國 立 交 通 大 學

電信工程研究所

博 士 論 文

(*N*,*K*)有限存取系統之最佳功率配置準則

Optimal Power Allocation for (*N*,*K*)-limited Access Channels

研 究 生：王士瑋

指導教授：陳伯寧 博士
王忠炫 博士

中 華 民 國 一百零一 年 四 月

(*N*,*K*)有限存取系統之最佳功率配置準則

Optimal Power Allocation for (*N*,*K*)-limited Access Channels

研究生：王士瑋　　Student: Shih-Wei Wang

指導教授：陳伯寧　博士　　Advisors: Dr. Po-Ning Chen

王忠炫　博士　　Dr. Chung-Hsuan Wang

國立交通大學

電信工程研究所

博士論文

A Dissertation

Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in
Communication Engineering
Hsinchu, Taiwan

2012 年 4 月

(*N*,*K*)有限存取系統之最佳功率配置準則

研究生：王士瑋　　指導教授：陳伯寧 博士

王忠炫 博士

國立交通大學電信工程研究所

摘　要

在本篇論文，我們探討一個有限存取的通訊傳輸系統，該系統由 N 個彼此獨立的平行通道所組成，且接收端在收到至少 K 個通道之完整傳輸訊號後才開始進行解碼。因為 N 個傳輸訊號中，允許僅有 K 個傳輸訊號到達接收端，此系統因而被稱為(*N*,*K*)有限存取通訊系統。對於(*N*,*K*)有限存取通訊系統，我們假定系統無法提供接收訊號個數之統計特性，因此該系統之最佳傳輸極限為任意 K 個通道所對應之交互資訊量(mutual information)之最小值；而找尋最佳功率配置使通道容量最大化便成為了一個最小值最大化的問題(max-min problem)。對於任意通道模型之(*N*,*K*)有限存取通訊系統，我們提出了一個系統化的快速搜尋演算法找尋其最佳功率配置，此演算法僅需解至多 K 個簡化後的最佳化問題，即可得到最佳功率配置，因此可大幅降低找尋最佳功率配置所需要之複雜度。此外，當通道模型為相加雜訊時，此演算法之搜尋步驟可進一步簡化成兩階段注水式功率配置。最後藉由兩階段注水式功率配置之概念，我們進一步探討如何判斷任意通道之『雜訊程度』(degree of noisiness)，並得到『雜訊程度』之定義應為交互資訊量之微分再取反函數所形成之複合函數。

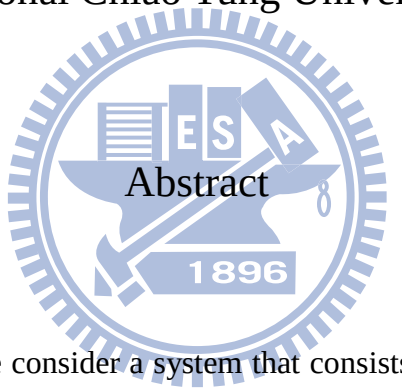
Optimal Power Allocation for (*N*,*K*)-limited Access Channels

Student：Shih-Wei Wang　　Advisors: Dr. Po-Ning Chen

Dr. Chung-Hsuan Wang

Institute of Communication Engineering

National Chiao Tung University



Abstract

In this dissertation, we consider a system that consists of *N* independent parallel channels, where the receiver starts to decode the information being transmitted when it has access to at least *K* of them. We refer to this system as the (*N*,*K*)-limited access channel. No prior knowledge for the distribution about which transmissions will be received is assumed. In addition, both the channel inputs and channel disturbances can be arbitrary, except that the mutual information function for each channel is assumed strictly concave with respect to the input power. Hence, the channel capacity below which the code rate is guaranteed to be attainable by a sequence of codes with vanishing error can be determined by the minimum mutual information among any *K* out of *N* channels. We then investigate the power allocation that maximizes this minimum mutual information subject to a total power constraint. As a result, the

optimal solution can be determined via a systematic algorithmic procedure by performing at most K single-power-sum-constrained maximizations. Based on this result, the close-form formula of the optimal power allocation for an (N,K)-limited access channel with channel inputs and additive noises respectively scaled from two independent and identically distributed random vectors of length N is subsequently established, and is shown to be well interpreted by a *two-phase water-filling* principle. Specifically, in the first *noise-power re-distribution phase*, the least N-K noise powers (equivalently, second moments) are first poured (as *noise water*) into a tank consisting of K interconnected unit-width vessels with solid base heights respectively equal to the remaining K largest noise powers. Afterwards those W vessels either with noise water inside or with solid base height equal to the new water surface level are subdivided into N-K+W vessels of rectangular shape with the same heights (as the water surface level) and widths in proportion to their noise powers. In the second *signal-power allocation phase*, the heights of vessel bases will be first either *lifted* or *lowered* according to the total signal power and channel mutual information functions, followed by the usual signal-power water-filling scheme. The two-phase water-filling interpretation then hints that the degree of "noisiness" for a general (possibly, non-additive and non-Gaussian) limited access channel might be identified by composing the derivative of the mutual information function with its inverse.

Acknowledgements

I would like to thank my advisor, Prof. Po-Ning Chen. Without his help and guidance, this thesis would not be completed. His strong insights in information theory leads me in a right way for my research, and his copious experiences teach me how to do research. I have learned many important things from him, not only in doing research but also in life. However, I am still sure I will never reach his art of doing research. I will follow the philosophies learned from him to face the challenges in my life. I am very lucky to be your student; thank you so much, Po-Ning!

I also would like to thank my co-advisor, Prof. Chung-Hsuan Wang. His many comments and suggestions in research and writing help me extremely. I will keep the suggestions provided by him in my mind.

Contents

List of Figures

Chapter 1

Introduction

A fundamental issue in multiple access channels is power allocation under a total power constraint. In the literature, the best known result in this subject is perhaps the water-filling power allocation principle obtained by maximizing the capacity of parallel additive white Gaussian noise (AWGN) channels [10]. An extension to additive color Gaussian noise channels has later been studied and was found to also follow the water-filling principle over the color spectra of the noises [8]. Recently, by characterizing the relationship between mutual information and minimum mean square error (MMSE) [11], the optimal power allocation for parallel AWGN channels with arbitrary input (possibly finite) has been established, resulting in a new graphical power allocation interpretation called the mercury/water-filling principle [18]. In light of this new finding, the optimal power allocations respectively for multi-user downlink orthogonal frequency-division multiplexing (OFDM) channels [19] and multiple-input-multiple-output (MIMO) channels [23] with arbitrary inputs in the presence of additive white Gaussian background noises are subsequently obtained and found to follow variations of mercury/water-filling principle.

Instead of assuming complete knowledge on channel statistics, a channel could have a number of states with unknown distribution. These channels are classified as *compound channels* as they are compounds of channels parameterized by their states [9], [14], [34].

Since the channel state of a compound channel is only known to be an element of some given set, its capacity below which the code rate is guaranteed to be attainable by a sequence of codes with vanishing error is then determined by the minimum mutual information among all stated channels. Different sets of channel states have been considered in the literature, and their respective optimal power allocations that maximize the minimum mutual information have been derived.

In [21], the states for an MIMO Gaussian compound channel are controlled by the fading parameter within an "isotropic" set, and the optimal power allocation that maximizes the minimum mutual information with respect to Gaussian inputs is shown to be uniform. In [31] and [33], the channel states for multiple-input-single-output (MISO) and multiple-input-multiple-output (MIMO) Gaussian compound channels are parameterized again by the channel fading but are now "ellipsoid" in nature, and the optimal strategy for power allocation becomes beamforming for Gaussian inputs. In [22], the authors model the channel state as the phase of the fading parameter in an MIMO Gaussian compound channel, and obtain that the covariance matrix of the Gaussian input that maximizes the capacity is diagonal. In [5], the channel capacity of MIMO Gaussian compound channels with partially known distribution in channel matrix is investigated. In [32], by considering a parallel Gaussian compound channel where the channel states are determined by the amplitudes of fading parameters, the power allocation that achieves a capacity lower bound obtained via Lagrange duality is proposed. When arbitrary inputs rather than Gaussian ones are considered for these existing results over compound channels, the new finding in [18] may lead to interesting extensions.

In this dissertation, we consider a compound channel with the channel state being a binary vector of length N. Although additive Gaussian noises are appropriate models for general physical channels, and thus are commonly assumed in the power allocation literature

(e.g., [18,19], [21–23,31–33]), experimental measurements in certain environments show that the ambient noise may be non-Gaussian distributed. These environments include indoor and urban radio channels [3] [27], underwater communication systems [20], power line channels [36] and digital subscriber lines [7]. We therefore assume that the channel disturbances can be arbitrary, not necessarily additive or Gaussian, and hence the results of the above literatures based on Gaussian compound channels cannot be applicable to our channel.

The channel states that we consider are decided according to whether or not the transmission signals can reach the receiver end. A straightforward scenario for this state model is a packet switched network, where packets can be lost during transmission [1]. In a highly mobile system, however, the transmission signals can also be missed by a moving mobile terminal. In certain situations, the receiver may still be required to recover the transmitted information from its partial receptions [15, 24, 25]. This raises the question of what the optimal power allocation principle will be for a compound channel with arbitrary input and partially delivered receptions. Notably, since the set of channel states we consider is no longer convex, the traditional techniques [5, 21, 22, 31–33] used to solve the power allocation problems based on a convex channel state space in compound channels cannot be applied and an alternative approach should be taken.

Specifically, among N individual transmissions, possibly parallelly or temporally, we assume that the receiver will begin to recover the information being transmitted when it has access to at least K of them. Since we assume the channel disturbances can be either non-additive or non-Gaussian, to find the optimal power allocation principle for this compound channel seems tricky. We then find that if the mutual information satisfies a certain concavity condition (cf. Assumption 1 in Chapter 2), the optimal power allocation can be obtained *algorithmically* by solving at most K Lagrange-multiplier maximizations (see Theorem 2). To demonstrate the value of the proposed algorithm in complexity reduction, comparison

between the proposed algorithm and a representative brute force method is discussed afterwards. Then, following the proposed algorithm, we further establish that when channel disturbances, in addition to independence, are reduced to being additive with distributions scaled from a common random variable, the optimal power allocation can be directly obtained from a *two-phase water-filling* process if the arbitrary channel inputs are given by the respective component variables in an independent and identically distributed (i.i.d.) random vector, multiplying by the square root of the allocated power. The two-phase water-filling interpretation then hints that the degree of "noisiness" for a general (possibly, non-additive and non-Gaussian) limited access compound channel might be identified by composing the derivative of the mutual information function with its inverse.

The rest of the dissertation is structured as follows. In Chapter 2, we introduce the channel model of the (N, K)-limited access channel considered in this paper as well as the corresponding channel capacity formula. Chapter 3 presents discussion regarding the properties of the optimal power allocation and the algorithm that determines the optimal power allocation. In Chapter 4, we simplify the channel model by further assuming that the channel inputs and additive noises are scaled respectively from two i.i.d. random vectors, which results in a *two-phase water-filling* graphical interpretation for optimal power allocation. In Chapter 5, following the notion of the *two-phase water-filling* interpretation, the degree of "noisiness" for a general limited access channel as well as the optimal power allocation in low- and high- power regimes are addressed. In Chapter 6, we conclude the dissertation and note some possible extensions.

Chapter 2

System Model for an (N, K)-Limited Access Channel

As shown in Figure 2.1, we consider a system that consists of N parallel channels with unit-power inputs adapted according to $\sqrt{\boldsymbol{p}} \triangleq [\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_N}]^{\mathtt{T}}$ satisfying $\sum_{i=1}^{N} p_i \leq P$. In this system, only a certain portion of channel outputs are guaranteed to be successfully received at the receiver end. The system however does not *a priori* know which outputs will be blocked or nullified, nor does the system have the knowledge of the statistics of these blockage. We can realize this assumption by introducing a set of auxiliary multiplicative coefficients $s_1, s_2, \ldots, s_N$ to the channel outputs, where the ith channel output is blocked or nullified when being multiplied by $s_i = 0$, and remains when the multiplicative constant s_i is equal to 1. It is assumed that by monitoring the channel activities, the receiver can perfectly tell the value of $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^{\mathtt{T}}$, where superscript "$\mathtt{T}$" is the matrix transpose operation.[1] Furthermore, $\boldsymbol{s}$ will remain unchanged within a codeword transmission period but may vary for different codeword blocks. The receiver will then decode the information based on the receptions $[\boldsymbol{s} \circ \boldsymbol{Y}_1, \boldsymbol{s} \circ \boldsymbol{Y}_2, \ldots, \boldsymbol{s} \circ \boldsymbol{Y}_n]$ if at least K out of N components of vector $\boldsymbol{s}$ are equal to one, where $\boldsymbol{Y}_i \triangleq [Y_{1,i}, Y_{2,i}, \ldots, Y_{N,i}]^{\mathtt{T}}$ are the channel outputs at time instance

[1] It has been remarked in [14, Thm. 1] that for compound discrete memoryless channels, the capacity remains unchanged even if the receiver knows nothing about $\boldsymbol{s}$. Therefore, for the channels considered in [14, Thm. 1], the result in this paper can also be applied without prior knowledge of $\boldsymbol{s}$.

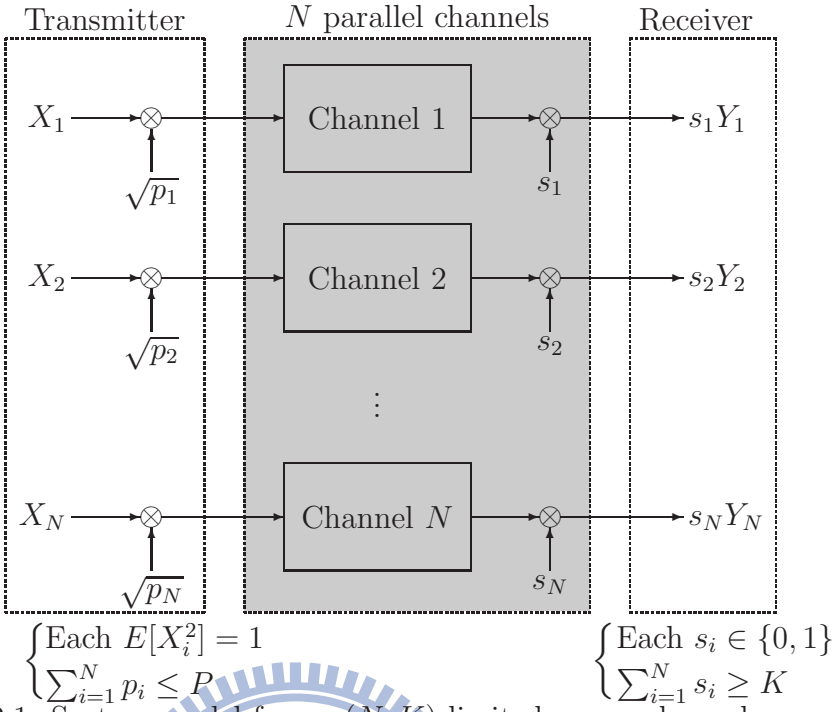


Figure 2.1: System model for an (N, K)-limited access channel.

i, n is the codeword length, and operator "$\circ$" denotes the matrix Hadamard product [17]. Conversely, the receiver will give up the decoding if $\sum_{i=1}^{N} s_i < K$. We thus refer to this channel model as an (N, K)-*limited access channel*.

In this setting, we are interested in the optimal power allocation $\boldsymbol{p}^* = [p_1^*, p_2^*, \ldots, p_N^*]^{\mathsf{T}}$ such that the minimum input-output mutual information subject to $\sum_{i=1}^{N} s_i \geq K$ is maximized. This quantity is generally regarded as the achievable rate under which the decoding error can be made arbitrarily small.[2]

Under the system model, the input-output mutual information can be in principle represented by

$$I(\sqrt{\boldsymbol{p}} \circ \boldsymbol{X}; \boldsymbol{s} \circ \boldsymbol{Y})$$

[2]Our focus in this paper is the decoding error given that $\sum_{i=1}^{N} s_i \geq K$, not the decoding error with respect to a statistically distributed $\boldsymbol{s}$. Note that since the statistics of $\boldsymbol{s}$ is assumed unknown, the latter (i.e., the expected probability of decoding error with respect to $\boldsymbol{s}$) actually cannot be established.

where $I(\cdot;\cdot)$ is the mutual information function and $\sqrt{\boldsymbol{p}} \triangleq [\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_N}]^{\mathsf{T}}$. Here, we overload the notation by denoting the channel output vector corresponding to one channel usage by $\boldsymbol{Y} \triangleq [Y_1, Y_2, \ldots, Y_N]^{\mathsf{T}}$, and likewisely denote the channel input vector for a single channel usage by $\boldsymbol{X} = [X_1, X_2, \ldots, X_N]^{\mathsf{T}}$. The achievable rate that guarantees a vanishing decoding error subject to $\sum_{i=1}^{N} s_i \geq K$ is therefore optimistically

$$\max_{\boldsymbol{X}} \max_{\left\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i\leq P\right\}} \min_{\left\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i\geq K\right\}} I(\sqrt{\boldsymbol{p}}\circ\boldsymbol{X};\boldsymbol{s}\circ\boldsymbol{Y}) \tag{2.1}$$

where $\Re_+$ is the set of nonnegative real numbers. If the parallel channels are independent in the sense that

$$\Pr(\boldsymbol{Y}|\sqrt{\boldsymbol{p}}\circ\boldsymbol{X}) = \prod_{i=1}^{N} \Pr(Y_i|\sqrt{p_i}X_i) \tag{2.2}$$

then the independence bound for mutual information yields that

$$I(\sqrt{\boldsymbol{p}}\circ\boldsymbol{X};\boldsymbol{s}\circ\boldsymbol{Y}) \leq \sum_{i=1}^{N} I(\sqrt{p_i}X_i; s_iY_i) = \sum_{i=1}^{N} s_i \cdot I(\sqrt{p_i}X_i; Y_i)$$

where the last equality follows from s_i being either 1 or 0. We can therefore focus on the optimal power allocation for independent input distributions, if the channel transition probability satisfies (2.2).

We next denote for convenience $f_i(p) \triangleq I(\sqrt{p}X_i; Y_i)$ for $1 \leq i \leq N$, and make the following assumption on these mutual information functions.

Assumption 1. For $1 \leq i \leq N$, $f_i(p)$ is continuous and strictly increasing for $p \geq 0$, and its first derivative, i.e.,

$$f_i'(p) \triangleq \frac{\partial f_i(p)}{\partial p}$$

exists and is continuous and strictly decreasing in $p \geq 0$, where we define $f_i'(0) \triangleq \lim_{p\downarrow 0} f_i'(p)$.[3]

[3]Since the mutual information function $f_i(p)$ is only defined for $p \geq 0$, its derivative at the origin cannot be defined under the usual mathematical principle, i.e., the derivative from the right equal to the derivative from the left. From the aspect of the optimization problem concerned in this work, we adopt $f_i'(0) \triangleq \lim_{p\downarrow 0} f_i'(p)$ as the "derivative" at the origin, specifically when zero power is considered to be allocated to channel i. See, for example, (4.15).

Generally speaking, the channels considered in Assumption 1 are supposed to have more available mutual information when more power is allocated, but the rate of increment is decreasing with respect to the power allotment. There are quite a few practical channels satisfying this assumption, such as antipodal binary-input AWGN channels with hard decision at receiver side, quaternary-input additive Laplace noise channels (cf. Example 1), scalar AWGN channels with arbitrary inputs [18], parallel AWGN channels with given independent inputs [18], and Gaussian fading channels with given inputs [19]. We will adopt Assumption 1 as a premise throughout the entire paper.

Under this assumption, it is clear that $f_i(p)$ is a strictly concave function of p with initial value $f_i(0) = I(0; Y_i) = 0$. Together with the fact that $f_i(p) \geq 0$ for $p \in \Re_+$, we can replace the two inequality constraints in (2.1) by their equality counterparts as

$$\max_{\left\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i\leq P\right\}}\min_{\left\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i\geq K\right\}}\sum_{i=1}^{N} s_i\cdot f_i(p_i) \tag{2.3}$$

$$=\max_{\left\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i= P\right\}}\min_{\left\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i= K\right\}}\sum_{i=1}^{N} s_i\cdot f_i(p_i) \tag{2.4}$$

for a given $\boldsymbol{X}$ that validates Assumption 1. In the next chapter, we will show that under Assumption 1, the maximization-minimization problem in (2.4) becomes algorithmically tractable.

Chapter 3

Analysis of the Optimal Power Allocation

This chapter presents the analysis for the optimization problem in (2.4). For $K = 1$, (2.4) can be simplified to

$$\max_{\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i=P\}} \min\{f_1(p_1), f_2(p_2), \cdots, f_N(p_N)\}.$$

It is thus straightforward that the optimal power allocation $\boldsymbol{p}^*$ satisfies

$$f_1(p_1^*) = f_2(p_2^*) = \cdots = f_N(p_N^*).$$

For $K = N$, the maximization-minimization power allocation problem reduces to a problem that requires only one maximization computation because $s_1 = s_2 = \ldots = s_N = 1$. Therefore, one can apply the Lagrange multipliers technique and Karuch-Kuhn-Tucker (KKT) condition to find the optimal power allocation [4]. However, for $1 < K < N$, a straight technique generally does not exist for this maximization-minimization problem. Nevertheless, we can find a necessary condition for the optimal power allocation such that the labor of examining all possible $\binom{N}{K}$ combinations of $\boldsymbol{s}$ satisfying $\sum_{i=1}^N s_i = K$ can be reduced as indicated in the next lemma.

Lemma 1. The optimal power allocation $\boldsymbol{p}^*$ for an (N, K)-limited access channel, where

$1 \le K \le N$, satisfies

$$f_{a_1}(p^*_{a_1}) \le f_{a_2}(p^*_{a_2}) \le \cdots \le f_{a_K}(p^*_{a_K}) = f_{a_{K+1}}(p^*_{a_{K+1}}) = \cdots = f_{a_N}(p^*_{a_N})$$

for some permutation $a_1, a_2, \ldots, a_N$ of sequence $1, 2, \ldots, N$.

Proof. Since the lemma trivially holds when $K = N$, we assume $K < N$ in the below proof. For the optimal power allocation $\boldsymbol{p}^*$, let $a_1, a_2, \ldots, a_N$ be a permutation of sequence $1, 2, \ldots, N$ satisfying

$$f_{a_1}(p^*_{a_1}) \le f_{a_2}(p^*_{a_2}) \le \cdots \le f_{a_K}(p^*_{a_K}) \le f_{a_{K+1}}(p^*_{a_{K+1}}) \le \cdots \le f_{a_N}(p^*_{a_N}).$$

We then have

$$\max_{\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i=P\}} \min_{\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i=K\}} \sum_{i=1}^{N} s_i \cdot f_i(p_i) = \sum_{i=1}^{K} f_{a_i}(p^*_{a_i}). \tag{3.1}$$

Suppose that there were some $j \ge K$ such that $f_{a_j}(p^*_{a_j}) < f_{a_{j+1}}(p^*_{a_{j+1}})$. Then, we can reduce $p^*_{a_{j+1}}$ down to

$$f^{(\mathtt{inv})}_{a_{j+1}}\left(f_{a_{j+1}}(p^*_{a_{j+1}}) - \delta\right)$$

where $f^{(\mathtt{inv})}_{a_{j+1}}$ is the inverse function[1] of $f_{a_{j+1}}$, and increase $p^*_{a_1}, p^*_{a_2}, \cdots, p^*_{a_j}$ respectively to

$$f^{(\mathtt{inv})}_{a_1}\left(f_{a_1}(p^*_{a_1}) + \Delta\right), \quad f^{(\mathtt{inv})}_{a_2}\left(f_{a_2}(p^*_{a_2}) + \Delta\right), \quad \cdots, \quad f^{(\mathtt{inv})}_{a_j}\left(f_{a_j}(p^*_{a_j}) + \Delta\right)$$

with positive δ and Δ satisfying

$$\sum_{i=1}^{j} f^{(\mathtt{inv})}_{a_i}\left(f_{a_i}(p^*_{a_i}) + \Delta\right) + f^{(\mathtt{inv})}_{a_{j+1}}\left(f_{a_{j+1}}(p^*_{a_{j+1}}) - \delta\right) = \sum_{i=1}^{j+1} p^*_{a_i}$$

and

$$0 < \Delta + \delta \le f_{a_{j+1}}(p^*_{a_{j+1}}) - f_{a_j}(p^*_{a_j}).$$

[1]In this paper, we use $f^{(\mathtt{inv})}$, instead of the usual f^{-1}, to denote the inverse function of f. This is to hopefully provide a clearer notational indication when the inverse of the first derivative f' is additionally required later, which will be denoted by $f'^{(\mathtt{inv})}$ in this work.

Note that the existence, continuity and strict monotonicity of $f_i^{(\mathtt{inv})}$ for $1 \leq i \leq N$ is guaranteed by Assumption 1. The new power assignment will clearly improve (3.1) up to

$$\sum_{i=1}^{K} f_{a_i}(p_{a_i}^*) + K\Delta.$$

A contradiction to the optimality of $\boldsymbol{p}^*$ is thus obtained. □

An immediate implication of Lemma 1 is that we can distinguish the optimal power allocation for an (N, K)-limited access channel into K disjoint cases. In other words, the condition

$$\max_{1\leq i\leq \ell-1} f_{a_i}(p_{a_i}^*) < f_{a_\ell}(p_{a_\ell}^*) = f_{a_{\ell+1}}(p_{a_{\ell+1}}^*) = \cdots = f_{a_N}(p_{a_N}^*) \tag{3.2}$$

is valid for exactly one value of ℓ in $\{1, 2, \ldots, K\}$. As a result, if the index set

$$\mathbb{A} \triangleq \{a_\ell, a_{\ell+1}, \cdots, a_N\}$$

in which their respective mutual information function values are equal to $\max_{1\leq i\leq N} f_i(p_i^*)$ is identified in advance, the maximization-minimization power allocation problem is simplified to a maximization problem as

$$\max_{\boldsymbol{p}\in\mathcal{P}(\mathbb{A})} \left\{ \sum_{i\notin\mathbb{A}} f_i(p_i) + (K + |\mathbb{A}| - N) \max_{1\leq j\leq N} f_j(p_j) \right\} \tag{3.3}$$

where

$$\mathcal{P}(\mathbb{A}) \triangleq \left\{ \boldsymbol{p} \in \Re_+^N : \begin{array}{l} (i)\ \sum_{i=1}^N p_i = P \\ (ii)\ f_i(p_i) < \max_{1\leq j\leq N} f_j(p_j) \text{ for } i \notin \mathbb{A} \\ (iii)\ f_i(p_i) = \max_{1\leq j\leq N} f_j(p_j) \text{ for } i \in \mathbb{A} \end{array} \right\}. \tag{3.4}$$

However, the direct identification of $\mathbb{A}$ without knowing $\boldsymbol{p}^*$ in advance is in general a challenge. The opposite, i.e., identifying $\mathbb{A}$ after determining $\boldsymbol{p}^*$, is more straightforward. In order to resolve the optimization problem, we propose in the following sections to first determine the *best* power allocation $\boldsymbol{p}^\diamond$ corresponding to a *conjectured* maximal-mutual-information index set, denoted by $\mathbb{B}$. Then we examine whether this conjecture is the

optimal one based on conditions we establish later. In case the conjectured $\mathbb{B}$ achieves only a suboptimal power allocation, a new round of maximization computation and follow-up examination will be launched based on a newly generated $\mathbb{B}$. Since the established conditions will help identifying one channel that is not in $\mathbb{A}$ at each round, the process will stop after $N - |\mathbb{A}| + 1$ iterations at which point $\boldsymbol{p}^*$ is obtained.

3.1 Determination of the best power allocation $\boldsymbol{p}^\diamond$ corresponding to a given index set $\mathbb{B}$

Based on a given index set $\mathbb{B}$, we transform the maximization-minimization problem into

$$\sup_{\boldsymbol{p}\in\mathcal{P}(\mathbb{B})} \left\{ \sum_{i\notin\mathbb{B}} f_i(p_i) + (K + |\mathbb{B}| - N) \max_{1\le j\le N} f_j(p_j) \right\} \tag{3.5}$$

where $\mathcal{P}(\mathbb{B})$ is defined the same as (3.4) except that $\mathbb{A}$ is replaced with $\mathbb{B}$. Since the given $\mathbb{B}$ may not be the optimal index set $\mathbb{A}$, the solution $\boldsymbol{p}^\diamond$ of the optimization problem defined in (3.5) could be at the boundary of $\mathcal{P}(\mathbb{B})$ in the sense that

$$f_i(p_i^\diamond) = \max_{1\le j\le N} f_j(p_j^\diamond) \quad \text{for some } i \notin \mathbb{B}.$$

For this reason, we use *supremum* instead of *maximum* in (3.5).

We next show that the third equality constraint in $\mathcal{P}(\mathbb{B})$ can be relaxed by incorporating the *aggregate mutual information function* that transforms the N-dimensional power allocation problem into an equivalent $(N - |\mathbb{B}| + 1)$-dimensional one.

Definition 1. The *aggregate mutual information function* $F_{\mathbb{B}}$ associated with a sequence of mutual information functions $\{f_i\}_{i\in\mathbb{B}}$ is defined through its inverse function[2] as follows:

$$F_{\mathbb{B}}^{(\mathtt{inv})}(y) \triangleq \sum_{i\in\mathbb{B}} f_i^{(\mathtt{inv})}(y) \quad \text{for } y \ge 0 \tag{3.6}$$

[2]For completeness, we define $f_i^{(\mathtt{inv})}(y) = \infty$ for $y \ge \omega_i \triangleq \lim_{p\to\infty} f_i(p)$ and $F_{\mathbb{B}}^{(\mathtt{inv})}(y) = \infty$ if one of $\{f_i^{(\mathtt{inv})}(y)\}_{i\in\mathbb{B}}$ is equal to ∞. Note that the inverse function value $F_{\mathbb{B}}(p)$ of function $F_{\mathbb{B}}^{(\mathtt{inv})}$ is always well defined for every $p \in \Re_+$ because each f_i is assumed to be a strictly increasing function, and $\lim_{p\to\infty} F_{\mathbb{B}}(p) = \min_{i\in\mathbb{B}} \omega_i$.

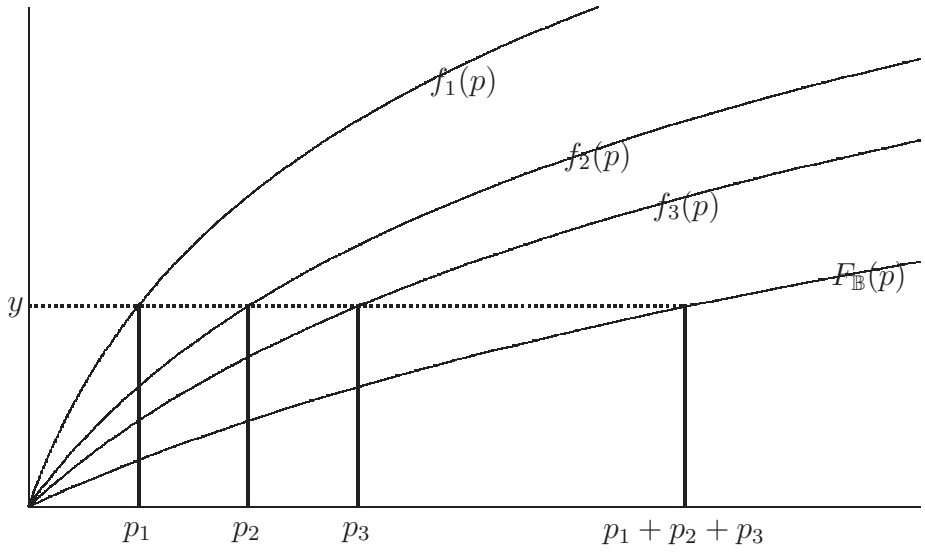


Figure 3.1: Graphical illustration of the aggregate mutual information function when $f_i(p) = \log(1 + p/\sigma_i^2)$ and $\sigma_i^2 = i$ for $i \in \mathbb{B} = \{1, 2, 3\}$.

provided that all the inverse functions exist (which is guaranteed by Assumption 1).

A graphical illustration of the aggregate mutual information function for $\mathbb{B} = \{1, 2, 3\}$ is given in Figure 3.1. In this figure, it is clear that

$$F_{\mathbb{B}}^{(\mathtt{inv})}(y) = f_1^{(\mathtt{inv})}(y) + f_2^{(\mathtt{inv})}(y) + f_3^{(\mathtt{inv})}(y) = p_1 + p_2 + p_3.$$

As a specific example, if $f_i(p) = \log(1 + p/\sigma_i^2)$ for some $\sigma_i^2 > 0$ and $1 \le i \le 3$, then

$$F_{\mathbb{B}}(p) = \log\left(1 + \frac{p}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}\right).$$

In terms of the aggregate mutual information function, we can simplify the constraints in $\mathcal{P}(\mathbb{B})$ in the following lemma, for which the proof is deferred to Appendix A.

Lemma 2. Fix an index set $\mathbb{B}$. The vector $\boldsymbol{p}^\diamond$ that achieves (3.5) satisfies

$$p_i^\diamond = \begin{cases} q_i^\diamond & \text{for } i \notin \mathbb{B} \\ f_i^{(\mathtt{inv})}\left(F_{\mathbb{B}}(q_{\mathbb{B}}^\diamond)\right) & \text{for } i \in \mathbb{B} \end{cases} \tag{3.7}$$

where the $(N-|\mathbb{B}|+1)$-dimensional vector $\boldsymbol{q}^{\diamond}$ achieves:

$$\sup_{\boldsymbol{q}\in\mathcal{Q}(\mathbb{B})}\left\{\sum_{i\notin\mathbb{B}} f_i(q_i)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}})\right\} \tag{3.8}$$

where

$$\mathcal{Q}(\mathbb{B})\triangleq\left\{\boldsymbol{q}=(\text{list of } q_i\ \forall i\notin\mathbb{B},\ q_{\mathbb{B}})\in\Re_+^{N-|\mathbb{B}|+1}:\ \begin{array}{l}(i)\ \sum_{i\notin\mathbb{B}} q_i+q_{\mathbb{B}}=P\\ (ii)\ f_i(q_i)<F_{\mathbb{B}}(q_{\mathbb{B}})\text{ for } i\notin\mathbb{B}\end{array}\right\}.$$

In addition, $\boldsymbol{q}^{\diamond}\in\mathcal{Q}(\mathbb{B})$ if, and only if, $\boldsymbol{p}^{\diamond}\in\mathcal{P}(\mathbb{B})$.

By reducing the number of constraints down to two in $\mathcal{Q}(\mathbb{B})$ in Lemma 2, we can further proceed to show that the inequality constraint in $\mathcal{Q}(\mathbb{B})$ is redundant in case $\boldsymbol{q}^{\diamond}\in\mathcal{Q}(\mathbb{B})$, as summarized in Theorem 1, for which the proof can be found in Appendix A.

Theorem 1. Given that $\boldsymbol{q}^{\diamond}\in\mathcal{Q}(\mathbb{B})$, the maximizer $\boldsymbol{q}^{\diamond}$ for (3.8) is equal to the maximizer $\tilde{\boldsymbol{q}}$ of the problem below:

$$\max_{\boldsymbol{q}\in\tilde{\mathcal{Q}}(\mathbb{B})}\left\{\sum_{i\notin\mathbb{B}} f_i(q_i)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}})\right\} \tag{3.9}$$

where

$$\tilde{\mathcal{Q}}(\mathbb{B})\triangleq\left\{\boldsymbol{q}\in\Re_+^{N-|\mathbb{B}|+1}:\sum_{i\notin\mathbb{B}} q_i+q_{\mathbb{B}}=P\right\}.$$

We summarize the notations we have used thus far as follows. The best power allocations for (3.8) and (3.9) with respect to a given $\mathbb{B}$ are denoted by $\boldsymbol{q}^{\diamond}=\boldsymbol{q}^{\diamond}(\mathbb{B})$ and $\tilde{\boldsymbol{q}}=\tilde{\boldsymbol{q}}(\mathbb{B})$, respectively. For convenience, we drop the dependence on $\mathbb{B}$ in their notational expressions. These two power allocations may not be equal unless $\boldsymbol{q}^{\diamond}\in\mathcal{Q}(\mathbb{B})$. Once $\mathbb{B}$ is taken to be the optimal $\mathbb{A}$ corresponding to the optimal power allocation $\boldsymbol{p}^*$ in the sense of (3.2), $\boldsymbol{p}^*$ can be derived from $\tilde{\boldsymbol{q}}$ (equivalently, $\boldsymbol{q}^{\diamond}$ since $\mathbb{B}=\mathbb{A}$ implies $\boldsymbol{q}^{\diamond}=\tilde{\boldsymbol{q}}$) through an assignment similar to (3.7). Such notational convention will be used throughout the paper. Notably, we will show in the next section that for finding the optimal power allocation $\boldsymbol{p}^*$, only the determination of $\tilde{\boldsymbol{q}}$ is required since the considered $\boldsymbol{q}^{\diamond}$ always belongs to $\mathcal{Q}(\mathbb{B})$. Hence, as

the optimal power allocation $\boldsymbol{p}^*$ is concerned, the computation of a general $\boldsymbol{q}^\diamond$ that may lie outside $\mathcal{Q}(\mathbb{B})$ is not necessary.

We conclude this section by pointing out that the maximization computation in (3.9) is now performed over the usual single power-sum constraint, and hence can be solved by treating $(K + |\mathbb{B}| - N)F_{\mathbb{B}}(\cdot)$ as the mutual information function of an auxiliary aggregate channel. Based on the result in Theorem 1, we are ready to present the algorithmic approach that helps identifying the optimal maximal-mutual-information index set $\mathbb{A}$ and the optimal power allocation $\boldsymbol{p}^*$.

3.2 Determination of the Optimal Maximal-Mutual-Information Index Set $\mathbb{A}$ and the Optimal Power Allocation $\boldsymbol{p}^*$

For an (N, K)-limited access channel, there are possibly $\sum_{\ell=1}^{K} \binom{N}{\ell-1}$ candidate index sets for the choices of $\mathbb{B}$ in Theorem 1, and it may be time-consuming to perform the optimization computation for (3.9) for each of them. The next theorem then shows that this time-consuming maximization labor can be reduced to only $N - |\mathbb{A}| + 1$.

Theorem 2. The optimal maximal-mutual-information index set $\mathbb{A}$ and the optimal power allocation $\boldsymbol{p}^*$ can be obtained through the following algorithmic procedure:

Step 1. Initialize $M = 1$ and $\mathbb{B}_1 = \{1, 2, \ldots, N\}$.

Step 2. Obtain the maximizer $\tilde{\boldsymbol{q}}_M$ for (3.9) by setting $\mathbb{B} = \mathbb{B}_M$, and calculate

$$\tilde{\boldsymbol{p}}_M = [\tilde{p}_{M,1}, \tilde{p}_{M,2}, \ldots, \tilde{p}_{M,N}]^{\mathtt{T}}$$

corresponding to the obtained $\tilde{\boldsymbol{q}}_M$ and the given $\mathbb{B}_M$ through an assignment similar to (3.7).

Step 3. Assign $\mathbb{B}_{M+1} = \mathbb{B}_M \setminus \{j_M\}$, where j_M is an index in $\mathbb{B}_M$ that satisfies

$$f'_{j_M}(\tilde{p}_{M,j_M}) = \min_{i \in \mathbb{B}_M} f'_i(\tilde{p}_{M,i}). \tag{3.10}$$

(If there are more than one index satisfying (3.10), just pick up any one of them as j_M.)

Step 4. If

$$(K - M)F'_{\mathbb{B}_{M+1}}\left(\sum_{i \in \mathbb{B}_{M+1}} \tilde{p}_{M,i}\right) \leq f'_{j_M}(\tilde{p}_{M,j_M}) \tag{3.11}$$

then set $\mathbb{A} = \mathbb{B}_M$ and $\boldsymbol{p}^* = \tilde{\boldsymbol{p}}_M$ and stop the algorithm; otherwise, set $M = M + 1$ and go to Step 2.

Proof. For better readability, we defer the detail of the proof to Appendix B and sketch only the key ideas here.

Following Lemma 2 and Theorem 1, we know that once $\mathbb{B}$ in (3.9) is taken to be $\mathbb{A}$, $\boldsymbol{p}^*$ can be derived from $\tilde{\boldsymbol{q}}$ through a similar assignment to (3.7). Hence, to confirm the proposed algorithm, it suffices to prove that when stop criterion (3.11) is first valid, the corresponding $\mathbb{B}_M$ is indeed equal to $\mathbb{A}$. The proof then requires the verification of the below two claims:

(a) If $|\mathbb{A}| < N$, then stop criterion (3.11) is violated and $\mathbb{A} \subseteq \mathbb{B}_{M+1}$ for $1 \leq M \leq N - |\mathbb{A}|$.

(b) If stop criterion (3.11) is violated for $1 \leq M \leq m$, then $m \leq N - |\mathbb{A}|$.

An immediate consequence of (b) is that if $|\mathbb{A}| = N$, then stop criterion (3.11) must be valid at $M = 1$ (because if stop criterion (3.11) is violated at $M = 1$, we would obtain $|\mathbb{A}| \leq N - 1$ from (b), a contradiction to $|\mathbb{A}| = N$). Hence, $\mathbb{A} = \mathbb{B}_1$ is obtained by the proposed algorithm. So, the proposed algorithm functions correctly when $|\mathbb{A}| = N$.

When $|\mathbb{A}| < N$, according to (a), we have

$$\mathbb{A} \subseteq \mathbb{B}_{N-|\mathbb{A}|+1} \quad (\subseteq \mathbb{B}_{N-|\mathbb{A}|} \subseteq \mathbb{B}_{N-|\mathbb{A}|-1} \subseteq \ldots \subseteq \mathbb{B}_2)$$

and stop criterion (3.11) is violated for every $1 \leq M \leq N-|\mathbb{A}|$. Together with the statement of (b), we obtain that stop criterion (3.11) must hold at $M = N - |\mathbb{A}| + 1$; otherwise, a contradiction as $m = N - |\mathbb{A}| + 1 \leq N - |\mathbb{A}|$ will be obtained from (b). Finally, we note that $\mathbb{A} \subseteq \mathbb{B}_{N-|\mathbb{A}|+1}$ and $|\mathbb{A}| = |\mathbb{B}_{N-|\mathbb{A}|+1}|$ jointly imply $\mathbb{A} = \mathbb{B}_{N-|\mathbb{A}|+1}$ Thus, the proposed algorithm also functions correctly when $|\mathbb{A}| < N$. The proof of Theorem 2 is therefore completed. □

We would like to point out that the algorithm in Theorem 2 will stop when (and usually before) M reaches K, because (3.11) trivially holds when $M = K$. This coincides with the definition of $\mathbb{A}$ in (3.2) that at most $K-1$ indices are outside $\mathbb{A}$. Our algorithm thus requires to solve at most K optimization problems in the form of (3.9).

We note that in general, there are $\sum_{\ell=1}^{K} \binom{N}{\ell-1}$ choices of $\mathbb{B}$ and only one of them is $\mathbb{A}$, and a straightforward method is to examine all of them. In comparison with our algorithm, the computation complexity of such a brute force method will be much higher when N and K are only moderately large. For example, consider an OFDM system, where there are 64 sub-carriers ($N = 64$) and at least 30 sub-carriers are required to be accessible ($K = 30$). The brute force method requires to examine $\sum_{\ell=1}^{K} \binom{N}{\ell-1} \approx 4.9097 \times 10^{18}$ maximizations in the form of (3.9), and yet, our algorithm only needs to consider at most $30 (= K)$ maximizations of the same form. Hence, the complexity reduction by the proposed algorithm is significant in this regard.

Theorem 2 indicates that given the first derivative of the marginal mutual information function $f_i(p) = I(\sqrt{p}X_i; Y_i)$ being positive, strictly decreasing and continuous in p for every $1 \leq i \leq N$ (i.e., Assumption 1), we can determine the optimal power allocation $\boldsymbol{p}^*$ for a spatially independent (N, K)-limited access channel with input $\sqrt{\boldsymbol{p}} \circ \boldsymbol{X}$ by performing $N - |\mathbb{A}| + 1$ maximizations in the sense of (3.9). In the next chapter, we will show that this maximization labor can be further reduced to one if the considered channels are corrupted by additive noises of the same family. Moreover, the resultant optimal power allocation can

be graphically interpreted by a *two-phase water-filling scheme*.

Chapter 4

Optimal Power Allocation over Additive Noise Channels

By additive noises of the same family, we mean that the relationship between channel inputs and outputs can be characterized by

$$Y_i = \sqrt{p_i}X_i + \sigma_i Z_i \quad \text{for } 1 \leq i \leq N \tag{4.1}$$

where $\{X_i\}_{i=1}^N$ and $\{Z_i\}_{i=1}^N$ are both i.i.d. complex random variables with unit second moments, and they are independent from each other. We then restrict our attention only to the case that Z_i is a continuous random variable[1] because Assumption 1 may fail when both X_i and Z_i are discrete. Notably, X_i often takes values in a finite alphabet (e.g., $\{\pm 1\}$) in practice. Specifically, when the intersection of two sets $\left\{\sqrt{p_i}x + \sigma_i z : P_{Z_i}(z) > 0\right\}$ and $\left\{\sqrt{p_i}\tilde{x} + \sigma_i z : P_{Z_i}(z) > 0\right\}$ is empty for every $x \neq \tilde{x}$ with $P_{X_i}(x) > 0$ and $P_{X_i}(\tilde{x}) > 0$, we have

$$f_i(p_i) = I(\sqrt{p_i}X_i; Y_i) = H(\sqrt{p_i}X_i) = H(X_i)$$

where $H(X_i)$ is the entropy of the channel input X_i [8]. This implies that in a discrete system, $f_i(p_i)$ can be equal to its maximum value $H(X_i)$ almost everywhere in p_i, in which case Assumption 1 is unquestionably violated.

[1] By a continuous random variable, we mean that its support can not be made finite or countable.

Observe that for continuous additive noises,

$$\begin{aligned} I(\sqrt{p_i}X_i; Y_i) &= h(Y_i) - h(Y_i|\sqrt{p_i}X_i) \\ &= h(Y_i) - h(\sqrt{p_i}X_i + \sigma_i Z_i|\sqrt{p_i}X_i) \\ &= h(\sigma_i \tilde{Y}_i) - h(\sigma_i Z_i) \qquad (4.2) \\ &= h(\tilde{Y}_i) - h(Z_i) \\ &= I\left(\frac{\sqrt{p_i}}{\sigma_i}X_i; \tilde{Y}_i\right) \end{aligned}$$

where $h(\cdot)$ is the differential entropy function [8], and (4.2) follows from the independence between X_i and Z_i, and $\tilde{Y}_i \triangleq (\sqrt{p_i}/\sigma_i)X_i + Z_i$. This immediately yields

$$f_i(p_i) = g\left(\frac{p_i}{\sigma_i^2}\right) \text{ for every } 1 \leq i \leq N \qquad (4.3)$$

with

$$g(\rho) \triangleq I(\sqrt{\rho}X_i; \sqrt{\rho}X_i + Z_i). \qquad (4.4)$$

Assumption 1 thus reduces to the single condition that function g is continuous and strictly increasing, and its first derivative exists and is continuous and strictly decreasing.

Based on this system setting, we show in the next theorem that the optimal power allocation $\boldsymbol{p}^*$ follows a *two-phase water-filling scheme*. Specifically, in the first phase (which we refer to as the *noise-power re-distribution phase*), the least $N - K$ noise powers among $\{\sigma_i^2\}_{i=1}^N$ will be first poured as *noise water* into a tank consisting of K interconnected vessels with solid base heights equal to the remaining K noise powers and with widths of unit length as shown in Figure 4.1(b). Afterwards those W vessels either with water inside or with solid base height equal to the water surface level will be subdivided into $N - K + W$ vessels of rectangular shape with the same heights (as the water surface level) and with widths in proportion to their noise powers (but the total volume remaining unchanged). As such, a tank with N vessels of proper heights and widths (corresponding to N channels) is ready for the second phase as exemplified in Figure 4.1(c). It is worth mentioning that after the first phase, the optimal maximal-mutual-information index set $\mathbb{A}$ has already been identified and

consists of the channel indices corresponding to the aforementioned W vessels and the least $N-K$ noise powers (hence, $|A| = W + N - K$).

In the second phase (which we refer to as the *signal-power allocation phase*), the heights of vessel bases will be first either *lifted* or possibly *lowered* according to total signal power P and function g as well as their current heights as shown in Figure 4.1(e). What follows, as exemplified in Figure 4.1(f), is the usual water-filling power allocation scheme. The pre-adjustment of base heights before water filling can be viewed as preparation for these vessels to be "capable" of supporting the water that is going to be poured in with amount P. As a result, the volume of water ended up in each vessel is exactly the power that should be allocated. Notably, for the special case that the noises $\{Z_i\}_{i=1}^N$ are complex Gaussian distributed, the heights of vessel bases can never be lowered in the pre-adjustment step; hence, a *mercury-filling* scheme before water pouring has been proposed to materialize the lifting of heights of vessel bases [18]. However, since the adjustment of heights of vessel bases generally can be in both *up* and *down* directions, the use of the name *mercury/water filling* may induce that the vessel bases should be lifted under general non-Gaussian additive noises; hence, we simply use the conventional name of water-filling in this work.

Theorem 3. Suppose that the information transmitted over an (N, K)-limited access channel is corrupted by additive noises of the same family characterized by (4.1), and the mutual information function $g(\rho)$ defined in (4.4) satisfies Assumption 1. Assume without loss of generality that

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_N^2.$$

Then, the optimal maximal-mutual-information index set $\mathbb{A}$ is given by

$$\mathbb{A} = \{\ell, \ell+1, \cdots, N\} \tag{4.5}$$

where

$$\ell \triangleq \min\left\{ i \in \{1, 2, \cdots, K\} \middle| \sigma_i^2 \leq \tilde{\sigma}_K^2 \text{ for every } 1 \leq i \leq K \right\} \tag{4.6}$$

and $\tilde{\sigma}_i^2 \triangleq \sigma_i^2 + [\lambda - \sigma_i^2]^+$ for $1 \le i \le K$ with λ chosen to satisfy $\sum_{i=1}^{K} [\lambda - \sigma_i^2]^+ = \sum_{i=K+1}^{N} \sigma_i^2$, and $[y]^+ \triangleq \max\{0, y\}$. The optimal power allocation $\boldsymbol{p}^*$ can therefore be obtained from $\boldsymbol{q}^*$ through an assignment similar to (3.7), where $\boldsymbol{q}^*$ is the maximizer for (3.9) with $\mathbb{B}$ equal to the above $\mathbb{A}$. In other words,

$$p_i^* = \begin{cases} q_i^* & \text{for } 1 \le i < \ell \\ \frac{\sigma_i^2}{\sum_{j=\ell}^{N} \sigma_j^2} \cdot q_{\mathbb{A}}^* & \text{for } \ell \le i \le N \end{cases} \tag{4.7}$$

with[2]

$$q_i^* = \left\{ \begin{array}{ll} \sigma_i^2 \cdot g'^{(\mathtt{inv})}\left(\nu\, \sigma_i^2\right) & \text{if } g'(\infty) < \nu\sigma_i^2 < g'(0) \\ 0 & \text{if } \nu\sigma_i^2 \ge g'(0) \end{array} \right\} \text{ for } 1 \le i < \ell \tag{4.8}$$

and

$$q_{\mathbb{A}}^* = \left(\sum_{j=\ell}^{N} \sigma_j^2 \right) \cdot g'^{(\mathtt{inv})} \left(\nu\, \frac{\sum_{j=\ell}^{N} \sigma_j^2}{K - \ell + 1} \right) \tag{4.9}$$

where $g'^{(\mathtt{inv})}$ is the inverse function of the first derivative g' of function g, and ν is chosen such that

$$\sum_{i=1}^{\ell-1} q_i^* + q_{\mathbb{A}}^* = P. \tag{4.10}$$

Proof. In terms of (4.3), the determination of j_M in (3.10) can be simplified to

$$\begin{aligned} j_M &= \arg\min_{i \in \mathbb{B}_M} f_i'(\tilde{p}_{M,i}) \\ &= \arg\min_{i \in \mathbb{B}_M} \frac{1}{\sigma_i^2} \cdot g'\left(\frac{\tilde{p}_{M,i}}{\sigma_i^2} \right) \\ &= \arg\min_{i \in \mathbb{B}_M} \frac{1}{\sigma_i^2} \cdot g'\left(g^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})) \right) \qquad (4.11) \\ &= \arg\max_{i \in \mathbb{B}_M} \sigma_i^2 \qquad (4.12) \end{aligned}$$

where (4.11) follows from

$$\tilde{p}_{M,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})) = \sigma_i^2 \cdot g^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})) \quad \text{for } i \in \mathbb{B}_M$$

[2]For notational convenience, we define $g'(\infty) \triangleq \lim_{\rho \uparrow \infty} g'(\rho)$ and note that $g'(\infty) = 0$ for most channels of practical interest such as channels with finite input alphabet. In the specific situation where $g'(\infty) > 0$, we point out that it is still unnecessary to consider the case of $\nu\sigma_i^2 \le g'(\infty)$ in (4.8) because the KKT condition requires $1/\sigma_i^2 \cdot g'(q_i^*/\sigma_i^2) \le \nu$; thus by the strict decreasingness of g', we have that $\nu\sigma_i^2 \ge g'(q_i^*/\sigma_i^2) \ge g'(P/\sigma_i^2) > g'(\infty)$ is always valid for finite total power P.

and (4.12) holds because $g'\left(g^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}))\right)$ is finite due to $\tilde{q}_{\mathbb{B}_M} > 0$ for $1 \le M \le N-|\mathbb{A}|+1$ (cf. (B.4c) and $\tilde{q}_{\mathbb{B}_1} = P$).

Condition (4.12) then gives that for $M = 1, 2, 3, \ldots,$

$$\mathbb{B}_M = \{M, M+1, \ldots, N\}. \tag{4.13}$$

Using (4.3) again simplifies stop criterion (3.11) to

$$(K-M)\sigma_M^2 \le \sum_{i=M+1}^{N} \sigma_i^2 \tag{4.14}$$

because

$$\frac{(K-M)}{f'_{j_M}(\tilde{p}_{M,j_M})} = \frac{(K-M)\sigma_M^2}{g'\left(g^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}))\right)}$$

and

$$\frac{1}{F'_{\mathbb{B}_{M+1}}\left(\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M,i}\right)} = \sum_{i\in\mathbb{B}_{M+1}} \frac{1}{f'_i(\tilde{p}_{M,i})} = \sum_{i\in\mathbb{B}_{M+1}} \frac{\sigma_i^2}{g'\left(g^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}))\right)}.$$

Then by definition of ℓ and the observation that the *noise water* level $\lambda = \tilde{\sigma}_K^2$, we have

$$\sigma_k^2 > \tilde{\sigma}_K^2 = \frac{1}{K-\ell+1}\sum_{j=\ell}^{N} \sigma_j^2 \quad \text{for } 1 \le k < \ell$$

which implies that in the above range of k,

$$(K-k)\sigma_k^2 \ge (K-\ell+1)\sigma_k^2 + \sum_{j=k+1}^{\ell-1} \sigma_j^2 > \sum_{j=\ell}^{N} \sigma_j^2 + \sum_{j=k+1}^{\ell-1} \sigma_j^2 = \sum_{j=k+1}^{N} \sigma_j^2.$$

Accordingly, (4.14) (equivalently, (3.11)) is violated for $1 \le M < \ell$. In addition, it can be verified that

$$\sigma_\ell^2 \le \tilde{\sigma}_K^2 = \frac{1}{K-\ell+1}\sum_{j=\ell}^{N} \sigma_j^2 = \frac{1}{K-\ell+1}\sigma_\ell^2 + \frac{1}{K-\ell+1}\sum_{j=\ell+1}^{N} \sigma_j^2$$

is exactly equivalent to the validity of (4.14) at $M = \ell$. Following the algorithm in Theorem 2, we can conclude from (4.13) that $\mathbb{A} = \{\ell, \ell+1, \ldots, N\}$.

The optimal power allocation $\boldsymbol{q}^*$ as well as its transformation to $\boldsymbol{p}^*$ follows the usual optimization process for (3.9) by setting $\mathbb{B} = \mathbb{A}$. Specifically, we can reduce (3.9) to

$$\begin{aligned}
&\max_{\{\boldsymbol{q}\in\Re_+^\ell:\sum_{i=1}^{\ell-1} q_i+q_{\mathbb{A}}=P\}} \left\{\sum_{i=1}^{\ell-1} f_i(q_i) + (K-\ell+1)F_{\mathbb{A}}(q_{\mathbb{A}})\right\} \\
&= \max_{\{\boldsymbol{q}\in\Re_+^\ell:\sum_{i=1}^{\ell-1} q_i+q_{\mathbb{A}}=P\}} \left\{\sum_{i=1}^{\ell-1} g\left(\frac{q_i}{\sigma_i^2}\right) + (K-\ell+1)g\left(\frac{q_{\mathbb{A}}}{\sum_{i=\ell}^N \sigma_i^2}\right)\right\} \\
&= \max_{\boldsymbol{q}\in\Re_+^\ell} \left\{\sum_{i=1}^{\ell-1} g\left(\frac{q_i}{\sigma_i^2}\right) + (K-\ell+1)g\left(\frac{q_{\mathbb{A}}}{\sum_{i=\ell}^N \sigma_i^2}\right) - \nu\left(\sum_{i=1}^{\ell-1} q_i + q_{\mathbb{A}} - P\right)\right\}
\end{aligned}$$

where in the above derivation, we apply $F_{\mathbb{A}}(p) = g\left(\frac{p}{\sum_{i\in\mathbb{A}}\sigma_i^2}\right)$, and ν is the Lagrange multiplier. Then, the Lagrange multipliers technique and KKT condition give that for $1 \le i < \ell$,

$$\begin{cases} \dfrac{1}{\sigma_i^2} g'\left(\dfrac{q_i^*}{\sigma_i^2}\right) - \nu = 0 & \text{if } q_i^* > 0 \\ \dfrac{1}{\sigma_i^2} g'(0) - \nu \le 0 & \text{if } q_i^* = 0 \end{cases} \tag{4.15}$$

and

$$\frac{K-\ell+1}{\sum_{j=\ell}^N \sigma_j^2} g'\left(\frac{q_{\mathbb{A}}^*}{\sum_{j=\ell}^N \sigma_j^2}\right) - \nu = 0 \quad \text{as } q_{\mathbb{A}}^* = \tilde{q}_{\mathbb{B}_{N-|\mathbb{A}|+1}} > 0 \text{ (see (B.4c))}$$

where ν is chosen to satisfy (4.10). The validity of (4.8) and (4.9) are therefore confirmed. The transformation from $\boldsymbol{q}^*$ to $\boldsymbol{p}^*$ can be derived as:

$$\begin{aligned}
p_i^* &= \begin{cases} q_i^* & \text{for } i \notin \mathbb{A} \\ f_i^{(\mathtt{inv})}\left(F_{\mathbb{A}}(q_{\mathbb{A}}^*)\right) & \text{for } i \in \mathbb{A} \end{cases} \\
&= \begin{cases} q_i^* & \text{for } i \notin \mathbb{A} \\ \sigma_i^2 \cdot g^{(\mathtt{inv})}\left(g\left(\frac{q_{\mathbb{A}}^*}{\sum_{i\in\mathbb{A}}\sigma_i^2}\right)\right) & \text{for } i \in \mathbb{A} \end{cases} \\
&= \begin{cases} q_i^* & \text{for } i \notin \mathbb{A} \\ \frac{\sigma_i^2}{\sum_{i\in\mathbb{A}}\sigma_i^2} \cdot q_{\mathbb{A}}^* & \text{for } i \in \mathbb{A}. \end{cases}
\end{aligned}$$

□

Several remarks can be made based on Theorem 3.

- First, it can be extended from Theorem 3 that as long as $\mathbb{A}$ is pre-determined, the maximization labor can always be reduced down to one. In the special case that the

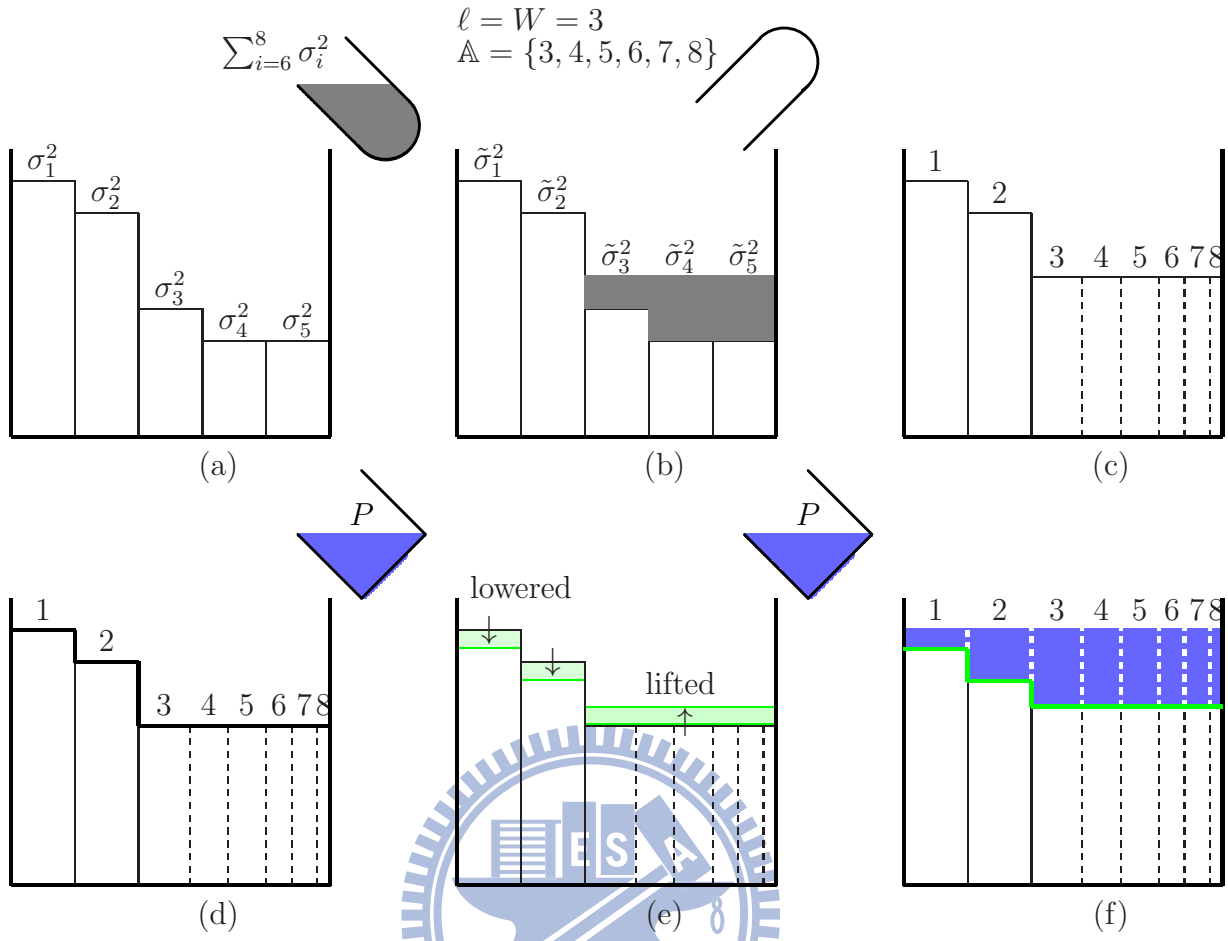


Figure 4.1: The graphical interpretation of the optimal *two-phase water-filling* power allocation for an $(8, 5)$-limited access channel with independent additive noises characterized by (4.1). In this figure, $[\sigma_1^2, \sigma_2^2, \cdots, \sigma_8^2] = [8, 7, 4, 3, 3, 2, 2, 1]$. Subfigures (a), (b) and (c) correspond to the *noise-power re-distribution* phase, while subfigures (d), (e) and (f) illustrate the *signal-power allocation* phase.

noises are additive and originated from the same family (as considered in this chapter), we can directly determine $\mathbb{A}$ in terms of (4.6).

- Secondly, when $\ell = 1$ (equivalently, $\mathbb{A} = \{1, 2, \ldots, N\}$), $\boldsymbol{p}^*$ can be determined without any maximization labor since we immediately have $q_{\mathbb{A}}^* = P$ by (4.10). In such a case, the optimal power allocation follows the *equal signal-to-noise ratio* (SNR) principle as

$$\frac{p_i^*}{\sigma_i^2} = \frac{P}{\sum_{j=1}^{N} \sigma_j^2} \quad \text{for every } 1 \le i \le N.$$

- Finally, the validity of Theorem 3 does not need to be restricted to channels with additive noises of the same family but can be extended to any (N, K)-limited access channel with marginal mutual information functions satisfying (4.3) for some function g that obeys Assumption 1. A straightforward example is the flat fading channels with known channel states at the receiver end, characterized by

$$Y_i = (\beta_i H_i)(\sqrt{p_i} X_i) + \sigma_i Z_i \quad \text{for } 1 \le i \le N \tag{4.16}$$

where $\{H_i\}_{i=1}^N$ is i.i.d. with unit second moment, and is independent of the channel input and additive noise. We then obtain $f_i(p_i) = g(\beta_i^2 p_i/\sigma_i^2)$ with $g(\rho) = I(\sqrt{\rho}X_i; \sqrt{\rho}H_i X_i + Z_i|H_i)$. Theorem 3 thus can be used to establish the optimal power allocation by treating σ_i^2/β_i^2 as the new noise power level.

An exemplified illustration of the two-phase water-filling scheme is depicted in Figure 4.1. Details are given below.

⟨*The noise-power re-distribution phase*⟩

Fig. 4.1(a) Set K vessels with widths of unit length and with base height of the ith vessel being σ_i^2 for $1 \le i \le K$. (Note that we assume $\sigma_1^2 \ge \sigma_2^2 \ge \cdots \ge \sigma_N^2$.)

Fig. 4.1(b) Pour in the "noise water" of amount $\sum_{j=K+1}^N \sigma_j^2$ and set $\tilde{\sigma}_i^2$ as the new water level of vessel i for $1 \le i \le K$. Let ℓ be the smallest integer among $\{1, 2, \ldots, K\}$ such that $\sigma_i^2 \le \tilde{\sigma}_K^2$ (cf. (4.6)). Assign $\mathbb{A} = \{\ell, \ell+1, \ldots, N\}$ and $W = K - \ell + 1$.

Fig. 4.1(c) Sub-divide the space of the last W vessels (i.e., $K-W+1, K-W+2, \ldots, K$) into $W + (N-K)$ new vessels of rectangular shape with base height the same as the water surface level and widths in proportion to σ_i^2 for $\ell \le i \le N$.

⟨*The signal-power allocation phase*⟩

Fig. 4.1(d) Retain the N vessels from the previous phase.

Fig. 4.1(e) Adjust the base height of the ith vessel to[3]

$$L_i(\nu) \triangleq \begin{cases} \sigma_i^2 \cdot G(\nu\sigma_i^2) & \text{for } 1 \le i < \ell \\ \tilde{\sigma}_K^2 \cdot G(\nu\tilde{\sigma}_K^2) & \text{for } \ell \le i \le N \end{cases} \tag{4.17}$$

where ν is the parameter chosen in Theorem 3 according to (4.10)*, and*

$$G(\zeta) \triangleq \begin{cases} \frac{1}{\zeta} - g'^{(\mathrm{inv})}(\zeta) & \text{if } g'(\infty) < \zeta < g'(0) \\ \frac{1}{g'(0)} & \text{if } \zeta \ge g'(0). \end{cases}$$

Fig. 4.1(f) Pour in the "signal water" of amount P. Then the volume of water in the ith vessel is the optimal power p_i^* to be allocated for channel i.

In the above procedure, the auxiliary function G will be reduced to what has been defined and identically denoted in [18, eq. (43)] for the *mercury* adjustment when $\{Z_i\}_{i=1}^N$ are i.i.d. complex Gaussian with unit variance. It can also be confirmed that for additive complex Gaussian noises and $K = N$, the induced mercury adjustment in [18] is exactly equal to that given by (4.17) by replacing the constant η therein with ν. Furthermore, when channel inputs $\{X_i\}_{i=1}^N$ are also independent and complex Gaussian distributed, the adjusted base heights in (4.17) are further reduced to the original noise variances, and the standard water-filling interpretation is resulted. We however found that the adjusted base heights may not be always greater than or equal to the original heights (as they should be for additive Gaussian noises). Thus, the intuition suggested by mercury-filling may not be applicable

[3]Since $1/\nu$ is the water level, (4.7) indicates that the base heights for unit-width vessels with indices $1 \le i < \ell$ should be given by

$$L_i(\nu) = \begin{cases} \frac{1}{\nu} - p_i^* & \left(< \frac{1}{\nu}\right) \quad \text{if } p_i^* > 0 \\ \frac{\sigma_i^2}{g'(0)} & \left(\ge \frac{1}{\nu}\right) \quad \text{if } p_i^* = 0 \end{cases} = \begin{cases} \sigma_i^2 \left(\frac{1}{\nu\,\sigma_i^2} - g'^{(\mathrm{inv})}\left(\nu\,\sigma_i^2\right)\right) & \text{if } g'(\infty) < \nu\sigma_i^2 < g'(0) \\ \sigma_i^2 \frac{1}{g'(0)} & \text{if } \nu\sigma_i^2 \ge g'(0). \end{cases}$$

A similar derivation can be made for vessels with indices $\ell \le i \le N$. We re-express the adjusted base heights in terms of an auxiliary function G in order to have a compatible formula to that in [18] when complex Gaussian additive noises are considered.

when the heights of vessel bases need to be lowered. We next give examples for both $K = N$ and $K < N$ to substantiate this finding.

Example 1 (Quaternary-input additive Laplace noise channels). Suppose that the i.i.d. channel inputs $\{X_i\}_{i=1}^N$ in (4.1) admit only four values with

$$\Pr\left[X_i = \frac{1+\imath}{\sqrt{2}}\right] = \Pr\left[X_i = \frac{1-\imath}{\sqrt{2}}\right] = \Pr\left[X_i = \frac{-1+\imath}{\sqrt{2}}\right] = \Pr\left[X_i = \frac{-1-\imath}{\sqrt{2}}\right] = \frac{1}{4}$$

and the complex zero-mean unit-variance i.i.d. additive noises $\{Z_i\}_{i=1}^N$ have marginal Laplace probability density function $\exp\left(-2\left(|\mathrm{Re}(z)| + |\mathrm{Im}(z)|\right)\right)$ for complex z, where $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ are the real and imaginary parts of z, respectively. The additive Laplace noise has been considered in many publications such as [2, 13, 16, 28–30, 35], and has been shown to be an appropriate model for, e.g., polarity detection [13], prediction error of image encoding [35] and communications at extremely low frequencies [2].

Assume $N = K = 4$, $P = 1.5$, and $[\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_4^2] = [1.2,\ 1.0,\ 0.4,\ 0.1]$. We can then derive as similarly to [12] that

$$g'(\rho) = \sqrt{\frac{2}{\rho}} \cdot \exp\left(-\sqrt{2\rho}\right) \cdot \mathrm{Gd}\left(\sqrt{2\rho}\right) \quad \text{for } \rho > 0$$

and

$$g'(0) \triangleq \lim_{\rho \downarrow 0} g'(\rho) = 2$$

where $\mathrm{Gd}(x) \triangleq 2 \cdot \arctan(e^x) - \pi/2$ is the Gudermannian function [6]. It can then be confirmed from Figure 4.2 that g' satisfies Assumption 1.

Since $N = K = 4$, we get $\tilde{\sigma}_i^2 = \sigma_i^2$ for $1 \le i \le 4$; hence, $\ell = 4$ and $\mathbb{A} = \{4\}$. We can then obtain numerically that

$$L_i(\nu) = \begin{cases} 1.00955 & \text{for } i = 1 \quad (< \tilde{\sigma}_1^2 = 1.2;\ \text{i.e., lowered}) \\ 0.94231 & \text{for } i = 2 \quad (< \tilde{\sigma}_2^2 = 1.0;\ \text{i.e., lowered}) \\ 0.80963 & \text{for } i = 3 \quad (> \tilde{\sigma}_3^2 = 0.4;\ \text{i.e., lifted}) \\ 0.95399 & \text{for } i = 4 \quad (> \tilde{\sigma}_4^2 = 0.1;\ \text{i.e., lifted}) \end{cases}$$

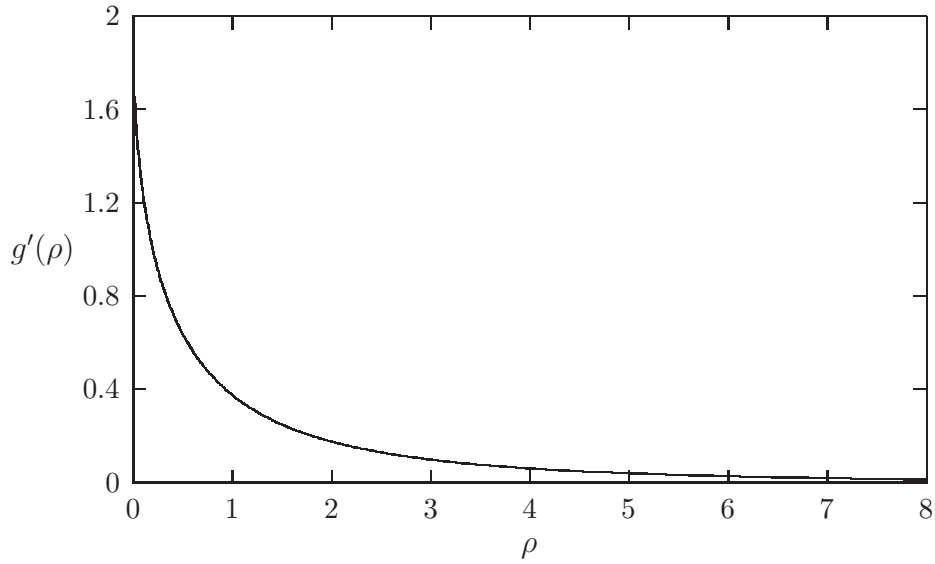


Figure 4.2: Function $g'(\rho)$ for quaternary-input additive Laplace noise channels.

where $\nu = 0.76695$ according to (4.10). Therefore, the base heights of the first two vessels are actually *lowered* rather than *lifted* as indicated above inside the parentheses. The optimal power allocation is given by

$$\boldsymbol{p}^* = [p_1^*,\ p_2^*,\ p_3^*,\ p_4^*] = [0.29432,\ 0.36156,\ 0.49424,\ 0.34988].$$

□

We next illustrate a situation with $K < N$.

Example 2. Following Example 1 but now using $K = 3$ and $P = 1$, we get $[\tilde{\sigma}_1^2,\ \tilde{\sigma}_2^2,\ \tilde{\sigma}_3^2] = [1.2,\ 1.0,\ 0.5]$; hence, $\ell = 3$ and $\mathbb{A} = \{3, 4\}$. We then obtain numerically that

$$L_i(\nu) = \begin{cases} 0.964300 & \text{for } i = 1 \quad (< \tilde{\sigma}_1^2 = 1.2; \text{i.e., lowered}) \\ 0.896885 & \text{for } i = 2 \quad (< \tilde{\sigma}_2^2 = 1.0; \text{i.e., lowered}) \\ 0.758525 & \text{for } i = 3 \quad (> \tilde{\sigma}_3^2 = 0.5; \text{i.e., lifted}) \\ 0.758525 & \text{for } i = 4 \quad (> \tilde{\sigma}_3^2 = 0.5; \text{i.e., lifted}) \end{cases}$$

where $\nu = 0.828795$ according to (4.10). Therefore, the base heights of the first two vessels are again *lowered* rather than *lifted*. Note that for vessels with indices in $\mathbb{A} = \{3, 4\}$,

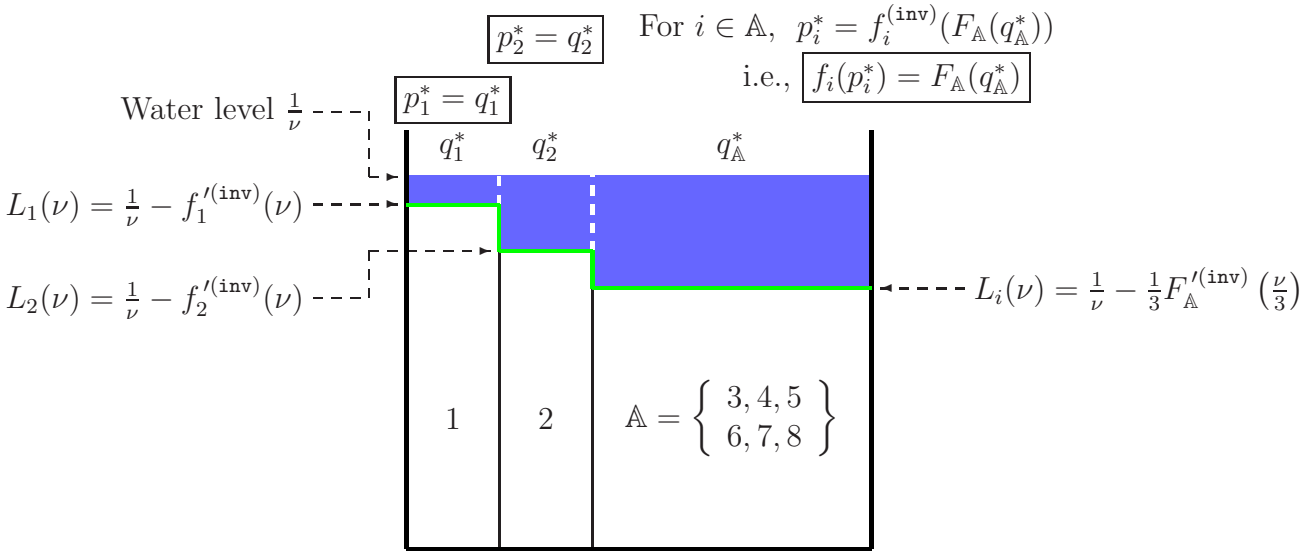


Figure 4.3: The graphical interpretation of the optimal power allocation for a general $(8,5)$-limited access channel. We assume $\mathbb{A}$ is known to be $\{3,4,5,6,7,8\}$ in order to facilitate its comparison (as an extension) with Figure 4.1(f).

their adjusted base heights should be equal and are determined by $\tilde{\sigma}_K^2$. The optimal power allocation is given by

$$\boldsymbol{p}^* = [p_1^*,\ p_2^*,\ p_3^*,\ p_4^*] = [0.242270,\ 0.309685,\ 0.358436,\ 0.089609].$$

□

Although the two-phase water-filling scheme cannot be extended to a general (N,K)-limited access channel (for which the channels may not be controlled by a common function g with single parameter σ^2), the resultant optimal power allocation $\boldsymbol{p}^*$ can still be graphically interpreted similarly to Figure 4.1(f). In particular, we can regard the tank to be structured by $N - |\mathbb{A}| + 1$ vessels, which have unit width except for the last one that is of width $K - N + |\mathbb{A}|$, as illustrated in Figure 4.3. The adjusted heights of vessel bases in their most

general form can be formulated by the following equations: if $i \notin \mathbb{A}$, then

$$L_i(\nu) \triangleq \begin{cases} \dfrac{1}{\nu} - f_i'^{(\mathtt{inv})}(\nu) & \text{if } f_i'(\infty) < \nu < f_i'(0) \quad (4.18a) \\ \dfrac{1}{f_i'(0)} & \text{if } \nu \geq f_i'(0) \quad (4.18b) \end{cases}$$

else (i.e., $i \in \mathbb{A}$)

$$L_i(\nu) \triangleq \frac{1}{\nu} - \frac{1}{K - N + |\mathbb{A}|} F_{\mathbb{A}}'^{(\mathtt{inv})}\left(\frac{\nu}{K - N + |\mathbb{A}|}\right). \quad (4.19)$$

It can be verified that taking function G into function L_i defined in (4.17) should assume the same form as (4.18a), (4.18b) and (4.19). From the above formula, it is clear that $1/\nu$ can be interpreted as the water level. Equations (4.18a) and (4.19) then reasonably imply that the optimal power allocation $\boldsymbol{q}^*$ satisfies $q_i^* = f_i'^{(\mathtt{inv})}(\nu)$ for $i \notin \mathbb{A}$ and $q_{\mathbb{A}}^* = F_{\mathbb{A}}'^{(\mathtt{inv})}\left(\frac{\nu}{K-N+|\mathbb{A}|}\right)$. The aggregate power $q_{\mathbb{A}}^*$ will then be re-distributed to those channels with indices in $\mathbb{A}$ according to *equal-mutual-information* principle, i.e., $f_i(p_i^*) = F_{\mathbb{A}}(q_{\mathbb{A}}^*)$ for every $i \in \mathbb{A}$. This *equal-mutual-information* principle is exactly the extension of *equal-SNR* principle for channels with additive noises of the same family. Moreover, when ν lies in the range specified in (4.18b) for some i, no power is allocated to the respective channel; hence, $p_i^* = q_i^* = 0$.

We close this chapter by the following observation. It may be worth knowing that for channels with additive noises of the same family, the optimal power allocation $\boldsymbol{p}^*$ can actually be determined directly by regarding the last W vessels as unit-width vessels with base height respectively equal to $\tilde{\sigma}_i^2$ for $K - W + 1 \leq i \leq K$ (cf. Figure 4.1(b)). This reduces the original problem to a power allocation problem over a (K, K)-limited access additive noise channel with effective noise powers $\{\tilde{\sigma}_i^2\}_{i=1}^K$. The resultant K-dimensional optimal power allocation $\boldsymbol{r}^* \triangleq [r_1^*, r_2^*, \ldots, r_K^*]$ is exactly the heights of water levels in each vessel. The desired optimal

power allocation $\boldsymbol{p}^*$ can then be given by

$$p_i^* = \begin{cases} r_i^* & \text{for } 1 \le i < \ell \\ \dfrac{\sigma_i^2}{\sum_{k=\ell}^{N} \sigma_k^2} \left(\sum_{j=\ell}^{K} r_j^* \right) & \text{for } \ell \le i \le N. \end{cases}$$

Although taking this aspect seems to save the effort of sub-dividing the vessels into ones with unequal widths in Figure 4.1(c), more effort can be saved if the last W vessels are aggregated as one. In other words, we can simply use a tank containing ℓ vessels, in which $\ell - 1$ of them have unit width and the remaining one has width W. We can then obtain the ℓ-dimensional optimal power allocation $\boldsymbol{q}^* = [q_1^*, q_2^*, \ldots, q_{\ell-1}^*, q_{\mathbb{A}}^*]$ through the water-filling scheme (cf. Figure 4.1(f)). The equal-SNR power allocation principle is subsequently applied to re-distribute $q_{\mathbb{A}}^*$ to $p_\ell^*, p_{\ell+1}^*, \ldots, p_N^*$ in proportion to $\sigma_\ell^2, \sigma_{\ell+1}^2, \ldots, \sigma_N^2$, respectively, as suggested in (4.7). This is actually what Theorem 3 implicitly indicates, which justifies the introduction of the aggregate mutual information that views the last $W + (N - K)$ channels as a single auxiliary channel.

Chapter 5

Implications from the Optimal Power Allocation

Theorem 2 indicates that the sequence of candidate maximal-mutual-information index sets $\mathbb{B}_1, \mathbb{B}_2, \mathbb{B}_3, \ldots$ can be identified via the determination of $j_1, j_2, j_3, \ldots$. In a sense, this sequence can be regarded as sorting the channels in their descending degrees of "noisiness," which can be supported by the result from Theorem 3, where the sequence of $j_1, j_2, j_3, \ldots$ coincides with $\sigma_{j_1}^2 \geq \sigma_{j_2}^2 \geq \sigma_{j_3}^2 \geq \cdots$.

For a general (N, K)-limited access channel in which the noises are not necessarily additive or scaled from the same family, can one identify such sequence through their mutual information functions? The next theorem may provide a guide along this direction of thinking. For simplification, all the proofs in this chapter are placed in Appendix C.

Theorem 4. For a general (N, K)-limited access channel, if

$$f'_{k_1}\left(f_{k_1}^{(\mathtt{inv})}(y)\right) \leq f'_{k_2}\left(f_{k_2}^{(\mathtt{inv})}(y)\right) \leq \cdots \leq f'_{k_N}\left(f_{k_N}^{(\mathtt{inv})}(y)\right) \quad \text{for all } y \geq 0$$

then $j_M = k_M$ for $M = 1, 2, 3, \ldots$.[1]

[1]When $\omega_i \triangleq \lim_{p\to\infty} f_i(p)$ is finite, the function $f'_{k_1}(f_{k_1}^{(\mathtt{inv})}(y))$ is clearly defined for $y < \omega_i$. For $y \geq \omega_i$, we define $f_i^{(\mathtt{inv})}(y) = \infty$ as emphasized in Footnote 2. This, together with the fact that $\omega_i < \infty$ and Assumption 1 jointly imply that $\lim_{p\to\infty} f'_i(p) = 0$, gives that $f'_i(f_i^{(\mathtt{inv})}(y)) = 0$ for $y \geq \omega_i$.

Here, regardless of the original goal of the determination of optimal power allocation, Theorem 4 (as an extension from Theorem 3) proposes a way to compare the degree of "noisiness" of general channels via their mutual information functions. For the additive noise channels of the same family, we have

$$f_i'\left(f_i^{(\mathtt{inv})}(y)\right) = \frac{1}{\sigma_i^2} g'\left(g^{(\mathtt{inv})}(y)\right).$$

Hence, the proposed ordering coincides with the general impression that the larger the σ_i^2, the noisier the ith channel is considered to be. To simplify the notation, we drop the parentheses between f_i' and $f_i^{(\mathtt{inv})}$ in the sequel.

For channels other than the one considered in Chapter 4, there could be no apparent winner between any two channels in the sense of $\{f_i' f_i^{(\mathtt{inv})}\}_{i=1}^N$. In other words, it could happen that

$$f_i' f_i^{(\mathtt{inv})}(y_1) > f_j' f_j^{(\mathtt{inv})}(y_1) \quad \text{but} \quad f_i' f_i^{(\mathtt{inv})}(y_2) < f_j' f_j^{(\mathtt{inv})}(y_2)$$

for two distinct y_1 and y_2 and two distinct i and j. As such, the sequence of $j_1, j_2, j_3, \ldots$ will become a function of the total signal power P. However, if a certain condition is satisfied, the pre-identification of the degrees of channel noisiness is still possible at two extreme situations: $P \to 0$ and $P \to \infty$, which we will respectively refer to as the low- and high-power regimes in later discussion.

Lemma 3.

1. If

$$\limsup_{y \downarrow 0} \operatorname{sgn}\left(f_i' f_i^{(\mathtt{inv})}(y) - f_j' f_j^{(\mathtt{inv})}(y)\right) \le 0 \quad \text{for every } 1 \le i < j \le N \tag{5.1}$$

then $j_i = i$ in the low-power regime, where sign function $\operatorname{sgn}(\rho)$ is equal to either 1, 0 or -1 depending on whether $\rho > 0$, $\rho = 0$ or $\rho < 0$.

2. If

$$\limsup_{y\uparrow\min\{\omega_i,\omega_j\}} \operatorname{sgn}\left(f_i' f_i^{(\mathtt{inv})}(y) - f_j' f_j^{(\mathtt{inv})}(y)\right) \le 0 \quad \text{for every } 1 \le i < j \le N \tag{5.2}$$

then $j_i = i$ in the high-power regime, provided that $\lim_{p\to\infty} f_i'(p) = 0$ for $1 \le i \le N$, where $\omega_i \triangleq \lim_{p\to\infty} f_i(p)$.

Since the input alphabet is usually finite for channels of practical interest, we have $\omega_i \triangleq \lim_{p\to\infty} f_i(p) \le H(X_i) < \infty$. This immediately validates the premise, i.e., $\lim_{p\to\infty} f_i'(p) = 0$, for condition (5.2) implying $j_i = i$ in the high-power regime. In other words, $\lim_{p\to\infty} f_i'(p) = 0$ is true for all finite-input channels. There however exists a certain kind of channels where $\omega_i = \infty$ while $\lim_{p\to\infty} f_i'(p) = 0$. An example is the Gaussian-input AWGN channel for which $f_i(p) = \log(1 + p/\sigma_i^2)$. We would like to emphasize that the inference regarding (5.2) still remains valid for channels with unbounded mutual information as long as $\lim_{p\to\infty} f_i'(p) = 0$.

Conditions (5.1) and (5.2) in Lemma 3 involve the examination of the limit supremum of function differences. The next corollary shows that their validity can be guaranteed by comparing the limiting behaviors of individual functions.

Corollary 1.

1. The validity of (5.1) for an (i, j) pair is certain if one of the three conditions below is satisfied:[2]

$$\begin{cases} f_i'(0) < f_j'(0) & \text{(5.3a)} \\ f_i'(0) = f_j'(0) \text{ and } f_i''(0) < f_j''(0) & \text{(5.3b)} \\ (\exists\ \delta > 0)\ f_i'(p) \le f_j'(p) \text{ for } 0 < p < \delta. & \text{(5.3c)} \end{cases}$$

2. The validity of (5.2) for an (i, j) pair is certain if

$$\omega_i = \lim_{p\to\infty} f_i(p) < \omega_j = \lim_{p\to\infty} f_j(p). \tag{5.4}$$

[2]The second derivative f_i'' at the origin is again defined as $f_i''(0) \triangleq \lim_{p\downarrow 0} f_i''(p)$.

We are now ready to illustrate an example that validates the sufficient conditions in Lemma 3 and Corollary 1.

Example 3 (Flat fading channels). Consider the fading channels characterized by (4.16), where the additive noises $\{Z_i\}_{i=1}^N$ are i.i.d. zero-mean complex Gaussian, but the channel inputs $\{X_i\}_{i=1}^N$ are no longer identically distributed. With the first three channel inputs being respectively BPSK, QPSK and 16-QAM and the remaining channel inputs being complex Gaussian signals, we obtain $j_i = i$ for $i = 1, 2, \ldots$ in both low- and high-power regimes by being given

$$\frac{\sigma_1^2}{\beta_1^2} = \frac{\sigma_2^2}{\beta_2^2} > \frac{\sigma_3^2}{\beta_3^2} = \frac{\sigma_4^2}{\beta_4^2} > \frac{\sigma_5^2}{\beta_5^2} \geq \frac{\sigma_6^2}{\beta_6^2}. \tag{5.5}$$

This can be verified as follows.

It has been derived in [19] that

$$f_i'(p) = E\left[\frac{\beta_i^2}{\sigma_i^2}|H_i|^2 \cdot \mathrm{MMSE}_i\left(p\frac{\beta_i^2}{\sigma_i^2}|H_i|^2\right)\right]$$

and

$$f_i'(0) = \lim_{p\to 0} f_i'(p) = \lim_{p\to 0} E\left[\frac{\beta_i^2}{\sigma_i^2}|H_i|^2 \cdot \mathrm{MMSE}_i\left(p\frac{\beta_i^2}{\sigma_i^2}|H_i|^2\right)\right] = E\left[\frac{\beta_i^2}{\sigma_i^2}|H_i|^2\right] = \frac{\beta_i^2}{\sigma_i^2}$$

where

$$\mathrm{MMSE}_i(\rho) = E\left[\left|X_i - E\left[X_i\middle|\sqrt{\rho}X_i + Z_i\right]\right|^2\right].$$

In the low-power regime, the order of those indices, where σ_i^2/β_i^2 are not equal, is thus confirmed by (5.3a). From [18], we know

$$\text{QPSK-MMSE}(\rho) = \text{BPSK-MMSE}(\rho/2) > \text{BPSK-MMSE}(\rho) \quad \text{for } \rho > 0.$$

Then

$$\begin{aligned} f_1'(p) - f_2'(p) &= E\left[\frac{\beta_1^2}{\sigma_1^2}|H_1|^2\mathrm{MMSE}_1\left(p\frac{\beta_1^2}{\sigma_1^2}|H_1|^2\right)\right] - E\left[\frac{\beta_2^2}{\sigma_2^2}|H_2|^2\mathrm{MMSE}_2\left(p\frac{\beta_2^2}{\sigma_2^2}|H_2|^2\right)\right] \\ &= \frac{\beta_1^2}{\sigma_1^2}E\left[|H_1|^2\left(\text{BPSK-MMSE}\left(p\frac{\beta_1^2}{\sigma_1^2}|H_1|^2\right) - \text{QPSK-MMSE}\left(p\frac{\beta_1^2}{\sigma_1^2}|H_1|^2\right)\right)\right] \\ &< 0 \end{aligned}$$

for $p > 0$. Then (5.3c) assures that $j_1 = 1$ and $j_2 = 2$ in the low-power regime, which somehow suggests that under equal effective noise power, QPSK modulations should be favored over BPSK modulations when the power budget is extremely tight. Since $j_3 = 3$ and $j_4 = 4$ in the low-power regime can be verified similarly, we omit their proof.

In the high-power regime, we first note that $\lim_{p\to\infty} f'(p) = 0$ for $1 \le i \le 6$. We then confirm $j_i = i$ for $1 \le i \le 3$ from

$$\lim_{p\to\infty} f_i(p) = \begin{cases} 2^{i-1} & i = 1, 2, 3 \\ \infty & i > 3 \end{cases} \tag{5.6}$$

according to (5.4). That $j_i = i$ for $4 \le i \le 6$ in the high-power regime can be substantiated by (5.2) and $f_i' f_i^{(\mathtt{inv})}(y) = \frac{\beta_i^2}{\sigma_i^2} \cdot g' g^{(\mathtt{inv})}(y)$, where

$$g(\rho) = E\big[\log\big(1 + \rho|H_i|^2\big)\big] \quad \text{and} \quad g'(\rho) = E\left[\frac{|H_i|^2}{1 + \rho|H_i|^2}\right].$$

□

After determining the sequence $j_1, j_2, j_3, \ldots$, the next task for finding the optimal power allocation is to determine $\mathbb{A}$. Recall that $\mathbb{A}$ is defined as the set of channels that have the largest mutual information for the optimal power assignment (see (3.2)). For channels corrupted with additive noises of the same family, $\mathbb{A}$ can be directly determined and has nothing to do with total power P. In more general cases, however, $\mathbb{A}$ depends on P. There is even no guarantee for its convergence in the low- or high-power regimes even if the monotonicity condition of $\{f_i' f_i^{(\mathtt{inv})}\}_{i=1}^N$ in Theorem 4 holds. This is because in terms of given $j_1, j_2, j_3, \ldots$, only sufficient conditions on the validity and violation of stop criterion (3.11) can be obtained as summarized in the next lemma.

Lemma 4. For the already pre-determined $j_1, j_2, j_3, \ldots$ and an integer M that is under examination in the algorithm of Theorem 2, we have the following logical statements to help determining $\mathbb{A}$. Again, in the high-power region, we assume that $\lim_{p\to\infty} f'(p) = 0$ for $1 \le i \le N$.

1. If

$$\limsup_{y\downarrow 0} \operatorname{sgn}\left(K - M - \sum_{i\in\mathbb{B}_{M+1}} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}\right) \le 0 \tag{5.7}$$

then stop criterion (3.11) holds in the low-power regime.

2. If

$$\liminf_{y\downarrow 0} \operatorname{sgn}\left(K - M - \sum_{i\in\mathbb{B}_{M+1}} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}\right) > 0 \tag{5.8}$$

then stop criterion (3.11) fails in the low-power regime.

3. If

$$\limsup_{y\uparrow\Omega(\mathbb{B}_M)} \operatorname{sgn}\left(K - M - \sum_{i\in\mathbb{B}_{M+1}} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}\right) \le 0 \tag{5.9}$$

then stop criterion (3.11) holds in the high-power regime, where

$$\Omega(\mathbb{B}_M) \triangleq \min_{i\in\mathbb{B}_M} \omega_i = \min_{i\in\mathbb{B}_M} \lim_{p\to\infty} f_i(p)$$

4. If

$$\liminf_{y\uparrow\Omega(\mathbb{B}_M)} \operatorname{sgn}\left(K - M - \sum_{i\in\mathbb{B}_{M+1}} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}\right) > 0 \tag{5.10}$$

then stop criterion (3.11) fails in the high-power regime.

Although conditions (5.7) and (5.8) are mutually exclusive, it is still possible that both are violated. In such a case, $\lim_{P\to 0} \mathbb{A}(P)$ may be indeterminate. Likewise, the same remark is applied to (5.9) and (5.10).

The conditions in Lemma 4 involve the examination of the limiting behavior of a difference between a constant $K - M$ and a sum of function ratios $\sum_{i\in\mathbb{B}_{M+1}} [f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)/f'_i f_i^{(\mathtt{inv})}(y)]$. We similarly derive sufficient conditions to (5.7)–(5.10) based on the limiting behavior of individual functions as shown in Corollary 2.

Corollary 2. Follow Lemma 4.

1. (5.7) is valid if

$$K - M < \sum_{i \in \mathbb{B}_{M+1}} \frac{f'_{j_M}(0)}{f'_i(0)}. \tag{5.11}$$

2. (5.8) is valid if

$$K - M > \sum_{i \in \mathbb{B}_{M+1}} \frac{f'_{j_M}(0)}{f'_i(0)}. \tag{5.12}$$

3. (5.9) is valid if

$$K - M < \sum_{i \in \mathbb{B}_{M+1}} \liminf_{y \uparrow \Omega(\mathbb{B}_M)} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}. \tag{5.13}$$

4. (5.10) is valid if

$$K - M > \sum_{i \in \mathbb{B}_{M+1}} \limsup_{y \uparrow \Omega(\mathbb{B}_M)} \frac{f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)}. \tag{5.14}$$

   Furthermore, if $j_1, j_2, j_3, \ldots$ are determined according to the condition in Lemma 3, i.e., (5.2), then (5.10) can be implied by

$$\omega_{j_M} = \lim_{p \to \infty} f_{j_M}(p) < \omega_{j_{M+1}} = \lim_{p \to \infty} f_{j_{M+1}}(p). \tag{5.15}$$

Example 4. Continue from Example 3 where we have determined $j_i = i$ for $i = 1, 2, \ldots$ for both low- and high-power regimes. Assume $K = 5$ and

$$\left[\frac{\sigma_1^2}{\beta_1^2}, \frac{\sigma_2^2}{\beta_2^2}, \frac{\sigma_3^2}{\beta_3^2}, \frac{\sigma_4^2}{\beta_4^2}, \frac{\sigma_5^2}{\beta_5^2}, \frac{\sigma_6^2}{\beta_6^2}\right] = [1, 1, 0.5, 0.5, 0.4, 0.4]\,.$$

Then, by $f'_i(0) = \beta_i^2/\sigma_i^2$ for $1 \le i \le 6$ and (5.11) and (5.12), we establish

$$\mathbb{A}\ (= \lim_{P \to 0} \mathbb{A}(P)\ )\ = \mathbb{B}_3 = \{3, 4, 5, 6\}$$

in the low-power regime.

In the high-power regime, (5.15) and (5.6) imply that $\mathbb{A} \subseteq \mathbb{B}_4 = \{4, 5, 6\}$. Examination of $M = 4$ with $\mathbb{B}_5 = \{5, 6\}$ gives

$$\sum_{i \in \mathbb{B}_5} \limsup_{y \uparrow \Omega(\mathbb{B}_4)} \frac{f'_{j_4} f_{j_4}^{(\mathtt{inv})}(y)}{f'_i f_i^{(\mathtt{inv})}(y)} = \sum_{i \in \mathbb{B}_5} \frac{\beta_4^2/\sigma_4^2}{\beta_i^2/\sigma_i^2} = 1.6 > K - M = 5 - 4 = 1.$$

Hence, (5.13) is valid, and so is the stop criterion of the algorithm. As a result, the algorithm in Theorem 2 will stop at $M = 4$ and $\mathbb{A}(= \lim_{P\to\infty} \mathbb{A}(P)) = \mathbb{B}_4 = \{4, 5, 6\}$. □

The two lemmas and two corollaries presented previously give the conditions under which $\mathbb{A}$ can be determined directly in the low- and high-power regimes. We can then compute the optimal power allocation $\boldsymbol{q}^*$ using the Lagrange multipliers technique and KKT condition in terms of the auxiliary aggregate channel associated with $\mathbb{A}$. After such step, the optimal power allocation p_i^* for a channel outside $\mathbb{A}$ equals the respective component q_i^* of $\boldsymbol{q}^*$, but the power allocation for channels inside $\mathbb{A}$ should be obtained by re-distributing the power $q_{\mathbb{A}}^*$ according to the equal mutual-information principle. Since we are concerned with the situation when P approaches either 0 or ∞, the optimal power $\boldsymbol{p}^*$ may as well approach the same limiting value. It is thus more meaningful to consider the ratio between the optimal power allocations of channel pairs in the low- and high-power regimes.

Lemma 5. After the determination of $\mathbb{A}$, the optimal power allocation q_i^* for the ith channel outside $\mathbb{A}$ asymptotically satisfies

$$\lim_{P\to 0} \frac{q_i^*}{q_{\mathbb{A}}^*} = \begin{cases} (K + |\mathbb{A}| - N)\dfrac{F_{\mathbb{A}}''(0)}{f_i''(0)} & \text{if } f_i'(0) = (K + |\mathbb{A}| - N)F_{\mathbb{A}}'(0) \\ 0 & \text{otherwise} \end{cases} \tag{5.16}$$

provided the second derivatives $f_i''(p)$ and $F_{\mathbb{A}}''(p)$ exist, $f_i''(0) \triangleq \lim_{p\to 0} f_i''(p) < \infty$, and $F_{\mathbb{A}}''(0) \triangleq \lim_{p\to 0} F_{\mathbb{A}}''(p) < \infty$. In addition, for two channels with indices i, j in $\mathbb{A}$, the re-distribution of $q_{\mathbb{A}}^*$ yields

$$\lim_{P\to 0} \frac{p_i^*}{p_j^*} = \frac{f_j'(0)}{f_i'(0)}$$

provided that $f_i'(0)$ and $f_j'(0)$ are both finite.

In the high-power regime, since we assume $\lim_{p\to\infty} f_i'(p) = 0$, we have $\limsup_{p\to\infty} f_i''(p) = 0$.[3] Discussion regarding the limiting power ratio between channel pair therefore cannot

[3]This can be seen from $0 \geq \int_\alpha^\beta f''(p)dp \geq f'(\beta) - f'(\alpha)$ as a consequence of the strict decreasingness of f_i' [26, p. 100] .

be stated in the same fashion as (5.16). The next observation then indicates that in the high-power regime, the power ratios between channel pairs are governed by the rate of convergences of $\{f_i'(p)\}_{i=1}^N$ at p large.

Observation 1.

1. For channels outside $\mathbb{A}$, the following statements hold.

   - If $F_{\mathbb{A}}'$ and each f_i' outside $\mathbb{A}$ vanish at a polynomial speed, i.e.,

   $$\lim_{q_{\mathbb{A}}\to\infty} q_{\mathbb{A}}^{m_{\mathbb{A}}}(K-N+|\mathbb{A}|)F_{\mathbb{A}}'(q_{\mathbb{A}}) = c_{\mathbb{A}} \quad \text{and} \quad \lim_{q_i\to\infty} q_i^{m_i} f_i'(q_i) = c_i \text{ for } i \notin \mathbb{A},$$

   where m_i, $m_{\mathbb{A}}$, c_i and $c_{\mathbb{A}}$ are all positive, or if $F_{\mathbb{A}}'$ and each f_i' outside $\mathbb{A}$ vanish at an exponential speed, i.e.,

   $$\lim_{q_{\mathbb{A}}\to\infty} q_{\mathbb{A}}^{m_{\mathbb{A}}} \log\left((K-N+|\mathbb{A}|)F_{\mathbb{A}}'(q_{\mathbb{A}})\right) = c_{\mathbb{A}} \quad \text{and} \quad \lim_{q_i\to\infty} q_i^{m_i} \log\left(f_i'(q_i)\right) = c_i \text{ for } i \notin \mathbb{A},$$

   where m_i, $m_{\mathbb{A}}$, c_i and $c_{\mathbb{A}}$ are all negative, then for $i, j \notin \mathbb{A}$,

   $$\lim_{P\to\infty} \frac{q_i^*}{q_j^*} = \begin{cases} 0 & \text{if } |m_i| > |m_j| \\ \left(\frac{c_i}{c_j}\right)^{1/m_i} & \text{if } |m_i| = |m_j| \\ \infty & \text{if } |m_i| < |m_j| \end{cases} \text{ and } \lim_{P\to\infty} \frac{q_i^*}{q_{\mathbb{A}}^*} = \begin{cases} 0 & \text{if } |m_i| > |m_{\mathbb{A}}| \\ \left(\frac{c_i}{c_{\mathbb{A}}}\right)^{1/m_i} & \text{if } |m_i| = |m_{\mathbb{A}}| \\ \infty & \text{if } |m_i| < |m_{\mathbb{A}}|. \end{cases}$$

   - If for $i, j \notin \mathbb{A}$, f_i' vanishes exponentially fast while f_j' and $F_{\mathbb{A}}'$ decay to zero at a polynomial speed, then

   $$\lim_{P\to\infty} \frac{q_i^*}{q_j^*} = \lim_{P\to\infty} \frac{q_i^*}{q_{\mathbb{A}}^*} = 0.$$

2. For channels i, j inside $\mathbb{A}$,

   $$\begin{cases} \lim_{P\to\infty} p_i^* = f_i^{(\mathtt{inv})}\left(\Omega(\mathbb{A})\right) & \text{if } \omega_i > \Omega(\mathbb{A}) = \min_{k\in\mathbb{A}} \omega_k \\ \lim_{P\to\infty} \frac{p_i^*}{p_j^*} = \lim_{y\uparrow\Omega(\mathbb{A})} \frac{f_j' f_j^{(\mathtt{inv})}(y)}{f_i' f_i^{(\mathtt{inv})}(y)} & \text{if } \omega_i = \omega_j = \Omega(\mathbb{A}). \end{cases}$$

There are certain channels with polynomially vanishing first derivatives in their mutual information functions. For example, the fading channels characterized in (4.16) satisfy $\lim_{q_i\to\infty} q_i f_i'(q_i) = 1$ when both channel inputs and additive noises are complex Gaussian. Examples for exponentially vanishing first derivatives in their mutual information functions are the AWGN channels with a finite channel input alphabet [18], for which $\lim_{q_i\to\infty} q_i^{-1} \log\left(f_i'(q_i)\right) = -d_i^2/(4\sigma_i^2)$, where d_i is the minimum distance between distinct channel inputs.

An interesting observation in the high-power regime is that the optimal power allocation p_i^* for a channel in $\mathbb{A}$ may be *bounded* even if the total power P goes to infinity. An available example can be constructed by re-assuming $K = 1$ in Example 3. Then, we have $\mathbb{A} = \{1, 2, \cdots, 6\}$, $\Omega(\mathbb{A}) = \min_{i\in\mathbb{A}} \omega_i = 1$ and $\lim_{P\to\infty} p_i^* = f_i^{(\mathtt{inv})}(\Omega(\mathbb{A})) = f_i^{(\mathtt{inv})}(1)$ for $1 \le i \le 6$, in which case $\lim_{P\to\infty} p_i^*$ is a finite positive number for $2 \le i \le 6$. It can be further verified that taking $K = 2$ in the same example gives $\mathbb{A} = \{2, 3, \cdots, 6\}$ in the high-power regime, which also results in $0 < \lim_{P\to\infty} p_i^* < \infty$ for $3 \le i \le 6$.

Chapter 6

Concluding Remarks

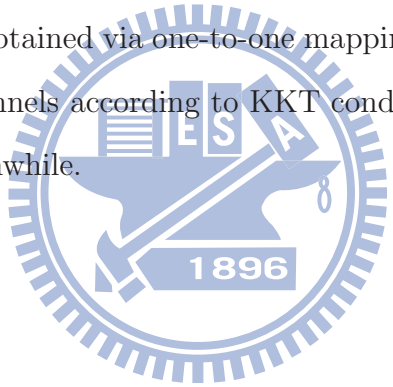
In this dissertation, we consider the (N, K)-limited access channel and establish an algorithmical procedure to find its optimal power allocation. The optimal power allocation obtained is not restricted to AWGN channels but can be applied to general channels with corresponding mutual information functions satisfying Assumption 1. For additive noises scaled from the same distribution family, finding the optimal power allocation is reduced to a simple two-phase water-filling process. This two-phase water-filling graphical interpretation can then be deduced to a general case, where the degrees of "channel noisiness" are in a sense implied by the composition functions $\{f_i' f_i^{(\mathrm{inv})}\}_{i=1}^N$ of the mutual information functions $\{f_i\}_{i=1}^N$. General behaviors of the optimal power allocation in low- and high-power regimes are also established. We would like to point out that the results in the work can be directly applied to a resource allocation problem associated with some "profit" functions $\{f_i\}_{i=1}^N$ as long as the problem is mathematically of the same form as (2.3). As such, the optimal resource allocation can be solved algorithmically, and sometimes directly if certain monotonicity conditions are satisfied. In addition, $\{f_i' f_i^{(\mathrm{inv})}\}_{i=1}^N$ now suggests the prioritized sequence of investments, i.e., the smaller the $f_i' f_i^{(\mathrm{inv})}$, the less profitable from the investment p_i.

One possible future work is to relax the independence assumption in (2.2) since a certain degree of dependence among channels may exist in practice. A good start would be to

investigate the additive color noise compound channel modeled by

$$Y_i = \sqrt{p_i} X_i + \sigma_i Z_i \quad \text{for } 1 \leq i \leq N$$

where $\{Z_i\}_{i=1}^{N}$ are dependent random variables. According to our preliminary study, a universal guideline is obtained for a group of permutation-invariant channels, in which the system mutual information remains unchanged when simultaneously permuting their power parameters and multiplicative coefficients (i.e., switching (p_i, σ_i^2) and (p_j, σ_j^2), also s_i and s_j, for channel i and j), that a channel with less power should have larger SNR. When all N channels belong to a permutation-invariant group, we also found that the optimal power allocation problem can be transformed to an equivalent problem for K parallel channels without limited access constraint via a water-filling noise-power-redistribution process, and then the optimal power allocation is obtained via one-to-one mapping from the power allocation solution of the K equivalent channels according to KKT condition. Further investigation along this direction might be worthwhile.



Appendix A

Proofs of Lemma 2 and Theorem 1

We first provide a simple property regarding the aggregate mutual information function. This property will be used in the proofs of both Theorems 1 and 2.

Property 1. If $\min_{i\in\mathbb{B}} p_i > 0$ and $f_i(p_i) = f_j(p_j)$ for every $i, j \in \mathbb{B}$, then

$$\frac{1}{F'_{\mathbb{B}}\left(\sum_{i\in\mathbb{B}} p_i\right)} = \sum_{i\in\mathbb{B}} \frac{1}{f'_i(p_i)}. \tag{A.1}$$

Proof. This is a direct consequence following (3.6) and the relation between the first derivative f' of a function f and its inverse $f^{(\mathtt{inv})}$ as:

$$\left.\frac{\partial f^{(\mathtt{inv})}(y)}{\partial y}\right|_{y=f(p)} = \frac{1}{f'(p)}$$

where in the above equation, f can be replaced by either any f_i with $1 \le i \le N$ or the aggregate mutual information function $F_{\mathbb{B}}$. □

Proof of Lemma 2. 1. For any $\boldsymbol{q} \in \mathcal{Q}(\mathbb{B})$, we can use the assignment in the lemma to obtain a corresponding $\boldsymbol{p}$, i.e.,

$$p_i = \begin{cases} q_i & \text{for } i \notin \mathbb{B} \\ f_i^{(\mathtt{inv})}(F_{\mathbb{B}}(q_{\mathbb{B}})) & \text{for } i \in \mathbb{B}. \end{cases} \tag{A.2}$$

Then, we have

$$f_i(p_i) = \begin{cases} f_i(q_i) & \text{for } i \notin \mathbb{B} \\ f_i\left(f_i^{(\mathtt{inv})}(F_{\mathbb{B}}(q_{\mathbb{B}}))\right) = F_{\mathbb{B}}(q_{\mathbb{B}}) & \text{for } i \in \mathbb{B} \end{cases}$$

and $f_i(p_i) = f_i(q_i) < F_{\mathbb{B}}(q_{\mathbb{B}})$ for $i \notin \mathbb{B}$ because q_i satisfies condition (ii) in $\mathcal{Q}(\mathbb{B})$. Thus, $\boldsymbol{p} \in \mathcal{P}(\mathbb{B})$ and

$$\sum_{i \notin \mathbb{B}} f_i(q_i) + (K + |\mathbb{B}| - N) F_{\mathbb{B}}(q_{\mathbb{B}}) = \sum_{i \notin \mathbb{B}} f_i(p_i) + (K + |\mathbb{B}| - N) \max_{1 \le j \le N} f_j(p_j).$$

Since the above derivation is true for any $\boldsymbol{q} \in \mathcal{Q}(\mathbb{B})$, we obtain that

$$\sup_{\boldsymbol{q} \in \mathcal{Q}(\mathbb{B})} \left\{ \sum_{i \notin \mathbb{B}} f_i(q_i) + (K + |\mathbb{B}| - N) F_{\mathbb{B}}(q_{\mathbb{B}}) \right\} \le \sup_{\boldsymbol{p} \in \mathcal{P}(\mathbb{B})} \left\{ \sum_{i \notin \mathbb{B}} f_i(p_i) + (K + |\mathbb{B}| - N) \max_{1 \le j \le N} f_j(p_j) \right\}. \quad \text{(A.3)}$$

2. Similarly, for any $\boldsymbol{p} \in \mathcal{P}(\mathbb{B})$, we can assign its corresponding $\boldsymbol{q}$ as

$$q_i = p_i \text{ for } i \notin \mathbb{B} \quad \text{and} \quad q_{\mathbb{B}} = \sum_{i \in \mathbb{B}} p_i. \quad \text{(A.4)}$$

Then, for $j \in \mathbb{B}$,

$$\begin{aligned} f_j(p_j) &= F_{\mathbb{B}}\left(F_{\mathbb{B}}^{(\mathtt{inv})}\left(f_j(p_j)\right)\right) \\ &= F_{\mathbb{B}}\left(\sum_{i \in \mathbb{B}} f_i^{(\mathtt{inv})}\left(f_j(p_j)\right)\right) \quad \text{(A.5)} \\ &= F_{\mathbb{B}}\left(\sum_{i \in \mathbb{B}} f_i^{(\mathtt{inv})}\left(f_i(p_i)\right)\right) \quad \text{(A.6)} \\ &= F_{\mathbb{B}}\left(\sum_{i \in \mathbb{B}} p_i\right) = F_{\mathbb{B}}\left(q_{\mathbb{B}}\right) \end{aligned}$$

where (A.5) follows the definition of the aggregate mutual information $F_{\mathbb{B}}$, and (A.6) holds because, according to condition (iii) in $\mathcal{P}(\mathbb{B})$, $f_i(p_i) = f_j(p_j)$ for every $i \in \mathbb{B}$. By applying conditions (ii) and (iii) in $\mathcal{P}(\mathbb{B})$, we obtain for $j \notin \mathbb{B}$,

$$f_j(q_j) = f_j(p_j) < \max_{1 \le i \le N} f_i(p_i) = \max_{i \in \mathbb{B}} f_i(p_i) = F_{\mathbb{B}}(q_{\mathbb{B}}).$$

Hence, $\boldsymbol{q} \in \mathcal{Q}(\mathbb{B})$ and

$$\sum_{i \notin \mathbb{B}} f_i(p_i) + (K + |\mathbb{B}| - N) \max_{1 \le j \le N} f_j(p_j) = \sum_{i \notin \mathbb{B}} f_i(q_i) + (K + |\mathbb{B}| - N) F_{\mathbb{B}}(q_{\mathbb{B}}).$$

Accordingly,

$$\sup_{\boldsymbol{p}\in\mathcal{P}(\mathbb{B})}\left\{\sum_{i\notin\mathbb{B}}f_i(p_i)+(K+|\mathbb{B}|-N)\max_{1\le j\le N}f_j(p_j)\right\}$$
$$\le\sup_{\boldsymbol{q}\in\mathcal{Q}(\mathbb{B})}\left\{\sum_{i\notin\mathbb{B}}f_i(q_i)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}})\right\}. \quad \text{(A.7)}$$

3. In summary, (A.3) and (A.7) jointly imply that equality holds for both inequalities, and the relation between maximizers $\boldsymbol{p}^\diamond$ and $\boldsymbol{q}^\diamond$ should follow (A.2) and (A.4).

□

Proof of Theorem 1. We first obtain from the definition of $F_{\mathbb{B}}$ and Property 1 that for $q>0$,

$$\frac{1}{F'_{\mathbb{B}}(q)}=\frac{1}{F'_{\mathbb{B}}\left(F_{\mathbb{B}}^{(\mathtt{inv})}(F_{\mathbb{B}}(q))\right)}=\frac{1}{F'_{\mathbb{B}}\left(\sum_{i\in\mathbb{B}}f_i^{(\mathtt{inv})}(F_{\mathbb{B}}(q))\right)}=\sum_{i\in\mathbb{B}}\frac{1}{f'_i\left(f_i^{(\mathtt{inv})}(F_{\mathbb{B}}(q))\right)}.$$

We can then infer from Assumption 1 that $F'_{\mathbb{B}}(q)$ is a positive, strict decreasing and continuous function for $q>0$, and $F'_{\mathbb{B}}(0)=\lim_{\rho\downarrow 0}F'_{\mathbb{B}}(\rho)$. This implies that

$$\sum_{i\in\mathbb{B}}f_i(q_i)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}}) \quad \text{(A.8)}$$

is strictly concave for $\boldsymbol{q}\in\tilde{\mathcal{Q}}(\mathbb{B})$. By definition of $\boldsymbol{q}^\diamond$ (cf. Lemma 2),

$$\sup_{\boldsymbol{q}\in\mathcal{Q}(\mathbb{B})}\left\{\sum_{i\notin\mathbb{B}}f_i(q_i)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}})\right\}=\sum_{i\notin\mathbb{B}}f_i(q_i^\diamond)+(K+|\mathbb{B}|-N)F_{\mathbb{B}}(q_{\mathbb{B}}^\diamond).$$

Then, the assumption that $\boldsymbol{q}^\diamond\in\mathcal{Q}(\mathbb{B})$ and the strict concavity of (A.8) together imply that $\boldsymbol{q}^\diamond$ is the unique global maximizer that maximizes (A.8) over $\tilde{\mathcal{Q}}(\mathbb{B})$. Hence, $\tilde{\boldsymbol{q}}=\boldsymbol{q}^\diamond$. □

Appendix B

Proof of Theorem 2

Two properties regarding the optimal maximal-mutual-information index set $\mathbb{A}$ and the optimal power allocation $\boldsymbol{p}^*$ must be established before our presenting the proof of Theorem 2.

Property 2. Fix $P > 0$. The optimal maximal-mutual-information index set $\mathbb{A}$ and the optimal power allocation $\boldsymbol{p}^*$ satisfy the following two properties:

1. For $j \notin \mathbb{A}$,
$$f_j'(p_j^*) \begin{cases} = (K + |\mathbb{A}| - N) F_{\mathbb{A}}'(\sum_{i \in \mathbb{A}} p_i^*) & \text{if } p_j^* > 0 \\ \leq (K + |\mathbb{A}| - N) F_{\mathbb{A}}'(\sum_{i \in \mathbb{A}} p_i^*) & \text{if } p_j^* = 0. \end{cases} \tag{B.1}$$
2. For any $i, j \in \{1, 2, \ldots, N\}$, if $f_i(p_i^*) < f_j(p_j^*)$, then
$$f_i'(p_i^*) \leq f_j'(p_j^*). \tag{B.2}$$

The first property indicates that the first derivative of the mutual information function attains the maximum value $(K + |\mathbb{A}| - N) F_{\mathbb{A}}'(\sum_{i \in \mathbb{A}} p_i^*)$ whenever its respective allocated power is positive. It also indicates that $q_{\mathbb{A}}^* = \sum_{i \in \mathbb{A}} p_i^* > 0$ for $P > 0$. The second property reveals that for the optimal power assignment, a larger mutual information cannot have a smaller first derivative. These two properties will be the basis of the prove-by-contradiction technique adopted in the proof of Theorem 2.

Proof of Property 2. We first observe that $\min_{j\in\mathbb{A}} p_j^* > 0$. This is because if $p_j^* = 0$ for some $j \in \mathbb{A}$, then (3.2) gives that $\max_{1\le i\le N} f_i(p_i^*) = f_j(p_j^*) = f_j(0) = 0$. Thus, $p_i^* = 0$ for every $1 \le i \le N$, and $P = \sum_{i=1}^N p_i^* = 0$, which contradicts the assumption that $P > 0$.

We next note that the Lagrange multipliers technique and KKT condition imply that the first derivatives of the mutual information functions achieve the maximum for those indices whose corresponding allocated powers are positive. Then by replacing $\mathbb{B}$ by $\mathbb{A}$ in (3.9) and by noting

$$q_{\mathbb{A}}^* = \sum_{i\in\mathbb{A}} p_i^* \ge \min_{i\in\mathbb{A}} p_i^* > 0,$$

we obtain that for $j \notin \mathbb{A}$,

$$p_j^* = q_j^* \quad \text{and} \quad f_j'(q_j^*) \begin{cases} = (K + |\mathbb{A}| - N)F_{\mathbb{A}}'(q_{\mathbb{A}}^*) & \text{if } q_j^* > 0 \\ \le (K + |\mathbb{A}| - N)F_{\mathbb{A}}'(q_{\mathbb{A}}^*) & \text{if } q_j^* = 0. \end{cases}$$

This completes the proof of (B.1).

The proof of (B.2) can be done as follows. Suppose

$$f_i(p_i^*) < f_j(p_j^*) \quad \text{and} \quad f_i'(p_i^*) \ge f_j'(p_j^*).$$

Then, $f_j(p_j^*) > 0$ and hence $p_j^* > 0$ by Assumption 1. By the continuity of functions f_i, f_i', f_j and f_j', there exists $0 < \Delta p < p_j^*$ such that

$$f_i(p_i^* + \Delta p) < f_j(p_j^* - \Delta p) \quad \text{and} \quad f_i'(p_i^* + \Delta p) > f_j'(p_j^* - \Delta p).$$

Hence,

$$\begin{aligned} f_i(p_i^* + \Delta p) + f_j(p_j^* - \Delta p) &> [f_i(p_i^*) + \Delta p \cdot f_i'(p_i^* + \Delta p)] + \left[f_j(p_j^*) - \Delta p \cdot f_j'(p_j^* - \Delta p)\right] \quad \text{(B.3)} \\ &= f_i(p_i^*) + f_j(p_j^*) + \Delta p \left[f_i'(p_i^* + \Delta p) - f_j'(p_j^* - \Delta p)\right] \\ &> f_i(p_i^*) + f_j(p_j^*) \end{aligned}$$

where (B.3) follows a relation derived from Assumption 1 that for every $1 \le k \le N$ and $0 < \Delta p < p$,

$$f_k'(p + \Delta p) < \frac{f_k(p + \Delta p) - f_k(p)}{\Delta p} < f_k'(p) < \frac{f_k(p) - f_k(p - \Delta p)}{\Delta p} < f_k'(p - \Delta p).$$

Noting that i must be outside $\mathbb{A}$ since $f_i(p_i^*) < f_j(p_j^*)$, we distinguish between two cases below: $j \notin \mathbb{A}$ and $j \in \mathbb{A}$.

First, if $j \notin \mathbb{A}$, then

$$\begin{aligned}
&\sum_{k\notin\mathbb{A}} f_k(p_k^*) + (K + |\mathbb{A}| - N) \max_{1\le m\le N} f_m(p_m^*) \\
&= \left(\sum_{k\notin\mathbb{A}\setminus\{i,j\}} f_k(p_k^*) + f_i(p_i^*) + f_j(p_j^*) \right) + (K + |\mathbb{A}| - N) \max_{1\le m\le N} f_m(p_m^*) \\
&< \sum_{k\notin\mathbb{A}\setminus\{i,j\}} f_k(p_k^*) + f_i(p_i^* + \Delta p) + f_j(p_j^* - \Delta p) + (K + |\mathbb{A}| - N) \max_{1\le m\le N} f_m(p_m^*)
\end{aligned}$$

which contradicts the optimality of $\boldsymbol{p}^*$. If however $j \in \mathbb{A}$, then

$$\begin{aligned}
&\sum_{k\notin\mathbb{A}} f_k(p_k^*) + (K + |\mathbb{A}| - N) \max_{1\le m\le N} f_m(p_m^*) \\
&= \left(\sum_{k\notin\mathbb{A}\setminus\{i\}} f_k(p_k^*) + f_i(p_i^*) \right) + \left(f_j(p_j^*) + (K + |\mathbb{A}| - N - 1) \max_{1\le m\le N} f_m(p_m^*) \right) \\
&< \sum_{k\notin\mathbb{A}\setminus\{i\}} f_k(p_k^*) + f_i(p_i^* + \Delta p) + f_j(p_j^* - \Delta p) + (K + |\mathbb{A}| - N - 1) \max_{1\le m\le N} f_m(p_m^*)
\end{aligned}$$

which again contradicts the optimality of $\boldsymbol{p}^*$. The proof of (B.2) is therefore completed. □

We are now ready to present the proof of Theorem 2.

Proof of Theorem 2. The proof of the theorem is divided into two parts.

In the forward part, we will show by induction that when $|\mathbb{A}| < N$,

$$\begin{cases}
\mathbb{A} \subseteq \mathbb{B}_{M+1} & \text{(B.4a)} \\
(K - M) F'_{\mathbb{B}_{M+1}} \left(\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M,i} \right) > f'_{j_M}(\tilde{p}_{M,j_M}) & \text{(B.4b)} \\
\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M+1,i} \ \left(= \tilde{q}_{\mathbb{B}_{M+1}} \right) > 0 & \text{(B.4c)}
\end{cases}$$

hold for every $1 \le M \le N - |\mathbb{A}|$. Condition (B.4b) then ensures that stop criterion (3.11) is violated for every $1 \le M \le m$ for some $m \in \{N - |\mathbb{A}|, N - |\mathbb{A}| + 1, \ldots, K - 1\}$ when $|\mathbb{A}| < N$; hence the algorithm will not stop before finding $\mathbb{A}$ and $\boldsymbol{p}^*$. Notably, the definition of $\mathbb{A}$ in (3.2) guarantees that $N - |\mathbb{A}| \le K - 1$; hence the set $\{N - |\mathbb{A}|, N - |\mathbb{A}| + 1, \ldots, K - 1\}$

can never be empty. Also, according to $\tilde{p}_{M+1,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{B}_{M+1}}(\tilde{q}_{\mathbb{B}_{M+1}}))$ from (3.7), (B.4c) is equivalent to $\tilde{p}_{M+1,i} > 0$ for every $i \in \mathbb{B}_{M+1}$.

After confirmation of the forward part, the converse part will subsequently be proved by induction, namely, if stop criterion (3.11) is violated for every $1 \le M \le m$, then $m \le N - |\mathbb{A}|$. An immediate consequence of the converse is that when $|\mathbb{A}| = N$, the stop criterion (3.11) must hold at $M = 1$ because the converse can be equivalently stated (by taking $m = 1$) as that $|\mathbb{A}| = N > N - 1$ implying the validity of (3.11) at $M = 1$.

Then the forward and converse parts together conclude that the smallest integer $M_{\min}$ that validates the stop criterion (3.11) is exactly $N - |\mathbb{A}| + 1$. The desired result $\mathbb{A} = \mathbb{B}_{M_{\min}}$ is therefore confirmed by deriving

$$|\mathbb{B}_{M_{\min}}| = N - M_{\min} + 1 = |\mathbb{A}|$$

and by applying (B.4a) (which has been proved to be valid for every $1 \le M \le N - |\mathbb{A}| = M_{\min} - 1$ in the forward part) to obtain

$$\mathbb{A} \subseteq \mathbb{B}_{M_{\min}} \left(\subseteq \mathbb{B}_{M_{\min}-1} \subseteq \cdots \subseteq \mathbb{B}_2 \right).$$

Note that stop criterion (3.11) trivially holds when $M = K$; hence, the above statement is applicable even at the extreme case that $|\mathbb{A}| = N - K + 1$.

A. Forward part: Under $|\mathbb{A}| < N$, (B.4a)–(B.4c) are valid for every $1 \le M \le N - |\mathbb{A}|$.

1. *Preliminary*:

   In comparison with the optimal power allocation $\boldsymbol{p}^*$, we define two index sets as follows:

$$\mathbb{I}_M \triangleq \{i \in \mathbb{B}_M : \tilde{p}_{M,i} < p_i^*\} \quad \text{and} \quad \mathbb{D}_M \triangleq \{i \in \mathbb{B}_M : \tilde{p}_{M,i} > p_i^*\}.$$

From their definitions, $\mathbb{I}_M$ consists of all indices in $\mathbb{B}_M$, corresponding to which $\tilde{p}_{M,i}$ needs to be *increased* to reach p_i^*. In contrast, for $i \in \mathbb{D}_M$, we shall *decrease* $\tilde{p}_{M,i}$ to achieve p_i^*.

We then claim that for the considered range $1 \leq M \leq N - |\mathbb{A}|$, $\mathbb{I}_M$ and $\mathbb{D}_M$ cannot be both empty, given that $\mathbb{A} \subseteq \mathbb{B}_M$. This is because if both of them were empty, then $\tilde{p}_{M,i} = p_i^*$ for every $i \in \mathbb{B}_M$. We then notice from (3.7) that function values $\{f_i(\tilde{p}_{M,i})\}$ are all equal for $i \in \mathbb{B}_M$. Also, by definition of $\mathbb{A}$ from (3.2), the set $\mathbb{A}$ should contain all indices whose respective function values $\{f_i(p_i^*)\}$ are equal. Thus, $A \subseteq \mathbb{B}_m$ and $\tilde{p}_{M,i} = p_i^*$ for $i \in \mathbb{B}_M$ immediately imply that $\mathbb{A} = \mathbb{B}_M$. Accordingly, $\mathbb{I}_M = \mathbb{D}_M = \emptyset$ can only occur when $M = N - |\mathbb{B}_M| + 1 = N - |\mathbb{A}| + 1$, which is outside the range $1 \leq M \leq N - |\mathbb{A}|$ that we consider here. The claim is thus validated.

2. *Validity of* (B.4a)–(B.4c) *when* $M = 1$:

Observe that $\mathbb{A}$ is always a subset of $\mathbb{B}_1 = \{1, 2, \ldots, N\}$, so we know from the claim in 1) *Preliminary* that $\mathbb{I}_1$ and $\mathbb{D}_1$ cannot be both empty. Based on this, we can further reason from

$$\sum_{i=1}^{N} \tilde{p}_{1,i} = \sum_{i=1}^{N} p_i^* = P$$

that $\mathbb{I}_1$ and $\mathbb{D}_1$ are both non-empty. Since

$$f_i(p_i^*) = \max_{1 \leq j \leq N} f_j(p_j^*) \quad \text{for } i \in \mathbb{A} \ (\subseteq \mathbb{B}_1)$$

and

$$f_i(\tilde{p}_{1,i}) \ \big(= F_{\mathbb{B}_1}(P) \big) \ < \max_{j \in \mathbb{B}_1} f_j(p_j^*) = \max_{1 \leq j \leq N} f_j(p_j^*) \quad \text{for } i \in \mathbb{B}_1 \ (\text{as } \mathbb{I}_1 \neq \emptyset) \qquad \text{(B.5)}$$

we can infer from the strict increasingness of functions $\{f_i\}$ that $\tilde{p}_{1,i} < p_i^*$ for every $i \in \mathbb{A}$, which indicates $\mathbb{A} \subseteq \mathbb{I}_1$.

We then claim and will prove by contradiction that j_1 obtained from

$$j_1 = \arg\min_{i \in \mathbb{B}_1} f_i'(\tilde{p}_{1,i}) \tag{B.6}$$

does not belong to $\mathbb{I}_1$, and therefore is not contained in $\mathbb{A}$. This will immediately yield $\mathbb{A} \subseteq \mathbb{B}_2 = \mathbb{B}_1 \setminus \{j_1\}$.

Suppose $j_1 \in \mathbb{I}_1$. Then the definition of $\mathbb{I}_1$ implies

$$\tilde{p}_{1,j_1} < p_{j_1}^*. \tag{B.7}$$

Because $\sum_{i=1}^N \tilde{p}_{1,i} = \sum_{i=1}^N p_i^* = P$, (B.7) further implies the existence of another index k (in $\mathbb{B}_1$) such that

$$\tilde{p}_{1,k} > p_k^*. \tag{B.8}$$

From (3.7) and (B.6), we respectively obtain

$$f_{j_1}(\tilde{p}_{1,j_1}) = f_k(\tilde{p}_{1,k}) = F_{\mathbb{B}_1}(P) \quad \text{and} \quad f_{j_1}'(\tilde{p}_{1,j_1}) \leq f_k'(\tilde{p}_{1,k}).$$

Then by (B.7), (B.8), the strict increasingness of functions f_{j_1} and f_k, and the strict decreasingness of functions f_{j_1}' and f_k', we have

$$f_{j_1}(p_{j_1}^*) > f_k(p_k^*) \quad \text{and} \quad f_{j_1}'(p_{j_1}^*) < f_k'(p_k^*)$$

which contradicts (B.2) in Property 2. Accordingly, $j_1 \notin \mathbb{I}_1$.

Next, we prove that (B.4b) is valid when $M = 1$. Using the prove-by-contradiction technique, we suppose

$$(K-1)F_{\mathbb{B}_2}'\left(\textstyle\sum_{i \in \mathbb{B}_2} \tilde{p}_{1,i}\right) \leq f_{j_1}'(\tilde{p}_{1,j_1}). \tag{B.9}$$

Since for every $1 \leq i \leq N$,

$$\tilde{p}_{1,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{B}_1}(P)) > 0 \quad (\text{as } P > 0)$$

and $\{f_i(\tilde{p}_{1,i})\}$ are all equal (to $F_{\mathbb{B}_1}(P)$), we obtain from Property 1 that

$$\frac{1}{F'_{\mathbb{B}_2}\left(\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right)} = \sum_{i\in\mathbb{B}_2}\frac{1}{f'_i(\tilde{p}_{1,i})} = \sum_{i\in\mathbb{A}}\frac{1}{f'_i(\tilde{p}_{1,i})} + \sum_{i\in\mathbb{B}_2\setminus\mathbb{A}}\frac{1}{f'_i(\tilde{p}_{1,i})}.$$

Accordingly,

$$\begin{aligned}
\sum_{i\in\mathbb{A}}\frac{1}{f'_i(\tilde{p}_{1,i})} &= \frac{1}{F'_{\mathbb{B}_2}\left(\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right)} - \sum_{i\in\mathbb{B}_2\setminus\mathbb{A}}\frac{1}{f'_i(\tilde{p}_{1,i})} \\
&\geq \frac{K-1}{f'_{j_1}(\tilde{p}_{1,j_1})} - \sum_{i\in\mathbb{B}_2\setminus\mathbb{A}}\frac{1}{f'_i(\tilde{p}_{1,i})} && \text{(B.10)} \\
&\geq \frac{K-1}{f'_{j_1}(\tilde{p}_{1,j_1})} - \sum_{i\in\mathbb{B}_2\setminus\mathbb{A}}\frac{1}{f'_{j_1}(\tilde{p}_{1,j_1})} && \text{(B.11)} \\
&= \frac{1}{f'_{j_1}(\tilde{p}_{1,j_1})}\left[(K-1) - |\mathbb{B}_2\setminus\mathbb{A}|\right] \\
&= \frac{1}{f'_{j_1}(\tilde{p}_{1,j_1})}\left[(K-1) - (|\mathbb{B}_2| - |\mathbb{A}|)\right] && \text{(B.12)} \\
&= \frac{1}{f'_{j_1}(\tilde{p}_{1,j_1})}(K + |\mathbb{A}| - N) && \text{(B.13)}
\end{aligned}$$

where (B.10) follows from (B.9), (B.11) is based on (B.6), (B.12) is due to $\mathbb{A}\subseteq\mathbb{B}_2$, and (B.13) is true because $|\mathbb{B}_2| = N-1$. Then based on $\mathbb{A}\subseteq\mathbb{I}_1$ and $j_1\notin\mathbb{I}_1$, we know $\tilde{p}_{1,i} < p_i^*$ for every $i\in\mathbb{A}$ and $\tilde{p}_{1,j_1}\geq p_{j_1}^*$. Hence, from the strict decreasingness of functions $\{f'_i\}$, (B.13) implies that

$$f'_{j_1}(p_{j_1}^*) > (K-N+|\mathbb{A}|)\frac{1}{\sum_{i\in\mathbb{A}}\frac{1}{f'_i(p_i^*)}} = (K-N+|\mathbb{A}|)F'_{\mathbb{A}}\left(\sum_{i\in\mathbb{A}}p_i^*\right)$$

which contradicts (B.1) in Property 2.

After proving (B.4a) and (B.4b) at $M=1$, what remains to be confirmed is the validity of (B.4c). Using the prove-by-contradiction technique, we first suppose (B.4c) were not true when $M=1$, i.e., $\sum_{i\in\mathbb{B}_2}\tilde{p}_{2,i}=0$. Then, we obtain

$$\tilde{p}_{2,j_1} = \sum_{i=1}^{N}\tilde{p}_{2,i} - \sum_{i\in\mathbb{B}_2}\tilde{p}_{2,i} = P > 0.$$

The KKT condition follows that

$$(K-1)F'_{\mathbb{B}_2}\left(\sum_{i\in\mathbb{B}_2}\tilde{p}_{2,i}\right) \le f'_{j_1}(\tilde{p}_{2,j_1}).$$

The strict decreasingness of functions $F'_{\mathbb{B}_2}$ and f'_{j_1}, together with the straightforward relations:

$$\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} \ge \sum_{i\in\mathbb{B}_2}\tilde{p}_{2,i} = 0 \quad \text{and} \quad \tilde{p}_{1,j_1} \le \tilde{p}_{2,j_1} = P,$$

implies

$$(K-1)F'_{\mathbb{B}_2}\left(\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right) \le f'_{j_1}(\tilde{p}_{1,j_1})$$

which contradicts (B.4b) at $M=1$. The proof of the case $M=1$ is then completed.

3. *Validity of* (B.4a)–(B.4c) *at* $M = m-1$ *implying their validity at* $M = m$ *for* $2 \le m \le N - |\mathbb{A}|$:

   Based on the premise that (B.4a) is true at $M = m-1$ (i.e., $\mathbb{A} \subseteq \mathbb{B}_m$), we know from the discussion in the *Preliminary* on page 51 that $\mathbb{I}_m$ and $\mathbb{D}_m$ cannot be both empty. So when $\mathbb{I}_m$ is not empty, we will show $\mathbb{A} \subseteq \mathbb{I}_m$ and $j_m \notin \mathbb{I}_m$ as similar to the proof at $M=1$, which immediately gives $\mathbb{A} \subseteq \mathbb{B}_{m+1} = \mathbb{B}_m \setminus \{j_m\}$. We will then show by contradiction that $\mathbb{I}_m = \emptyset$ and $\mathbb{D}_m \neq \emptyset$ can never occur if (B.4a) and (B.4c) are both true at $M = m-1$. The desired result, i.e., $\mathbb{A} \subseteq \mathbb{B}_m$ implies $\mathbb{A} \subseteq \mathbb{B}_{m+1}$ for every $2 \le m \le N - |\mathbb{A}|$, is therefore verified.

   Case 1) $\mathbb{I}_m \neq \emptyset$.

   Since

$$f_i(p_i^*) = \max_{1\le j\le N} f_j(p_j^*) \quad \text{for } i \in \mathbb{A}\ (\subseteq \mathbb{B}_m)$$

and

$$f_i(\tilde{p}_{m,i})\ \left(= F_{\mathbb{B}_m}(\tilde{q}_{\mathbb{B}_m})\right)\ < \max_{j\in\mathbb{B}_m} f_j(p_j^*) \le \max_{1\le j\le N} f_j(p_j^*) \quad \text{for } i \in \mathbb{B}_m\ (\text{as } \mathbb{I}_m \neq \emptyset)$$

we can infer from the strict increasingness of functions $\{f_i\}$ that $\tilde{p}_{m,i} < p_i^*$ for every $i \in \mathbb{A}$, which indicates $\mathbb{A} \subseteq \mathbb{I}_m$.

In order to prove $j_m \notin \mathbb{I}_m$, we need to first show the existence of an index $k \in \mathbb{B}_m$ such that $\tilde{p}_{m,k} \geq p_k^*$. This can be proved by contradiction. Suppose no such index exists in $\mathbb{B}_m$ (i.e., $\mathbb{I}_m = \mathbb{B}_m$). Because $\sum_{i=1}^N \tilde{p}_{m,i} = \sum_{i=1}^N p_i^* = P$, there must exist an index u outside $\mathbb{B}_m$, satisfying $\tilde{p}_{m,u} > p_u^*$. Since $p_u^* \geq 0$, we know $\tilde{p}_{m,u}$ must be strictly positive. Then we can derive by the KKT condition (i.e., the first derivatives of the mutual information functions with positive allocated powers should achieve the maximum) that

$$(K-m+1)F'_{\mathbb{B}_m}\left(\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i}\right) = f'_u(\tilde{p}_{m,u}) \tag{B.14}$$

where equality in (B.14) follows from the validity of (B.4c) at $M = m-1$, i.e., $\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i} > 0$. Since the validity of (B.4c) at $M = m-1$ is equivalent to $\tilde{p}_{m,i} > 0$ for $i \in \mathbb{B}_m$, and function values $\{f_i(\tilde{p}_{m,i})\}$ are all equal also for $i \in \mathbb{B}_m$, we have from Property 1 that

$$\begin{aligned}&\frac{1}{(K-m+1)}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(\tilde{p}_{m,i})}\\ &= \frac{1}{(K-m+1)}\cdot\frac{1}{F'_{\mathbb{B}_m}\left(\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i}\right)} = \frac{1}{f'_u(\tilde{p}_{m,u})}.\end{aligned} \tag{B.15}$$

Based on (B.15), we can further reason from

$$\tilde{p}_{m,i} < p_i^* \text{ for } i \in \mathbb{I}_m = \mathbb{B}_m, \quad \tilde{p}_{m,u} > p_u^*$$

and the strict decreasingness of functions $\{f'_i\}$ that

$$\frac{1}{(K-m+1)}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(p_i^*)} > \frac{1}{f'_u(p_u^*)}. \tag{B.16}$$

Using again that $p_i^* > \tilde{p}_{m,i}$ for $i \in \mathbb{B}_m = \mathbb{I}_m$, we know that $p_i^* > 0$ for $i \in \mathbb{B}_m$. Property

(B.1) then leads to

$$\begin{aligned}
&\frac{1}{K-m+1}\sum_{i\in\mathbb{B}_m}\frac{1}{f_i'(p_i^*)}\\
&=\frac{1}{K-m+1}\left(\sum_{i\in\mathbb{A}}\frac{1}{f_i'(p_i^*)}+\sum_{i\in\mathbb{B}_m\setminus\mathbb{A}}\frac{1}{f_i'(p_i^*)}\right)\\
&=\frac{1}{K-m+1}\left(\frac{1}{F_{\mathbb{A}}'(\sum_{i\in\mathbb{A}}p_i^*)}+\sum_{i\in\mathbb{B}_m\setminus\mathbb{A}}\frac{1}{(K+|\mathbb{A}|-N)}\frac{1}{F_{\mathbb{A}}'(\sum_{i\in\mathbb{A}}p_i^*)}\right)\\
&=\frac{1}{(K+|\mathbb{A}|-N)}\frac{1}{F_{\mathbb{A}}'(\sum_{i\in\mathbb{A}}p_i^*)}\\
&\le\frac{1}{f_u'(p_u^*)}
\end{aligned}$$

where the last inequality cannot be replaced by an equality because p_u^* may be zero, and in the above derivation, we have implicitly applied the validity of (B.4a) at $M=m-1$ to obtain $|\mathbb{B}_m\setminus\mathbb{A}|=|\mathbb{B}_m|-|\mathbb{A}|=N-m+1-|\mathbb{A}|$. A contradiction to (B.16) is thus obtained. Accordingly, $I_m\neq\mathbb{B}_m$, and hence confirmation of the existence of $k\in\mathbb{B}_m$ such that $\tilde{p}_{m,k}\ge p_k^*$ is completed.

We can now proceed to prove that $j_m\notin\mathbb{I}_m$ by contradiction. Suppose $j_m\in\mathbb{I}_m$; hence, $\tilde{p}_{m,j_m}<p_{j_m}^*$. Then following four observations below:

(a) $f_{j_m}(\tilde{p}_{m,j_m})=f_k(\tilde{p}_{m,k})$,

(b) $f_{j_m}'(\tilde{p}_{m,j_m})=\min_{i\in\mathbb{B}_m}f_i'(\tilde{p}_{m,i})\le f_k'(\tilde{p}_{m,k})$,

(c) the strict increasingness of functions f_{j_m} and f_k, and

(d) the strict decreasingness of functions f_{j_m}' and f_k',

we have

$$f_{j_m}(p_{j_m}^*)>f_k(p_k^*)\quad\text{and}\quad f_{j_m}'(p_{j_m}^*)<f_k'(p_k^*)$$

which contradicts (B.2) in Property 2. Accordingly, $j_m\notin\mathbb{I}_m$. This finishes the proof of $\mathbb{A}\subseteq\mathbb{B}_{m+1}=\mathbb{B}_m\setminus\{j_m\}$ given $\mathbb{I}_m\neq\emptyset$.

Case 2) $\mathbb{I}_m = \emptyset$ but $\mathbb{D}_m \neq \emptyset$.

Our goal is to show that this case can never happen if (B.4a) and (B.4c) are valid at $M = m-1$. Since $\mathbb{I}_m = \emptyset$ and $\mathbb{D}_m \neq \emptyset$, and $\sum_{i=1}^N \tilde{p}_{m,i} = \sum_{i=1}^N p_i^* = P$, there exists an index u outside $\mathbb{B}_m$, satisfying $\tilde{p}_{m,u} < p_u^*$. Thus, the KKT condition implies that

$$(K-m+1)F'_{\mathbb{B}_m}\left(\textstyle\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i}\right) \geq f'_u(\tilde{p}_{m,u}) \tag{B.17}$$

where the above inequality follows the premise that (B.4c) is true at $M = m-1$. Since the function values $\{f_i(\tilde{p}_{m,i})\}$ are all equal for $i \in \mathbb{B}_m$, we have from (B.17) and Property 1 that

$$\frac{1}{(K-m+1)}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(\tilde{p}_{m,i})} = \frac{1}{(K-m+1)}\frac{1}{F'_{\mathbb{B}_m}\left(\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i}\right)} \leq \frac{1}{f'_u(\tilde{p}_{m,u})}. \tag{B.18}$$

Based on (B.18), we can then reason from

$$\tilde{p}_{m,i} \geq p_i^* \text{ for } i \in \mathbb{B}_m, \quad \tilde{p}_{m,u} < p_u^*$$

and also the strict decreasingness of functions $\{f'_i\}$ that

$$\frac{1}{(K-m+1)}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(p_i^*)} < \frac{1}{f'_u(p_u^*)}. \tag{B.19}$$

By noticing that $p_u^* > \tilde{p}_{m,u}$ implies $p_u^* > 0$, we can derive using (B.1) in Property 2 that

$$\begin{aligned}
&\frac{1}{(K-m+1)}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(p_i^*)} \\
&= \frac{1}{(K-m+1)}\left(\sum_{i\in\mathbb{A}}\frac{1}{f'_i(p_i^*)} + \sum_{i\in\mathbb{B}_m\setminus\mathbb{A}}\frac{1}{f'_i(p_i^*)}\right) \\
&\geq \frac{1}{(K-m+1)}\left(\frac{1}{F'_{\mathbb{A}}(\sum_{i\in\mathbb{A}}p_i^*)} + \sum_{i\in\mathbb{B}_m\setminus\mathbb{A}}\frac{1}{(K+|\mathbb{A}|-N)}\frac{1}{F'_{\mathbb{A}}(\sum_{i\in\mathbb{A}}p_i^*)}\right) \\
&= \frac{1}{(K+|\mathbb{A}|-N)}\frac{1}{F'_{\mathbb{A}}(\sum_{i\in\mathbb{A}}p_i^*)} \\
&= \frac{1}{f'_u(p_u^*)}
\end{aligned}$$

where we again implicitly use $|\mathbb{B}_m \setminus \mathbb{A}| = |\mathbb{B}_m| - |\mathbb{A}| = N - m + 1 - |\mathbb{A}|$ (i.e., $\mathbb{A} \subseteq \mathbb{B}_m$) in the above derivation. A contradiction to (B.19) is thus obtained. Therefore, $\mathbb{I}_m = \emptyset$ and $\mathbb{D}_m \neq \emptyset$ cannot occur if (B.4a) and (B.4c) are valid at $M = m - 1$.

After the completion of the proof for $\mathbb{A} \subseteq \mathbb{B}_{m+1}$, we next prove that (B.4b) is valid at $M = m$. Using the prove-by-contradiction technique, we suppose (B.4b) were not true at $M = m$. Then, using the just proved $\mathbb{A} \subseteq \mathbb{B}_{m+1}$, the validity of (B.4c) at $M = m - 1$ (i.e., $\tilde{p}_{m,i} > 0$ for $i \in \mathbb{B}_m$), and the observation that the function values $\{f_i(\tilde{p}_{m,i})\}$ are all equal for $i \in \mathbb{B}_{m+1}$, we obtain from Property 1 that

$$\frac{1}{F'_{\mathbb{B}_{m+1}}\left(\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m,i}\right)} = \sum_{i\in\mathbb{B}_{m+1}} \frac{1}{f'_i(\tilde{p}_{m,i})} = \sum_{i\in\mathbb{A}} \frac{1}{f'_i(\tilde{p}_{m,i})} + \sum_{i\in\mathbb{B}_{m+1}\setminus\mathbb{A}} \frac{1}{f'_i(\tilde{p}_{m,i})}.$$

Accordingly,

$$\begin{aligned}
\sum_{i\in\mathbb{A}} \frac{1}{f'_i(\tilde{p}_{m,i})} &= \frac{1}{F'_{\mathbb{B}_{m+1}}\left(\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m,i}\right)} - \sum_{i\in\mathbb{B}_{m+1}\setminus\mathbb{A}} \frac{1}{f'_i(\tilde{p}_{m,i})} \\
&\geq \frac{K-m}{f'_{j_m}(\tilde{p}_{m,j_m})} - \sum_{i\in\mathbb{B}_{m+1}\setminus\mathbb{A}} \frac{1}{f'_i(\tilde{p}_{m,i})} && \text{(B.20)} \\
&\geq \frac{K-m}{f'_{j_m}(\tilde{p}_{m,j_m})} - \sum_{i\in\mathbb{B}_{m+1}\setminus\mathbb{A}} \frac{1}{f'_{j_m}(\tilde{p}_{m,j_m})} && \text{(B.21)} \\
&= \frac{1}{f'_{j_m}(\tilde{p}_{m,j_m})}[(K-m) - |\mathbb{B}_{m+1}\setminus\mathbb{A}|] \\
&= \frac{1}{f'_{j_m}(\tilde{p}_{m,j_m})}[(K-m) - (|\mathbb{B}_{m+1}| - |\mathbb{A}|)] && \text{(B.22)} \\
&= \frac{1}{f'_{j_m}(\tilde{p}_{m,j_m})}(K + |\mathbb{A}| - N) && \text{(B.23)}
\end{aligned}$$

where (B.20) follows from the assumed violation of (B.4b) at $M = m$, (B.21) is based on $j_m = \arg\min_{i\in\mathbb{B}_m} f'_i(\tilde{p}_{m,i})$, (B.22) is due to $\mathbb{A} \subseteq \mathbb{B}_{m+1}$, and (B.23) is true because $|\mathbb{B}_{m+1}| = N - m$. Since $\mathbb{A} \subseteq \mathbb{I}_m$ and $j_m \notin \mathbb{I}_m$, we know $\tilde{p}_{m,i} < p^*_i$ for $i \in \mathbb{A}$ and $\tilde{p}_{m,j_m} \geq p^*_{j_m}$. Hence, from the strict decreasingness of functions $\{f'_i\}$, (B.23) implies that

$$f'_{j_m}(p^*_{j_m}) > (K - N + |\mathbb{A}|)\frac{1}{\sum_{i\in\mathbb{A}} \frac{1}{f'_i(p^*_i)}} = (K - N + |\mathbb{A}|)F'_{\mathbb{A}}(\textstyle\sum_{i\in\mathbb{A}} p^*_i).$$

A contradiction to (B.1) is thus obtained.

After proving (B.4a) and (B.4b) at $M = m$, what remains to confirm is the validity of (B.4c). We require the next inequality to proceed:

$$(K-m)F'_{\mathbb{B}_{m+1}}\left(\sum_{i\in\mathbb{B}_{m+1}}\tilde{p}_{m,i}\right) > \max_{j\notin\mathbb{B}_{m+1}} f'_j(\tilde{p}_{m,j}) \tag{B.24}$$

which can be proved as follows. Given that (B.4c) is true at $M = m-1$ (i.e., $\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i} > 0$), we know from the KKT condition that

$$(K-m+1)F'_{\mathbb{B}_m}\left(\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i}\right) \geq \max_{j\notin\mathbb{B}_m} f'_j(\tilde{p}_{m,j}). \tag{B.25}$$

We can then derive

$$\begin{aligned}
&\frac{1}{(K-m)F'_{\mathbb{B}_{m+1}}\left(\sum_{i\in\mathbb{B}_{m+1}}\tilde{p}_{m,i}\right)} = \frac{1}{(K-m)}\sum_{i\in\mathbb{B}_{m+1}}\frac{1}{f'_i(\tilde{p}_{m,i})} \\
&= \frac{1}{K-m+1}\left(\sum_{i\in\mathbb{B}_{m+1}}\frac{1}{f'_i(\tilde{p}_{m,i})} + \frac{1}{K-m}\sum_{i\in\mathbb{B}_{m+1}}\frac{1}{f'_i(\tilde{p}_{m,i})}\right) \\
&= \frac{1}{K-m+1}\left(\sum_{i\in\mathbb{B}_{m+1}}\frac{1}{f'_i(\tilde{p}_{m,i})} + \frac{1}{(K-m)}\frac{1}{F'_{\mathbb{B}_{m+1}}(\sum_{i\in\mathbb{B}_{m+1}}\tilde{p}_{m,i})}\right) \\
&< \frac{1}{K-m+1}\left(\sum_{i\in\mathbb{B}_{m+1}}\frac{1}{f'_i(\tilde{p}_{m,i})} + \frac{1}{f'_{j_m}(\tilde{p}_{m,j_m})}\right) \qquad \text{(B.26)} \\
&= \frac{1}{K-m+1}\sum_{i\in\mathbb{B}_m}\frac{1}{f'_i(\tilde{p}_{m,i})} \\
&= \frac{1}{(K-m+1)}\frac{1}{F'_{\mathbb{B}_m}(\sum_{i\in\mathbb{B}_m}\tilde{p}_{m,i})} \\
&\leq \min_{j\notin\mathbb{B}_m}\left\{\frac{1}{f'_j(\tilde{p}_{m,j})}\right\} \qquad \text{(B.27)}
\end{aligned}$$

where (B.26) follows from the validity of (B.4b) at $M = m$, and (B.27) is a consequence of (B.25). By applying the validity of (B.4b) at $M = m$ again, we have

$$\begin{aligned}
(K-m)F'_{\mathbb{B}_{m+1}}\left(\sum_{i\in\mathbb{B}_{m+1}}\tilde{p}_{m,i}\right) &> \max\left\{f_{j_m}(\tilde{p}_{m,j_m}), \max_{j\notin\mathbb{B}_m} f'_j(\tilde{p}_{m,j})\right\} \\
&= \max_{j\notin\mathbb{B}_{m+1}} f'_j(\tilde{p}_{m,j}).
\end{aligned}$$

Inequality (B.24) is thus proved.

We proceed to prove (B.4c) by contradiction. Suppose (B.4c) were not true at $M = m$ (i.e., $\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m+1,i} = 0$). Then by the assumed validity of (B.4c) at $M = m-1$ (i.e., $\tilde{q}_{\mathbb{B}_m} = \sum_{i\in\mathbb{B}_m} \tilde{p}_{m,i} > 0$), we have

$$\tilde{p}_{m,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{B}_m}(\tilde{q}_{\mathbb{B}_m})) > 0 \quad \text{for } i \in \mathbb{B}_m$$

and hence

$$\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m,i} > \sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m+1,i} = 0 \tag{B.28}$$

because $\mathbb{B}_{m+1} \subset \mathbb{B}_m$. Inequality (B.28) and $\sum_{i=1}^N \tilde{p}_{m,i} = \sum_{i=1}^N \tilde{p}_{m+1,i} = P$ then implies

$$\sum_{j\notin\mathbb{B}_{m+1}} \tilde{p}_{m,i} < \sum_{j\notin\mathbb{B}_{m+1}} \tilde{p}_{m+1,i}$$

which immediately indicates that there exists $k \notin \mathbb{B}_{m+1}$ such that

$$\tilde{p}_{m,k} < \tilde{p}_{m+1,k} \quad \text{and} \quad \tilde{p}_{m+1,k} > 0. \tag{B.29}$$

The KKT condition thus follows that

$$(K-m)F'_{\mathbb{B}_{m+1}}\left(\textstyle\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m+1,i}\right) \le f'_k(\tilde{p}_{m+1,k}).$$

The strict decreasingness of functions $F'_{\mathbb{B}_{m+1}}$ and f'_k, together with (B.28) and (B.29), implies

$$(K-m)F'_{\mathbb{B}_{m+1}}\left(\textstyle\sum_{i\in\mathbb{B}_{m+1}} \tilde{p}_{m,i}\right) < f'_k(\tilde{p}_{m,k})$$

which contradicts (B.24). The proof of the forward part is thus completed.

B. Converse part: If the stop criterion (3.11) is violated for every $1 \le M \le m$, where $m \in \{1, 2, \ldots, K-1\}$, then $|\mathbb{A}| \le N - m$.

We now prove the converse part by induction.

1. *Validity of the converse statement at $m = 1$:*

It suffices to prove that $|\mathbb{A}| = N$ cannot be true if (3.11) is violated at $M = 1$, and we prove this by contradiction. Suppose $|A| = N$. Then, $\tilde{\boldsymbol{p}}_1 = \boldsymbol{p}^*$ is the optimal power allocation, and for every $1 \le i \le N$, $\tilde{p}_{1,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{A}}(P)) > 0$ because $P > 0$.

Observe that the violation of (3.11) at $M = 1$ tells that

$$(K-1)F'_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right) > f'_{j_1}(\tilde{p}_{1,j_1}).$$

By the continuity and strict decreasingness of functions $F'_{\mathbb{B}_2}$ and f'_{j_1}, there exists $0 < \Delta p < \tilde{p}_{1,j_1}$ such that

$$(K-1)F'_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p\right) > f'_{j_1}(\tilde{p}_{1,j_1} - \Delta p).$$

Hence, using an argument similar to (B.3) yields

$$\begin{aligned}
&(K-1)F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p\right) + f_{j_1}(\tilde{p}_{1,j_1} - \Delta p)\\
&> (K-1)\left[F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right) + \Delta p \cdot F'_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p\right)\right]\\
&\quad + \left[f_{j_1}(\tilde{p}_{1,j_1}) - \Delta p \cdot f'_{j_1}(\tilde{p}_{1,j_1} - \Delta p)\right]\\
&= (K-1)F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right) + f_{j_1}(\tilde{p}_{1,j_1})\\
&\quad + \Delta p\left[(K-1)F'_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p\right) - f'_{j_1}(\tilde{p}_{1,j_1} - \Delta p)\right]\\
&> (K-1)F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i}\right) + f_{j_1}(\tilde{p}_{1,j_1}).
\end{aligned} \tag{B.30}$$

Consequently, another power allocation

$$\hat{p}_j \triangleq \begin{cases} \tilde{p}_{1,j} - \Delta p & \text{if } j = j_1\\ f_j^{(\mathtt{inv})}\left(F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p\right)\right) & \text{otherwise}\end{cases}$$

which satisfies the power-sum constraint:

$$\begin{aligned}
\sum_{j=1}^{N}\hat{p}_j &= \hat{p}_{j_1} + \sum_{j\in\mathbb{B}_2}\hat{p}_j\\
&= (\tilde{p}_{1,j_1} - \Delta p) + \sum_{j\in\mathbb{B}_2} f_j^{(\mathtt{inv})}\left(F_{\mathbb{B}_2}(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p)\right)\\
&= (\tilde{p}_{1,j_1} - \Delta p) + F_{\mathbb{B}_2}^{(\mathtt{inv})}\left(F_{\mathbb{B}_2}(\textstyle\sum_{i\in\mathbb{B}_2}\tilde{p}_{1,i} + \Delta p)\right)\\
&= (\tilde{p}_{1,j_1} - \Delta p) + \left(\textstyle\sum_{j\in\mathbb{B}_2}\tilde{p}_{1,j} + \Delta p\right) = \sum_{j=1}^{N}\tilde{p}_{1,j} = P
\end{aligned}$$

will give that for every $j \in \mathbb{B}_2$,

$$\begin{aligned} f_j(\hat{p}_j) &= F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2} \tilde{p}_{1,i} + \Delta p\right) \\ &> F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2} \tilde{p}_{1,i}\right) \\ &= F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2} f_i^{(\mathtt{inv})}(F_{\mathbb{A}}(P))\right) \\ &= F_{\mathbb{B}_2}\left(F_{\mathbb{B}_2}^{(\mathtt{inv})}(F_{\mathbb{A}}(P))\right) \\ &= F_{\mathbb{A}}(P) \\ &= f_{j_1}(\tilde{p}_{1,j_1}) > f_{j_1}(\hat{p}_{j_1}) = f_{j_1}(\tilde{p}_{1,j_1} - \Delta p). \end{aligned}$$

The above inequality and (B.30) then jointly imply that for some $u \in \mathbb{B}_2$,

$$\begin{aligned} \min_{\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i=K\}} \textstyle\sum_{j=1}^N s_j f_j(\hat{p}_j) &= (K-1) f_u(\hat{p}_u) + f_{j_1}(\hat{p}_{j_1}) \\ &= (K-1) F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2} \tilde{p}_{1,i} + \Delta p\right) + f_{j_1}(\tilde{p}_{1,j_1} - \Delta p) \\ &> (K-1) F_{\mathbb{B}_2}\left(\textstyle\sum_{i\in\mathbb{B}_2} \tilde{p}_{1,i}\right) + f_{j_1}(\tilde{p}_{1,j_1}) = K \cdot F_{\mathbb{A}}(P) \\ &= \min_{\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i=K\}} \textstyle\sum_{j=1}^N s_j f_j(\tilde{p}_{1,j}). \end{aligned}$$

This indicates that $\tilde{\boldsymbol{p}}_1$ cannot be the optimal power allocation $\boldsymbol{p}^*$. The desired contradiction is thus obtained.

2. *Validity of the converse statement at $m = k+1$ implying its validity at $m = k$ for $2 \le k \le K-1$:*

   We are given that the stop criterion (3.11) is violated for every $1 \le M \le k-1$ and have already confirmed that $|A| \le N-k+1$. Now, since the stop criterion is violated again for $M = k$, we should then prove that $|A| \ne N-k+1$, which immediately implies the desired $|A| \le N-k$. We use the prove-by-contradiction technique.

   Suppose $|\mathbb{A}| = N-k+1$. Then, $\tilde{\boldsymbol{p}}_k = \boldsymbol{p}^*$ is the optimal power allocation, since we have already proven in the forward part that $\mathbb{A} \subseteq \mathbb{B}_{N-|\mathbb{A}|+1} = B_k$, which together with

$$|\mathbb{A}| = N-k+1 = |\mathbb{B}_k|$$

results in $\mathbb{A} = \mathbb{B}_k$. By definition of $\mathbb{A}$ in (3.2), we get

$$\max_{i\notin\mathbb{B}_k} \{f_i(\tilde{p}_{k,i})\} < f_j(\tilde{p}_{k,j}) \text{ for } j \in \mathbb{B}_k = \mathbb{A}. \tag{B.31}$$

Inequality (B.31) then implies that $f_{j_k}(\tilde{p}_{k,j_k}) > 0$; hence, $\tilde{p}_{k,j_k} > 0$. Observe that the violation of (3.11) at $M = k$ tells that

$$(K-k)F'_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i}\right) > f'_{j_k}(\tilde{p}_{k,j_k}).$$

By the continuity and strict decreasingness of functions $F'_{\mathbb{B}_{k+1}}$ and f'_{j_k}, there exists $0 < \Delta p < \tilde{p}_{k,j_k}$ such that

$$(K-k)F'_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i} + \Delta p\right) > f'_{j_k}(\tilde{p}_{k,j_k} - \Delta p).$$

Hence, through the same procedure as (B.30), we obtain

$$\begin{aligned}(K-k)F_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i} + \Delta p\right) + f_{j_k}(\tilde{p}_{k,j_k} - \Delta p)& \\ > (K-k)F_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i}\right) + f_{j_k}(\tilde{p}_{k,j_k}).& \quad \text{(B.32)}\end{aligned}$$

Consequently, another power allocation

$$\hat{p}_j \triangleq \begin{cases} \tilde{p}_{k,j} & \text{if } j \notin \mathbb{B}_k \\ \tilde{p}_{k,j} - \Delta p & \text{if } j = j_k \\ f_j^{(\text{inv})}(F_{\mathbb{B}_{k+1}}(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i} + \Delta p)) & \text{if } j \in \mathbb{B}_{k+1} \end{cases}$$

which satisfies the power-sum constriant:

$$\sum_{j=1}^{N}\hat{p}_j = \sum_{j\notin\mathbb{B}_k}\tilde{p}_{k,j} + (\tilde{p}_{k,j_k} - \Delta p) + \left(\sum_{j\in\mathbb{B}_{k+1}}\tilde{p}_{k,j} + \Delta p\right) = \sum_{j=1}^{N}\tilde{p}_{k,j} = P$$

will give that for every $j \in \mathbb{B}_{k+1}$,

$$\begin{aligned} f_j(\hat{p}_j) &= F_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i} + \Delta p\right) \\ &> F_{\mathbb{B}_{k+1}}\left(\sum_{i\in\mathbb{B}_{k+1}}\tilde{p}_{k,i}\right) = f_{j_k}(\tilde{p}_{k,j_k}) > \max\left\{f_{j_k}(\hat{p}_{j_k}), \max_{i\notin\mathbb{B}_k} f_i(\tilde{p}_{k,i})\right\} \quad \text{(B.33)}\end{aligned}$$

where the last strict inequality in (B.33) follows the strict increasingness of function

f_{j_k} and (B.31). The above inequality and (B.32) then jointly imply

$$\begin{aligned}
&\min_{\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i=K\}} \textstyle\sum_{j=1}^N s_j f_j(\hat{p}_j) \\
&= (K-k)F_{\mathbb{B}_{k+1}}\left(\textstyle\sum_{i\in\mathbb{B}_{k+1}} \hat{p}_{k,i}\right) + f_{j_k}(\hat{p}_{k,j_k}) + \sum_{i\notin\mathbb{B}_k} f_i(\hat{p}_{k,i}) \\
&= (K-k)F_{\mathbb{B}_{k+1}}\left(\textstyle\sum_{i\in\mathbb{B}_{k+1}} \tilde{p}_{k,i} + \Delta p\right) + f_{j_k}(\tilde{p}_{k,j_k} - \Delta p) + \sum_{i\notin\mathbb{B}_k} f_i(\tilde{p}_{k,i}) \\
&> (K-k)F_{\mathbb{B}_{k+1}}\left(\textstyle\sum_{i\in\mathbb{B}_{k+1}} \tilde{p}_{k,i}\right) + f_{j_k}(\tilde{p}_{k,j_k}) + \sum_{i\notin\mathbb{B}_k} f_i(\tilde{p}_{k,i}) \\
&= \min_{\{\boldsymbol{s}\in\{0,1\}^N:\sum_{i=1}^N s_i=K\}} \textstyle\sum_{j=1}^N s_j f_j(\tilde{p}_{k,j}).
\end{aligned}$$

This indicates that $\tilde{\boldsymbol{p}}_k$ cannot be the optimal power allocation $\boldsymbol{p}^*$. The desired contradiction is thus obtained.

□

Appendix C

Proofs of the Theorem, Lemmas and Corollaries in Chapter 5

C.1 Proof of Theorem 4

By noting $\tilde{p}_{M,i} = f_i^{(\text{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}))$ for $i \in \mathbb{B}_M$ and letting $y = F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})$, (3.10) can be equivalently written as

$$f'_{j_M} f_{j_M}^{(\text{inv})}(y) = \min_{i \in \mathbb{B}_M} f'_i f_i^{(\text{inv})}(y)$$

Thus the condition in Theorem 4 implies that $j_M = k_M$ for $M = 1, 2, 3, \ldots$.

C.2 Proof of Lemma 3

1. (5.1) implies the existence of $\delta > 0$ such that

$$f'_1 f_1^{(\text{inv})}(y) \leq f'_2 f_2^{(\text{inv})}(y) \leq \cdots \leq f'_N f_N^{(\text{inv})}(y) \quad \text{for } 0 < y < \delta. \tag{C.1}$$

   Thus following similar proof of Theorem 4, we have $j_i = i$ when the total power P is less than $\min_{1 \leq i \leq N} f_i^{(\text{inv})}(\delta)$.

2. We first show that the allocated power $\tilde{q}_{\mathbb{B}_M}$ will go to infinity as $P \to \infty$.

Recall that the KKT condition gives

$$(K-N+|\mathbb{B}_M|)F'_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})\begin{cases}=\nu & \text{if } \tilde{q}_{\mathbb{B}_M}>0\\ \le \nu & \text{if } \tilde{q}_{\mathbb{B}_M}=0\end{cases} \tag{C.2}$$

and for $i \notin \mathbb{B}_M$,

$$f'_i(\tilde{q}_{M,i})\begin{cases}=\nu & \text{if } \tilde{q}_{M,i}>0\\ \le \nu & \text{if } \tilde{q}_{M,i}=0\end{cases} \tag{C.3}$$

where the Lagrange multiplier ν is chosen such that $\sum_{i\notin\mathbb{B}_M}\tilde{q}_{M,i}+\tilde{q}_{\mathbb{B}_M}=P$. This then implies that $\tilde{q}_{\mathbb{B}_M}$ will go to infinity as $P\to\infty$, because if there exists a sequence $P_1,P_2,P_3,\ldots$, such that $\lim_{k\to}P_k=\infty$ and

$$\sup_{k\ge 1}\tilde{q}_{\mathbb{B}_M}\left(=\sup_{k\ge 1}\tilde{q}_{\mathbb{B}_M}(P_k)\right)<\infty$$

then we can use $\sum_{i\notin\mathbb{B}_M}\tilde{q}_{M,i}+\tilde{q}_{\mathbb{B}_M}=P$ to obtain that

$$\lim_{k\to\infty}\sum_{i\notin\mathbb{B}_M}\tilde{q}_{M,i}\left(=\lim_{k\to\infty}\sum_{i\notin\mathbb{B}_M}\tilde{q}_{M,i}(P_k)\right)=\infty$$

which then implies the existence of sequence $m_1,m_2,m_3,\ldots\notin\mathbb{B}_M$ (where $m_k=m_k(P_k)$) such that $\tilde{q}_{M,m_k}>0$ or equivalently $f'_{m_k}(\tilde{q}_{M,m_k})=\nu$ for all sufficiently large k and $\tilde{q}_{M,m_k}\to\infty$ as $k\to\infty$. Since $\lim_{k\to\infty}f'_i(\tilde{q}_{M,m_k})=0$ for every $i\notin\mathbb{B}_M$, we have $\lim_{k\to\infty}\nu=0$; hence, $\lim_{k\to\infty}(K-N+|\mathbb{B}_M|)F'_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})=0$ by (C.2), which is a contradiction to $\sup_{k\ge1}\tilde{q}_{\mathbb{B}_M}<\infty$. As a consequence, $\tilde{q}_{\mathbb{B}_M}$ diverges to infinity for every sequence $P_1,P_2,P_3,\ldots$, and therefore, $\lim_{P\to\infty}\tilde{q}_{\mathbb{B}_M}$ exists and is equal to ∞.

Next, we observe that (5.2) implies

$$\omega_1\le\omega_2\le\cdots\le\omega_N \tag{C.4}$$

because if $\omega_i>\omega_j$ for some $1\le i<j\le N$, then

$$\limsup_{y\uparrow\min\{\omega_i,\omega_j\}}\operatorname{sgn}\left(f'_if_i^{(\mathtt{inv})}(y)-f'_jf_j^{(\mathtt{inv})}(y)\right)=1$$

which is a contradiction to (5.2). Then (5.2) and (C.4) togetherly imply that for $1 \leq i < N$, there exists $\delta_i > 0$ such that for $\omega_i - \delta_i \leq y \leq \omega_i$,

$$f_i' f_i^{(\mathtt{inv})}(y) \leq f_{i+u}' f_{i+u}^{(\mathtt{inv})}(y) \quad \text{for every } 1 \leq u \leq N-i. \tag{C.5}$$

In addition, by noting $\tilde{p}_{M,i} = f_i^{(\mathtt{inv})}(F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}))$ for $i \in \mathbb{B}_M$ and letting $y = F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M})$, (3.10) can be equivalently written as

$$f_{j_M}' f_{j_M}^{(\mathtt{inv})}(y) = \min_{i \in \mathbb{B}_M} f_i' f_i^{(\mathtt{inv})}(y). \tag{C.6}$$

Consider $M = 1$. Since $\lim_{P\to\infty} \tilde{q}_{\mathbb{B}_1} = \infty$, we obtain from $\lim_{q\to\infty} F_{\mathbb{B}_M}(q) = \min_{i\in\mathbb{B}_M} \omega_i$ that $F_{\mathbb{B}_1}(\tilde{q}_{\mathbb{B}_1})$ will lie in $[\omega_1 - \delta_1, \omega_1]$ as P sufficiently large. Condition (C.5) (with setting $i = 1$) and (C.6) then jointly imply $j_1 = 1$ as P sufficiently large. We can repeat the procedure by further enlarging P (whenever necessary) to make $F_{\mathbb{B}_2}(\tilde{q}_{\mathbb{B}_2}) \in [\omega_2 - \delta_2, \omega_2]$, and obtain $j_2 = 2$ in the high-power regime. A similar argument can be applied to obtain $j_M = M$ for $M = 3, 4, \ldots$.

C.3 Proof of Corollary 1

1. That (5.3c) implies (5.1) is obvious. From (5.3a), we can infer by the continuity of f_j' that there exists $\gamma > 0$ such that $f_i'(0) < f_j'(\gamma)$. So for $0 < y < f_j(\gamma)$, we have $0 < f_j^{(\mathtt{inv})}(y) < \gamma$ by $f_j(0) = 0$ and the strictly increasingness of f_j. Hence,

$$f_i' f_i^{(\mathtt{inv})}(y) < f_i'(0) < f_j'(\gamma) < f_j' f_j^{(\mathtt{inv})}(y)$$

where the first and last strict inequalities follow the strict decreasingness of f_i' and f_j', respectively. Confirmation of (5.3a) implying (5.1) is then completed. It remains to verify that (5.3b) implies (5.1). By definition, $f_i''(0) < f_j''(0)$ implies the existence of $\delta > 0$ such that $f_i''(p) < f_j''(p)$ for $0 < p < \delta$, which together with $f_i'(0) = f_j'(0)$ implies $f_i'(p) < f_j'(p)$ for $0 < p < \delta$. Thus (5.3b) implies (5.3c), which in turns implies (5.1).

2. Since $\omega_i < \omega_j$, we have

$$\lim_{y\to\omega_i} f_i' f_i^{(\mathtt{inv})}(y) = 0 \quad \text{and} \quad \lim_{y\to\omega_i} f_j' f_j^{(\mathtt{inv})}(y) = f_j' f_j^{(\mathtt{inv})}(\omega_i) > 0.$$

Thus, (5.2) is valid.

C.4 Proof of Lemma 4

We first observe that (3.11) can be re-written as

$$(K-M) \le \frac{f_{j_M}'(\tilde{p}_{M,j_M})}{F_{\mathbb{B}_{M+1}}'\left(\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M,i}\right)} = \sum_{i\in\mathbb{B}_{M+1}} \frac{f_{j_M}'(\tilde{p}_{M,j_M})}{f_i'(\tilde{p}_{M,i})} = \sum_{i\in\mathbb{B}_{M+1}} \frac{f_{j_M}' f_{j_M}^{(\mathtt{inv})}(y)}{f_i' f_i^{(\mathtt{inv})}(y)}$$

where $y = F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}) = f_i(\tilde{p}_{M,i})$ for $i \in \mathbb{B}_M$, and (B.4c) guarantees $\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M,i} > 0$ and hence $F_{\mathbb{B}_{M+1}}'\left(\sum_{i\in\mathbb{B}_{M+1}} \tilde{p}_{M,i}\right) < \infty$. Based on this observation and noting

$$F_{\mathbb{B}_M}(\tilde{q}_{\mathbb{B}_M}) \to \begin{cases} 0 & \text{when } P \to 0 \\ \Omega(\mathbb{B}_M) & \text{as } \tilde{q}_{\mathbb{B}_M} \to \infty \text{ when } P \to \infty \end{cases}$$

this lemma becomes straightforward.

C.5 Proof of Corollary 2

1. Re-write (5.11) as

$$(K-M) F_{\mathbb{B}_{M+1}}'(0) < f_{j_M}'(0). \tag{C.7}$$

Then, a similar proof for (5.3a) can be used to prove (5.11) implying (5.7). Note the validity of (5.11) implicitly indicates the finiteness of $\min_{i\in\mathbb{B}_{M+1}} f_i'(0)$ because $\min_{i\in\mathbb{B}_{M+1}} f_i'(0) = \infty$ would fail (5.11); so $F_{\mathbb{B}_{M+1}}'(0) < \infty$.

2. That (5.12) is a sufficient condition for (5.8) can be proved in the same way as item 1) by reversing the order of inequality (C.7); hence, we omit it.

3. Since (5.13) implies

$$\begin{aligned}\limsup_{y\to\Omega(\mathbb{B}_M)}\left(K-M-\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\right)&=(K-M)-\liminf_{y\to\Omega(\mathbb{B}_M)}\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\\&\le(K-M)-\sum_{i\in\mathbb{B}_{M+1}}\liminf_{y\to\Omega(\mathbb{B}_M)}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\\&<0\end{aligned}$$

we have

$$\limsup_{y\to\Omega(\mathbb{B}_M)}\operatorname{sgn}\left(K-M-\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\right)=-1$$

which validates (5.9).

4. Again, since (5.14) implies

$$\begin{aligned}\liminf_{y\to\Omega(\mathbb{B}_M)}\left(K-M-\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\right)&=(K-M)-\limsup_{y\to\Omega(\mathbb{B}_M)}\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\\&\ge(K-M)-\sum_{i\in\mathbb{B}_{M+1}}\limsup_{y\to\Omega(\mathbb{B}_M)}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\\&>0\end{aligned}$$

we have

$$\liminf_{y\to\Omega(\mathbb{B}_M)}\operatorname{sgn}\left(K-M-\sum_{i\in\mathbb{B}_{M+1}}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}\right)=1$$

which validates (5.10).

Finally, if $j_1, j_2, j_3, \ldots$ are determined according to condition (5.2) in Lemma 3, then

$$\begin{cases}\omega_{j_1}\le\omega_{j_2}\le\omega_{j_3}\le\cdots & \text{(C.8a)}\\ \Omega(\mathbb{B}_M)=\omega_{j_M} & \text{(C.8b)}\end{cases}$$

where (C.8a) has been proved in (C.4), and (C.8b) is a consequence of (C.8a). Based on these results, we can derive from (5.15) that for $i\in\mathbb{B}_{M+1}$,

$$\limsup_{y\to\Omega(\mathbb{B}_M)}\frac{f'_{j_M}f_{j_M}^{(\mathtt{inv})}(y)}{f'_if_i^{(\mathtt{inv})}(y)}=0, \qquad \text{(C.9)}$$

because $\omega_{j_M} < \omega_{j_{M+1}}$ implies $f'_{j_M} f_{j_M}^{(\mathtt{inv})}(y)$ approaching zero and $f'_i f_i^{(\mathtt{inv})}(y)$ being bounded away from zero as y approaching $\Omega(\mathbb{B}_M) = \omega_{j_M}$. The proof is thus completed.

C.6 Proof of Lemma 5

Recall that the KKT condition gives

$$(K - N + |\mathbb{A}|)F'_{\mathbb{A}}(q^*_{\mathbb{A}}) = \nu \quad \text{and} \quad \text{for } i \notin \mathbb{A}, f'_i(q^*_i) \begin{cases} = \nu & \text{if } q^*_i > 0 \\ \leq \nu & \text{if } q^*_i = 0 \end{cases} \tag{C.10}$$

where we have used the fact that $q^*_{\mathbb{A}} = q^*_{\mathbb{B}_{N-|\mathbb{A}|+1}} > 0$ (see (B.4c)), and the Lagrange multiplier ν is chosen such that $\sum_{i \notin \mathbb{A}} q^*_i + q^*_{\mathbb{A}} = P$. We then distinguish among the below three cases:

Case 1: $f'_i(0) > (K - N + |\mathbb{A}|)F'_{\mathbb{A}}(0)$ *for some* $i \notin \mathbb{A}$. In this case, there exists $\gamma > 0$ such that

$$f'_i(q) > (K - N + |\mathbb{A}|)F'_{\mathbb{A}}(0) \quad \text{for every } 0 < q < \gamma.$$

Hence, when $0 < P < \gamma$, we have

$$\nu \geq f'_i(q^*_i) \geq f'_i(P) > (K - N + |\mathbb{A}|)F'_{\mathbb{A}}(0) > (K - N + |\mathbb{A}|)F'_{\mathbb{A}}(q^*_{\mathbb{A}})$$

where the last strict inequality follows the strict decreasingness of $F'_{\mathbb{A}}$. This contradicts (C.10); hence, Case 1 cannot happen.

Case 2: $f'_i(0) < (K + |\mathbb{A}| - N)F'_{\mathbb{A}}(0)$ *for some* $i \notin \mathbb{A}$.

In this case, there exists $\gamma > 0$ such that

$$f'_i(0) < (K + |\mathbb{A}| - N)F'_{\mathbb{A}}(q) \quad \text{for every } 0 < q < \gamma.$$

Hence, when $0 < P < \gamma$, we have

$$f'_i(q^*_i) \leq f'_i(0) < (K + |\mathbb{A}| - N)F'_{\mathbb{A}}(P) \leq (K + |\mathbb{A}| - N)F'_{\mathbb{A}}(q^*_{\mathbb{A}}) = \nu$$

which implies $q_i^* = 0$ for any $0 < P < \gamma$.

Case 3: $f_i'(0) = (K + |\mathbb{A}| - N)F_{\mathbb{A}}'(0)$ *for some* $i \in \mathbb{A}$.

In this case, $q_i^* > 0$ for $P > 0$ fixed, because if $q_i^* = 0$, we obtain

$$\nu \geq f_i'(q_i^*) = f_i'(0) = (K + |\mathbb{A}| - N)F_{\mathbb{A}}'(0) > F_{\mathbb{A}}'(q_{\mathbb{A}}^*)$$

which then contradicts (C.10). Based on $q_i^* > 0$ and $q_{\mathbb{A}}^* > 0$ for $P > 0$, we derive from (C.10) that

$$\begin{aligned}
\lim_{P\to 0} \frac{q_i^*}{q_{\mathbb{A}}^*} &= \lim_{\nu \uparrow f_i'(0)} \frac{f_i'^{(\mathtt{inv})}(\nu)}{F_{\mathbb{A}}'^{(\mathtt{inv})}(\nu/(K - N + |\mathbb{A}|))} \\
&= \lim_{\nu \uparrow f_i'(0)} \frac{\frac{1}{f_i'' f_i'^{(\mathtt{inv})}(\nu)}}{\frac{1}{(K-N+|\mathbb{A}|)\cdot F_{\mathbb{A}}'' F_{\mathbb{A}}'^{(\mathtt{inv})}(\nu/(K-N+|\mathbb{A}|))}} \\
&= \frac{\frac{1}{f_i''(0)}}{\frac{1}{(K-N+|\mathbb{A}|)F_{\mathbb{A}}''(0)}} \\
&= (K - N + |\mathbb{A}|)\frac{F_{\mathbb{A}}''(0)}{f_i''(0)}.
\end{aligned} \tag{C.11}$$

The proof for (5.16) is completed.

We now turn to the power allocations for channels i, j in $\mathbb{A}$. By

$$f_i(p_i^*) = f_j(p_j^*) = F_{\mathbb{A}}(q_{\mathbb{A}}^*) \ (= y) \tag{C.12}$$

we derive

$$\lim_{P\to 0} \frac{p_i^*}{p_j^*} = \lim_{y\to 0} \frac{f_i^{(\mathtt{inv})}(y)}{f_j^{(\mathtt{inv})}(y)} = \lim_{y\to 0} \frac{1/f_i' f_i^{(\mathtt{inv})}(y)}{1/f_j' f_j^{(\mathtt{inv})}(y)} = \frac{1/f_i'(0)}{1/f_j'(0)}. \tag{C.13}$$

C.7 Proof of Observation 1

The result for 1) is a direct consequence of the given rates of convergence, and the result for 2) follows similarly to (C.13). Hence, we omit them.

Bibliography

[1] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoded transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[2] S. L. Bernstein, M. L. Burrows, J. E. Evans, A. S. Griffiths, D. A. Mcneill, C. W. Niessen, I. Richer, D. P. White, and D. K. Willim, "Long-range communications at extremely low frequencies," in *Proc. IEEE*, vol. 62, no. 3, pp. 292–312, Mar. 1974.

[3] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sept. 1993.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[5] C. D. Charalambous, S. Z. Denic, and C. Constantinou, "Capacity of the class of MIMO channels with Incomplete CDI–Properties of mutual Information for a class of channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3725–3734, Aug. 2009.

[6] G. R. Clements and L. T. Winson, *Manual of Mathematics and Mechanics.* Maryland: Wildside Press, 2008.

[7] J. W. Cook, "Wide-band impulsive noise survey of the access network," *BT Technical Journal*, vol. 11, no. 3, pp. 155–162, Jul. 1993.

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

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[10] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[11] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1283, Apr. 2005.

[12] ——, "Additive non-Gaussian noise channels: Mutual information and conditional mean estimation," in *Proc. 2005 IEEE Int. Symp. information Theory*, Adelaide, Australia, Sept. 2005, pp. 719–723.

[13] M. Kanefsky and J. B. Thomas, "On polarity detection schemes with non-Gaussian inputs," *Journal of the Franklin Institute*, vol. 280, no. 2, pp. 120–138, Aug. 1965.

[14] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[15] I-C. Lee, C.-S. Chang, and C.-M. Lien, "On the throughput of multicasting with incremental forward error correction," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 900–918, Mar. 2005.

[16] T.-H. Li and K.-S. Song, "Estimation of the parameters of sinusoidal signals in non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp.62–72, Jan. 2009.

[17] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[18] A. Lozano, A. M. Tulino and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[19] ——, "Optimum power allocation for multiuser OFDM with arbitrary signal constellations," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 828–837, May 2008.

[20] D. Middleton, "Channel modeling and threshold signal processing in underwater acoustics: An analytical overview," *IEEE J. Ocean. Eng.*, vol. OE-12, no. 1, pp. 4–28, Jan. 1987.

[21] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, Jul. 2003.

[22] M. Payaro, A. Wiesel, J. Yuan, and M. A. Lagunas, "On the capacity of linear vector Gaussian channels with magnitude knowledge and phase uncertainty," in *31st Int. Conf. on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006, pp. IV–IV.

[23] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary Inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[24] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions—Part I: (n, k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[25] ——, "n-channel symmetric multiple descriptions—Part II: An achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1377–1392, Apr. 2005.

[26] H. L. Royden, *Real Analysis*, 3rd ed., New York: Prentice-Hall, 1988.

[27] M. G. Sánchez, A. V. Alejos, and I. Cuiñas, "Urban wide-band measurement of the UMTS electromagnetic environment," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1014–1022, Jul. 2004.

[28] G. Shevlyakov and K. Kiseon, "Robust minimax detection of a weak signal in noise with a bounded variance and density value at the center of symmetry," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1206–1211, Mar. 2006.

[29] M. W. Thompson and H.-S. Chang, "Coherent detection in Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 452–461, Apr. 1994.

[30] M. W. Thompson, D. R. Halverson, and G. L. Wise, "Robust detection in nominally Laplace noise," *IEEE Trans. Commun.* vol. 42, no. 2/3/4, pp. 1651–1660, Feb./Mar./Apr. 1994.

[31] A. Wiesel, Y. C. Eldar, and S. Shamai, "Beamforming maximizes the compound capacity in MISO channels," in *Int. ITG Workshop on Smart Antennas WSA 2005*, Duisburg, Germany, Apr. 2005.

[32] ——, "Robust power allocation for maximizing the compound capacity," in *Proc. NEWCOM-ACoRN Joint Workshop* Vienna, Austria, Sept. 2006.

[33] ——, "Optimization of the MIMO compound capacity," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 3, pp. 1094–1101, Mar. 2007.

[34] J. Wolfowitz, *Coding Theorems of Information Theory,* 3rd ed. Berlin, Germany: Springer-Verlag, 1978.

[35] J. Woods and S. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1278–1288, Oct. 1986.

[36] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broadband powerline communications," *IEEE Trans. Electromagn. Compat.*, vol. 44, no. 1, pp. 249–258, Feb. 2002.

About the Author

Shih-Wei Wang was born in Pingtung, R.O.C., in 1981. He received the B.S. and M.S. degrees in electrical engineering from the National Central University, Chungli, Taiwan, in 2003 and 2005, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan. He held a visiting position with Queen's University, Kingston, Canada, in 2010. His research interests lie in information theory and convex optimization.