{(k_b)}) : \sum_{i \in \mathcal{T}} R^{(i)} \leq I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}), \forall \mathcal{T} \subseteq [k_b] \right\} \tag{3.36}$$

where

$$I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) \triangleq I((V_{(j-1)b+i} : i = 1, \ldots, b, j \in \mathcal{T}); Y_{\mathcal{T}}) \tag{3.37}$$

is the maximum achievable sum-rate of the $|\mathcal{T}|$-sender Gaussian MISO MAC

$$Y_{\mathcal{T}} = \sum_{j \in \mathcal{T}} \sum_{i=1}^{b} w_{(j-1)b+i} V_{(j-1)b+i} + Z. \tag{3.38}$$

Here the inputs $(V_{(j-1)b+i} : i = 1, \ldots, b, j \in \mathcal{T})$ are i.i.d. according to $\mathcal{P}$, $w_{(j-1)b+i}$ is a constant channel gain associated with the $i$-th antenna of the $j$-th sender, and the noise $Z$ is distributed according to $\mathcal{N}(0, \sigma_z^2)$. For instance, when $k_b = 2$, $\mathcal{T} = \{1, 2\}$, and $\mathcal{P} = \mathcal{N}(0, \sigma_a^2)$,

$$I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) = \frac{1}{2} \log\left(1 + (\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2)\frac{\sigma_a^2}{\sigma_z^2}\right).$$

From (3.36), the symmetric maximum achievable rate (SMAR) of CMMAC is obtained as

$$C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w}) = \min_{\mathcal{T} \subseteq [k]} \left[ \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) \right]. \tag{3.39}$$

**Equivalent MISO MAC (EMMAC) and Connection to CMMAC**

Let us define the EMMAC problem. The measurement model (3.32) can be alternatively represented by

$$\mathbf{Y} = A_{S_1}\mathbf{w}_1 + \cdots + A_{S_{k_b}}\mathbf{w}_{k_b} + \mathbf{Z} \tag{3.40}$$

where $A_i = (\mathbf{A}_{(i-1)b+1}, \ldots, \mathbf{A}_{ib}) \in \mathbb{R}^{n\times b}$ is the $i$-th block of columns of $A$ of which elements are i.i.d. re. Contrasting (3.40) to (3.35), we can view the support recovery problem as the following communication problem: There are $k_b$ senders and a common receiver. Each sender has $b$ transmit antennas and the common receiver has a single receive antenna. All the senders share the common codebook $A = \{A_1, \ldots, A_{m_b}\}$ where $A_i \in \mathbb{R}^{n\times b}$ is a MISO codeword of which elements are i.i.d. realizations of the probability distribution $\mathcal{P}$. The MISO codeword index chosen by sender $i$ and the corresponding MISO codeword are $S_i$ and $A_{S_i}$, respectively. The common rate, $R^{(b)} = (\log m_b)/n$. Upon receiving $\mathbf{Y}$, the receiver estimates the *set* of MISO codeword indices transmitted by the $k_b$ senders, $\{\hat{S}_1, \ldots, \hat{S}_k\}$. The channel gains $\mathbf{w}$ are unknown to the receiver.

Let $R^{(b)}_{\max}$ denote the maximum $R^{(b)}$ at which the probability of error, $\mathsf{P}\{\{\hat{S}_1, \ldots, \hat{S}_{k_b}\} \neq \{S_1, \ldots, S_{k_b}\}\}$, diminishes as $m \to \infty$. Since determining the set of MISO codeword indices in the EMMAC problem is identical to recovering the block support in the original problem, the asymptotic sufficient and necessary condition for exact support recovery is given by

$$\frac{\log m_b}{n} = R^{(b)}_{\max}. \tag{3.41}$$

The distinctive features of EMMAC compared to CMMAC are *recovery of the MISO codeword index set*, *common codebook*, and *unknown MISO channel gains at the receiver*. The effect of the differences on the asymptotic performance limits are essentially the same as in the case of scalar-sparse signals, discussed in Section 3.2.2. Consequently, $R^{(b)}_{\max} = C^{(b)}_{\text{sym}}(\mathcal{P}, \mathbf{w})$, which explicitly relates the performance limits of support recovery of block-sparse signals to the SMAR of CMMAC.

### 3.3.3 Main Results

Now, we provide sufficient and necessary conditions for asymptotically reliable support recovery of block-sparse signals. Note that we consider the support recovery of a sequence of block-sparse signals generated with the same signal value vector $\mathbf{w}$. In particular, we assume that $k$ (or $k_b$) is fixed. The proofs of the next two theorems are presented in Appendices 3.5 and 3.6, respectively.

**Theorem 3** *If*

$$\limsup_{m\to\infty} \frac{\log m}{n_m} < C_{sym}^{(b)}(\mathcal{P}, \mathbf{w}) \tag{3.42}$$

*where* $C_{sym}^{(b)}(\mathcal{P}, \mathbf{w})$ *is given by (3.39), then there exists a sequence of support recovery maps* $\{d^{(m)}\}_{m=k}^{\infty}$, $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{[m]}$, *such that*

$$\lim_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq supp(\mathbf{X}(\mathbf{w}, \mathbf{S}, b))\} = 0. \tag{3.43}$$

**Theorem 4** *If*

$$\limsup_{m\to\infty} \frac{\log m}{n_m} > C_{sym}^{(b)}(\mathcal{P}, \mathbf{w}), \tag{3.44}$$

*then for any sequence of support recovery maps* $\{d^{(m)}\}_{m=k}^{\infty}$, $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{[m]}$,

$$\liminf_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq supp(\mathbf{X}(\mathbf{w}, \mathbf{S}), b)\} > 0. \tag{3.45}$$

Theorems 3 and 4 together indicate that $\eta_b \triangleq n/(\log m) = 1/C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w})$ is the normalized sufficient and necessary (NSN) number of measurements for exact support recovery of block-sparse signals. The constant $C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w})$ explicitly captures the role of each parameter: the number of nonzero blocks, block size, SNR of each nonzero block $\gamma_i^{(b)}$, and distribution of the measurement matrix $\mathcal{P}$.

Significance of the main results in information theory discussed in Section 3.2.4 is straightforwardly extended to the case of block-sparse signals. We can readily derive a more general result, by slightly modifying our proofs, that the optimal achievable rate region of the Gaussian MISO MAC with an arbitrary input distribution does not decrease when the channel gains are not known to the receiver and no channel training procedure is allowed.

### 3.3.4 Benefit of Block-Sparsity Structure

One of the most interesting questions in the problem of support recovery of block-sparse signals is how much the NSN number can be reduced by exploiting the known block-sparsity structure, compared to treating the block-sparse signal as being scalar-sparse. In fact, there are three factors that can potentially reduce the NSN number.

(i) *Increased SNR*: In EMMAC, total transmit power per sender increases linearly with the number of transmit antennas (or block size). Thus, the SNR of a MISO codeword $\gamma_i^{(b)}$ increases accordingly. This is different from a conventional comparison [98] between single-input single-output (SISO) and MISO in communication theory, where total transmit power is assumed to be the same, regardless of the number of transmit antennas.

(ii) *Reduced effective number of nonzero elements*: The signal support is estimated not by the individual element but by the block of elements.

(iii) *Diversity*: Suppose that correlation among magnitudes of each nonzero element $|w_1|, \ldots, |w_k|$ is small and $\gamma_1, \ldots, \gamma_k$ are low. Then the NSN number is mainly determined by $|w_{\min}| = \min\{|w_1|, \ldots, |w_k|\}$ when the block-sparse signal is treated as being scalar-sparse signal (see (3.11)). However, if we appropriately exploit the block-sparsity structure (see (3.39)), the NSN number is determined by $|w_{\min}^{(b)}| = \min\{\|\mathbf{w}_1\|, \ldots, \|\mathbf{w}_{k_b}\|\}$. Note that $|w_{\min}^{(b)}|$ can be much larger than $|w_{\min}|$ due to diversity effect.

We further explore in which environment and how much the above factors take effect. To elaborate, we assume that $\mathcal{P} = \mathcal{N}(0, \sigma_a^2)$.

**The increased SNR and the reduced effective number of nonzero elements are beneficial in the low and the high SNR regime, respectively**

Consider block-sparse signals $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ with $k_{b_1}$ and $k_{b_2}$ nonzero blocks ($k_{b_1} \neq k_{b_2}$), respectively. Assume that $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ have the same block size $b$ and the same $l_2$-norm for each nonzero block, i.e., $\|\mathbf{w}_1\| = \|\mathbf{w}_2\| = \cdots$. Note that

the total number of nonzero elements are different between the two signals. Let $\eta_1$ and $\eta_2$ be the NSN number for $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, respectively. Let $\gamma = \|\mathbf{w}_1\|^2\sigma_a^2/\sigma_z^2$ denote the SNR of each nonzero block. Then,

$$\eta_2/\eta_1 \approx k_{b_2}/k_{b_1} \text{ for high } \gamma \text{ and} \tag{3.46}$$

$$\eta_1 \approx \eta_2 \propto 1/\gamma \text{ for low } \gamma. \tag{3.47}$$

The proof is presented in Appendix 3.7. The above result indicates that the NSN number is mainly determined by the number of nonzero blocks in the high SNR regime and is mainly determined by the SNRs of each nonzero block in the low SNR regime.

**Minimum block-sparsity gain and diversity effect**

Consider a block-sparse signal $\mathbf{X}^{(1)}$. Let $\eta_1$ denote the NSN number when $\mathbf{X}^{(1)}$ is treated as a scalar-sparse signal, i.e., $\eta_1 = 1/C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ where $C_{\text{sym}}(\mathcal{P}, \mathbf{w})$ is given by (3.11). Let $\eta_2$ denote the NSN number when the block-sparsity structure is properly utilized, i.e., $\eta_2 = 1/C^{(b)}_{\text{sym}}(\mathcal{P}, \mathbf{w})$ where $C^{(b)}_{\text{sym}}(\mathcal{P}, \mathbf{w})$ is given by (3.39). Then,

$$\eta_2 \leq \frac{1}{b}\eta_1. \tag{3.48}$$

The proof is presented in Appendix 3.8. Thus, we can reduce the number of measurements for asymptotically reliable support recovery by at least '1/(block size)' if we properly utilize the block-sparsity structure, compared to ignoring the the structure. It can be readily shown that equality in (3.48) holds when $|w_1| = \cdots = |w_k|$, i.e., the benefit of block-sparsity structure is minimal in case the nonzero values have the same magnitude. In contrast, as discussed above, if the correlation in the magnitudes of the nonzero values is small, $\eta_2/\eta_1$ can be much less that $1/b$ due to diversity effect. However, it can be easily shown that for any $\mathbf{w}$, $\lim_{\sigma_z^2 \to 0} \frac{\eta_2}{\eta_1} = \frac{1}{b}$, which implies that the gain from block-sparsity structure is no more than 1/(block size) when the noise power is small, regardless of the correlation in the magnitudes of the nonzero values.

## 3.4 Concluding Remarks

In this paper, we have developed asymptotic performance limits of recovering the support of block-sparse signals (including scalar-sparse signals as a special class) in a noisy setting, through a random measurement matrix of which elements are i.i.d. realizations of a given probability distribution. For the case when the number of nonzero blocks is fixed, sharp sufficient and necessary conditions for asymptotically reliable support recovery have been derived as a function of the input–output mutual information of plural point-to-point communication channels. The results reveal the role of the distribution of the random measurement matrix as well as the benefit of the known block-sparsity structure.

Our work rely on an important connection between the problem of support recovery of block-sparse signals and the problem of communication over MISO MAC, generalizing the idea originally introduced in [81]. We provide a clear connection between the two problems, which helps us understand the original support recovery problem in more depth. In particular, this interpretation has the potential to deal with various theoretical and practical issues related to sparse signal recovery. Some examples of potential applications were addressed in [81, 82]. Two more interesting potential directions are discussed below.

(i) *Design of the measurement matrix for block-sparse signals with correlated nonzero elements within a block*: In this paper, we have considered the random measurement matrix whose elements are i.i.d. and therefore all the columns are independent. This is optimal for the case when the values within each nonzero block are uncorrelated. This is also the best approach even for the case when the values within each nonzero block are correlated, but the correlation coefficient is unknown, although a genie could find the optimal measurement matrix depending on the exact values of nonzero elements, i.e., could implement transmit beamforming so that the sub-codewords (different columns associated with a block) are coherently combined, providing 3 dB SNR gain compared to independent sub-codewords. In contrast, if the correlation information is available, we can improve the performance by appropri-

ately designing the measurement matrix taking into account the correlation information. This problem was briefly addressed in [71], conjectured to be a difficult design problem. According to the interpretation of the support recovery of block-sparse signals as a communication over MISO MAC, this can be regarded as the problem of designing the MISO codewords with spatially correlated MISO channels. One can develop optimal design strategies for the measurement matrix by applying the results for correlated MISO systems (e.g., [99]).

(ii) *Adaptive measurement matrix*: Another interesting potential application is the problem of adaptive measurement matrix [100, 101], variously referred to as adaptive sensing and dictionary learning, where the measurement matrix is not deterministic but adaptive to the noisy measurements available. According to our information-theoretic analytical framework, this problem can be viewed as communication over Gaussian MAC with feedback [102, 103]. Therefore, one can explore this connection and leverage the techniques available in communication with feedback (e.g., [104]) to come up with adaptive algorithms as well as the performance bounds.

Finally, we conclude this paper by leaving it as future work to extend our approach to the problem of support recovery of block-sparse signals with multiple measurement vectors [70, 82, 105, 106], which can be interpreted as the problem of communication over multiple-input multiple-output (MIMO) MAC.

The text of this chapter, in part, is a reprint of the paper, H. Kwon, B. Rao, and Y. Kim, "Limits on Support Recovery of Sparse signals: Measurement Matrix and Block-Sparsity", in preparation for submission. The dissertation author is the primary researcher and author, and the co-authors contributed to or supervised the research which forms the basis of this chapter.

## 3.5 Proof of Theorems 1 and 3

Recall that Theorem 3 is a generalized result of Theorem 1 since scalar-sparse signals are a special class of block-sparse signals, i.e., block size $b = 1$. Thus,

we only present the proof of Theorem 3. We assume that the receiver applies the distance decoding [92] to obtain an estimate of codeword index set. To bound the probability of decoding error, we use Sanov's theorem [91, Section 11.4, Theorem 11.4.1] that gives a bound on the probability of observing an atypical sequence of samples from a given probability distribution.

Fix $\epsilon \in \mathbb{R} > 0$. First, form an estimate $\hat{\rho}$ of $\|\mathbf{w}\|$ as

$$\hat{\rho} \triangleq \sqrt{\frac{\left|\frac{1}{n}\|\mathbf{Y}\|^2 - \sigma_z^2\right|}{\sigma_a^2}}. \tag{3.49}$$

For $r, \zeta > 0$, let $\mathcal{Q} = \mathcal{Q}(r, \zeta)$ be a minimal set of points in $\mathbb{R}^k$ satisfying the following properties:

i) $\mathcal{Q} \subseteq \mathcal{B}_k(r)$, where $\mathcal{B}_k(r)$ is the $k$-dimensional hypersphere of radius $r$, i.e., $\mathcal{B}_k(r) \triangleq \{\mathbf{b} : \mathbf{b} \in \mathbb{R}^k, \|\mathbf{b}\| = r\}$,

ii) For any $\mathbf{b} \in \mathcal{B}_k(r)$, there exists $\hat{\mathbf{w}} \in \mathcal{Q}$ such that $\|\hat{\mathbf{w}} - \mathbf{b}\| \leq \frac{\zeta}{2}$.

The following properties [81] are useful:

**Lemma 2** *1)*

$$\lim_{m\to\infty} \mathsf{P}\left(\exists \hat{\mathbf{W}} \in \mathcal{Q}(\hat{\rho}, \zeta) \text{ such that } \|\hat{\mathbf{W}} - \mathbf{w}\| < \zeta\right) = 1.$$

*2)* $q(r, \zeta) \triangleq |\mathcal{Q}(r, \zeta)|$ *is monotonically non-decreasing in* $r$ *for fixed* $\zeta$.

Given $\hat{\rho}$ and $\epsilon$, fix $\mathcal{Q} = \mathcal{Q}(\hat{\rho}, \epsilon)$. Declare $d(\mathbf{Y}) = \{\hat{s}_1, \hat{s}_2, ..., \hat{s}_{k_b}\} \subseteq [m_b]$ is the recovered block-support of the signal, if it is the unique set of indices such that

$$\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k_b} A_{\hat{s}_j}\hat{\mathbf{W}}_j\right\|^2 \leq \sigma_z^2 + \epsilon^2\sigma_a^2 \tag{3.50}$$

for some $\hat{\mathbf{W}} = (\hat{\mathbf{W}}_1^\intercal, \ldots, \hat{\mathbf{W}}_{k_b}^\intercal)^\intercal \in \mathcal{Q}$ with $\hat{\mathbf{W}}_j \in \mathbb{R}^{b\times 1}$. Here $A_{\hat{s}_j} \in \mathbb{R}^{n\times b}$ is the $\hat{s}_j$-th block of columns of $A$. If there is none or more than one such set, pick an arbitrary set of $k_b$ indices.

Next, we analyze the average probability of error

$$\mathsf{P}(\mathcal{E}) = \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq \{S_1, ..., S_{k_b}\}\}.$$

We assume without loss of generality that $S_j = j$ for $j = 1, 2, ..., k_b$, which gives

$$\mathbf{Y} = \sum_{j=1}^{k_b} A_j \mathbf{w}_j + \mathbf{Z}$$

for some $\mathbf{w} = (\mathbf{w}_1^\intercal, \ldots, \mathbf{w}_{k_b}^\intercal)^\intercal$. Define the event

$$\mathcal{E}_{s_1,s_2,...,s_{k_b}} \triangleq \Bigg\{ \exists \hat{\mathbf{W}} \in \mathcal{Q} \text{ and } \{s'_1, s'_2, ..., s'_{k_b}\} = \{s_1, s_2, ..., s_{k_b}\} \\ \text{such that } \frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j} \hat{\mathbf{W}}_j \right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \Bigg\}.$$

Then

$$\begin{aligned} \mathsf{P}(\mathcal{E}) &= \mathsf{P}\left( \mathcal{E}^c_{1,2,...,k_b} \cup \left( \bigcup_{s_1 < \cdots < s_{k_b} : \{s_1,...,s_{k_b}\} \neq [k_b]} \mathcal{E}_{s_1,s_2,...,s_{k_b}} \right) \right) \\ &\leq \mathsf{P}\left( \mathcal{E}^c_{\text{aux}} \cup \mathcal{E}^c_{1,2,...,k_b} \cup \left( \bigcup_{s_1 < \cdots < s_{k_b} : \{s_1,...,s_{k_b}\} \neq [k_b]} (\mathcal{E}_{s_1,s_2,...,s_{k_b}} \cap \mathcal{E}_{\text{aux}}) \right) \right) \\ &\leq \mathsf{P}(\mathcal{E}^c_{\text{aux}}) + \mathsf{P}(\mathcal{E}^c_{1,2,...,k_b}) + \sum_{s_1 < \cdots < s_{k_b} : \{s_1,...,s_{k_b}\} \neq [k_b]} \mathsf{P}(\mathcal{E}_{s_1,s_2,...,s_{k_b}} \cap \mathcal{E}_{\text{aux}}) \quad (3.51) \end{aligned}$$

where $\mathcal{E}_{\text{aux}} \triangleq \{\hat{\rho} - \|\mathbf{w}\| \in (-\epsilon, \epsilon)\}$.

We now bound the terms in (3.51). First, by the LLN, $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}^c_{\text{aux}}) = 0$. Next, we consider $\mathsf{P}(\mathcal{E}^c_{1,2,...,k_b})$. Note that, for any $\hat{\mathbf{W}} \in \mathcal{Q}$,

$$\begin{aligned} &\frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^{k_b} A_j \hat{\mathbf{W}}_j \right\|^2 = \frac{1}{n} \left\| \sum_{j=1}^{k_b} A_j \mathbf{w}_j + \mathbf{Z} - \sum_{j=1}^{k_b} A_j \hat{\mathbf{W}}_j \right\|^2 \\ &= \frac{1}{n} \sum_{j=1}^{k_b} \sum_{l=1}^{k_b} \sum_{i=1}^{b} (w_{(j-1)b+i} - \hat{W}_{(j-1)b+i})(w_{(l-1)b+i} - \hat{W}_{(l-1)b+i}) \mathbf{A}^\intercal_{(j-1)b+i} \mathbf{A}_{(l-1)b+i} \\ &+ \frac{2}{n} \sum_{j=1}^{k_b} \sum_{i=1}^{b} (w_{(j-1)b+i} - \hat{W}_{(j-1)b+i}) \mathbf{A}^\intercal_{(j-1)b+i} \mathbf{Z} + \frac{1}{n} \|\mathbf{Z}\|^2. \quad (3.52) \end{aligned}$$

By applying the weak law of large numbers (LLN) to each term in (3.52) and using Lemma 2-1), we have

$$\lim_{m\to\infty} \mathsf{P}\left( \exists \hat{\mathbf{W}} \in \mathcal{Q} \text{ s.t. } \frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^{k_b} A_j \hat{\mathbf{W}}_j \right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \right) = 1$$

which implies that $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}^c_{1,2,...,k_b}) = 0$.

Next, we consider $\mathsf{P}(\mathcal{E}_{s_1,s_2,...,s_{k_b}} \cap \mathcal{E}_{\text{aux}})$ for $\{s_1, s_2, ..., s_{k_b}\} \neq [k_b]$. For notational simplicity, define $\xi \triangleq \sigma_z^2 + \epsilon^2\sigma_a^2$, $\mathcal{T} \triangleq \{s_1, s_2, ..., s_{k_b}\} \cap [k_b]$, and $\mathcal{T}^c \triangleq \{s_1, s_2, ..., s_{k_b}\}\backslash\mathcal{T}$. For any permutation $(s'_1, s'_2, ..., s'_{k_b})$ of $\{s_1, s_2, ..., s_{k_b}\}$ and any $\hat{\mathbf{W}} \in \mathcal{Q}$,

$$
\begin{aligned}
&\mathsf{P}\left(\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j}\hat{\mathbf{W}}_j\right\|^2 \leq \xi \middle| \mathcal{E}_{\text{aux}}\right) \\
&= \mathsf{P}\left(\frac{1}{n}\left\|\sum_{j=1}^{k_b}\sum_{i=1}^{b} w_{(j-1)b+i}\mathbf{A}_{(j-1)b+i} + \mathbf{Z} - \sum_{j=1}^{k_b}\sum_{i=1}^{b} \hat{W}_{(j-1)b+i}\mathbf{A}_{s'_{(j-1)b+i}}\right\|^2 \leq \xi \middle| \mathcal{E}_{\text{aux}}\right) \\
&= \mathsf{P}\left(\frac{1}{n}\left\|\sum_{j\in[k_b]\backslash\mathcal{T}}\sum_{i=1}^{b} w_{(j-1)b+i}\mathbf{A}_{(j-1)b+i}\right.\right. \\
&- \left(\sum_{s'_j\in\mathcal{T}^c}\sum_{i=1}^{b} \hat{W}_{(j-1)b+i}\mathbf{A}_{s'_{(j-1)b+i}} + \sum_{s'_j\in\mathcal{T}}\sum_{i=1}^{b} (\hat{W}_{(j-1)b+i} - w_{s'_{(j-1)b+i}})\mathbf{A}_{s'_{(j-1)b+i}}\right) \\
&\left.\left. + \mathbf{Z}\right\|^2 \leq \xi \middle| \mathcal{E}_{\text{aux}}\right).
\end{aligned} \tag{3.53}
$$

Define $\mathcal{T}' \triangleq [k_b]\backslash\mathcal{T}$. Consider the $|\mathcal{T}'|$-sender CMMAC

$$
Y_{\mathcal{T}'} = \sum_{j\in\mathcal{T}'}\sum_{i=1}^{b} w_{(j-1)b+i}V_{(j-1)b+i} + Z. \tag{3.54}
$$

Thus, $V_l$ $(l = 1, \ldots, k)$ are i.i.d. according to $\mathcal{P}$ and $Z \sim \mathcal{N}(0, \sigma_z^2)$. Define

$$
V_{\mathcal{T}'} \triangleq \sum_{j\in\mathcal{T}'}\sum_{i=1}^{b} w_{(j-1)b+i}V_{(j-1)b+i}. \tag{3.55}
$$

Let $\mathcal{V}_{\mathcal{T}'}$ denote the set of all possible values of $V_{\mathcal{T}'}$, which depends on $\mathcal{P}$. Let $Q(v)$, $Q(y)$, $Q(y|v)$, and $Q(v, y)$ denote the marginal distributions, the conditional distribution, and the joint distribution of $V_{\mathcal{T}'}$ and $Y_{\mathcal{T}'}$, respectively. Note that $Y^n_{\mathcal{T}'}$ corresponds to

$$
\sum_{j\in[k_b]\backslash\mathcal{T}}\sum_{i=1}^{b} w_{(j-1)b+i}\mathbf{A}_{(j-1)b+i} + \mathbf{Z} \tag{3.56}
$$

in (3.53). Also define

$$V_{\hat{W}} \triangleq \sum_{s'_j \in \mathcal{T}^c} \sum_{i=1}^{b} \hat{W}_{(j-1)b+i} V_{s'_{(j-1)b+i}} + \sum_{s'_j \in \mathcal{T}} \sum_{i=1}^{b} (\hat{W}_{(j-1)b+i} - w_{s'_{(j-1)b+i}}) V_{s'_{(j-1)b+i}} \quad (3.57)$$

where $V_l$ are i.i.d. $\sim \mathcal{P}$. Let $Q_{\hat{W}}(v)$ denote the distribution of $V_{\hat{W}}$. Note that $V^n_{\mathcal{T}'}$ and $V^n_{\hat{W}}$ are independent. Also note that

$$\frac{1}{n}\left\| Y^n_{\mathcal{T}'} - V^n_{\hat{W}} \right\|^2 = \frac{\log Q(y^n_{\mathcal{T}'} | v^n_{\hat{W}})}{\log e} + c. \quad (3.58)$$

for some constant $c$ since $Y^n_{\mathcal{T}'}$ is the output sequence of the CMMAC $Y_{\mathcal{T}'} = V_{\mathcal{T}'} + Z$. Thus, the probability in (3.53) is bounded by the probability that $(V^n_{\hat{W}},\ Y^n_{\mathcal{T}'}) \sim \Pi_{i=1}^n Q_{\hat{W}}(v_i) Q(y_i)$ are jointly $\epsilon'$-typical with respect to $Q(v, y)$, where $\epsilon'$ depends only on $\epsilon$ and tends to zero as $\epsilon \to 0$. By applying Sanov's theorem [91, Section 11.4, Theorem 11.4.1], we have

$$\mathsf{P}\left( \frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j} \hat{\mathbf{W}}_j \right\|^2 \le \xi \middle| \mathcal{E}_{\text{aux}} \right) \le 2^{-n\{D(p(v,y)^* || Q_{\hat{W}}(v) Q(y)) - \delta_n\}}, \quad (3.59)$$

where $D(p||q)$ denotes the relative entropy [91] of $p$ with respect to $q$ (also known as the Kullback-Leibler divergence), $\delta_n$ tends to zero as $n \to \infty$,

$$p(v, y)^* = \arg \min_{p(v,y) \in \mathcal{D}^{\epsilon'}} D(p(v, y) || Q_{\hat{W}}(v) Q(y)), \quad (3.60)$$

and

$$\mathcal{D}^{\epsilon'} = \{ p(v, y) : |p(v, y) - Q(v, y)| \le \epsilon' Q(v, y),\ \forall (v, y) \in \mathcal{U}_{\mathcal{T}'} \times \mathbb{R} \}.$$

Since $\min_{p(v,y) \in \mathcal{D}^{\epsilon'}} D(p(v, y) || Q_{\hat{W}}(v) Q(y))$ increases as $\epsilon'$ decreases and $\lim_{\epsilon' \to 0} \mathcal{D}^{\epsilon'} = Q(v, y)$, we have

$$\begin{aligned} \mathsf{P}\left( \frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j} \hat{\mathbf{W}}_j \right\|^2 \le \xi \middle| \mathcal{E}_{\text{aux}} \right) &\le 2^{-n\{D(p(v,y)^* || Q_{\hat{W}}(v) Q(y)) - \delta_n\}} \\ &\le 2^{-n\{D(Q(v,y) || Q_{\hat{W}}(v) Q(y)) - \delta_{\epsilon'} - \delta_n\}} \\ &\le 2^{-n\{I(V_{\mathcal{T}'}; Y_{\mathcal{T}'}) - \delta_{\epsilon'} - \delta_n\}}, \end{aligned} \quad (3.61)$$

where $\delta_{\epsilon'}$ tends to zero as $\epsilon' \to 0$, $I(V_{\mathcal{T}'}; Y_{\mathcal{T}^c})$ is the input–output mutual information of the AWGN channel $Y_{\mathcal{T}'} = V_{\mathcal{T}'} + Z$ (or the maximum achievable sum-rate

of the $|\mathcal{T}'|$-sender CMMAC (3.54)), and the last inequality follows from the fact that

$$D(Q(v,y)||Q_{\hat{W}}(v)Q(y)) = I(V_{\mathcal{T}'};Y_{\mathcal{T}'}) + D(Q(v)||Q_{\hat{W}}(v)) \tag{3.62}$$

and $D(Q(v)||Q_{\hat{W}}(v)) \geq 0$.

Hence, by the union of events bound, for $\{s_1, s_2, ..., s_{k_b}\} \neq [k_b]$,

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_{s_1,s_2,...,s_{k_b}}|\mathcal{E}_{\text{aux}}) \\
&\leq \sum_{\{s'_1,...,s'_{k_b}\}=\{s_1,...,s_{k_b}\}} \mathsf{P}\left(\exists \hat{\mathbf{W}} \in \mathcal{Q} \text{ s.t. } \frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j}\hat{\mathbf{W}}_j\right\|^2 \leq \xi \middle| \mathcal{E}_{\text{aux}}\right) \\
&\leq \sum_{\{s'_1,...,s'_{k_b}\}=\{s_1,...,s_{k_b}\}} \sum_{\hat{\mathbf{W}}\in\mathcal{Q}} \mathsf{P}\left(\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k_b} A_{s'_j}\hat{\mathbf{W}}_j\right\|^2 \leq \xi \middle| \mathcal{E}_{\text{aux}}\right) \\
&\leq k_b! \cdot |\mathcal{Q}| \cdot 2^{-n\{I(V_{\mathcal{T}'};Y_{\mathcal{T}'})-\delta_{\epsilon'}-\delta_n\}}.
\end{aligned}$$

Furthermore, conditioned on $\mathcal{E}_{\text{aux}}$, $\hat{\rho} < \|\mathbf{w}\| + \epsilon$ and hence $|\mathcal{Q}| \leq q(\|\mathbf{w}\| + \epsilon, \epsilon)$ by Lemma 2-2). Thus,

$$\mathsf{P}(\mathcal{E}_{s_1,s_2,...,s_k} \cap \mathcal{E}_{\text{aux}}) \leq k_b! \cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \cdot \leq 2^{-n\{I(V_{\mathcal{T}'};Y_{\mathcal{T}'})-\delta_{\epsilon'}-\delta_n\}}. \tag{3.63}$$

Note that the probability upper-bound (3.63) depends on $s_1, ..., s_{k_b}$ only through $\mathcal{T}$. Grouping the $\binom{m_b-k_b}{k_b-|\mathcal{T}|}$ events $\{\mathcal{E}_{s_1,s_2,...,s_{k_b}} \cap \mathcal{E}_{\text{aux}}\}$ with the same $\mathcal{T}$,

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}) \\
&\leq \mathsf{P}(\mathcal{E}^c_{\text{aux}}) + \mathsf{P}(\mathcal{E}^c_{1,2,...,k_b}) + \sum_{\mathcal{T}\subset[k_b]} \binom{m_b - k_b}{k_b - |\mathcal{T}|} \\
&\qquad \cdot k_b! \cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \cdot 2^{-n\{I(V_{\mathcal{T}'};Y_{\mathcal{T}'})-\delta_{\epsilon'}-\delta_n\}} \\
&\leq \mathsf{P}(\mathcal{E}^c_{\text{aux}}) + \mathsf{P}(\mathcal{E}^c_{1,2,...,k_b}) + k_b! \cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \\
&\qquad \cdot \sum_{\mathcal{T}\subset[k_b]} 2^{(k_b-|\mathcal{T}|)\log m_b} \cdot 2^{-n\{I(V_{\mathcal{T}'};Y_{\mathcal{T}'})-\delta_{\epsilon'}-\delta_n\}} \\
&= \mathsf{P}(\mathcal{E}^c_{\text{aux}}) + \mathsf{P}(\mathcal{E}^c_{1,2,...,k_b}) + k_b! \cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \\
&\qquad \cdot \sum_{\mathcal{T}\subseteq[k_b]} 2^{|\mathcal{T}|\log m_b} \cdot 2^{-n\{I(V_{\mathcal{T}};Y_{\mathcal{T}})-\delta_{\epsilon'}-\delta_n\}}
\end{aligned}$$

which tends to zero as $m_b \to \infty$, if

$$\limsup_{m_b\to\infty} \frac{\log m_b}{n_{m_b}} < \frac{1}{|\mathcal{T}|} 2^{-n\{I(V_{\mathcal{T}};Y_{\mathcal{T}})-\delta_{\epsilon'}\}} \tag{3.64}$$

for all $\mathcal{T} \subseteq [k_b]$. Since $\epsilon > 0$ is arbitrarily chosen, the proof is complete.

## 3.6 Proof of Theorems 2 and 4

Since Theorem 4 is a generalized result of Theorem 2, we only present the proof of Theorem 4. We show that if

$$\lim_{m\to\infty} \overline{P}_e^{(m_b)} = 0 \tag{3.65}$$

where $\overline{P}_e^{(m_b)} = \mathsf{P}\{d^{(m_b)}(A\mathbf{X}+\mathbf{Z}) \neq \text{supp}(\mathbf{X}(\mathbf{w},\mathbf{S},b))\}$, then

$$\limsup_{m\to\infty} \frac{\log m}{n_m} \leq C_{\text{sym}}^{(b)}(\mathcal{P},\mathbf{w}). \tag{3.66}$$

The main techniques to prove the above statement include Fano's inequality and the properties of entropy.

For any $\mathcal{T} \subseteq [k_b]$, denote the tuple of random variables $(S_l : l \in \mathcal{T})$ by $S(\mathcal{T})$. From Fano's inequality [91], we have

$$\begin{aligned} H(S(\mathcal{T})|\mathbf{Y},A) &\leq H(S_1,...,S_{k_b}|\mathbf{Y},A) \\ &\leq \log k_b! + H(\{S_1,...,S_{k_b}\}|\mathbf{Y},A) \\ &\leq \log k_b! + \overline{P}_e^{(m_b)} \log \binom{m_b}{k_b} + 1. \end{aligned} \tag{3.67}$$

On the other hand, by a basic permutation argument,

$$H(S(\mathcal{T})|S(\mathcal{T}^c),A) = \log\left(\prod_{q=0}^{|\mathcal{T}|-1}(m_b-(k_b-|\mathcal{T}|)-q)\right) = |\mathcal{T}|\log m_b - n\epsilon_{1,n} \tag{3.68}$$

where $\mathcal{T}^c \triangleq [k_b]\backslash\mathcal{T}$ and

$$\epsilon_{1,n} \triangleq \frac{1}{n}\log\left(m_b^{|\mathcal{T}|}/\prod_{q=0}^{|\mathcal{T}|-1}(m_b-(k_b-|\mathcal{T}|)-q)\right) \tag{3.69}$$

which tends to zero as $n \to \infty$. Hence, combining (3.67) and (3.68), we have

$$\begin{aligned}
|\mathcal{T}| \log m_b &= H(S(\mathcal{T})|S(\mathcal{T}^c), A) + n\epsilon_{1,n} \\
&= I(S(\mathcal{T}); \mathbf{Y}|S(\mathcal{T}^c), A) + H(S(\mathcal{T})|\mathbf{Y}, S(\mathcal{T}^c), A) + n\epsilon_{1,n} \\
&\leq I(S(\mathcal{T}); \mathbf{Y}|S(\mathcal{T}^c), A) + H(S(\mathcal{T})|\mathbf{Y}, A) + n\epsilon_{1,n} \qquad (3.70) \\
&\leq I(S(\mathcal{T}); \mathbf{Y}|S(\mathcal{T}^c), A) + \log k_b! + \overline{P}_e^{(m_b)} \log \binom{m_b}{k_b} + 1 + n\epsilon_{1,n} \\
&= \sum_{i=1}^{n} I(Y_i; S(\mathcal{T})|Y_1^{i-1}, S(\mathcal{T}^c), A) + \log k_b! + \overline{P}_e^{(m_b)} \log \binom{m_b}{k_b} + 1 + n\epsilon_{1,n} \qquad (3.71) \\
&\leq nI_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) + \log k_b! + \overline{P}_e^{(m_b)} \log \binom{m_b}{k_b} + 1 + n\epsilon_{1,n} \qquad (3.72)
\end{aligned}$$

where (3.70) follows the fact that conditioning reduces entropy, (3.71) follows the chain rule of mutual information [91], and $I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w})$ is defined in (3.37). Therefore,

$$\limsup_{m_b \to \infty} \frac{\log m_b}{n_m} - \frac{\log k_b! + \overline{P}_e^{(m_b)} \log \binom{m_b}{k_b} + 1 + n_m\epsilon_{1,n_m}}{|\mathcal{T}|n_m} \leq \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w})$$

for all $\mathcal{T} \subseteq [k_b]$. Due to the fact that $\log \binom{m_b}{k} \leq k \log m_b$, we have

$$\limsup_{m_b \to \infty} \frac{(1 - k_b\overline{P}_e^{(m_b)}/|\mathcal{T}|) \log m_b}{n_m} - \frac{\log k_b! + n_m\epsilon_{1,n_m} + 1}{|\mathcal{T}|n_m} \leq \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w})$$

for all $\mathcal{T} \subseteq [k_b]$. Since $\lim_{m_b \to \infty} \overline{P}_e^{(m_b)} = 0$, we reach the conclusion

$$\limsup_{m \to \infty} \frac{\log m}{n_m} \leq \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w})$$

for all $\mathcal{T} \subseteq [k_b]$, which completes the proof.

## 3.7 Proof of Property 1

It can be easily shown that if $\|\mathbf{w}_1\|^2 = \cdots = \|\mathbf{w}_k\|^2$, then

$$C_{\text{sym}}^{(b)}(\mathcal{P}, \mathbf{w}) = \min_{\mathcal{T} \subseteq [k_b]} \left[ \frac{1}{|\mathcal{T}|} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) \right] = \frac{1}{k_b} I_{\mathcal{T}}^{(b)}(\mathcal{P}, \mathbf{w}) \qquad (3.73)$$

with $\mathcal{T} = [k_b]$. Therefore, we have $\eta_i = k_{b_i}/\log\,(1 + k_i\gamma)$, for $i = 1, 2$. We can easily derive the result in the high and the low SNR regimes, respectively.

## 3.8 Proof of Property 2

$$\eta_1 = \frac{1}{C_{\text{sym}}(\mathcal{P}, \mathbf{w})} = \frac{1}{\min_{\mathcal{T}\subseteq[k]}\left[\frac{1}{|\mathcal{T}|} I_{\mathcal{T}}(\mathcal{P}, \mathbf{w})\right]} \tag{3.74}$$

and

$$\eta_2 = \frac{1}{C^{(b)}_{\text{sym}}(\mathcal{P}, \mathbf{w})} = \frac{1}{\min_{\mathcal{T}\subseteq[k_b]}\left[\frac{1}{|\mathcal{T}|} I^{(b)}_{\mathcal{T}}(\mathcal{P}, \mathbf{w})\right]}. \tag{3.75}$$

Define $\mathcal{S}_1 \triangleq \{\frac{1}{|\mathcal{T}|} I_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) : \mathcal{T} \subseteq [k]\}$ and $\mathcal{S}_2 \triangleq \{\frac{1}{|\mathcal{T}|} I^{(b)}_{\mathcal{T}}(\mathcal{P}, \mathbf{w}) : \mathcal{T} \subseteq [k_b]\}$. If $a \in \mathcal{S}_2$, then $b \cdot a \in \mathcal{S}_1$. Using the inequality, $\min[a_1, a_2] \leq 1/b \min[b \cdot a_1, b \cdot a_2, a_3, a_4], \forall a_i \in \mathbb{R}$, we have

$$\eta_2 \leq \frac{1}{b}\eta_1. \tag{3.76}$$

# Chapter 4

# Conclusion

This dissertation investigates the performance of communication and compressed sensing systems with emphasis on non-Gaussian practical distributions for the channel input and the measurement matrix. In particular, we have studied (1) how much is the performance loss from a non-Gaussian signal distribution compared to the optimal Gaussian distribution, especially for the signal distributions commonly used in practice?, (2) what are the conditions under which a non-Gaussian distribution performs close to the Gaussian distribution?, and (3) when the actual signal distribution is non-Gaussian, how much is the performance loss from the algorithm that is optimal in case the signal distribution is Gaussian, compared to the solution optimal for the given distribution? To answer these questions, we have investigated two problems, power allocation over fading channel and support recovery of block-sparse signals.

First, we have considered the problem of power allocation over fading channels with QAM inputs. We have studied the performance of three power allocation schemes: (1) waterfilling (WF) that is optimal when the channel inputs are Gaussian-distributed, (2) mercury/water-filling (MWF) that is optimal in our setting, and (3) uniform power allocation with Thresholding (UPAT) that is a simple and practical solution with reduced feedback requirement. We have shown that WF and the optimal UPAT perform close to the optimal MWF as long as the constellation size is properly chosen so that the performance is not limited by too small constellation size. In addition, we have studied the practical solution, UPAT

in more depth. A constellation size selection rule for the optimal UPAT has been derived with the aim of providing a good compromise between performance and complexity. The proposed rule is: minimize the constellation size while achieving the maximum diversity. We have also analyzed the gain of the optimal UPAT over the uniform power allocation that evenly assigns the total power across the blocks. The gain in average transmit power increases without bound as the number of independent fading blocks, $B$ increases, but rapidly increases in dB scale only in the low $B$ regime. Finally, we have developed a simple method to determine the threshold value for UPAT. Compared to the optimal UPAT, the proposed method significantly reduces the computational complexity with minimal performance loss. We leave it as future work to extend our result to fading distributions other than Rayleigh and to multiple-user scheduling scenarios as in orthogonal frequency division multiple access (OFDMA) systems.

Next, we have studied the asymptotic performance limits of reliably recovering the support of block-sparse signals (including scalar-sparse signals as a special class) through an arbitrarily distributed random measurement matrix in the Gaussian noise scenario. We have developed a new perspective from which the problem of recovering of signal support of block-sparse signals is viewed as a problem of communication over multiple input single output (MISO) multiple access channel (MAC). Based on this connection, we have established an information-theoretic analytical framework to unearth the performance limits in the support recovery of block-sparse signals. The new perspective also leads to the opportunity of leveraging the rich results and insights available in information theory to help understand the performance limits of block-sparse signal recovery. We have also derived sharp sufficient and necessary conditions for asymptotically reliable support recovery in terms of the signal dimension, the number of nonzero blocks, the block size, the number of measurements, the distribution of the random measurement matrix, and signal-to-noise ratio (SNR) of each nonzero block. Based on the result, the loss from a non-Gaussian measurement matrix compared to the Gaussian measurement matrix can be significant when both the alphabet size and the number of nonzero elements are small and SNR is large. Finally, we have identified and discussed

three factors of block-sparse signals that can reduce the number of measurement required for reliable support recovery of block-sparse signals, increased SNR, Reduced effective number of nonzero elements, and diversity. Our approach that the support recovery problem is interpreted as a communication problem can be extended to the problem of support recovery of block-sparse signals with multiple measurement vectors, which can be interpreted as the problem of communication over multiple-input multiple-output (MIMO) MAC.

# Bibliography


[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 2006.

[2] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, 1999.

[3] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, 1997.

[4] A. Lozano, A. M. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with rrbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, July 2006.

[5] D. Dardari, "Ordered subcarrier selection algorithm for OFDM-based high-speed WLANs," *IEEE Trans. Wireless Communi.*, vol. 3, no. 5, Sep. 2004.

[6] H. Kwon and B. D. Rao, "Uniform bit and power allocation with subcarrier selection for coded OFDM systems," in *Proc. IEEE VTC Spring*, 2011.

[7] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Power allocation for block-fading channels with arbitrary input constellations," *IEEE Trans. Wireless Communi.*, vol. 8, no. 5, 2009.

[8] J. T. Aslanis, "Coding for communication channels with memory," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1993.

[9] P. S. Chow, "Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1993.

[10] B. Schein and M. Trott, "Sub-optimal power spectra for colored Gaussian channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 1997.

[11] W. Yu and J. M. Cioffi, "Constant-power waterfilling: performance bound and low-complexity implementation," *IEEE Trans. Communi.*, vol. 54, no. 1, Jan. 2006.

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[13] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol.43, no. 2, pp. 359-378, May 1994.

[14] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, Nov. 1999.

[15] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, 1982.

[16] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, 1992.

[17] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[18] T. F. Wong, "Numerical calculation of symmetric capacity of Rayleigh fading channel with BPSK/QPSK," *IEEE Communi. Letters*, vol. 5, no. 8, 2001.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[20] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, Apr. 2005.

[21] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998.

[22] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[23] F. Tosato and P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," in *Proc. IEEE ICC*, vol. 2, 2002.

[24] A. Guillén i Fàbregas and G. Caire, "Coded modulation in the block-fading channel: coding theorems and code construction," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, 2006.

[25] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Outage exponents of block-fading channels with power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, 2010.

[26] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 2009.

[27] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic Press, 2000.

[28] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, Nov. 2010.

[29] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Proces. Mag.*, Jan. 2013.

[30] 3GPP LTE Physical Layer specifications, *TS 36.211-36.216*.

[31] IEEE Std 802.16m-2011, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Broadband Wireless Access Systems - Amendment 3: Advanced Air Interface, 2011.

[32] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[33] E. J. Candes, "Compressive sampling," *Proc. Int. Congress of Mathematicians*, pp. 1433–1452, 2006.

[34] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted norm minimization algorithm," *IEEE Trans. Sig. Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[35] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *J. Electroencephalog. Clinical Neurophysiol.*, vol. 95, pp. 231-251, 1995.

[36] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," *Proc. ICASSP*, pp. 1885-1888, 1998.

[37] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 83–91, 2008.

[38] S. D. Cabrera and T. W. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 4, pp. 842-851, 1991.

[39] Y. Jin and B. D. Rao, "Algorithms for robust linear regression by exploiting the connection to sparse signal recovery," *Proc. ICASSP*, 2010.

[40] W. C. Chu, *Speech coding algorithms*. Wiley-Interscience, 2003.

[41] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Communications*, vol. 50, pp. 374–377, 2002.

[42] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," *Proc. CISS*, 2008.

[43] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 8, pp. 508-518, 2000.

[44] B. D. Rao and B. Song, "Adaptive filtering algorithms for promoting sparsity," *Proc. ICASSP*, pp. 361–364, 2003.

[45] D. Guo, "Neighbor discovery in ad hoc networks as a compressed sensing problem," *Presented at Information Theory and Application Workshop, UCSD*, 2009.

[46] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig. Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[47] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, 1993.

[48] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[49] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIREV*, vol. 43, no. 1, pp. 129–159, 2001.

[50] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *JMLR*, 2001.

[51] D. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presense of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[52] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[53] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, 2006.

[54] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[55] P. Zhao and B. Yu, "On model selection consistency of lasso," *Journal of Machine Learning Research*, vol. 7, pp. 2541-2563, 2006

[56] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *preprint*, 2006.

[57] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 22312242, Oct. 2004.

[58] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 62366255, Sep. 2011.

[59] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the Reconstruction of Block-Sparse Signals with an Optimal Number of Measurements," *IEEE Trans. Sig. Process.*, Aug. 2009.

[60] Z. Ben-Haim and Y. C. Eldar, "Near-Oracle Performance of Greedy Block-Sparse Estimation Techniques From Noisy Measurements," *IEEE J. Sel. Topics Signal Process.*, Sep. 2011.

[61] Z. Zhang, T. Jung, S. Makeig, B. Rao , "Compressed sensing for energy-efcient wireless telemonitoring of non-invasive fetal ECG via block sparse bayesian learning," *IEEE Trans. on Biomedical Engineering*, vol. 60, no. 2, pp. 300-309, 2013.

[62] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, pp. 14-30, 2001.

[63] D. Wipf and S. Nagarajan, "A unified bayesian framework for MEG/EEG source imaging," *NeuroImage*, pp. 947-966, 2008.

[64] Z. Zhang, T. Jung, S. Makeig, B. Rao , "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *IEEE Trans. on Biomedical Engineering, Special Issue on Health Informatics and Personalized Medicine*, vol. 60, no. 1, pp. 221-224, 2013.

[65] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, Mar. 2009.

[66] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, Apr. 2009.

[67] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, no. 1, pp.37-52, 1967.

[68] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *Electr. Eng. Dept., TechnionIsrael Institute of Technology, Haifa, CCIT Report*, 751 Dec-09, EE Pub no. 1708; arXiv 0912.2495, 2009.

[69] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, Jun. 2008.

[70] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302-5316. 2009.

[71] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Sig. Process.*, Jun. 2010.

[72] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, arXiv:1104.0654v3 arXiv. 0912.2495, 2009.

[73] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Selected Topics in Sig. Process.*, vol. 5, no. 5, pp. 912-926, 2011.

[74] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," to appear in *IEEE Trans. Sig. Process.*.

[75] Z. Tian and G. Giannakis, "Compressed sensing for wideband cognitive radios," *Proc. ICASSP*, pp. 1357-1360, 2007.

[76] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the lasso," Ann. Statist., vol. 34, no. 3, pp. 14361462, 2006.

[77] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, no. 12, pp. 5728–5741, Dec. 2009.

[78] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, Jun. 2010.

[79] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[80] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[81] Y. Jin, Y. Kim, and B. Rao, "Limits on support recovery of sparse signals via multiple-access communication techniques," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, Dec. 2011.

[82] Y. Jin and B. Rao, "Support recovery of sparse signals in the presence of multiple measurement vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, May 2013.

[83] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[84] E. Candés, and M. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Magazine*, Mar. 2008.

[85] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 28452862, Nov. 2001.

[86] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 25582567, Nov. 2002.

[87] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 21972202, Mar. 2003.

[88] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 255261, Jan. 2006.

[89] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," in *Constructive Approximation*. Springer, New York, 2008.

[90] J. Haupt, L. Applebaum, and R. Nowak, "On the Restricted Isometry of deterministically subsampled Fourier matrices," *Information Sciences and Systems (CISS)*, Mar. 2010.

[91] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.

[92] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 1520–1529, 1996.

[93] A. E. Gamal and Y. Kim, *Network Information Theory*. Cambridge, 2013.

[94] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[95] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, 2000.

[96] G. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998.

[97] G. Caire and K. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[98] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[99] E. Jorswieck and H. Boche, "Optimal Transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Sig. Processing*, vol. 52, no. 12, Dec. 2004.

[100] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 23462356, Jun. 2008.

[101] M. Jafari and M. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, Sep. 2011.

[102] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, Jul. 1984.

[103] J. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, Sep. 1987.

[104] Y. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, Jan. 2010.

[105] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Processing*, vol. 53, no. 7, pp. 2477-2488, 2005.

[106] Z. Zhang and B. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," *Proc. ICASSP*, 2010.