August 2016

Contributions On Theory And Practice For Multi-Mission Wireless Systems

Yu Zhao
Syracuse University

ABSTRACT

The field of wireless systems has long been an active research area with various applications. Recently much attention has been given to multi-mission wireless systems that combine capabilities including information sensing, data processing, energy harvesting as well as the traditional data communication. This dissertation describes our endeavor in addressing some of the research challenges in multi-mission wireless systems, including the development of fundamental limits of such multi-mission wireless systems and effective technologies for improved performance.

The first challenge addressed in this dissertation is how to handle interference, which is encountered in almost all wireless systems involving multiple nodes, an attribute shared by most multi-mission systems. To deepen our understanding on the impact of interference, we study a class of Gaussian interference channels (GICs) with mixed interference. A simple coding scheme is proposed based on Sato's non-naïve frequency division. The achievable region is shown to be equivalent to that of Costa's *noiseberg* region for the one-sided Gaussian interference channel. This allows for an indirect proof that this simple achievable rate region is indeed equivalent to the Han-Kobayashi (HK) region with Gaussian input and with time sharing for this class of Gaussian interference channels with mixed interference.

Optimal power management strategies are then investigated for a remote estimation system with an energy harvesting sensor. We first establish the asymptotic optimality of uncoded transmission for such a system under Gaussian assumption. With the aim of minimizing the mean squared error (MSE) at the receiver, optimal power allocation policies are proposed under various assumptions with regard to the knowledge at the transmitter and the receiver as well as battery storage capacity. For the case where non-causal side information (SI) of future harvested energy is available and battery storage is unlimited, it is shown that

the optimal power allocation amounts to a simple 'staircase-climbing' procedure, where the power level follows a non-decreasing staircase function. For the case where battery storage has a finite capacity, the optimal power allocation policy can also be obtained via standard convex optimization techniques. Dynamic programming is used to optimize the allocation policy when causal SI is available. The issue of unknown transmit power at the receiver is also addressed. Finally, to make the proposed solutions practically more meaningful, two heuristic schemes are proposed to reduce computational complexity.

Related to the above remote sensing problem, we provide an information theoretic formulation of a multi-functioning radio where communication between nodes involves transmission of both messages and source sequences. The objective is to study the optimal coding trade-off between the rate for message transmission and the distortion for source sequence estimation. For point-to-point systems, it is optimal to simply split total capacity into two components, one for message transmission and one for source transmission. For the multi-user case, we show that such separation-based scheme leads to a strictly suboptimal rate-distortion trade-off by examining the simple problem of sending a common source sequence and two independent messages through a Gaussian broadcast channel.

Finally we study the design of a practical multi-mission wireless system - the dual-use of airborne radio frequency (RF) systems. Specifically, airborne multiple-input-multiple-output (MIMO) communication systems are leveraged for the detection of moving targets in a typical airborne environment that is characterized by the lack of scatterers. With uniform linear arrays (ULAs), angular domain decomposition of channel matrices is utilized and target detection can be accomplished by detection of change in the resolvable paths in the angular domain. For both linear and nonlinear arrays, Doppler frequency analysis can also be applied and the change in frequency components indicates the presence of potential airborne targets. Nonparametric detection of distribution changes is utilized in both approaches.

CONTRIBUTIONS ON THEORY AND PRACTICE FOR MULTI-MISSION WIRELESS SYSTEMS

by

Yu Zhao

B.S.(Electrical Engineering), Beijing Institute of Technology, 2009
M.S.(Applied Statistics), Syracuse University, 2014

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
August 2016

ACKNOWLEDGMENTS

My deepest gratitude goes to my advisor, Dr. Biao Chen. Thank you for introducing me to the research field and opening my world to many fascinating new impressions. I have truly enjoyed the opportunity to work under your guidance and to learn from your profound knowledge. Your continuous support, patience and encouragement has made my PhD experience productive and stimulating. I could not have imagined having a better advisor.

Besides my advisor, I thank the rest of my committee - (in alphabetic order) Dr. Yingbin Liang, Dr. Kishan G. Mehrotra, Dr. Ruixin Niu, Dr. Lixin Shen, and Dr. Pramod K. Varshney. Thank you for taking the time to review my work and offering insightful suggestions.

I am grateful to Dr. Pinyuen Chen, Dr. Makan Fardad, and all the other professors from whom I have taken courses. You have helped me build a solid technical background for my doctoral research and broadened my horizons in many ways.

Thanks to my fellow labmates and friends, Wei, Minna, Fangfang, Ge, Kapil, Pengfei, Fangrong, Shengyu, Yang and Tiexing. I will miss the time that we spent together in group meetings, whiteboard discussions and weekend hangouts, etc. The past years, without you, would have been a very lonely journey.

My final but everlasting gratitude goes to my family, who have been a constant source of love and strength. It is only because of their unconditional support, that I have had the chances to pursue my dreams. Heartfelt thanks to my boyfriend, Jielong. Thank you for holding my hand through tough times in this PhD adventure.

Your cheerful attitudes towards life have always been a great mental support.

I would also like to acknowledge the generous support of the National Science Foundation under award CCF0905320 , the Air Force Research Laboratory under award FA8750-15-1-0045, the Air Force Office of Scientific Research under award FA9550-16-1-0077.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Background

The field of wireless systems has been one of the most vibrant research areas for over a century, starting around 1896 when Marconi made the first demonstration of his wireless telegraph machine for the British government. Since then, many types of wireless systems have flourished, including but not limited to cellular networks, radar systems, communication satellites, etc.

Driven by an explosively increasing demand, the research thrust over the past few decades has led to a richer set of perspectives and tools on how to design wireless systems that simultaneously perform multiple missions. For example, wireless sensor networks that integrate the technologies of information sensing, data processing and transmission are widely applied [1–4]. Some of them operate on energy harvested from the environment through various sources such as solar cells and vibration absorption devices [5–7]. There has been an increasing interest in designing systems involving multi-functioning electromagnetic nodes as well. For example, passive radars that leverage the existence of radio waves for communications have attracted much attention in recent years [8–10]. Also, there have been existing radar systems that have integrated communication capability [11–14].

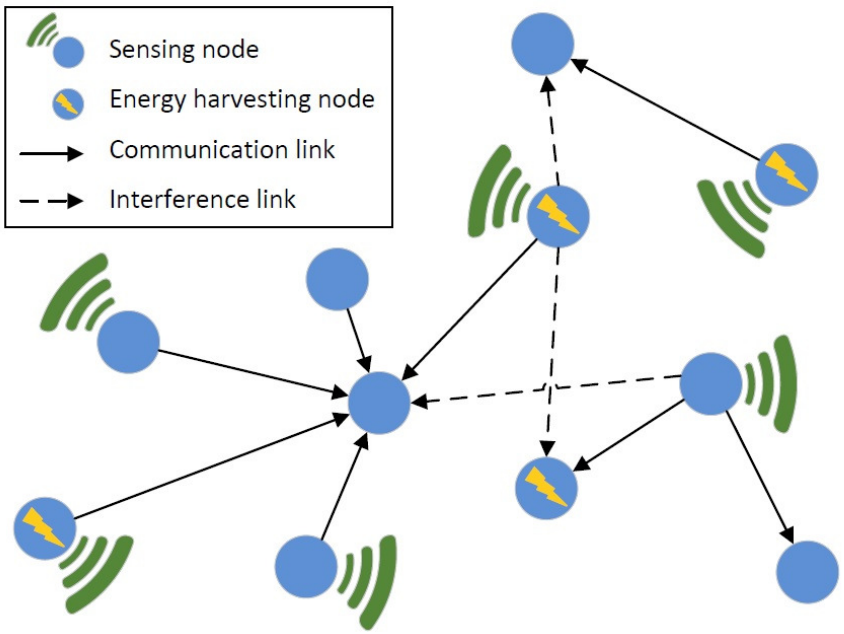


Fig. 1.1: A multi-mission wireless system.

Another example is the WiFi positioning systems which provide indoor localization service based on measurement of the intensity of the WiFi signals [15–17]. Fig. 1.1 shows an example of a multi-mission wireless system that combines various functions including sensing, communication and energy harvesting, etc.

In many of the current wireless systems, interference is inevitable as communication networks have evolved dramatically in terms of both size and data rate and multiple users often need to share, hence contend for, a common medium. It is thus essential to understand the impact of interference on system performance and to explore technologies for effective interference management. In most existing systems, interference is dealt with either by interference avoidance where the communication links are orthogonalized in time or frequency, or by treating interference as noise, often assisted by power control at the transmitters. These approaches are typically not optimal, and often lead to the loss of degrees of freedom.

In Chapter 2, we consider the interference channel (IC) which is a basic model for studying the theoretical limits of communication rate with interference. It mathemati-

cally abstracts the situation when multiple transmitters communicate with their intended receivers while generating interference to unintended receivers. Specifically, we propose a simple coding scheme for a class of Gaussian interference channels (GICs) with mixed interference and prove that it is equivalent to the Han-Kobayashi (HK) region with Gaussian input and with time sharing, which is the largest achievable region for the class of GICs.

Another challenge for designing wireless systems has to do with power management. Recent developments in hardware design have empowered many wireless networks to support themselves by harvesting energy from nature through various sources such as solar cells, vibration absorption devices, among others, and store excess energy for future use. Unlike traditional battery-powered systems, where transmission is often subject to a constant power constraint, the energy available to an energy harvesting system typically fluctuates in time and is often modeled as a random process. This introduces additional difficulty in system design, and in particular, how to allocate power across time for improved system performance.

In Chapter 3 we consider a remote estimation system with the transmitter powered by energy harvesting devices. The central question we try to answer is how to determine transmit power across time under various assumptions as to what is known at the transmitter and/or receiver: whether the transmitter has non-causal side information (SI) of future harvested energy, whether the battery has unlimited storage capacity, and whether the receiver knows the transmit power *a priori*. The clairvoyant case, namely the transmitter has non-causal knowledge of future harvested energy, the battery capacity is infinite, and the receiver knows the exact transmit power, serves as a benchmark for performance comparison. We will then replace these idealized assumptions with more realistic ones and provide solutions to power allocation under each scenario. Performance evaluation will be conducted to identify conditions under which the optimal performance in terms of minimum mean square error (MMSE) with realistic assumptions is close to that of the clairvoyant case, therefore providing guidance on system design. Heuristic schemes are also proposed

to alleviate the computational burden.

Following discussions on interference and power management, we present in Chapter 4 a simple information theoretic formulation for multi-mission wireless systems. Specifically, we consider communication systems where simultaneous transmission of messages and source sequences is required. Using primarily Gaussian channels as examples, the central question we address is the following: is it optimal to treat the channel as simply a bit pipe where message and source encoding divide up the bandwidth? For the point to point system, we show that this is indeed the case. That is, the optimal scheme simply splits the total channel capacity into two components, one for message communication and the other for source transmission. The same conclusion, however, no longer holds for multi-user systems. Using a simple Gaussian broadcast channel example, we show that such a separation approach will lead to a strictly suboptimal rate-distortion trade-off.

As a practical application of such multi-mission wireless systems, we then consider integration of sensing capability into airborne multiple-input-multiple-output (MIMO) communication systems. While MIMO communications is known to be the most promising technologies for improving the spectral efficiency of wireless communications, there exist significant challenges in adopting MIMO communications in an airborne environment. Chief among them is the lack of scatterers in airborne RF channels which appears to limit the potential of MIMO - channel matrices may not support the intended capacity gain where communication channels are dominated by the line-of-sight (LOS) components. However, it was pointed out in [18] that the lack of scattering can be largely compensated by the large aperture of the airborne transceivers. This phenomenon is reminiscent of the distributed antenna arrays in [19, Chapter 7] which lead to significant capacity gain even with only LOS between antenna elements at the transmitter and the receiver.

The fact that an airborne radio frequency (RF) environment is scarce in scatterers has actually opened up opportunities for expanding functions of the MIMO RF asset. Specifically, the presence of an airborne target may alter the transmission channels, as reflected in

various domains depending on the target characteristics. As such, by prudent processing at the receiver, one may be able to extract useful information of airborne targets which often act as strong reflectors/scatterers of the MIMO channel; such expanded functions can often be accomplished without compromising the communication task at hand. In Chapter 5, we investigate the dual-use of airborne MIMO and propose target detections methods that exploit the scarcity of scatterers in an airborne environment.

1.2 Thesis Organization

The main contributions of this dissertation are presented as follows.

In Chapter 2, a simple encoding scheme based on Sato's non-naïve frequency division is proposed for a class of Gaussian interference channels with mixed interference. The achievable region is shown to be equivalent to that of Costa's *noiseberg* region for the one-sided Gaussian interference channel. This allows for an indirect proof that this simple achievable rate region is indeed equivalent to the Han-Kobayashi (HK) region with Gaussian input and with time sharing for this class of Gaussian interference channels with mixed interference.

In Chapter 3, optimal transmit power allocation strategies are proposed for a remote estimation system, where energy can be harvested from the environment and buffered in a battery for future use. We first establish the optimality of uncoded transmission for such a system. With the aim of minimizing the mean squared error (MSE) at the receiver, two types of side information available to the transmitter are considered: causal SI (energy harvested in the past) and non-causal SI (energy harvested in the past, present and future). For the case where non-causal SI is available and battery storage is unlimited, it is shown that the optimal power allocation amounts to a simple 'staircase-climbing' procedure, where the power level follows a non-decreasing staircase function. For the case where battery storage has a finite capacity, the optimal power allocation policy can also be obtained via standard

convex optimization techniques. Dynamic programming is used to optimize the allocation policy when causal SI is available. The issue of unknown transmit power at the receiver is also addressed. Finally, to make the proposed solutions practically more meaningful, two heuristic schemes are proposed to reduce computational complexity.

In Chapter 4, we consider a wireless system with multi-functioning radios: communication between nodes involve transmissions of both messages and source sequences. For point-to-point systems, this amounts to a simple trade-off between message transmission and source transmission: an optimal strategy is to split total capacity into two components, one for message transmission and one for source transmission as long as the message and the source sequence are independent of each other. For the multi-user case, we show that this is no longer the case by examining the simple problem of sending a common source sequence and two independent messages through a Gaussian broadcast channel.

Chapter 5 investigates the dual-use of airborne MIMO. In addition to communications, received signals are also used for target detection that exploits the scarcity of scatterers in an airborne environment. Nonparametric target detection methods are proposed based on peak detection as well as Euclidean distance between sample observations in transform domains. For the case of uniform linear arrays, informative statistics are obtained through estimation of angular domain channel matrix. For the case of nonlinear arrays, parameters are selected based on Doppler frequency analysis on the received signal.

CHAPTER 2
THE HAN-KOBAYASHI REGION FOR A CLASS OF GAUSSIAN INTERFERENCE CHANNELS

2.1 Channel Model And Literature Review

Interference channel models a scenario where multiple transmitters communicate with their intended receivers in a shared medium while generating interference to the unintended receivers. A two-user GIC in its standard form can be represented as

$$\begin{aligned} Y_1 &= X_1 + bX_2 + Z_1, \\ Y_2 &= aX_1 + X_2 + Z_2, \end{aligned} \tag{2.1}$$

where X_1 and X_2 are the input signals and are subject to respective power constraints P_1 and P_2; Y_1 and Y_2 are the received signals; Z_1 and Z_2 are Gaussian noises of unit variance and are independent of the inputs X_1 and X_2. This model is depicted in Fig. 2.1.

Despite decades of intensive research, the characterization of the capacity region for a two-user interference channel is an open problem except for the strong and very strong

interference cases [20–22]. To date, the largest achievable rate region is the celebrated Han-Kobayashi region that employs rate splitting at the transmitters and simultaneous decoding at the receivers [22]. The general HK region is defined as

$$\mathcal{R}_{\mathcal{HK}} = \text{closure of} \bigcup_{Z \in \mathcal{P}(Z)} \mathcal{R}(Z) \tag{2.2}$$

where $\mathcal{P}(Z)$ is the set of all $Z = QU_1W_1U_2W_2X_1X_2Y_1Y_2 \in \mathcal{P}(Z)$ such that

- U_1, W_1, U_2, W_2 are conditionally independent given the time sharing random variable Q, where $\| Q \| \leq 11$;
- $X_1 = f_1(U_1W_1|Q)$, $X_2 = f_2(U_2W_2|Q)$ where f_i is an arbitrary deterministic function for $i = 1, 2$;
- $p(y_1y_2|x_1x_2)$ is the channel transition probability.

In Eq. (2.2), $\mathcal{R}(Z)$ is the set of all achievable (R_1, R_2) such that

$$\begin{aligned} R_1 &= S_1 + T_1 \\ R_2 &= S_2 + T_2, \end{aligned}$$

and S_1, T_1, S_2, T_2 are defined in [22, Eq. (3.2) - Eq. (3.15)] which are included as follows

$$\begin{aligned} S_1 &\leq I(U_1; Y_1|W_1W_2Q) \\ T_1 &\leq I(W_1; Y_1|U_1W_2Q) \\ T_2 &\leq I(W_2; Y_1|U_1W_1Q) \\ S_1 + T_1 &\leq I(U_1W_1; Y_1|W_2Q) \\ S_1 + T_2 &\leq I(U_1W_2; Y_1|W_1Q) \\ T_1 + T_2 &\leq I(W_1W_2; Y_1|U_1Q) \\ S_1 + T_1 + T_2 &\leq I(U_1W_1W_2; Y_1|Q) \end{aligned}$$

$$
\begin{aligned}
S_2 &\leq I(U_2; Y_2|W_1W_2Q) \\
T_2 &\leq I(W_2; Y_2|U_2W_1Q) \\
T_1 &\leq I(W_1; Y_2|U_2W_2Q) \\
S_2 + T_2 &\leq I(U_2W_2; Y_2|W_1Q) \\
S_2 + T_1 &\leq I(U_2W_1; Y_2|W_2Q) \\
T_1 + T_2 &\leq I(W_1W_2; Y_2|U_2Q) \\
S_2 + T_1 + T_2 &\leq I(U_2W_1W_2; Y_2|Q).
\end{aligned}
$$

Not surprisingly, for those ICs whose capacity regions are completely characterized, it is without an exception that the capacity region coincides with the HK region. However, the general HK region involves a time sharing variable Q that makes its evaluation intractable.

For the Gaussian interference channel, additional power constraints are imposed on the distribution of X_1 and X_2, i.e.,

$$
\begin{aligned}
\mathbb{E}(X_1^2) &\leq P_1 \\
\mathbb{E}(X_2^2) &\leq P_2.
\end{aligned}
$$

Besides the time sharing variable, another difficulty for characterization of the HK region of GICs is the input distribution. While for all the cases where the capacity results are known for a GIC, the optimal input distribution is invariably Gaussian, it is not yet known (or proven) that such is the case for the general GIC. In order to describe the HK region with a Gaussian codebook, denoted by $\mathcal{R}_{\mathcal{HK}}$, we first define a region $\mathcal{R}_0(P_1, P_2, \alpha, \beta)$ as

the collection of all rate pairs (R_1, R_2) satisfying [23]

$$\begin{aligned}
R_1 &\leq \psi_1 = \gamma\left(\frac{P_1}{1+b\beta P_2}\right) \\
R_2 &\leq \psi_2 = \gamma\left(\frac{P_2}{1+a\alpha P_1}\right) \\
R_1+R_2 &\leq \psi_3 = \min\{\psi_{31},\psi_{32},\psi_{33}\} \\
2R_1+R_2 &\leq \psi_4 = \gamma\left(\frac{P_1+b(1-\beta)P_2}{1+b\beta P_2}\right)+\gamma\left(\frac{\alpha P_1}{1+b\beta P_2}\right)+\gamma\left(\frac{\beta P_2+a(1-\alpha)P_1}{1+a\alpha P_1}\right) \\
R_1+2R_2 &\leq \psi_5 = \gamma\left(\frac{P_2+a(1-\alpha)P_1}{1+a\alpha P_1}\right)+\gamma\left(\frac{\beta P_2}{1+a\alpha P_1}\right)+\gamma\left(\frac{\alpha P_1+b(1-\beta)P_2}{1+b\beta P_2}\right)
\end{aligned}$$

where $0 \leq \alpha, \beta \leq 1$ and

$$\begin{aligned}
\psi_{31} &= \gamma\left(\frac{P_1+b(1-\beta)P_2}{1+b\beta P_2}\right)+\gamma\left(\frac{\beta P_2}{1+a\alpha P_1}\right) \\
\psi_{32} &= \gamma\left(\frac{\alpha P_1}{1+b\beta P_2}\right)+\gamma\left(\frac{P_2+a(1-\alpha)P_1}{1+a\alpha P_1}\right) \\
\psi_{33} &= \gamma\left(\frac{\beta P_2+a(1-\alpha)P_1}{1+a\alpha P_1}\right)+\gamma\left(\frac{\alpha P_1+b(1-\beta)P_2}{1+b\beta P_2}\right)
\end{aligned}$$

where $\gamma(x)$ is defined as $\frac{1}{2}\log(1+x)$. The region $\mathcal{R}_0(P_1, P_2, \alpha, \beta)$ can be rewritten in a matrix form

$$\mathcal{R}_0(P_1, P_2, \alpha, \beta) = \{\mathbf{R} | A\mathbf{R} \leq \Psi(P_1, P_2, \alpha, \beta)\}$$

where

$$\begin{aligned}
\mathbf{R}^t &= [R_1, R_2] \\
\Psi^t &= [\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5]
\end{aligned}$$

and

$$A^t = \begin{bmatrix} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix}.$$

Now the region $\mathcal{R}_{\mathcal{HK}}$ can be defined as a region obtained from enlarging $\mathcal{R}_0(P_1, P_2, \alpha, \beta)$ through time sharing, i.e.,

$$A\mathbf{R} \leq \sum_{i=1}^{q} \lambda_i \Psi(P_{1i}, P_{2i}, \alpha_i, \beta_i)$$

where $q \in \mathbb{N}$ and

$$\begin{aligned} \sum_{i=1}^{q} \lambda_i P_{1i} &\leq P_1 \\ \sum_{i=1}^{q} \lambda_i P_{2i} &\leq P_2 \\ \sum_{i=1}^{q} \lambda_i &= 1 \\ \lambda_i &\geq 0 \end{aligned}$$

for $0 \leq \alpha_i, \beta_i \leq 1$ and $i \in \{1, 2, \cdots, q\}$.

There has been recent progress in obtaining computable subregion of the HK achievable region using Sato's non-naïve frequency division [24]. For the one-sided GIC (denoted as ZGIC) shown in Fig. 2.2(a), Motahari and Khandani established that such a non-naïve frequency division scheme achieves the HK region with Gaussian input [25]. In this case, $\mathcal{R}_0$ can be represented as all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq \gamma(P_1) \\ R_2 &\leq \gamma\left(\frac{P_2}{1 + a\alpha P_1}\right) \\ R_1 + R_2 &\leq \gamma\left(\frac{a(1-\alpha)P_1 + P_2}{1 + a\alpha P_1}\right) + \gamma(\alpha P_1). \end{aligned}$$

Most recently, Costa introduced the so-called *noiseberg* scheme which uses water filling to achieve optimal power sharing between two orthogonal dimensions [26]. It turns out, as shown in the next section, that this simple noiseberg scheme achieves precisely the same

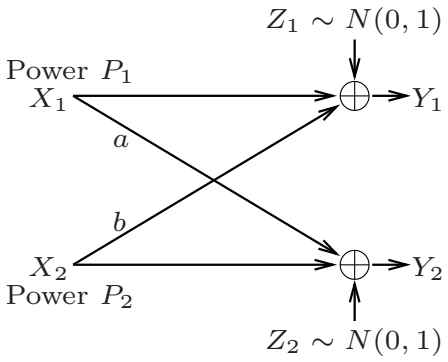
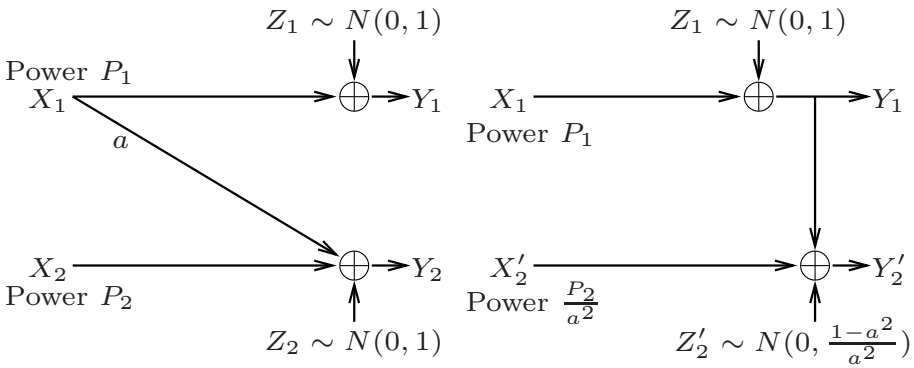


Fig. 2.1: Two-user Gaussian interference channel.

HK region with Gaussian input.

This chapter focuses on GICs with mixed interference (MGIC) and with $ab \geq 1$, $a \leq 1$ and $b \geq 1$ (cf. Fig. 2.1 and Eq. (2.1)). We describe a simple coding scheme that combines the noiseberg scheme with that of simultaneous decoding at the receiver experiencing strong interference. The obtained rate region is subsequently shown to coincide with the HK region with Gaussian input.



(a) ZGIC (b) Degraded GIC

Fig. 2.2: One-sided Gaussian interference channel and an equivalent degraded GIC.

2.2 Noiseberg Region For the ZGIC

In this section we consider the degraded GIC shown in Fig. 2.2(b), which is equivalent to the ZGIC with $a < 1$ in Fig. 2.2(a) [27].

The noiseberg region, denoted by $\mathcal{R}_{\mathcal{N}}$ and introduced by Costa in [26] for a ZGIC with weak interference ($a < 1$ in Fig. 2.2(a)) is the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq \bar{\lambda} R_{1\bar{\lambda}} + \lambda R_{1\lambda}, \\ R_2 &\leq \bar{\lambda} R_{2\bar{\lambda}}, \end{aligned}$$

where

$$R_{1\bar{\lambda}} \leq \frac{1}{2}\log\left(1 + \frac{P_{1A}}{\bar{\lambda}}\right) + \frac{1}{2}\log\left(1 + \frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1 + a^2\frac{P_{1A}}{\bar{\lambda}} + \frac{P_2}{\bar{\lambda}}}\right), \tag{2.3}$$

$$R_{2\bar{\lambda}} \leq \frac{1}{2}\log\left(1 + \frac{\frac{P_2}{\bar{\lambda}}}{1 + a^2\frac{P_{1A}}{\bar{\lambda}}}\right), \tag{2.4}$$

$$R_{1\lambda} \leq \frac{1}{2}\log\left(1 + \frac{P_{1B}}{\lambda}\right), \tag{2.5}$$

and the power limits P_{1A}, P_{1B} and P_{1C} are determined by two parameters h and λ such that

$$\begin{aligned} \frac{P_{1A}}{\bar{\lambda}} &= P_1 - \frac{P_2\lambda}{a^2\bar{\lambda}} - \lambda\min\left\{h, \frac{1-a^2}{a^2}\right\} - \max\left\{0, h - \frac{1-a^2}{a^2}\right\}, \\ \frac{P_{1B}}{\lambda} &= P_1 + \frac{P_2}{a^2} + \bar{\lambda}\min\left\{h, \frac{1-a^2}{a^2}\right\}, \\ \frac{P_{1C}}{\bar{\lambda}} &= \max\left\{0, h - \frac{1-a^2}{a^2}\right\} \end{aligned}$$

where $0 \leq \lambda \leq 1$ and $h \geq 0$.

Costa [26] showed that $\mathcal{R}_{\mathcal{N}}$ is achievable for the ZGIC with weak interference by a coding scheme that uses a two-band non-naïve frequency division multiplexing (FDM) with

water filling for optimal power allocation between the two subbands. The coding scheme, as well as its achievable region, involves two parameters $0 \leq \lambda \leq 1$ and $h \geq 0$. They vary over the admissible region as shown in Fig. 2.3, resulting in different transmission schemes depending on the values of the parameters. The parameter λ determines how to allocate the frequency band.

- The λ subband is reserved for the communication between transmitter 1 and receiver 1.
- The $\bar{\lambda}$ subband is shared by both transceiver pairs and the corresponding coding scheme is determined by the other parameter h.

As the noise Z_2 does not affect the transmission of X_1, water filling allows the overall power level in the λ-subband to be raised above that of the $\bar{\lambda}$-subband, with part of the noise spectrum of Z_2 floating above the signal level. This phenomenon, i.e., the existence of difference in heights of power spectrum for the two subbands is referred to as the *noiseberg*. The parameter h is defined as the height of total power density in the λ-subband above that of X_2's power density in the $\bar{\lambda}$-subband. Different h values divide the admissible region for the parameter pairs into two regions, each employing a different coding scheme in the $\bar{\lambda}$-subband:

Multiplex region This corresponds to $h \leq \frac{1-a^2}{a^2}$. As shown in Fig. 2.4, Z_2' prevents user 1's power from spilling over to the $\bar{\lambda}$-band thus no rate-splitting is involved. Receiver 2 decodes W_1 first, subtracts it and decodes W_2.

Overflow region This corresponds to $h > \frac{1-a^2}{a^2}$. As shown in Fig. 2.5, water-filling of user 1's power occurs as the power spills over from the λ-subband to the $\bar{\lambda}$-subband. The encoding scheme in the $\bar{\lambda}$ subband thus involves rate splitting for W_1: a common message W_{1c} with power P_{1c} decoded by both receivers and a private message W_{1p} with power P_{1A} decoded only by receiver 1. Receiver 2 decodes W_{1c} first, subtracts it, and decodes W_2, all the while treating W_{1p} as noise.

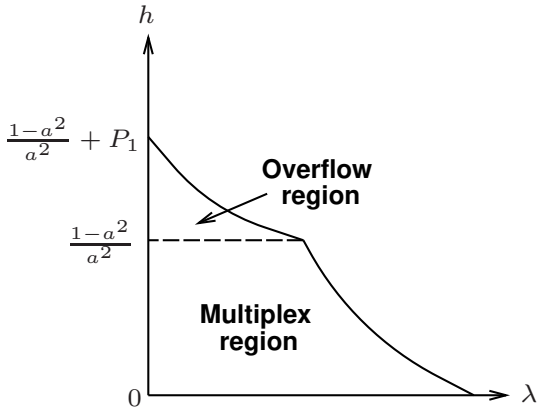


Fig. 2.3: Admissible region for (λ, h).

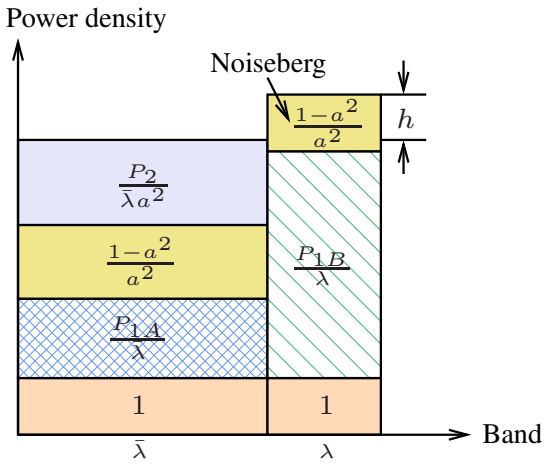


Fig. 2.4: Multiplex region.

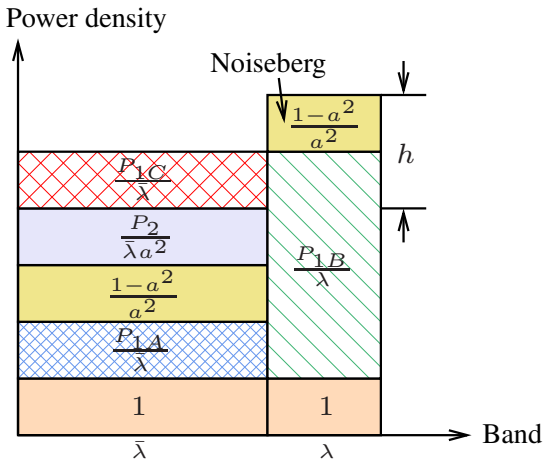


Fig. 2.5: Overflow region.

It is remarkable that such a simple transmission scheme turns out to achieve precisely the same HK region with time sharing and with Gaussian input.

Theorem 2.2.1. *For the weak ZGIC,* $\mathcal{R}_{\mathcal{N}} = \mathcal{R}_{\mathcal{HK}}$.

Proof. Motahari and Khandani showed that for the ZGIC, the non-naïve FDM region, denoted by $\mathcal{R}_{\mathcal{FDM}}$, is equivalent to $\mathcal{R}_{\mathcal{HK}}$, whose boundary points can be characterized by the optimization problem [25, Eq. (151)]. It suffices to verify the equivalence between $\mathcal{R}_{\mathcal{N}}$ and $\mathcal{R}_{\mathcal{FDM}}$.

We start by considering water filling in the two-band FDM applied to the degraded GIC shown in Fig. 2.2(b). First, we split W_1 into private message W_{1p} with power constraint P_{1p} and common message W_{1c} with power constraint P_{1c} such that $P_{1p} + P_{1c} = P_1$. Power allocation into λ and $\bar{\lambda}$ subbands is done in the following order. First, P_{1p} is allocated to the two subbands in an arbitrary way. On top of that, P_2 is allocated to the two subbands via waterfilling. As Y_2' sees additional noise Z_2', P_2 is allocated on top of Z_2' (see, e.g., Fig. 2.6(d)). Finally, P_{1c} is allocated to the two subbands, again, using waterfilling.

Depending on P_{1p} and its allocation between the two subbands, there are four possible power allocation outcomes of this scheme, as shown in Fig. 2.6. Since the scenarios illustrated in Fig. 2.6(a)(b)(c) are equivalent to noiseberg cases, it remains to argue that

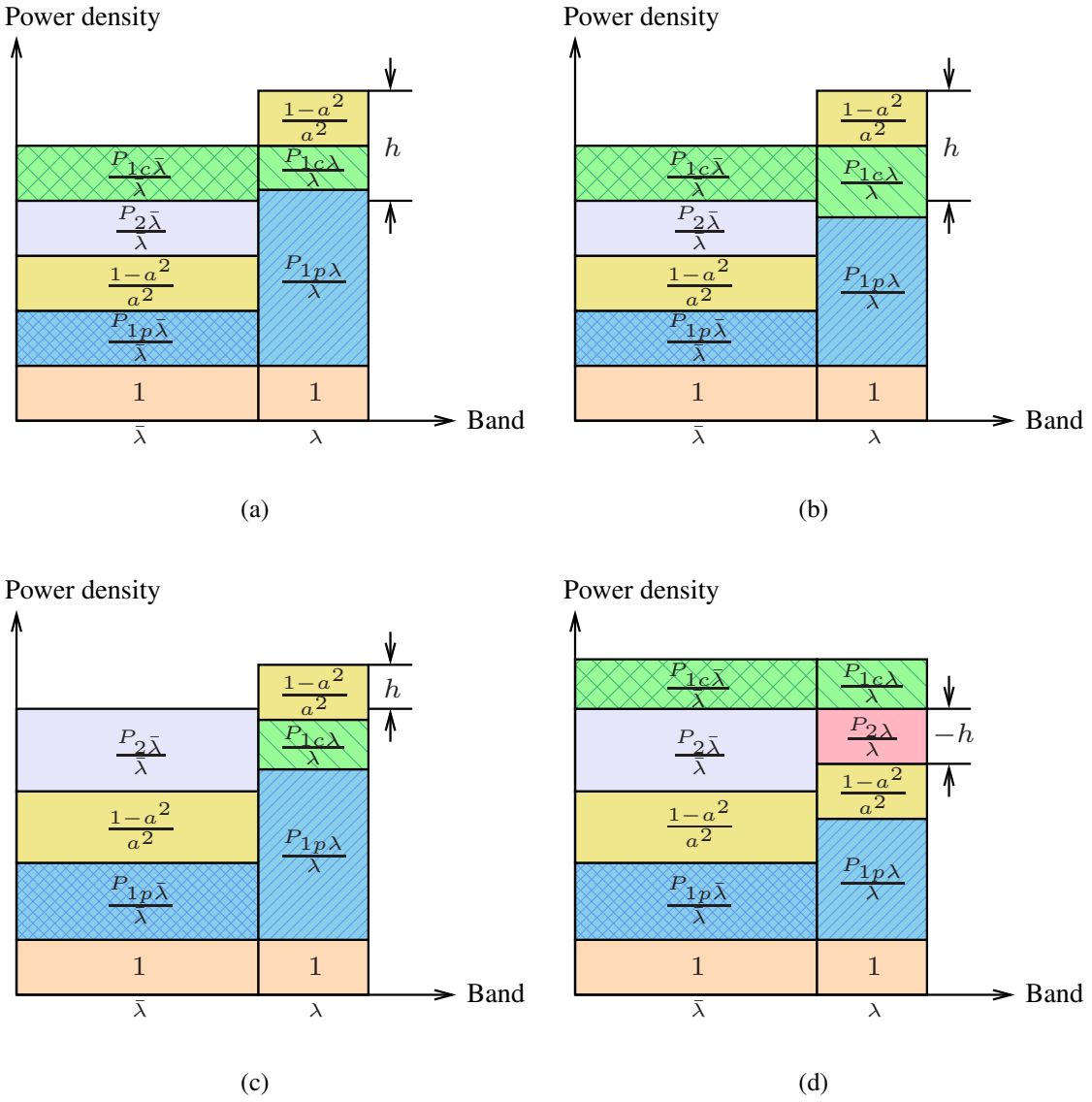


Fig. 2.6: Possible power allocation outcomes of the two-band non-naïve FDM scheme with ZGIC.

the power allocation scheme with flat spectrum top as in Fig. 2.6(d) is not optimal. This is because the achievable rates under such a scheme are formed by convex combinations of points on the curve of associated broadcast channel capacity, as the flat top requires $\frac{P_{1c\bar{\lambda}}}{\bar{\lambda}} = \frac{P_{1c\lambda}}{\lambda}$. As the broadcast channel capacity curve is convex, we can only achieve the points on the chord, which are dominated by the points on the curve corresponding to the scheme with no frequency division. Thus they are not optimal.

Next we generalize this conclusion to three-band FDM. Similarly we argue that the power-bandwidth allocation schemes with two or more subbands sharing the same flat top are not optimal. If two subbands are occupied solely by the transmission of X_1, they can be merged into one and this reduces to one of the noiseberg scenarios. □

2.3 Achievable Region For a Class of MGIC

For the MGICs with $0 < a \leq 1$, $b \geq 1$ and $ab \geq 1$, the HK region with Gaussian input can be simplified to be the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq \sum_{i=1}^{q} \lambda_i \left\{ \frac{1}{2} \log(1 + P_{1i}) \right\}, \\ R_2 &\leq \sum_{i=1}^{q} \lambda_i \left\{ \frac{1}{2} \log \left(1 + \frac{P_{2i}}{1 + a^2 \alpha_i P_{1i}} \right) \right\}, \\ R_1 + R_2 &\leq \sum_{i=1}^{q} \lambda_i \left\{ \frac{1}{2} \log \left(1 + \frac{P_{2i} + a^2 \bar{\alpha}_i P_{1i}}{1 + a^2 \alpha_i P_{1i}} \right) + \frac{1}{2} \log(1 + \alpha_i P_{1i}) \right\}, \end{aligned}$$

where $q \in \mathbb{N}$, $i \in \{1, \ldots, q\}$, $0 \leq \alpha_i \leq 1$, , $\bar{\alpha}_i = 1 - \alpha_i$, $\lambda_i \geq 0$, $\sum_{i=1}^{q} \lambda_i = 1$, $\sum_{i=1}^{q} \lambda_i P_{1i} \leq P_1$ and $\sum_{i=1}^{q} \lambda_i P_{2i} \leq P_2$. We refer to $\mathcal{R}_{\mathcal{HK}}$ as the above HK region with Gaussian input. The difficulty in using this region is largely due to the presence of the time sharing variable.

We now describe a simple transmission scheme for a MGIC with $ab \geq 1$. The scheme resembles the noiseberg scheme as it also utilizes the two-band non-naïve FDM. Specifically, in the λ-subband, only transmitter 1 transmits and receiver 1 decodes W_1 with a rate constraint defined in (2.5).

In the $\bar{\lambda}$-subband, transmitter 2 does not use rate splitting. Transmitter 1, on the other hand, employs two encoding schemes depending on the value of h. The corresponding decoding schemes are also different. We describe them in details below:

Multiplex region This corresponds to $h \leq \frac{1-a^2}{a^2}$. Sequential decoding is used at both

receivers. Receiver 1 first decodes W_2, which leads to the constraint

$$R_{2\bar{\lambda}} \leq \frac{1}{2}\log\left(1+\frac{b^2P_2/\bar{\lambda}}{1+P_{1A}/\bar{\lambda}}\right). \tag{2.6}$$

Subsequently, X_2 is subtracted from Y_1 and W_1 is decoded with constraint (2.3). Receiver 2 decodes W_2 with constraint (2.4). Comparing (2.4) and (2.6), we see that (2.6) is redundant. Therefore sequential decoding in the multiplex region achieves $\mathcal{R}_{\mathcal{N}}$.

Overflow region This corresponds to $h > \frac{1-a^2}{a^2}$. Receiver 1 employs simultaneous decoding while receiver 2 still uses sequential decoding. For receiver 1, let S_1, T_1 and T_2 be, respectively, the rates of transmitter 1's private message W_{1p}, common message W_{1c} and W_2. Then $R_{1\bar{\lambda}} = S_1 + T_1$ and $R_{2\bar{\lambda}} = T_2$. Evaluation of error probability will give us

$$\begin{aligned}
S_1 &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right), && (2.7)\\
T_1 &\leq \frac{1}{2}\log\left(1+\frac{P_{1C}}{\bar{\lambda}}\right), && (2.8)\\
T_2 &\leq \frac{1}{2}\log\left(1+\frac{b^2P_2}{\bar{\lambda}}\right), && (2.9)\\
S_1+T_1 &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}\right), && (2.10)\\
S_1+T_2 &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right), && (2.11)\\
T_1+T_2 &\leq \frac{1}{2}\log\left(1+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right), && (2.12)\\
S_1+T_1+T_2 &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right). && (2.13)
\end{aligned}$$

As receiver 2 decodes W_{1C} and W_2 sequentially, there will be two constraints

$$\begin{aligned}
T_1 &\leq \frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right), && (2.14)\\
T_2 &\leq \frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right). && (2.15)
\end{aligned}$$

Fourier-Motzkin elimination on (2.7)-(2.15) gives us

$$
\begin{aligned}
R_{1\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}\right), && (2.16)\\
R_{1\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right), && (2.17)\\
R_{2\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right), && (2.18)\\
R_{1\bar{\lambda}}+R_{2\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right), && (2.19)\\
R_{1\bar{\lambda}}+R_{2\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right), && (2.20)\\
R_{1\bar{\lambda}}+2R_{2\bar{\lambda}} &\leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right). && (2.21)
\end{aligned}
$$

Then the achievable rate region in overflow region is

$$
\begin{aligned}
R_1 &\leq \bar{\lambda}R_{1\bar{\lambda}}+\lambda R_{1\lambda},\\
R_2 &\leq \bar{\lambda}R_{2\bar{\lambda}},
\end{aligned}
\qquad (2.22)
$$

where $R_{1\bar{\lambda}}$, $R_{2\bar{\lambda}}$ and $R_{1\lambda}$ satisfy (2.16)-(2.21) and (2.5). Let $\mathcal{R}_{\mathcal{O}}$ denote this region. In Appendix A we provide a simplification of $\mathcal{R}_{\mathcal{O}}$ and prove that it is equivalent to $\mathcal{R}_{\mathcal{N}}$.

Combining the results in both the multiplex and overflow regions, we conclude that $\mathcal{R}_{\mathcal{N}}$ is achievable for the MGICs with $ab > 1$. Thus we have,

Theorem 2.3.1. *For the MGICs with $ab > 1$, $\mathcal{R}_{\mathcal{N}}$ associated with the ZGIC obtained by removing the interfering link with gain b is achievable.*

Corollary 1. *For the MGICs with $ab > 1$, $\mathcal{R}_{\mathcal{N}} = \mathcal{R}_{\mathcal{HK}}$.*

Proof. On the one hand, with Gaussian input, $\mathcal{R}_{\mathcal{HK}}$ of the MGIC is a subset of that of the ZGIC. On the other hand, $\mathcal{R}_{\mathcal{N}} \subseteq \mathcal{R}_{\mathcal{HK}}$ for the MGIC in general. But Theorem 1 states that $\mathcal{R}_{\mathcal{N}} = \mathcal{R}_{\mathcal{HK}}$ for the ZGIC. This establishes the equivalence between $\mathcal{R}_{\mathcal{N}}$ and $\mathcal{R}_{\mathcal{HK}}$ for the MGICs with $ab > 1$. □

This is consistent with [25, Theorem 12] that states for the MGICs with $ab > 1$, $\mathcal{R}_{\mathcal{HK}}$ is equivalent to that of the corresponding ZGIC.

2.4 Summary

This chapter established the equivalence between Costa's noiseberg region and the HK region with Gaussian input for the ZGIC. For the MGICs with $ab > 1$, an achievable rate region was developed by rate splitting and a mixture of sequential and simultaneous decoding. By comparing the new region to the noiseberg region of the ZGIC, we proved that it is a simplification of the HK region for the MGICs with $ab > 1$ and with Gaussian codebook.

CHAPTER 3
OPTIMAL POWER MANAGEMENT FOR REMOTE ESTIMATION WITH AN ENERGY HARVESTING SENSOR

3.1 Motivation And Literature Review

Conventional wireless and embedded systems often involve nodes that are powered by batteries. In many situations, however, periodically replacing or recharging batteries may be inconvenient or even impossible due to various physical restrictions. Thus for applications where sustainable and autonomous operation is expected, energy supply often becomes a severe bottleneck. Much effort has been devoted to the design of how to efficiently utilize the limited battery power to maximize the life span of the system [28–30].

With the recent technological breakthrough in renewable energy, an alternative approach has emerged that promises to supplement or even supplant the traditional battery power: harvesting energy from the ambient environment which enables much more sustainable operations [5–7]. There are existing wireless networks that operate solely or primarily on energy harvested from their operating environments through various sources such

as solar cells and vibration absorption devices. As hardware advances lead to much more efficient energy conversion and storage, it is expected that energy harvesting will play an increasingly important role in many power limited wireless systems.

However, utilizing harvested energy also presents many unique challenges, chief among them is the fluctuation of the energy level in time as dictated by the characteristics of the energy source. Further complicating the design issue is the finite energy storage capacity in most systems. This is in sharp contrast with the traditional power supply that is either constant over time or has a fixed total amount of energy for battery powered systems. The problem of particular interest to the present work is to devise optimal power allocation strategy over time given the unpredictable nature of the energy supply and a finite energy storage constraint. Additionally, we consider progressively more realistic assumptions to ensure that the proposed solutions are practically meaningful.

There have been many recent works on power management for energy harvesting communication and estimation systems. In [31–34], optimal transmission power and scheduling were studied for throughput maximization of wireless communication systems. In [35], throughput optimal power management policies were developed while maintaining the stability of data queues. Separately, minimizing outage probability instead of maximizing throughput was used as the design metric for power allocation [36]. In [37], distortion minimization for a sensor node communicating over a fading channel with delay constraint was studied. The problem of energy allocation over source acquisition/compression and transmission was addressed in [38]. The authors in [39] and [40] considered power allocation in multi-sensor remote estimation systems.

Most closely related to the present work is the remote estimation problem studied in [41], where an energy harvesting sensor observes and transmits the state of a discrete-time finite-valued Markov chain or a multi-dimensional Gaussian source. The sensor may not communicate all the time depending on the currently available energy level and the estimator relies on messages received from the sensor to produce real-time estimates of the

source state. Optimal communication scheduling and estimation strategies were developed, which jointly minimize an expected sum of communication and distortion costs over a finite time horizon. Invariably, in the absence of non-causal information at the transmitter about future energy arrival, a key ingredient among all those works is the use of dynamic programming (DP) [42] for obtaining power allocation.

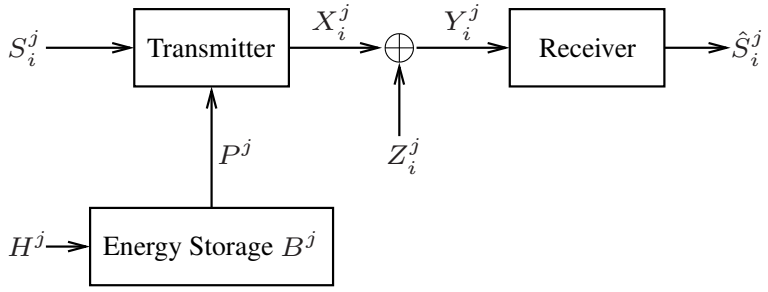


Fig. 3.1: Illustration of an energy harvesting estimation system.

In this work, we consider a remote estimation system with the transmitter powered by energy harvesting devices. The system is illustrated in Fig. 3.1 where the transmitter observes the source sequence S_i^j and has access to energy source H^j through a storage device. Both the signals and the additive noises are assumed to be independent and identically distributed (i.i.d.) Gaussian sequences and are independent of each other. Different from [41] where each transmission consumes 1 unit of energy and the sensor may choose to stay silent to save energy for future communication, our model assumes that the uncertainty at the estimator comes from the channel noise and the time-varying transmit power as dictated by the power allocation strategy and the energy harvesting sequence H^j. The objective is to minimize the mean squared error (MSE) at the receiver, averaged over the entire sequence. A natural transmission scheme for the Gaussian model is an uncoded transmission scheme where the transmitter simply amplifies the observed the sequence with the receiver implementing MMSE estimator [43] [44]. Indeed, for this remote estimation system and under a *constant* power constraint, uncoded transmission is known to be an optimal strategy [54].

The central question we try to answer in the present work is what is the optimal trans-

mission scheme at the transmitter and the associated power allocation policy over time under progressively more realistic assumptions with regard to the knowledge at the transmitter and/or the receiver as well as the storage capacity:

- whether the transmitter has non-causal side information (SI) of future harvested energy;
- whether the receiver knows the transmit power *a priori*;
- whether the storage device has finite or infinite capacity.

The clairvoyant case, namely the transmitter has non-causal knowledge of future harvested energy and the receiver knows the exact transmit power and the battery has infinite storage capacity, serves as a benchmark for performance comparison. It is shown that the optimal power allocation scheme for the clairvoyant case is a 'staircase-climbing' procedure, where the power level follows a non-decreasing staircase-like function. We will then replace these idealized assumptions with progressively more realistic ones and provide solutions to power allocation under each scenario. In particular, with causal side information, dynamic programming is used to obtain the power allocation scheme that minimizes the average MSE at the receiver. Heuristic approaches motivated by the structure of the optimal solutions are then developed that incur much less computational complexity compared with the DP solution. Performance evaluation will be conducted to identify conditions under which the achieved performance in terms of minimum mean square error (MMSE) with realistic assumptions is close to that of the clairvoyant case, therefore providing useful guidance to system design.

3.2 System Model

While the transmission of input sequence occurs 'instantaneously' using uncoded transmission (i.e., amplify and forward), we assume that energy harvesting and power alloca-

tion occur in a block-wise fashion: within each block the transmitter uses a given average transmit power, and energy harvested at the current block is assumed to be available for transmission during future blocks. Let the entire sequence consist of K blocks, indexed by $j = 1, 2, \cdots, K$. The sequence within each block is indexed using $i = 1, 2, \cdots, N$, i.e., each block is comprised of N observations. Let S_i^j denote the ith sample of the jth block of the observation sequence, which is assumed to be i.i.d. Gaussian with mean zero and variance σ_S^2. Thus, the energy harvested at the end of the jth block will be available for transmissions at the $(j+1)$th block. The energy storage is subject to a capacity limit, denoted by B_{max}.

For estimating an i.i.d. Gaussian sequence over an additive white Gaussian noise (AWGN) channel, it was known that uncoded transmission is optimal for minimizing the MSE under an average power constraint [46]. The optimality of such an uncoded transmission carries over to the case where the transmitter is subject to energy harvesting constraint where the harvested energy is a stationary ergodic sequence.

Lemma 3.2.1. *For estimating an i.i.d Gaussian sequence over an AWGN channel with $B_{max} \to \infty$, uncoded transmission is asymptotically optimal for minimizing the MSE, where the harvested energy is a stationary ergodic sequence and $K \to \infty$.*

Proof. See Appendix B. □

This optimality, along with its simplicity in implementation and analysis, motivate the use of uncoded transmission throughout this work; thus the transmitted signal corresponding to S_i^j is expressed as

$$X_i^j = \sqrt{\frac{P^j}{\sigma_S^2}} S_i^j, \tag{3.1}$$

where P^j is the average power constraint in block j. The transmitted signal goes through a

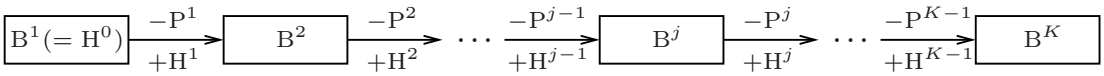


Fig. 3.2: Blockwise energy harvesting.

Gaussian channel with output

$$Y_i^j = X_i^j + Z_i^j,$$

where Z_i^j is an i.i.d. Gaussian noise sequence with mean zero and variance σ_Z^2 and is independent of the transmitted signal. The receiver output $\hat{S}_i^j$ denotes the estimate of the source sequence S_i^j.

As shown in Figs. 3.1 and 3.2, the blockwise energy harvesting and the transmission scheme is modeled by the following parameters:

- P^j denotes the average amount of energy to be expended during block j. To ease our notation, we have implicitly assumed that the energy is normalized by the block length, thus P^j represents both the average power at block j as well as the energy expended during block j.

- H^j denotes the average amount of energy harvested during block j. We assume that energy is replenished at the end of each block, hence H^j is not available until the beginning of block $j+1$.

- B^j indicates the energy level available at the beginning of block j. It varies linearly as long as the storage limit B_{max} is not exceeded, i.e.,

$$B^{j+1} = \min\left(B^j - P^j + H^j, B_{\text{max}}\right),$$

for $j = 1, \cdots, K-1$.

We denote the initially stored energy as B^1 or H^0 and assume

$$0 \leq B^1 = H^0 \leq B_{\max}.$$

This initial energy may come from the traditional battery source or energy harvested prior to initial transmission. The energy harvesting constraint is imposed on each block in the sense that the expended energy can not exceed the currently stored energy, i.e.,

$$P^j \leq B^j,$$

for any $j = 1, \cdots, K$. The block length N is assumed to be sufficiently large such that the average power constraint is satisfied.

The objective is to minimize the MSE averaged over the entire sequence

$$D = \frac{1}{KN} \sum_{j=1}^{K} \sum_{i=1}^{N} \left(\hat{S}_i^j - S_i^j \right)^2. \tag{3.2}$$

It is thus apparent from the above model that the achievable MSE depends on the *a priori* knowledge at both the transmitter and the receiver, i.e., whether the transmitter knows the non-causal side information about the harvested energy and whether the receiver knows the transmit power at each block. We start with the case where the transmitter has non-causal side information about harvested energy.

3.3 Non-Causal SI With Infinite $B_{\max}$

Non-causal SI is said to be available if the transmitter has prior knowledge of the entire harvested energy sequence $\left\{H^0, H^1, \cdots, H^{K-1}\right\}$ before transmission begins. This corresponds to the idealized assumption of fully predictable environment whose solution gives a lower bound on the MSE performance under more realistic conditions. To develop more

insight on the structure of the optimal power allocation scheme, we assume in this section that the battery storage has unlimited capacity, i.e., $B_{\max} \to \infty$. The case of finite $B_{\max}$ will be studied in Section 3.4.

3.3.1 Transmit Power Known to Receiver

From (3.1), if the receiver knows the transmit power, it can construct the optimal MMSE estimator

$$\hat{S}_i^j = \frac{\sqrt{P^j}\sigma_S Y_i^j}{\sigma_Z^2 + P^j}, \tag{3.3}$$

and the corresponding MSE is

$$\mathbb{E}[D] \;=\; \frac{1}{K}\sum_{j=1}^{K} \frac{\sigma_S^2\sigma_Z^2}{\sigma_Z^2 + P^j} \triangleq \frac{1}{K}\sum_{j=1}^{K} h_1(P^j), \tag{3.4}$$

where the expectation is taken with respect to the observation sequence and the channel noise sequence for a given transmit power sequence $\{P^j, j = 1, \cdots, K\}$. Then the optimal power allocation can be obtained by solving the following optimization problem

$$\text{minimize} \quad \mathbb{E}[D] \tag{3.5}$$

$$\text{subject to} \quad P^j \geq 0$$

$$\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \leq 0, \quad \text{for} \quad j = 1, \cdots, K. \tag{3.6}$$

It is easy to verify that the above problem is convex and satisfies the Slater's condition [47]. Hence, the Lagrange duality method can be used to obtain the global optimum. The Lagrangian associated with this problem is

$$\mathcal{L} = \frac{1}{K}\sum_{j=1}^{K} \frac{\sigma_S^2\sigma_Z^2}{\sigma_Z^2 + P^j} - \sum_{j=1}^{K} \mu_j P^j + \sum_{j=1}^{K} \lambda_j \left(\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \right).$$

Using the Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} P^j &\geq 0, \\ \mu_j &\geq 0, \\ \mu_j P^j &= 0, \\ \sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k &\leq 0, \\ \lambda_j &\geq 0, \qquad (3.7) \\ \lambda_j \left(\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \right) &= 0, \qquad (3.8) \\ \frac{\partial \mathcal{L}}{\partial P^j} = \frac{-\sigma_S^2 \sigma_Z^2}{(\sigma_Z^2 + P^j)^2} - \mu_j + \sum_{k=j}^{K} \lambda_k &= 0, \end{aligned}$$

we obtain the optimal solution as

$$P^j = \max\left(\nu_j - \sigma_Z^2, 0\right), \qquad (3.9)$$

where

$$\nu_j = \frac{\sigma_S \sigma_Z}{\sqrt{\sum_{k=j}^{K} \lambda_k}}. \qquad (3.10)$$

Comparing (3.9) with the solution to the standard water-filling problem (e.g., Example 5.2 in [47]), one can immediately recognize the similarity except for a time varying water level in the present problem and a time varying ground level in [47]. Specifically,

- the varying 'ground level' in the original water-filling problem becomes flat with uniform height σ_Z^2;
- the optimal 'water level' is not a constant, but changes over blocks.

As we shall see, the solution to equation (3.9) can be interpreted as a staircase climbing

process (also referred to in [33] as staircase water-filling). Define the following for convenience:

- If the power level changes after block t, i.e., $\nu_t \neq \nu_{t+1}$, $t = 1, \cdots, K$, then block t is defined to be a *transition block*. From (3.10), it is clear that the change of power level (i.e., climbing to the next level) occurs only when the corresponding $\lambda_k > 0$. Block K is also a transition block, as we define $\nu_{K+1} = \infty$.

- Let $\mathcal{S} = \{t_1, t_2, \cdots, t_{|S|}\}$ be a sequence that contains all the indices of the transition blocks. From the assumption $\nu_{K+1} = \infty$ it follows that $t_{|S|} = K$.

- If $t_i, t_{i+1} \in \mathcal{S}$, then the blocks indexed by $\{t_i + 1, \cdots, t_{i+1}\}$ are collectively referred to as the ith *transition interval*.

We can now describe the structure of the optimal power allocation scheme.

- The power level is non-decreasing over the blocks, i.e.,

$$\nu_{t_1} \leq \nu_{t_2}, \quad \text{if} \quad t_1 < t_2 \quad \text{and} \quad t_1, t_2 \in \mathcal{S}.$$

  This can be easily seen from (3.7) and (3.10). Given that σ_Z^2 is a constant, from (3.9), it is clear that the optimal P_j behaves like staircase climbing: P_j remains constant within each transition interval but increases (climbs to the next level) after each transition block.

- The available battery storage is depleted at the end of a transition block, i.e.,

$$P^t = B^t \quad \text{and} \quad B^{t+1} = H^t, \quad \text{if} \quad t \in \mathcal{S}.$$

  Thus at the beginning block of each transition interval, the available energy comes entirely from the energy harvested during the previous block. This can be observed from the definition of level (3.10) and the complementary slackness condition (3.8).

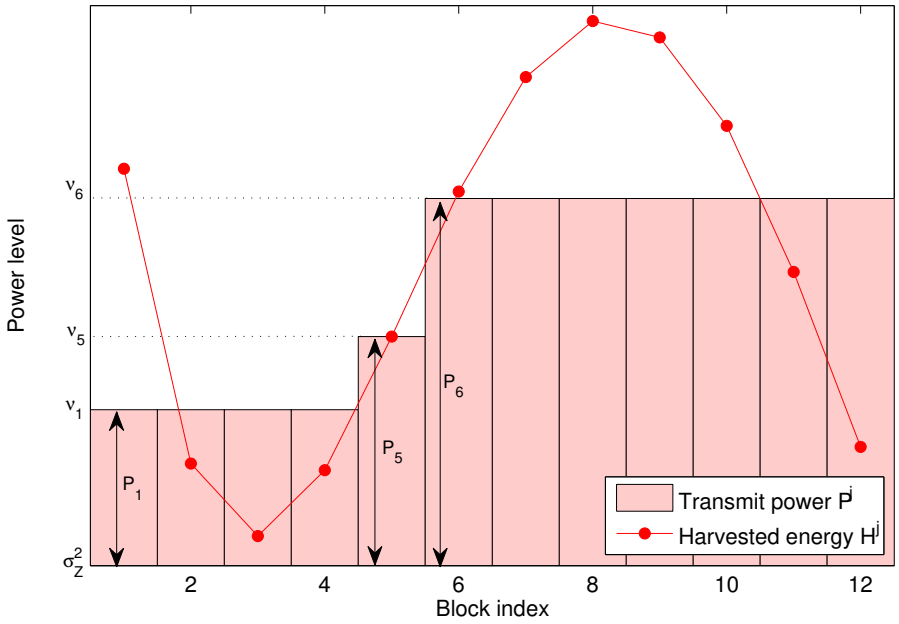


Fig. 3.3: Structure of the optimal power allocation structure with non-causal SI and infinite $B_{\max}$.

- The optimal power allocation can be obtained by considering individual transition intervals separated by transition blocks. Within each interval that consists of blocks $\{t_i + 1, \cdots, t_{i+1}\}$ where $t_i, t_{i+1} \in \mathcal{S}$, the entire available energy, i.e., $\sum_{j=t_i}^{t_{i+1}-1} H^j$ is equally allocated to each block. This follows naturally from the previous property that energy harvested in any transition interval is completely depleted.

Fig. 3.3 illustrates the structure of the optimal power allocation scheme. The curve represents the amount of energy harvested over the blocks, while the shaded bars correspond to the transmit power. Even though the amount of harvested energy fluctuates, the optimal transmit power forms an ascending staircase-like function. Apparently, this is only feasible when the transmitter knows the entire harvested energy sequence *a priori*.

Thus the optimal power allocation scheme amounts to identifying all the transition blocks as power allocation between two neighboring transition blocks can be trivially solved, i.e., a simple average of all available energy within the transition interval. A backward-search procedure (Algorithm 2 in [33]) can be implemented to find the optimal

transition blocks.

3.3.2 Transmit Power Not Known to Receiver

If the transmit power is not readily available, the receiver may use an estimate of the transmit power in the MMSE estimator. Choose the following $\hat{P}^j$ to be the estimate of the transmit power in block j

$$\hat{P}^j = \max\left(\frac{1}{N}\sum_{i=1}^{N}(Y_i^j)^2 - \sigma_Z^2, 0\right). \tag{3.11}$$

The corresponding MSE, denoted by $\tilde{D}$, is easily shown in Appendix C to be

$$\tilde{D} = \frac{1}{K}\sum_{j=1}^{K}\frac{\left(\hat{P}^j\sigma_Z^2 + \left(\sigma_Z^2 + \hat{P}^j - \sqrt{\hat{P}^j P^j}\right)^2\right)\sigma_S^2}{(\sigma_Z^2 + \hat{P}^j)^2} \triangleq \frac{\sigma_S^2}{K}\sum_{j=1}^{K} f_j(\hat{P}^j). \tag{3.12}$$

Thus we now formulate the optimization problem with $\mathbb{E}_{\hat{P}^j}[\tilde{D}]$ as the objective function

$$\text{minimize} \quad \mathbb{E}_{\hat{P}^j}[\tilde{D}] \tag{3.13}$$

$$\text{subject to} \quad P^j \geq 0$$

$$\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \leq 0, \quad \text{for} \quad j = 1, \cdots, K. \tag{3.14}$$

Since the statistical characterization of $\hat{P}^j$ does not admit a closed-form expression of $\mathbb{E}_{\hat{P}^j}[\tilde{D}]$, we consider its approximation which is asymptotically accurate as N becomes large. To begin, we first replace $\hat{P}^j$ in (3.11) using the following estimate

$$\tilde{P}^j = \frac{1}{N}\sum_{i=1}^{N}(Y_i^j)^2 - \sigma_Z^2. \tag{3.15}$$

Notice that as long as $P^j > 0$, $\tilde{P}^j = \hat{P}^j$ with probability approaching 1 as N becomes

large. Define Δ to be the bias of this estimate, i.e.,

$$\tilde{P}^j = P^j + \Delta.$$

It is straightforward to verify that

$$\begin{aligned} \mathbb{E}[\Delta] &= 0, \\ \mathbb{E}[\Delta^2] &= \frac{2}{N}(\sigma_Z^2 + P^j)^2. \end{aligned}$$

Then it follows from Taylor series approximation that

$$\begin{aligned} \mathbb{E}_{\tilde{P}^j}[\tilde{D}] &\doteq \frac{\sigma_S^2}{K}\sum_{j=1}^{K}\mathbb{E}\left[f_j(P^j) + f_j'(P^j)\Delta + \frac{f_j''(P^j)}{2}\Delta^2\right] \\ &= \frac{\sigma_S^2}{K}\sum_{j=1}^{K}\left(f_j(P^j) + f_j'(P^j)\mathbb{E}[\Delta] + \frac{f_j''(P^j)}{2}\mathbb{E}[\Delta^2]\right) \\ &= \frac{\sigma_S^2}{K}\sum_{j=1}^{K}\left(\frac{\sigma_Z^2}{\sigma_Z^2 + P^j} + \frac{(P^j)^2 + 14P^j\sigma_Z^2 + \sigma_Z^4}{2NP^j(\sigma_Z^2 + P^j)}\right) \qquad (3.16) \\ &\triangleq \frac{1}{K}\sum_{j=1}^{K} h_2(P^j). \qquad (3.17) \end{aligned}$$

where $f_j'(\cdot)$ and $f_j''(\cdot)$ denote the first and second derivatives of $f_j(\cdot)$.

Substituting (3.16) into the objective function, we rewrite the optimization problem (3.13) as

$$\begin{aligned} \text{minimize} \quad & \frac{\sigma_S^2}{K}\sum_{j=1}^{K}\left(\frac{\sigma_Z^2}{\sigma_Z^2 + P^j} + \frac{(P^j)^2 + 14P^j\sigma_Z^2 + \sigma_Z^4}{2NP^j(\sigma_Z^2 + P^j)}\right) \qquad (3.18) \\ \text{subject to} \quad & P^j \geq 0 \qquad (3.19) \\ & \sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \leq 0, \quad \text{for} \quad j = 1, \cdots, K. \qquad (3.20) \end{aligned}$$

It is straightforward to verify that (3.18) is convex in P^j by showing that its second or-

der derivative is positive. Thus the Lagrange duality method can be applied to obtain the solution. The Lagrangian associated with this problem is

$$\begin{aligned}\mathcal{L} &= \frac{\sigma_S^2}{K}\sum_{j=1}^{K}\left(\frac{\sigma_Z^2}{\sigma_Z^2+P^j}+\frac{(P^j)^2+14P^j\sigma_Z^2+\sigma_Z^4}{2NP^j(\sigma_Z^2+P^j)}\right)\\ &\quad -\sum_{j=1}^{K}\mu_j P^j+\sum_{j=1}^{K}\lambda_j\left(\sum_{k=1}^{j}P^k-\sum_{k=0}^{j-1}H^k\right).\end{aligned}$$

and the KKT conditions are

$$\begin{aligned} P^j &\geq 0\\ \mu_j &\geq 0\\ \mu_j P^j &= 0\\ \sum_{k=1}^{j}P^k-\sum_{k=0}^{j-1}H^k &\leq 0\\ \lambda_j &\geq 0\\ \lambda_j\left(\sum_{k=1}^{j}P^k-\sum_{k=0}^{j-1}H^k\right) &= 0\\ \frac{\partial\mathcal{L}}{\partial P^j}=-\frac{\sigma_S^2\sigma_Z^2\left((2N+13)(P^j)^2+2P^j\sigma_Z^2+\sigma_Z^4\right)}{2KN(P^j)^2(P^j+\sigma_Z^2)^2}-\mu_j+\sum_{k=j}^{K}\lambda_k &= 0.\end{aligned}$$

Since P^j appears in the denominator of (3.18), constraint (3.19) can be further tightened as

$$P^j > 0.$$

Then $\mu_j = 0$ and we obtain

$$\sum_{k=j}^{K}\lambda_k=\frac{\sigma_S^2\sigma_Z^2}{K}\frac{(2N+13)(P^j)^2+2P^j\sigma_Z^2+\sigma_Z^4}{2N(P^j)^2(\sigma_Z^2+P^j)^2}\triangleq g(P^j). \tag{3.21}$$

From (3.21), it is difficult to compute the inverse of $g(\cdot)$ and solve for P^j. However, the derivatives of both $g(\cdot)$ and its inverse $g^{-1}(\cdot)$ are negative, which means that P^j is

monotonically decreasing in $\sum_{k=j}^{K} \lambda_k$. Comparing with (3.9) and (3.10) where P^j is also a decreasing function of $\sum_{k=j}^{K} \lambda_k$, we observe that the optimal power allocation can, again, be obtained through a staircase climbing procedure, but with a flat bottom of height 0.

3.4 Non-Causal SI With Finite $B_{\max}$

We now consider the non-causal SI case with finite $B_{\max}$, i.e., the battery storage has a finite capacity. The infinite $B_{\max}$ assumption used in the previous section may not be applicable for some energy harvesting systems, where the battery storage capacity is not large enough compared with incoming energy.

We first consider the case where the receiver knows the transmit power. With finite and arbitrary battery storage capacity $B_{\max}$, it is possible that in certain blocks harvested energy exceeds the battery storage limit. In this case, the total transmit power is strictly less than the total amount of harvested energy over the K blocks, which suggests that the energy harvesting constraint (3.6) in optimization problem (3.5) is loose. Therefore we replace it with $P^j \leq B^j$ and reconstruct the optimization problem as follows to find optimal power allocation

$$\text{minimize} \quad \frac{1}{KN} \sum_{j=1}^{K} \sum_{i=1}^{N} \mathbb{E}[(\hat{S}_i^j - S_i^j)^2] \triangleq \frac{1}{K} \sum_{j=1}^{K} h(P^j) \tag{3.22}$$

$$\text{subject to} \quad 0 \leq P^j \leq B^j, \quad \text{for} \quad j = 1, \cdots, K, \tag{3.23}$$

where $h(\cdot) = h_1(\cdot)$ in (3.4) and

$$B^j = \begin{cases} B^1, & j = 1, \\ \min\left(B^{j-1} - P^{j-1} + H^{j-1}, B_{\max}\right), & j = 2, \cdots, K. \end{cases}$$

The constraints in (3.23) can be decomposed into

$$\begin{aligned} P^j &\geq 0, \\ \sum_{k=1}^{j} P^k &\leq \sum_{k=0}^{j-1} H^k, \\ \sum_{m=j-q}^{j} P^m &\leq B_{\text{max}} + \sum_{m=j-q}^{j-1} H^m, \end{aligned} \tag{3.24}$$

for $q = 0, \cdots, j-1$ and $j = 1, \cdots, K$.

Clearly, the new problem (3.22) is convex. Comparing with the optimization problem (3.5) formed under infinite B_{max} assumption, we observe that the only difference lies in the additional set of linear constraints in (3.24). Thus the Lagrange duality method can, again, be applied to obtain the global optimum. The Lagrangian associated with this problem is

$$\begin{aligned} \mathcal{L} &= \frac{1}{K} \sum_{j=1}^{K} \frac{\sigma_S^2 \sigma_Z^2}{\sigma_Z^2 + P^j} - \sum_{j=1}^{K} \mu_j P^j + \sum_{j=1}^{K} \lambda_j \left(\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \right) \\ &\quad + \sum_{j=1}^{K} \sum_{q=0}^{j-1} \gamma_{jq} \left(\sum_{m=j-q}^{j} P^m - B_{\text{max}} - \sum_{m=j-q}^{j-1} H^m \right). \end{aligned}$$

Using the KKT conditions,

$$\begin{aligned}
P^j &\geq 0 \\
\mu_j &\geq 0 \\
\mu_j P^j &= 0 \\
\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k &\leq 0 \\
\lambda_j &\geq 0 \\
\lambda_j \left(\sum_{k=1}^{j} P^k - \sum_{k=0}^{j-1} H^k \right) &= 0 \\
\gamma_{jq} &\geq 0 \\
\sum_{m=j-q}^{j} P^m - B_{\max} - \sum_{m=j-q}^{j-1} H^m &\leq 0 \\
\gamma_{jq} \left(\sum_{m=j-q}^{j} P^m - B_{\max} - \sum_{m=j-q}^{j-1} H^m \right) &= 0 \\
\frac{\partial \mathcal{L}}{\partial P^j} = \frac{-\sigma_S^2 \sigma_Z^2}{K(\sigma_Z^2 + P^j)^2} - \mu_j + \sum_{k=j}^{K} \lambda_k + \sum_{k=j}^{K} \sum_{q=0}^{k-1} \gamma_{kq} &= 0,
\end{aligned}$$

we obtain

$$P^j = \max \left(\frac{\sigma_S \sigma_Z}{\sqrt{K} \sqrt{\sum_{k=j}^{K} \lambda_k + \sum_{k=j}^{K} \sum_{q=0}^{k-1} \gamma_{kq}}} - \sigma_Z^2, 0 \right). \tag{3.25}$$

Compared with (3.9), we see that the effect of finite $B_{\max}$ on the solution is only observed through γ_{kq} in (3.25). In particular, if the constraints in (3.24) are strictly satisfied, the optimal power allocation scheme reduces to the one in (3.9) with the same structure as described in Section 3.3. This is the case where the energy harvested at any given block does not result in the breach of the storage limit, thus the solution is equivalent to setting $B_{\max}$ to infinity. However, when the constraints in (3.24) are satisfied with equality, the monotonicity of the power level as well as other properties may no longer hold.

For the case where the receiver does not know the transmit power, suitable forms of

estimators can be used to replace P^j in (3.3) and construct the estimate of the sequence S_i^j. For example, if we choose the estimator in (3.15) and apply the same approximation as proposed in the previous section, the convex problem (3.22) can again be applied to find the optimal power allocation policy by letting $h(\cdot) = h_2(\cdot)$ in (3.17).

3.5 Causal SI

We now consider the more realistic causal SI case, i.e., the transmitter only has knowledge to energy harvested in the past. The non-causal SI case discussed in the previous section is often time too idealized to be applicable for a real energy harvesting system where future available energy typically fluctuates and is hard to predict. Nevertheless, for most applications, the available energy from ambient sources does exhibit certain degree of correlation, e.g., solar power supply rarely experiences abrupt changes within a certain time window. As a crude approximation, we now model the harvested energy H^j as a first-order stationary Markov process over j, i.e., given $H^0 = h^0$,

$$\mathcal{P}\left(H^{K-1} \mid H^0 = h^0\right) = \prod_{j=2}^{K-1} \mathcal{P}\left(H^j \mid H^{j-1}\right) \mathcal{P}\left(H^1 \mid H^0 = h^0\right).$$

In addition, we also impose the more realistic assumption that the battery storage capacity $B_{\max}$ is finite.

We consider first the case where the receiver knows the transmit power. The following problem is constructed to find the optimal power allocation

$$\begin{aligned} \text{minimize} \quad & \frac{1}{KN} \sum_{j=1}^{K} \sum_{i=1}^{N} \mathbb{E}[(\hat{S}_i^j - S_i^j)^2] \triangleq \frac{1}{K} \sum_{j=1}^{K} h(P^j) \\ \text{subject to} \quad & 0 \le P^j \le B^j \\ & B^j = \min\left(B^{j-1} + H^{j-1} - P^{j-1}, B_{\max}\right), \quad \text{for} \quad j = 1, \cdots, K, \end{aligned} \tag{3.26}$$

where $h(\cdot) = h_1(\cdot)$ in (3.4).

In general, this problem cannot be solved by independently minimizing the MSE in each individual block due to energy harvesting constraints: how much energy dispensed in the current and previous blocks affects how much energy that is available for future blocks. As a simple example, consider the case of $K = 2$. The problem simplifies to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\left(\frac{\sigma_S^2\sigma_Z^2}{\sigma_Z^2+P^1}+\frac{\sigma_S^2\sigma_Z^2}{\sigma_Z^2+P^2}\right) \\ \text{subject to} \quad & 0 \le P^1 \le B^1 \\ & 0 \le P^2 \le B^2 \\ & B^1 = H^0 \\ & B^2 = \min\left(B^1+H^1-P^1, B_{\max}\right). \end{aligned} \tag{3.27}$$

Clearly, when deciding the transmit power P^1 for the first block, we should consider not only its contribution to the MSE of the first block, but also its impact on the transmit power P^2, as it affects the stored energy B^2 available for the second block. However, at the beginning of block 1, the other factor that determines B^2 - harvested energy H^1 - is not known yet, which makes it impossible to evaluate the precise impact of P^1 on B^2. Thus directly optimizing P^1 at the beginning of block 1 is not feasible. Instead, we can

- first, look for the optimal P^{2*} that minimizes the expected MSE of block 2 given all possible P^1; in this example, the optimal strategy would be to simply use up the energy supply available during the second block, i.e., $P^{2*} = B^2$ for any P^1;
- then, optimize P^1 with P^2 replaced by P^{2*} as a function of P^1.

The above procedure is in fact a simple example of dynamic programming. For the general problem with causal SI, DP can indeed be used to solve for the optimal power allocation problem. Specifically, given the initial energy storage $B^1 = H^0$, the optimal power allocation policy can be obtained by recursively computing $J_K, \cdots, J_1$ based on

Bellman's equation [42], where

$$J_K(H^{K-1}, B^K) = \min_{0 \le P^K \le B^K} h(P^K) = h(B^K), \tag{3.28}$$

$$J_j(H^{j-1}, B^j) = \min_{0 \le P^j \le B^j} \left(h(P^j) + \bar{J}_{j+1}(H^{j-1}, B^j - P^j) \right), \tag{3.29}$$

and

$$\bar{J}_{j+1}(H^{j-1}, x) = \mathbb{E}_{\tilde{H}^j} \left[J_{j+1} \left(\tilde{H}^j, \min \left(B_{\max}, x + \tilde{H}^j \right) \right) \mid H^{j-1} \right], \tag{3.30}$$

for $j = 1, \cdots, K-1$, where $\tilde{H}^j = \mathbb{E}[H^j | H^{j-1}]$ is the expected harvested energy in the jth block.

In (3.30), $\bar{J}_{j+1}$ indicates the expected MSE contributed from blocks $\{j+1, \cdots, K\}$. As expected, the solution ensures that all power is depleted at block K. As for the intermediate blocks, the optimal power allocation policy can be interpreted as a tradeoff between the current reward and the future reward and is determined by the past harvested energy and allocated power. The resulting transmit power P^j is no longer a non-decreasing sequence as in the non-causal SI case.

For the case where the receiver does not have knowledge of the transmit power, we can again replace P^j with suitably constructed estimates. For example, replacing P^j with the same estimator as proposed in Section 3.3, the optimization problem (3.26) can also be used to solve for the optimal power allocation policy by letting $h(\cdot) = h_2(\cdot)$ in (3.17). Again, DP can be applied to obtain the optimal solution and we omit the details here.

In order to evaluate the complexity of DP, we define ϵ to be the step size of the search for optimal transmit power. Then, DP involves roughly $\mathcal{O}\left(\frac{KB_{\max}^3}{\epsilon^3}\right)$ arithmetic operations [48], which results in long execution time especially with fine step sizes. This motivates us to look for heuristic approaches for power allocation that resemble the optimal power allocation schemes yet with significantly reduced complexity. In addition, the DP formulation requires backward recursion thus relies on the finite horizon assumption (i.e., $K < \infty$).

A perhaps more natural solution would allow infinite horizon operation with reasonable performance and complexity.

3.6 Heuristic Schemes

In this section, we propose two heuristic schemes for the case where causal SI is available at the transmitter and the receiver does not have knowledge of the transmit power. Both schemes have significantly reduced complexity compared with DP. The first one, referred to as myopic DP, is capable of avoiding the recursive structure introduced by the energy harvesting constraints in DP. Instead of considering the expected MSE contributed from all future blocks, it only involves the immediate future reward - expected MSE from the next block, which significantly alleviates the computational complexity when K is large. The second scheme, which we simply refer to as an equalizing scheme, is motivated by the observation that the power allocation profile obtained with non-causal SI is flat within each transition interval and change only occurs when energy is completely depleted. It works in a way such that the running average of the total consumed energy is close to the expectation of the harvested energy per block. Both schemes involve only computations in a forward manner and thus can be applied to the cases with infinite time horizon, i.e. $K \to \infty$. This is a significant operational advantage compared with the DP approach that is restricted to a finite time horizon due to its recursive structure.

3.6.1 Myopic DP

Recall that in the original DP approach, there are two terms in the objective function (3.29) for block j, $j = 1, \cdots, K-1$ - current reward $h(P^j)$ and future expected reward $\bar{J}_{j+1}(H^{j-1}, B^j - P^j)$, where $\bar{J}$ represents the expected MSE contributed from all future blocks. Such nested structure often requires high computational complexity, which becomes a major obstacle in practical implementation. One way to simplify the computation

is to replace the $\bar{J}$ function by

$$\bar{J}'_{j+1}(H^{j-1}, x) = \mathbb{E}_{\tilde{H}^j}\left[\min_{0\leq P^{j+1}\leq B^{j+1}} h(P^{j+1}) \mid H^{j-1}\right], \tag{3.31}$$

where

$$B^{j+1} = \min\left(B^j - P^j + \tilde{H}^j, B_{\text{max}}\right).$$

In essence, this simplified DP approach takes a myopic view of the future where it only involves the immediate future payoff (i.e., that of the next block instead of the entire future blocks). This myopic approach leads to the simplification of (3.31) into the following form

$$\bar{J}'_{j+1}(H^{j-1}, x) = \mathbb{E}_{\tilde{H}^j}\left[h\left(\min\left(B^j - P^j + \tilde{H}^j, B_{\text{max}}\right)\right) \mid H^{j-1}\right].$$

This simplification is possible because $h(\cdot)$ is a non-decreasing function hence the optimal P^j in (23) always takes the maximum possible value, i.e., a greedy strategy for the energy consumption in the next block. This removes the recursive structure of the program and greatly eases the computation burden. We note that while this approach determines the current power consumption by *assuming* a greedy approach for the next block in terms of energy consumption, this does not necessarily imply that the actual power allocation will always result in a greedy approach - the actual power consumed at the next block will be determined in subsequent steps of the algorithm.

3.6.2 An Equalizing Scheme

This scheme is motivated by the simple observation that for the non-causal SI case, the optimal power allocation tends to smooth out power distribution to the extent possible - it is always flat within each transition interval and change only occurs when energy is completely depleted. For sufficiently large K, it is reasonable to assume that the total

available energy is approximately $K\mathbb{E}[H^j]$, i.e., the average energy consumed for each block j is approximately $\mathbb{E}[H^j]$. As such, one may strive to determine the transmit power in a way such that the running average of consumed energy per block up to the current block, say, block j, is $\mathbb{E}[H^j]$, i.e., choose P^j such that $\sum_{k=1}^{j} P^k = j\mathbb{E}[H^j]$ if $B^j \geq P^j$; otherwise we choose $P^j = B^j$. While with high probability the power levels at different blocks are not identical, such a naive scheme leads to an allocation scheme that results in a rather smooth transmit power profile that exhibits only small fluctuation.

3.7 Numerical Results

In this section, we present some numerical results to compare the performance of all the previously discussed schemes, namely the staircase climbing algorithm for non-causal SI case as well as DP and the two heuristic schemes for causal SI case. Denote D_1 and D_2 to be the MSE (defined in (3.2)) achieved at the receiver with and without knowledge of the transmit power, respectively. They serve as indicators of system performance and are plotted with varying channel noise variance σ_Z^2 in all the following figures. Another important parameter is the block length N, as it has a significant impact on the accuracy of transmit power estimation and therefore on the overall performance. We fix the total number of blocks $K = 12$ and choose the harvested energy H^j to be i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$, for $j = 0, 1, \cdots, 11$.

In Fig. 3.4, we fix $\sigma_S^2 = 1$ and $N = 10^4$ and plot D_1 with σ_Z^2 from -20dB to 20dB to examine the effect of different battery storage capacity B_{max} on the performance with non-causal SI. While it is apparent that the MSE increases as noise variance increases, the estimation performance improves with increasing storage capacity. Indeed, with $B_{\text{max}} = 12$, the performance is identical to that of the case with infinite storage capacity.

In Fig. 3.5, we fix $N = 10^4$ and vary $\sigma_S^2 \in \{10, 10^2, 10^4\}$ and σ_Z^2 from -20dB to 20dB to examine D_1, i.e., the corresponding MSE when the receiver knows the transmit power.

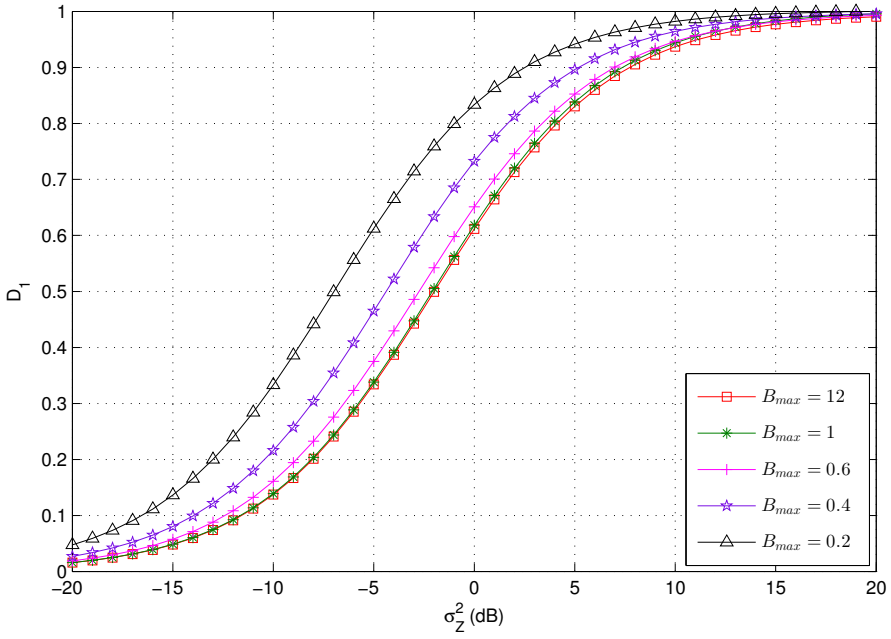


Fig. 3.4: D_1 vs σ_Z^2 with non-causal SI and $N = 10^4$, $K = 12$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, 1, \cdots, 11$.

As expected, the distortion with both causal SI and non-causal SI increases with increasing noise variance. The MSE with causal SI always dominates that with non-causal SI although the difference between the two is very small.

Fig. 3.6 shows the performance when the knowledge of transmit power is not available to the receiver. The signal variance σ_S^2 is fixed to be 1. It can be seen, again, that D_2 increases with increasing channel noise variance σ_Z^2 and the MSE obtained with causal SI is always higher than that with non-causal SI. Here we also compare the performance with two different block lengths, i.e., $N = 10^2$ and $N = 10^5$. As expected, smaller N results in larger distortion, since its estimate of the transmit power is less accurate than that of larger N.

Figs. 3.7-3.9 illustrate the average MSE for the case where the transmitter has only causal SI of future harvested energy and the receiver does not know the transmit power. We fix the signal variance $\sigma_S^2 = 1$ and the battery storage capacity $B_{\text{max}} = 1$. The D_2 curves of DP and the two heuristic schemes in Figs. 3.7-3.9 correspond to the average

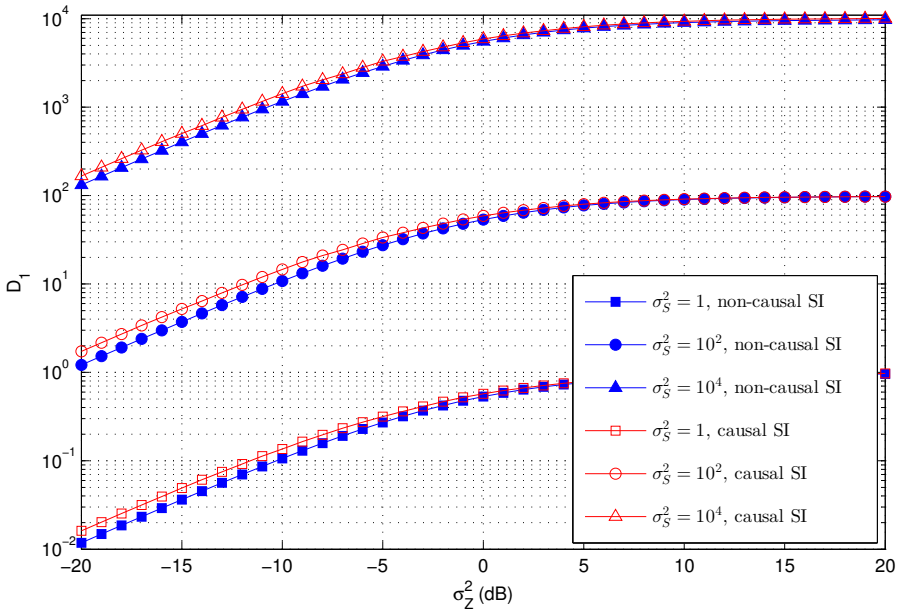


Fig. 3.5: D_1 vs σ_Z^2 with $N = 10^4$, $K = 12$, $B_{\max} = 1$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, 1, \cdots, 11$.

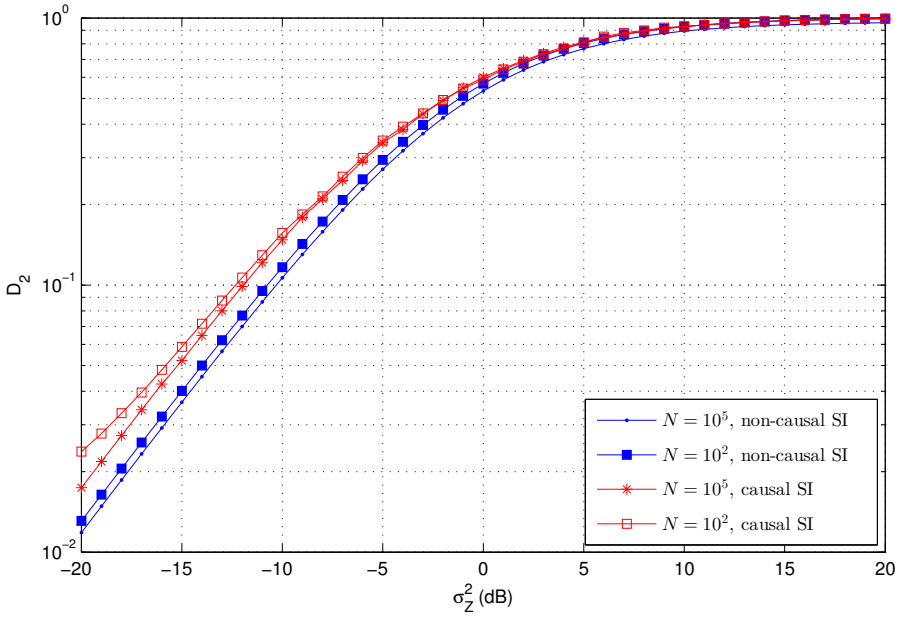


Fig. 3.6: D_2 vs σ_Z^2 with $\sigma_S^2 = 1$, $K = 12$, $B_{\max} = 1$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, \cdots, 11$.

MSE of 100 randomly generated sequences $\{H^j, j = 0, \cdots, 11\}$ for $N = 10^5$, $N = 100$ and $N = 10$, respectively. We also plot the optimal MSE curves when non-causal SI is

available as benchmarks for performance comparison. In Fig. 3.7, it can be observed that both heuristic schemes as well as DP have almost the same MSE when $N = 10^5$; in Fig. 3.8 when $N = 100$, the curves corresponding to DP and myopic DP still overlap each other while the performance of the equalizing scheme is considerably worse; when $N = 10$ in Fig. 3.9, the MSE of DP is smaller than that of the two heuristic schemes, especially when σ_Z^2 is large.

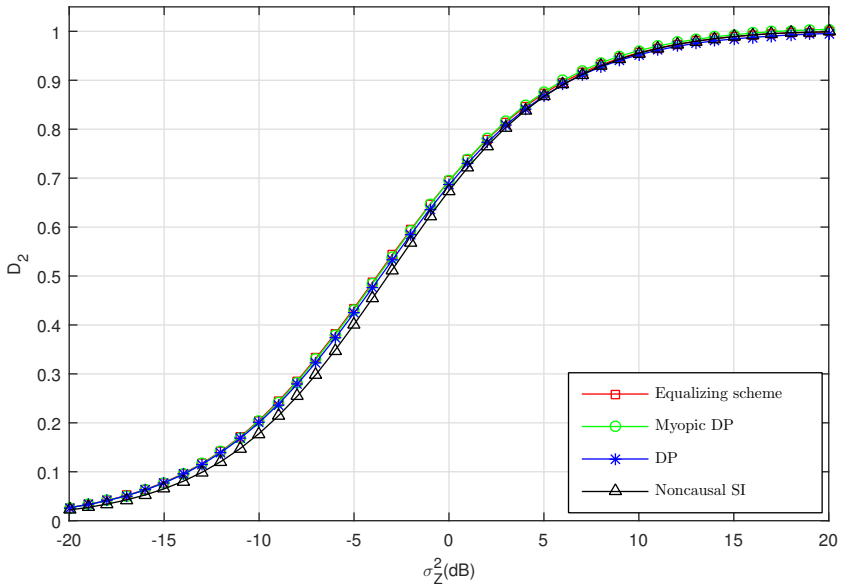


Fig. 3.7: Average D_2 with 100 randomly generated inputs $\{H^j\}$, $N = 10^5$, $K = 12$, $B_{\max} = 1$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, \cdots, 11$.

Summarizing, the MSE obtained for the case of non-causal SI is always lower than that with causal SI; for the causal SI case where the receiver does not know the transmit power, larger block length N results in lower MSE. The performance of heuristic schemes is also close to that of DP with large N, since the estimation of transmit power is more accurate with larger block length.

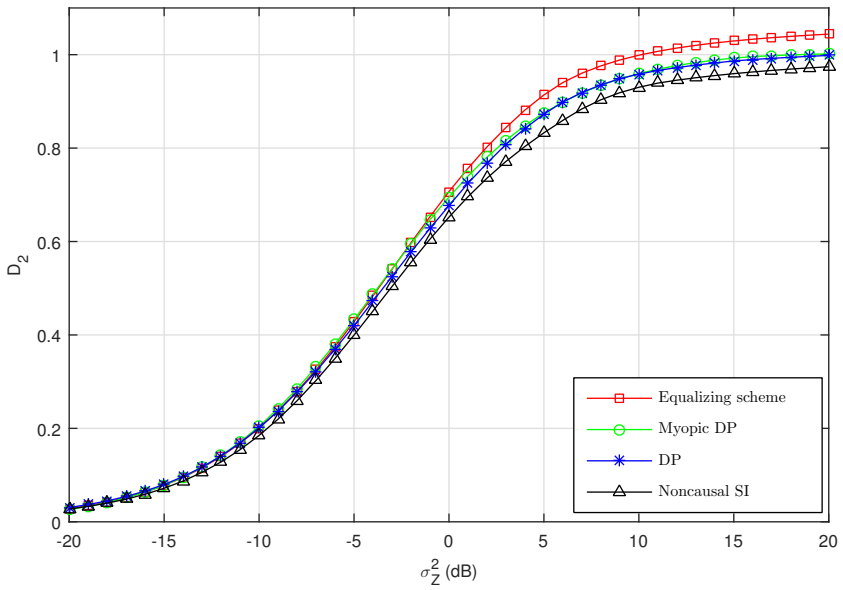


Fig. 3.8: Average D_2 with 100 randomly generated inputs $\{H^j\}$, $N = 100$, $K = 12$, $B_{\max} = 1$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, \cdots, 11$.

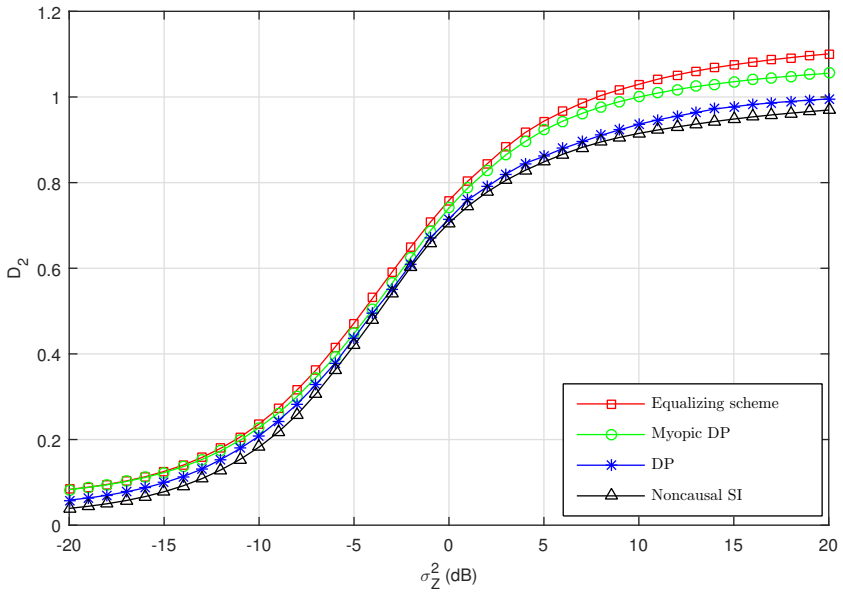


Fig. 3.9: Average D_2 with 100 randomly generated inputs $\{H^j\}$, $N = 10$, $K = 12$, $B_{\max} = 1$ and H^j i.i.d. uniformly distributed on $\{.1, .2, \cdots, .9\}$ for $j = 0, \cdots, 11$.

3.8 Summary

We have studied the power allocation problem for an estimation system where the estimator observes noisy output of the amplified source sequence. The transmitter relies on

energy harvested from its operating environment and we considered various cases of transmitter and receiver side information: whether the transmitter has non-causal knowledge of the harvested energy and whether the receiver/estimator has the knowledge of the transmit power. For the clairvoyant case when both transmitter and estimator have complete side information and the battery storage capacity is infinite, the optimal power allocation is shown to be reminiscent of a staircase climbing process where energy is evenly distributed and completely depleted between any neighboring transition blocks. If the battery storage has a finite capacity, the optimal power allocation can also be obtained via standard convex optimization algorithms. For the case with causal side information at the transmitter, the optimal power allocation can be solved via dynamic programming for a finite number of blocks. The case where the receiver does not have knowledge of transmit power is addressed by resorting to simple estimators that are asymptotically accurate as the block length increases.

Two heuristic approaches have also been proposed. Besides of greatly reduced computational complexities, both algorithms can be carried out in a forward manner thus do not require a finite number of blocks K as is the case for dynamic programming. Numerical simulations are conducted to compare the performance of the proposed approaches and it was found that when the block length N is large, the MSE's obtained with the heuristic schemes are close to that with DP; when N is small, DP performs considerably better than the heuristic schemes.

CHAPTER 4
CAPACITY THEOREMS FOR MULTI-FUNCTIONING RADIO

4.1 Problem Formulation And Literature Review

This chapter gives a simple information theoretic formulation for multi-functioning radios. Specifically, we consider communication systems where communication involves transmission of both messages and source sequences. The objective is to find the optimal trade-off between the rate of reliable message transmission and the distortion for the source sequence estimation. Using primarily Gaussian channels as examples, the central question we address is the following: is it optimal to treat the channel as simply a bit pipe where message and source encoding divide up the bandwidth? For the point to point system, we show that this is indeed the case. That is, it is optimal to simply split the channel capacity into two components and separately code the message and the source sequence. The same conclusion, however, no longer holds for multi-user systems. Using a simple Gaussian broadcast channel example, we show that such a separation approach will lead to a strictly suboptimal rate-distortion trade-off.

We emphasize here that the problem is different from the classic source-channel coding

problem where source sequences are to be communicated via noisy channels. The model in the present work involves both source sequences an messages that need to be transmitted via noisy channels. Closely related to this problem is the problem of state information transmission over state-dependent channels [49–52]. In [49], Sutivong *et al* studied the problem of communication over a state-dependent Gaussian channel where the receiver is not only interested in decoding the message but also wants to estimate the channel state. It was shown that the optimal trade-off between the information rate and the mean squared error (MSE) of state estimation is achieved by a power sharing scheme via Costa's dirty paper coding [53] and simple state amplification. This problem was extended to a non-Gaussian setting in [50] and to Gaussian broadcast channels in [51]. In [52], Choudhuri *et al* studied the problem of state information transmission over a state-dependent discrete memoryless channel, where the state is known strictly causally at the transmitter.

4.2 Point-to-Point Channels

In this section, we consider a communication system shown in Fig. 4.2, where the transmitter wishes *to simultaneously* send a message W and a source sequence S^n to the receiver over an additive Gaussian channel.

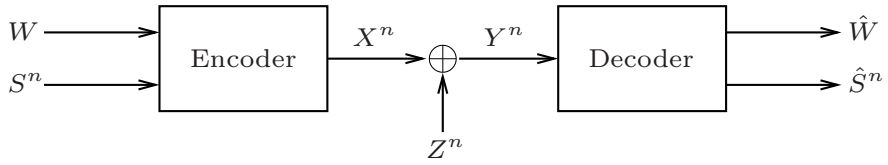


Fig. 4.1: Simultaneous transmission of a message and a source sequence over a point-to-point channel with additive Gaussian noise.

Let S_i denote the ith sample of the observation sequence of length n, which is assumed to be independent and identically distributed (i.i.d.) Gaussian with mean zero and variance Q. Let W denote the message to be communicated, which is uniformly distributed on

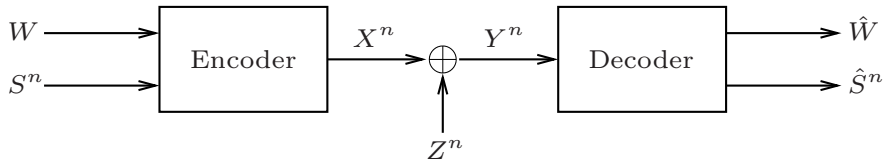


Fig. 4.2: Simultaneous transmission of a message and a source sequence over a point-to-point channel with additive Gaussian noise.

$\left\{1, 2, \cdots, 2^{nR}\right\}$ and is independent of S^n. Based on the message index W and the sequence S^n, the transmitter chooses $X^n\left(W, S^n\right)$ subject to an average power constraint P and sends it through a Gaussian channel with output

$$Y^n = X^n\left(W, S^n\right) + Z^n$$

where Z^n is an i.i.d. Gaussian noise sequence with mean zero and variance N and is independent of the transmitted signal. Upon receiving Y^n, the receiver decodes the message $\hat{W}\left(Y^n\right) \in \left\{1, 2, \cdots, 2^{nR}\right\}$ and reconstructs the source sequence $\hat{S}^n\left(Y^n\right)$. The error probability of message decoding and MSE distortion of sequence estimation are given by

$$P_e = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} Pr\left(\hat{W}(Y^n) \neq m | W = m\right)$$

and

$$d(S^n, \hat{S}^n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\left|\hat{S}_i - S_i\right|^2\right]$$

respectively. More formally, the problem is defined as follows.

Definition 1. *A $\left(2^{nR}, n\right)$ code consists of an encoder map*

$$X^n: \quad \left\{1, 2, \cdots, 2^{nR}\right\} \times \mathbb{R}^n \to \mathbb{R}^n$$

with codewords satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] \leq P,$$

where P is the average power constraint, and decoder maps

$$\begin{aligned}\hat{W} &: \quad \mathbb{R}^n \to \left\{1, 2, \cdots, 2^{nR}\right\} \\ \hat{S}^n &: \quad \mathbb{R}^n \to \mathbb{R}^n.\end{aligned}$$

Definition 2. *An (R, D) pair is said to be achievable if there exist a sequence of $(2^{nR}, n)$ codes such that the probability of error $P_e \to 0$ and the MSE distortion $d(S^n, \hat{S}^n) \leq D$ for each block length n.*

The objective is to characterize the optimal (R, D) trade-off region, which is given by the closure of the convex hull of all achievable (R, D) pairs.

Theorem 4.2.1. *For the point-to-point channel with additive Gaussian noise, the optimal (R, D) trade-off region is given by*

$$D \geq 2^{2R}\frac{NQ}{N+P} \tag{4.1}$$

where $0 \leq R \leq C$ and C denotes the channel capacity

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right).$$

Fig. 4.3 is an illustration of the achievable (R, D) region. Notice that one can equivalently rewrite (4.1) to be

$$R + \frac{1}{2}\log\frac{Q}{D} \leq C.$$

Thus the interpretation is clear: the total channel capacity is divided into two parts, one for source transmission and the other for message transmission.

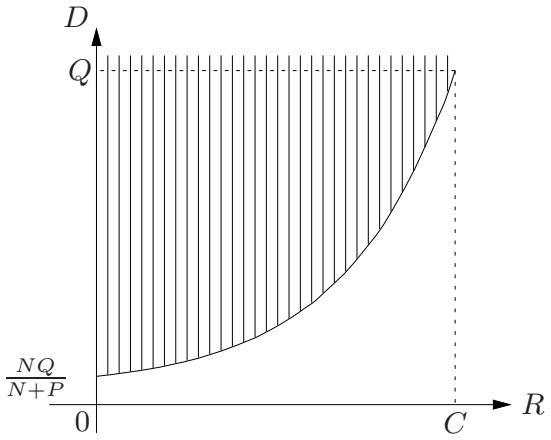


Fig. 4.3: Optimal (R, D) trade-off region for simultaneously transmitting a message and a source sequence over a Gaussian point-to-point channel.

4.2.1 Proof of Achievability

We propose a simple power sharing scheme that achieves the (R, D) region in (4.1). For some $0 \leq \lambda \leq 1$, the transmitter allocates λP and $(1-\lambda)P$ to encode W and S^n. In particular, rate-distortion code [55] is applied for S^n and the channel input is given by

$$X^n = X_W^n + X_S^n,$$

where X_W^n and X_S^n denote the codewords for W and S^n respectively. The receiver first recovers the source sequence S^n and subtracts codeword X_S^n, and then decodes the message W, with corresponding (R, D) pair satisfying

$$R \leq \frac{1}{2}\log\left(1+\frac{\lambda P}{N}\right) \triangleq R(\lambda, P) \tag{4.2}$$

$$D \geq Q\frac{\lambda P + N}{P+N} \triangleq D(\lambda, P). \tag{4.3}$$

Eliminating λ, we obtain the achievable region in (4.1). In addition, the above achievable scheme also implies separation from the operational sense: the two encodings (that of source and message) are carried out independently and codewords are superimposed at the transmitter.

Remark 1. *The (R, D) region in (4.1) can also be achieved by applying uncoded transmission [54] for the source sequence S^n with*

$$X_S^n = \sqrt{\frac{(1-\lambda)P}{Q}} S^n$$

and dirty paper coding [53] for the message W treating X_S^n as state.

4.2.2 Proof of Converse

We now establish the converse by equivalently showing that there exists a $0 \leq \lambda_\mu \leq 1$ as a function of μ such that any achievable (R, D) pair satisfies

$$R + \frac{\mu}{2} \log \frac{Q}{D} \leq R(\lambda_\mu, P) + \frac{\mu}{2} \log \frac{Q}{D(\lambda_\mu, P)} \tag{4.4}$$

for all $\mu \geq 0$, where $R(\lambda_\mu, P)$ and $D(\lambda_\mu, P)$ are defined in (4.2) and (4.3). We start from showing

$$\begin{aligned}
& \frac{1}{n} I(S^n; Y^n) \\
= \; & \frac{1}{n} \left(h(S^n) - h(S^n | Y^n) \right) \\
\overset{(a)}{=} \; & \frac{1}{n} \Big(h(S^n) - h(S^n - \hat{S}^n (Y^n) \, | Y^n) \Big) \\
\overset{(b)}{\geq} \; & \frac{1}{n} \Big(h(S^n) - h(S^n - \hat{S}^n (Y^n)) \Big) \\
\overset{(c)}{=} \; & \frac{1}{n} \left(\sum_{i=1}^{n} h(S_i) - h(S^n - \hat{S}^n (Y^n)) \right) \\
\overset{(d)}{\geq} \; & \frac{1}{n} \sum_{i=1}^{n} \Big(h(S_i) - h(S_i - \hat{S}_i) \Big) \\
= \; & \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \log(2\pi e Q) - h(S_i - \hat{S}_i) \right) \\
\overset{(e)}{\geq} \; & \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \log(2\pi e Q) - \frac{1}{2} \log \Big(2\pi e \mathbb{E}(S_i - \hat{S}_i)^2 \Big) \right) \\
\overset{(f)}{\geq} \; & \frac{1}{2} \log(2\pi e Q) - \frac{1}{2} \log \left(2\pi e \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(S_i - \hat{S}_i)^2 \right) \\
= \; & \frac{1}{2} \log \frac{Q}{D}
\end{aligned} \tag{4.5}$$

where

(a) is because $\hat{S}^n(Y^n)$ is a function of Y^n;

(b) is because conditioning reduces entropy;

(c) follows from chain rule and the i.i.d. assumption of S^n;

(d) follows from chain rule and conditioning reduces entropy;

(e) is because Gaussian distribution maximizes entropy for fixed variance;

(f) is due to Jensen's inequality.

Then, we can bound the weighted sum on the left hand side of (**??**) by

$$
\begin{aligned}
& R + \frac{\mu}{2} \log \frac{Q}{D} \\
\leq\ & R + \frac{\mu}{n} I(S^n; Y^n) \\
=\ & \mu \left(R + \frac{1}{n} I(S^n; Y^n) \right) + (1-\mu) R \\
\overset{(a)}{=}\ & \frac{\mu}{n} \left(h(W) + I(S^n; Y^n) \right) + \frac{1-\mu}{n} h(W) \\
\overset{(b)}{=}\ & \frac{\mu}{n} \left(h(W|S^n) + I(S^n; Y^n) \right) + \frac{1-\mu}{n} h(W|S^n) \\
\overset{(c)}{\leq}\ & \frac{\mu}{n} \left(I(W; Y^n|S^n) + I(S^n; Y^n) \right) + \frac{1-\mu}{n} I(W; Y^n|S^n) + \epsilon_n \\
\overset{(d)}{\leq}\ & \frac{\mu}{n} \left(I(X^n; Y^n|S^n) + I(S^n; Y^n) \right) + \frac{1-\mu}{n} I(X^n; Y^n|S^n) + \epsilon_n \\
=\ & \frac{\mu}{n} I(X^n, S^n; Y^n) + \frac{1-\mu}{n} I(X^n; Y^n|S^n) + \epsilon_n \\
=\ & \frac{\mu}{n} \left(h(Y^n) - h(Y^n|X^n, S^n) \right) \\
& + \frac{1-\mu}{n} \left(h(Y^n|S^n) - h(Y^n|X^n, S^n) \right) + \epsilon_n \\
\overset{(e)}{\leq}\ & \frac{\mu}{n} \left(\sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i|Y^{i-1}, X^n, S^n) \right) \\
& + \frac{1-\mu}{n} \left(\sum_{i=1}^{n} h(Y_i|S_i) - \sum_{i=1}^{n} h(Y_i|Y^{i-1}, X^n, S^n) \right) + \epsilon_n \\
=\ & \frac{\mu}{n} \sum_{i=1}^{n} \left(h(Y_i) - h(Z_i) \right) + \frac{1-\mu}{n} \sum_{i=1}^{n} \left(h(Y_i|S_i) - h(Z_i) \right) + \epsilon_n \\
=\ & \frac{1}{n} \sum_{i=1}^{n} \left(\mu h(Y_i) + (1-\mu) h(Y_i|S_i) - h(Z_i) \right) + \epsilon_n
\end{aligned} \tag{4.6}
$$

where

(a) is because W is uniformly distributed on $\{1, 2, \cdots, 2^{nR}\}$;

(b) is because W and S^n are independent;

(c) follows from Fano's inequality;

(d) follows from data processing inequality;

(e) follows from chain rule and conditioning reduces entropy.

In (4.6), for each i and $0 \leq \mu \leq 1$, the sum of the first two terms can be further bounded in the following

$$\begin{aligned}
& \mu h(Y_i) + (1-\mu) h(Y_i|S_i) \\
\overset{(a)}{\leq} & \frac{\mu}{2} \log\left(2\pi e \mathbb{E}Y_i^2\right) + (1-\mu) h(Y_i|S_i) \\
= & \frac{\mu}{2} \log\left(2\pi e \mathbb{E}Y_i^2\right) + (1-\mu) h(Y_i - \gamma S_i|S_i) \\
\overset{(b)}{\leq} & \frac{\mu}{2} \log\left(2\pi e \mathbb{E}Y_i^2\right) + (1-\mu) h(Y_i - \gamma S_i) \\
\overset{(c)}{\leq} & \frac{\mu}{2} \log\left(2\pi e \mathbb{E}Y_i^2\right) + \frac{1-\mu}{2} \log\left(2\pi e \left(\mathbb{E}Y_i^2 - \frac{(\mathbb{E}S_iY_i)^2}{\mathbb{E}S_i^2}\right)\right) \\
\overset{(d)}{\leq} & \frac{\mu}{2} \log\left(2\pi e (P_i + N)\right) + \frac{1-\mu}{2} \log\left(2\pi e (\lambda_i P_i + N)\right)
\end{aligned}$$

where

(a) is because Gaussian distribution maximizes differential entropy for fixed variance;

(b) is because conditioning reduces entropy;

(c) is obtained by letting

$$\gamma = \frac{\mathbb{E}S_iY_i}{\mathbb{E}S_i^2}$$

and it follows that

$$\mathbb{E}(Y_i - \gamma S_i)^2 = \mathbb{E}Y_i^2 - \frac{(\mathbb{E}S_iY_i)^2}{\mathbb{E}S_i^2};$$

(d) is because Gaussian distribution maximizes differential entropy. Note that equality holds when X_i is jointly Gaussian with S_i and Z_i, i.e., by letting

$$X_i = X_{W,i} + \sqrt{\frac{(1-\lambda_i)P_i}{Q}} S_i$$

where $X_{W,i} \sim N(0, \lambda_i P_i)$ with λ_i chosen such that the covariance of (S_i, X_i, Z_i, Y_i) is the same as that in the code.

Then we continue the chain of inequalities from (4.6)

$$\begin{aligned}
& R + \frac{\mu}{2}\log\frac{Q}{D} \\
\leq\ & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mu}{2}\log\left(2\pi e(P_i+N)\right) + \frac{1-\mu}{2}\log\left(2\pi e(\lambda_i P_i + N)\right) - \frac{1}{2}\log\left(2\pi e N\right)\right) + \epsilon_n \\
=\ & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\log\left(\frac{P_i+N}{N}\right)^{\mu} + \frac{1}{2}\log\left(\frac{\lambda_i P_i + N}{N}\right)^{1-\mu}\right) + \epsilon_n \\
\overset{(a)}{\leq}\ & \frac{1}{2}\log\left(\frac{P+N}{N}\right)^{\mu} + \frac{1}{2}\log\left(\frac{\lambda P + N}{N}\right)^{1-\mu} + \epsilon_n \\
=\ & \frac{1}{2}\log(1+\frac{\lambda P}{N}) + \frac{\mu}{2}\log\frac{P+N}{\lambda P + N} + \epsilon_n \\
\overset{(b)}{\leq}\ & \frac{1}{2}\log(1+\frac{\lambda_\mu P}{N}) + \frac{\mu}{2}\log\frac{P+N}{\lambda_\mu P + N} + \epsilon_n \\
=\ & R(\lambda_\mu, P) + \frac{\mu}{2}\log\frac{Q}{D(\lambda_\mu, P)} + \epsilon_n
\end{aligned}$$

where

(a) is by Jensen's inequality with λ chosen such that

$$\lambda P = \frac{1}{n}\sum_{i=1}^{n}\lambda_i P_i$$

and it follows that

$$(1-\lambda)P = \frac{1}{n}\sum_{i=1}^{n}(1-\lambda_i)P_i;$$

(b) is by choosing λ_μ that maximizes the expression.

Thus (4.4) is proved and this completes the converse proof.

4.3 Broadcast channels

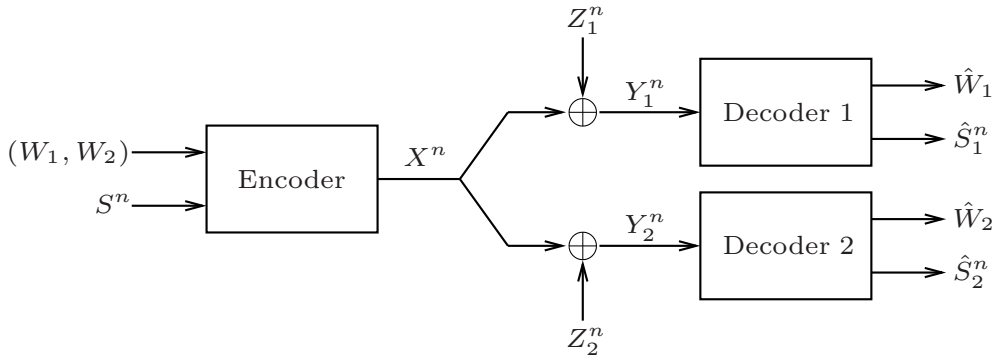


Fig. 4.4: Simultaneous transmission of two messages and a source sequence over a Gaussian broadcast channel.

In this section, we study the problem of simultaneously communicating messages and source sequences via Gaussian broadcast channels. As shown in Fig. 4.4, S^n denotes an i.i.d. Gaussian source sequence with mean zero and variance Q. Let $W_1 \in \{1, 2, \cdots, 2^{nR_1}\}$ and $W_2 \in \{1, 2, \cdots, 2^{nR_2}\}$ denote two independent messages which are assumed to be uniformly distributed and independent of S^n. Based on the source sequence S^n and message indices (W_1, W_2), the transmitter sends X^n subject to an average power constraint P and transmits it over a Gaussian broadcast channel with outputs

$$Y_k^n = X^n + Z_k^n$$

for $k = 1, 2$, where Z_k^n is an i.i.d zero-mean Gaussian noise sequence. The variance of Z_k is denoted as N_k, and without loss of generality we assume $N_1 \leq N_2$. Upon observing channel output Y_k^n, receiver k decodes the message $\hat{W}_k$ and reconstructs the source sequence $\hat{S}_k^n$ within certain MSE distortion. Formally, we have the following problem.

Definition 3. *A $(2^{nR_1}, 2^{nR_2}, n)$ code consists of an encoder map*

$$X^n : \quad \left\{1, 2, \cdots, 2^{nR_1}\right\} \times \left\{1, 2, \cdots, 2^{nR_2}\right\} \times \mathbb{R}^n \to \mathbb{R}^n$$

with codewords satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] \leq P,$$

where P is the average power constraint, and decoder maps

$$\begin{aligned} \hat{W}_k &: \quad \mathbb{R}^n \to \left\{1, 2, \cdots, 2^{nR_k}\right\} \\ \hat{S}_k^n &: \quad \mathbb{R}^n \to \mathbb{R}^n \end{aligned}$$

for $k = 1, 2$.

Definition 4. *A tuple (R_1, R_2, D_1, D_2) is said to be achievable if there exist a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that for each block length n, the probability of error for message decoding $P_{e,k} \to 0$ and the MSE of sequence estimation $d(S^n, \hat{S}_k^n) \leq D_k$ for $k = 1, 2$.*

Note that the minimum D_k achievable is given by

$$D_k^{\min} \triangleq Q\frac{N_k}{P + N_k}$$

where $k = 1, 2$ when there is no message transmission involved. On the other hand, if the distortion requirements at both receivers are relaxed to the variance of the source sequence, i.e., $D_1 > Q$ and $D_2 > Q$, the transmitter can ignore the source sequence and the problem reduces to channel coding. Therefore, in the following theorem, we characterize the optimal (R_1, R_2, D_1, D_2) trade-off region, i.e., the collection of all the achievable (R_1, R_2, D_1, D_2) tuples where $D_k \geq D_k^{\min}$ and $\min_k D_k \leq Q$ for $k = 1, 2$.

Theorem 4.3.1. *For the broadcast channel with additive Gaussian noise, the optimal*

(R_1, R_2, D_1, D_2) *trade-off region is given by*

$$\begin{aligned}
\frac{1}{2}\log\frac{Q}{D_1} &\leq \frac{1}{2}\log\left(1+\frac{\beta P}{(1-\alpha)P+N_1}\right) && (4.7)\\
\frac{1}{2}\log\frac{Q}{D_1}+R_1 &\leq \frac{1}{2}\log\left(1+\frac{(1-\alpha+\beta)P}{N_1}\right) && (4.8)\\
R_2 &\leq \frac{1}{2}\log\left(1+\frac{(\alpha-\beta)P}{(1-\alpha+\beta)P+N_2}\right) && (4.9)\\
\frac{1}{2}\log\frac{Q}{D_2}+R_2 &\leq \frac{1}{2}\log\left(1+\frac{\alpha P}{(1-\alpha)P+N_2}\right) && (4.10)
\end{aligned}$$

where $0 \leq \beta \leq \alpha \leq 1$.

4.3.1 Proof of Achievability

We describe a power sharing scheme that achieves the (R_1, R_2, D_1, D_2) region given by (4.7) - (4.10). For some $0 \leq \beta \leq \alpha \leq 1$, the transmitter allocates $(1-\alpha)P, (\alpha-\beta)P$ and βP to encode W_1, W_2 and X_S^n, respectively, and the channel input is given by

$$X^n = X_{W_1}^n + X_{W_2}^n + X_S^n.$$

In particular, an uncoded scheme is used for transmission of S^n with $X_S^n = \sqrt{\frac{\beta P}{Q}}S^n$ and dirty paper coding [53] is applied for W_1, treating X_S^n as known state. Receiver 2 estimates S^n and W_2 treating $X_{W_1}^n$ as interference. Receiver 1 first decodes W_2 and subtracts $X_{W_2}^n$, and then estimates S^n and W_1. By Theorem 4.2.1 we see that the error probability can be made arbitrarily small and the distortion requirement can be satisfied at both receivers, i.e., $P_{e,k} \to 0$ and $d(S^n, \hat{S}_k^n) \leq D_k$ for $k = 1, 2$, if (4.7) - (4.10) hold.

4.3.2 Proof of Converse

We now sketch the proof of converse by showing that there exist $0 \leq \beta \leq \alpha \leq 1$ such that any achievable (R_1, R_2, D_1, D_2) tuple satisfies (4.7) - (4.10). As there is no receiver

cooperation, we can assume a physically degraded Gaussian broadcast channel.

We start from (4.7). By Fano's inequality,

$$\begin{aligned}\frac{1}{2}\log\frac{Q}{D_1} &\leq \frac{1}{n}I(S^n;Y_1^n|W_2)\\ &= \frac{1}{n}h(Y_1^n|W_2)-\frac{1}{n}h(Y_1^n|W_2,S^n). \end{aligned} \tag{4.11}$$

Since

$$h(Y_1^n)\geq h(Y_1^n|W_2)\geq h(Y_1^n|W_2,S^n)=h(Z_1^n),$$

there must exist $0\leq\beta\leq\alpha\leq 1$ such that

$$h(Y_1^n|W_2,S^n)=\frac{n}{2}\log\left(2\pi e((1-\alpha)P+N_1)\right), \tag{4.12}$$

$$h(Y_1^n|W_2)=\frac{n}{2}\log\left(2\pi e((1-\alpha+\beta)P+N_1)\right). \tag{4.13}$$

Plugging (4.12) and (4.13) into (4.11), we obtain (4.7).

Next we prove (4.8). By Fano's inequality,

$$\begin{aligned}\frac{1}{2}\log\frac{Q}{D_1}+R_1 &\leq \frac{1}{n}I(W_1,S^n;Y_1^n|W_2)\\ &= \frac{1}{n}h(Y_1^n|W_2)-\frac{1}{n}h(Z_1^n). \end{aligned} \tag{4.14}$$

Plugging (4.13) into (4.14), we obtain (4.8).

For (4.9) and (4.10), by Fano's inequality we have

$$R_2 \leq \frac{1}{n}\left(\frac{n}{2}\log(2\pi e(P+N_2))-h(Y_2^n|W_2)\right) \tag{4.15}$$

$$\frac{1}{2}\log\frac{Q}{D_2}+R_2 \leq \frac{1}{n}\left(\frac{n}{2}\log(2\pi e(P+N_2))-h(Y_2^n|W_2,S^n)\right). \tag{4.16}$$

By the degradedness assumption, we let $Y_2=Y_1+\tilde{Z}_2$, where $\tilde{Z}_2$ is Gaussian with mean

zero and variance $N_2 - N_1$ and is independent of everything else. Then by the entropy power inequality [56] [57]

$$\begin{aligned} h(Y_2^n|W_2) &\geq \frac{n}{2}\log\left(2^{\frac{2}{n}h(Y_1^n|W_2)} + 2^{\frac{2}{n}h(\tilde{Z}_2|W_2)}\right) \\ &= \frac{n}{2}\log\left(2\pi e((1-\alpha+\beta)P + N_2)\right) \qquad (4.17) \\ h(Y_2^n|W_2, S^n) &\geq \frac{n}{2}\log\left(2^{\frac{2}{n}h(Y_1^n|W_2,S^n)} + 2^{\frac{2}{n}h(\tilde{Z}_2|W_2,S^n)}\right) \\ &= \frac{n}{2}\log\left(2\pi e((1-\alpha)P + N_2)\right). \qquad (4.18) \end{aligned}$$

Plugging (4.17) into (4.15) and (4.18) into (4.16), we obtain (4.9) and (4.10). Therefore the converse proof is completed.

4.3.3 Discussions

Clearly, when $D_1 \geq D_2$ the constraint in (4.7) is redundant as receiver 1 sees a better channel than receiver 2 does.

For the case where $D_1 < D_2$, consider a simple separation scheme as follows. Given that a Gaussian source is successively refinable [58], the transmitter encodes the source sequence S^n into two parts X_S^n and $X_S'^n$, which are then superimposed with the message codewords, resulting in the channel input as

$$X^n = X_{W_1}^n + X_S'^n + X_{W_2}^n + X_S^n.$$

Let γP be the fraction of power allocated to $X_{W_2}^n$ and X_S^n while $X_{W_1}^n$ and $X_S'^n$ share the rest $(1-\gamma)P$, where $0 \leq \gamma \leq 1$. Given the above encoding scheme, it is easy to obtain the following achievable region

$$\begin{aligned} \frac{1}{2}\log\frac{D_2}{D_1} + R_1 &\leq \frac{1}{2}\log\left(1 + \frac{(1-\gamma)P}{N_1}\right) \triangleq C_1 \qquad (4.19) \\ \frac{1}{2}\log\frac{Q}{D_2} + R_2 &\leq \frac{1}{2}\log\left(1 + \frac{\gamma P}{(1-\gamma)P + N_2}\right) \triangleq C_2. \qquad (4.20) \end{aligned}$$

Note that (C_1, C_2) gives the boundary of the capacity region of Gaussian broadcast channel [55] by varying γ from 0 to 1. Hence this achievable region can be interpreted as a separation of two parts - channel coding for the messages and successive refinement coding of the source sequence. Fig. 4.5 illustrates this approach where separation is in the sense of dividing up the rate pair (C_1, C_2) into two parts, one for message transmission and the other for source transmission.

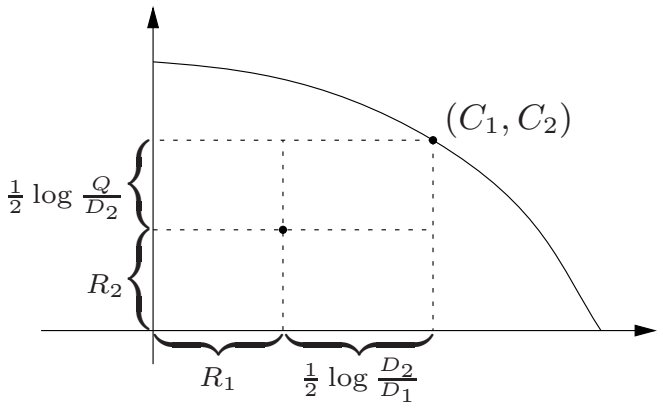


Fig. 4.5: Separation approach for simultaneously transmitting two messages and a source sequence over a Gaussian broadcast channel.

Recall that for the point to point case in Section 4.2, it is optimal to split the total rate into two parts and separately code the message and the sequence. However, it is easy to show that the region specified by (4.19) and (4.20) is strictly suboptimal for Gaussian broadcast channels.

Adding (4.19) and (4.20), we obtain

$$\frac{1}{2}\log\frac{Q}{D_1} + R_1 + R_2 \leq \frac{1}{2}\log\left(\frac{((1-\gamma)P + N_1)(P+N_2)}{N_1((1-\gamma)P+N_2)}\right). \tag{4.21}$$

On the other hand, Fourier-Motzkin elimination on (4.7) - (4.10) gives

$$\frac{1}{2}\log\frac{Q}{D_1} + R_1 + R_2 \leq \frac{1}{2}\log\left(\frac{((1-\alpha+\beta)P + N_1)(P+N_2)}{N_1((1-\alpha+\beta)P+N_2)}\right). \tag{4.22}$$

Choose $\gamma = \alpha - \beta$ so that (4.21) and (4.22) are identical. Then comparing (4.10) and

(4.20), it is clear that the separation scheme is suboptimal.

The strict sub-optimality for the above simple scheme can be attributed to the fact that a digital scheme is not optimal for sending a common Gaussian source through a Gaussian broadcast channel [59]. On the other hand, the optimal achievable scheme for this simple model requires a coupling of the channel and source coding through the use of dirty paper coding. It is clear that a simple superposition scheme (which amounts to operationally a separation scheme) is not optimal for this case even if one uses analog transmission for the source sequence.

4.4 Summary

This chapter provided complete solutions for the problems of simultaneously transmitting messages and source sequences over Gaussian point-to-point and broadcast channels. For point-to-point channels, rate splitting and separate coding of message and sequence is optimal. The same conclusion is no longer valid for the multi-user system as illustrated using a simple Gaussian broadcast channel example. Future work will focus on more complicated model (e.g., sending two independent sequences over a Gaussian broadcast channel) to gain a better understanding of the potential interplay between source and message transmissions.

CHAPTER 5
DUAL-USE OF AIRBORNE MIMO SYSTEMS

5.1 Motivation And Related Technologies

Air target detection has long been an active area of research with various defense and commercial applications. For example, MIMO radar, which employs multiple orthogonal transmit waveforms and has the ability to jointly process signals received at multiple antennas, is known to be a promising technology for airborne target detection [60]. Compared with standard phased-array radars, MIMO radars exploit the spatial and waveform diversity to provide higher angular resolution and improved target parameter identifiability [61–66].

Another attractive technology is passive radar, which is essentially a receiver-only radar that uses non-cooperative sources of illumination in the environment [8–10, 67–70]. Compared to active radar systems, passive radars are virtually undetectable to surveillance receivers and there is no constraint in spectrum allocation. Passive radars that leverage the existence of radio waves for communications have received much recent attention [8–10]. For example, FM radio, TV broadcast, satellite-borne signals and Global System for Mobile communication (GSM) signals have already been used as radar waveforms.

In this chapter, we consider yet another alternative: dual-use airborne MIMO systems where airborne MIMO communications are leveraged for target detection. Different from the MIMO radar, the primary task of the deployed system is for communications, hence the waveform at the transmit side is designed for MIMO communications instead of target detection. On the other hand, such a system differs from passive radars in that one has a richer knowledge of the transmitted signals that can be leveraged for improved target detection performance.

Specifically, we consider the dual use of airborne MIMO systems where airborne MIMO communications are leveraged for target detection. While the change due to the presence of new airborne targets often manifests itself directly in the change in the channel matrices (or their estimates), we utilize instead the change of channel characteristics in transform domains where airborne targets may exhibit distinct features. Two particular transform domains are explored in this chapter: 1) for uniform linear arrays (ULA) [19], angular domain channel decomposition is utilized - the lack of scattering leads to sparse angular domain channel representations and the presence of new targets is often manifested as new angular domain components; 2) for arbitrary array configuration in which directional information is hard to retrieve, Doppler domain decomposition can be explored as mobility (both of the airborne targets and of the receiver) induces Doppler components that are often distinct for different paths.

Different from the traditional MIMO channels that are rich in scattering, airborne MIMO channels do not admit accurate statistical models. This lack of statistical channel modeling carries over to the transform domains. As such, we adopt nonparametric approaches in detecting the change in channel realization. Recognizing the complexity of this problem, we consider offline methods to identify the change points (if present) in a sequence of sample observations.

In this chapter, we will first introduce a nonparametric offline change detection algorithm which is a special case of the E-divisive method [73]. As it is computationally simple

and capable of detecting any type of changes in multivariate distributions, we apply this method to statistics extracted based on peak detection in both angular and Doppler domains. Motivated by this change detector, we develop another change detection method which is directly applicable to sample observations from the angular and Doppler domain and thus preserves more information on the possible presence of airborne targets.

5.2 Nonparametric Detection For Distribution Change

Change detection concerns the inference of a change in distribution for a sequence of temporally or spatially ordered observations. We consider the offline version where sample size is fixed and retrospective analysis of an entire sequence is performed. Specifically, denote the observations to be $X_i \in \mathbb{R}^p$ where $i = 1, 2, \cdots, T$. The null hypothesis is that there is no change and all the observations come from the same multi-variate distribution, while the alternative hypothesis is that there exists a single change point τ which partitions the sequence into two sets following different distributions, i.e.,

$$\begin{aligned} \mathrm{H}_0 \quad &: \quad X_i \sim F_0, \quad \text{for} \quad i = 1, \cdots, T, \\ \mathrm{H}_1 \quad &: \quad X_i \sim F_0, \quad \text{for} \quad i = 1, \cdots, \tau, \\ &\qquad X_j \sim F_1, \quad \text{for} \quad j = \tau + 1, \cdots, T. \end{aligned}$$

In general, change detection has been performed in either parametric or nonparametric settings [74]. Parametric techniques often rely heavily on likelihood functions and thus require knowledge on the underlying distributions [75–77]. Nonparametric alternatives are applicable to a wider range of real-world problems as they maintain a desired detection performance without assuming any knowledge on distribution [78–82].

Although research on distribution-free approaches has increased in recent years, the majority of existing studies focus on monitoring changes in location or scale parameters

[80, 81]. As for detection of arbitrary changes in distribution, many methods consider empirical density functions, which are difficult to generalize to multivariate samples [82].

In this work, we adopt a nonparametric approach which is a special case of the E-Divisive technique, where inference is performed in a manner that simultaneously identifies the number and the location of change points based on hierarchical clustering [73]. This detector is capable of detecting any distributional change within a sequence of independent multivariate observations and does not make any assumption on the distributions beyond the existence of the αth moment, for some $0 < \alpha < 2$.

Consider the divergence measure $E(\alpha)$ which is based on Euclidean distances

$$E(\alpha) = 2 * E|X - Y|^{\alpha} - E|X - X'|^{\alpha} - E|Y - Y'|^{\alpha}.$$

It is shown in [83] that if $0 < \alpha < 2$, then $E(\alpha) \geq 0$ and equality holds if and only if X and Y are identically distributed. In fact, $E(\alpha)$ is equivalent to another divergence measure

$$D(\alpha) = \int |\phi_0(t) - \phi_1(t)|^2 \, w(t, \alpha) dt$$

where $\phi(\cdot)$ denote the characteristic function and the weight function is given by

$$w(t, \alpha) = \left(\frac{2\pi^{p/2}\Gamma(1 - \alpha/2)}{\alpha 2^{\alpha}\Gamma((p + \alpha)/2)} |t|^{p+\alpha} \right)^{-1}$$

if $E(|X|^{\alpha} + |Y|^{\alpha}) < \infty$ for some $0 < \alpha < 2$ [83, Theorem 2]. Furthermore, an empirical divergence measure $\hat{E}(\alpha, \tau)$ converges to $E(\alpha)$ almost surely as the sample size goes to

infinity [84], where

$$\begin{aligned}\hat{E}(\alpha,\tau) &= \frac{2}{\tau(T-\tau)}\sum_{i=1}^{\tau}\sum_{j=\tau+1}^{T-\tau}|X_i-X_j|^{\alpha} \\ &\quad -\binom{\tau}{2}^{-1}\sum_{1\le i<i'\le\tau}|X_i-X_{i'}|^{\alpha} \\ &\quad -\binom{(T-\tau)}{2}^{-1}\sum_{\tau+1\le j<j'\le T}|X_j-X_{j'}|^{\alpha}.\end{aligned}$$

Now we consider a scaled version of the empirical divergence measure which leads to a consistent estimator of change location [73, Theorem 4]

$$\hat{Q}(\alpha,\tau)=\frac{\tau(T-\tau)}{T}\hat{E}(\alpha).$$

The most probable change location of a sequence of sample observations can thus be estimated by searching for the partition that maximizes the scaled sample divergence of two sets of sample observations, i.e.,

$$\hat{\tau}=\text{argmax}_{\tau}\hat{Q}(\alpha,\tau).$$

The scaled sample divergence corresponding to the estimated change point is then compared to a threshold γ to determine whether there exists a change point, i.e.,

$$\begin{aligned}\mathrm{H}_0 &: \hat{Q}(\alpha,\hat{\tau})\le\gamma, \\ \mathrm{H}_1 &: \hat{Q}(\alpha,\hat{\tau})>\gamma.\end{aligned}$$

5.3 Angular Domain Target Detection

In this section, we discuss the selection of informational statistics in the angular domain for detection of changes caused by airborne targets for uniform linear arrays. We first

consider peak detection on the estimated angular domain channel matrices and feed the peak values as input to the nonparametric change detector introduced in Section 5.2. Then another change detection method is developed based directly on the Euclidean distance of adjacent angular domain channel matrices. This approach preserves more information on the possible presence of targets and can be applied to air-to-ground communication when only side of the communicating parties (either on the transmitter side or on the receiver side) is a uniform linear array.

5.3.1 Angular Domain Channel Representation

Consider the narrowband MIMO channel

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

where $\mathbf{x} \in \mathbb{C}^{n_t}$, $\mathbf{y} \in \mathbb{C}^{n_r}$ and $\mathbf{w} \sim CN(0, \sigma_n^2 \mathbf{I}_{n_r})$ denote the transmitted signal, received signal and white Gaussian noise respectively at a symbol time. The n_t transmit and n_r receive antennas are placed in uniform linear arrays of lengths L_t and L_r normalized by the wavelength λ_c of the passband transmitted signal. The normalized separation between transmit and receive antennas are $\Delta_t = L_t/(n_t - 1)$ and $\Delta_r = L_r/(n_r - 1)$. Assume that in addition to the LOS path there are possibly other scarcely distributed paths reflected by the airplane itself or targets in the air. The ith path has attenuation a_i, angle of departure ϕ_{ti} and angle of arrival ϕ_{ri}, for $i = 1, \cdots, q$. The channel matrix $\mathbf{H}$ can be modeled as

$$\mathbf{H} = \sum_{i=1}^{q} a_i^b \mathbf{e}_r(\cos(\phi_{ri}))\mathbf{e}_t^*(\cos(\phi_{ti}))$$

where

$$a_i^b = a_i \sqrt{n_t n_r} e^{-\frac{2\pi j d_i}{\lambda_c}}$$

and $\mathbf{e}_r(\cdot)$ and $\mathbf{e}_t(\cdot)$ are the steering vectors at the receiver and the transmitter, defined as

$$\begin{aligned}
\mathbf{e}_r(\Omega) &= \frac{1}{\sqrt{n_r}} \begin{bmatrix} 1 \\ e^{-2\pi j \Delta_r \Omega} \\ \vdots \\ e^{-2\pi j (n_r-1) \Delta_r \Omega} \end{bmatrix} \\
\mathbf{e}_t(\Omega) &= \frac{1}{\sqrt{n_t}} \begin{bmatrix} 1 \\ e^{-2\pi j \Delta_t \Omega} \\ \vdots \\ e^{-2\pi j (n_t-1) \Delta_t \Omega} \end{bmatrix}.
\end{aligned}$$

The distance of path i is denoted as d_i.

The transmitted and received signals can be transformed into the angular domain

$$\mathbf{x^a} = \mathbf{U_t}^* \mathbf{x}$$

$$\mathbf{y^a} = \mathbf{U_r}^* \mathbf{y}$$

where $\mathbf{U_t}$ and $\mathbf{U_r}$ are unitary matrices defined by the transmitted and received unit spatial signatures

$$\begin{aligned}
\mathbf{U_t} &= \left(\mathbf{e_t}(0), \mathbf{e_t}(\frac{1}{L_t}), \cdots, \mathbf{e_t}(\frac{n_t - 1}{L_t}) \right) \\
\mathbf{U_r} &= \left(\mathbf{e_r}(0), \mathbf{e_r}(\frac{1}{L_r}), \cdots, \mathbf{e_r}(\frac{n_r - 1}{L_r}) \right).
\end{aligned}$$

Clearly we have an equivalent representation of the channel in the angular domain

$$\mathbf{y^a} = \mathbf{H^a x^a} + \mathbf{w^a}$$

where

$$\mathbf{H^a} = \mathbf{U_r}^* \mathbf{H} \mathbf{U_t}$$

and

$$\mathbf{w^a} = \mathbf{U_r}^* \mathbf{w}.$$

Particularly, the (k, l)th element of the angular domain channel matrix is given by

$$h_{kl}^a = \sum_{i=1}^{q} a_i^b \left(\mathbf{e_r}(k/L_r)^* \mathbf{e}_r(\cos(\phi_{ri}))\right) \left(\mathbf{e}_t(\cos(\phi_{ti}))^* \mathbf{e_t}(l/L_t)\right).$$

We then define the kth bin ($k = 1, \cdots, n_r$) at the receiver angular domain to contain all paths whose receive directional cosine is within a window of width $1/L_r$ around k/L_r, i.e., $\frac{k-0.5}{L_r} < \cos(\phi_{ri}) < \frac{k+0.5}{L_r}$. Similarly, bins can be formed at the transmitter angular domain labelled with $l = 1, \cdots, n_t$ [19, Chapter 7].

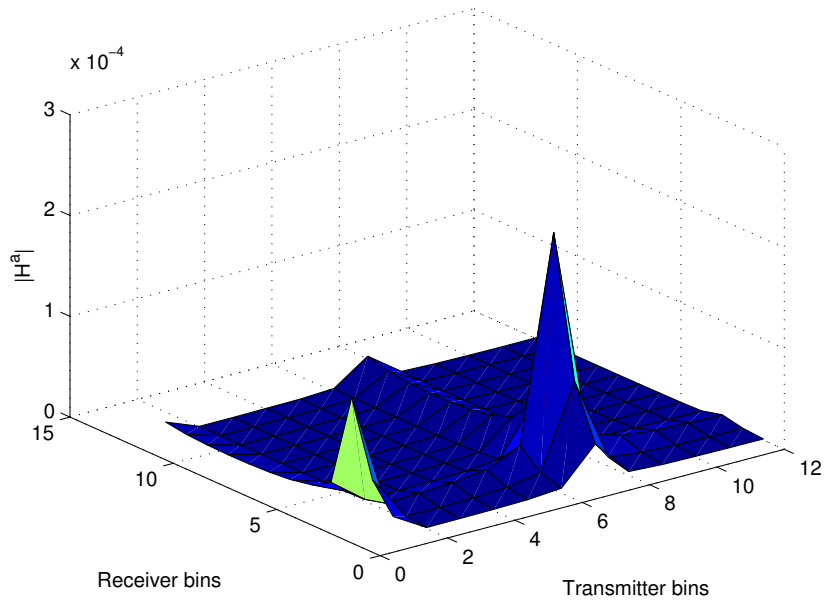


Fig. 5.1: An example of $\mathbf{H^a}$ with LOS path and a target-reflected path.

Fig. 5.1 shows an example of an angular domain channel matrix, where the angular spaces at both the transmitter and the receiver are divided into 12 bins. Each bin contains the set of all paths that roughly depart from a certain transmit angle and arrive in a certain receive angle. In this particular case, the channel is modeled by an LOS path plus a reflected path, thus the matrix demonstrates small angular spreads at both the transmitter and the receiver.

5.3.2 Peak-Detection-Based Target Detection

Since airborne MIMO communication is scarce of strong scatterers, physical paths may be resolvable from angular domain representation of the channel. Specifically, each distinct path may correspond to a distinct peak in the angular domain matrix. Consider a sequence of estimated angular domain matrices whose amplitudes are denoted as $\{\hat{\mathbf{H}}_1^a, \hat{\mathbf{H}}_2^a, \cdots, \hat{\mathbf{H}}_T^a\}$. Each physical path corresponds to a distinct peak which either stays in the same position throughout the sequence, or moves very slowly on the grid. The presence of a target manifests itself as, hence can be detected as a new peak in the estimated angular domain channel matrices.

Motivated by this observation, we propose a change detection method based on peak detection over the estimated angular domain channel matrices. The highest p peak values are constantly monitored and fed to the change detecter as input. Specifically, an element in an estimated channel matrix is considered to be a peak if its absolute value is greater than all the 8 neighbors. Note that the first and last rows or columns are also considered adjacent for peak detection, because the angular space is circular.

5.3.3 Euclidean-Distance-Based Target Detection

Since the onset of airborne targets often introduces new peaks on the estimated angular domain matrix, it is likely to contribute to increased distance between adjacent frames in the sequence of channel matrices. We then consider a target detection method based

directly on the distance. Let $|\hat{\mathbf{H}}^a_\tau - \hat{\mathbf{H}}^a_{\tau-1}|$ denote the Euclidean distance between two consecutive channel matrices, where $|\cdot|$ is the Frobenius norm.

Note that the movement of an existing peak (caused by the mobility of the airborne platform) can also contribute to a large distance between two consecutive frames of angular domain channel matrices. Such movement, however, is usually limited to its immediate neighbor of the angular domain in between frames. For example, assume channel estimation is carried out every 0.01s. Suppose a target 200m away from a 12-element ULA is traveling at 500 m/s along a path that is perpendicular to the LOS path between itself and the ULA. Then the maximum angular domain movement is 0.2 rad which is far below the angular domain resolution $\pi/6$ rad. In order to reduce the distance contributed by the movement of an existing peak, we smooth out the matrices by replacing each entry with the average of itself and the adjacent ones. After pre-processing, all peaks are blurred out and the slow movement of existing peaks will not cause significant changes between adjacent frames. Denote the pre-processed matrices to be $\{\tilde{\mathbf{H}}^a_1, \tilde{\mathbf{H}}^a_2, \cdots, \tilde{\mathbf{H}}^a_T\}$.

Another factor that may affect the Euclidean distance between adjacent frames is the noise introduced by channel estimation. To reduce it, we consider, instead, the distance between time-averaged frames, i.e.,

$$D_w(\tau) = \left| \frac{1}{w} \sum_{i=\tau}^{\tau+w-1} \tilde{\mathbf{H}}^a_i - \frac{1}{w} \sum_{i=\tau-w}^{\tau-1} \tilde{\mathbf{H}}^a_i \right|$$

for $\tau = w+1, \cdots, T+w-1$, where the window size w is chosen such that the effect of noise is averaged out while sufficient peak information is preserved. The most probable change point can then be estimated by searching for the time that maximizes the distance, i.e.,

$$\hat{\tau} = \text{argmax}_\tau D_w(\tau). \tag{5.1}$$

The distance corresponding to the estimated change point is then compared with a threshold

γ to determine whether there exists a change, i.e.,

$$\begin{aligned} H_0 &: D_w(\hat{\tau}) \leq \gamma, \\ H_1 &: D_w(\hat{\tau}) > \gamma. \end{aligned} \tag{5.2}$$

In fact, this method can also be applied to the scenario where there is only one uniform linear antenna array. Without loss of generality, we assume that the receive antenna is a ULA while the transmit antenna is nonlinear. In this case, the channel matrix is given by

$$\mathbf{H} = \sum_{i=1}^{q} a_i \sqrt{n_r} \mathbf{e}_r(\cos(\phi_{ri}))[e^{-2\pi j d_{1,i}/\lambda_c}, \cdots, e^{-2\pi j d_{n_t,i}/\lambda_c}]$$

where $d_{l,i}$ is the length of the ith physical path from the lth transmit antenna to the receive antenna array, for $l = 1, \cdots, n_t$. Then the angular domain channel matrix can be obtained by multiplying the Hermitian conjugate of the received unit spatial signature from the left, i.e.,

$$\mathbf{H^a} = \mathbf{U_r}^* \mathbf{H}.$$

The (k, l)th entry of the angular domain channel matrix $\mathbf{H^a}$ is given by

$$h^a_{kl} = \sum_{i=1}^{q} a_i \sqrt{n_r} \left(\mathbf{e_r}(k/L_r)^* \mathbf{e}_r(\cos(\phi_{ri}))\right) e^{-2\pi j d_{l,i}/\lambda_c},$$

where $\mathbf{e_r}(k/L_r)^* \mathbf{e}_r(\cos(\phi_{ri}))$ is only significant if the receive directional cosine is within a window of width $1/L_r$ around k/L_r.

The one-dimensional angular resolution can then be exploited by taking the sum along the row dimension of every estimated angular domain channel matrix which results in a sample vector $\mathbf{h}^a_\tau$ for $\tau = 1, \cdots, T$. Denote the distance between adjacent sample observa-

tions as

$$D(\tau) = |\mathbf{h}_{\tau}^{a} - \mathbf{h}_{\tau-1}^{a}|$$

where $\tau = 2, \cdots, T$. A similar decision rule can be obtained as (5.1) and (5.2).

5.4 Doppler Domain Target Detection

Since it is generally difficult to obtain the steering vector for a nonlinear array, the angular domain model is not directly applicable. In this section, we consider target detection based on Doppler frequency analysis. Assume that each path corresponds to a distinct Doppler frequency ω_i. Then the (k, l)th entry of the channel matrix at time n is given by

$$\mathbf{H}_{kl}[n] = \sum_{i=1}^{q} h_{kli} e^{j\omega_i n}$$

where h_{kli} indicates the gain contributed by the ith path between the lth transmit antenna and the kth receive antenna

$$h_{kli} = a_i e^{-2\pi j d_{kli}/\lambda_c}$$

and d_{kli} is length of the ith path. Then, the received signal on the kth antenna at time n is given by

$$\begin{aligned} y_k[n] &= \sum_{l=1}^{n_t} \sum_{i=1}^{q} h_{kli} e^{j\omega_i n} x_l[n] + w_k[n] \\ &= \sum_{i=1}^{q} \left(e^{j\omega_i n} \sum_{l=1}^{n_t} h_{kli} x_l[n] \right) + w_k[n]. \end{aligned}$$

Since the phase of the transmitted signal is known from decoding, it can be compensated at the receiver. The Doppler frequency components are thus directly reflected in the spectrum

of the received signal on each antenna. Therefore, target presence can be detected by identifying change points in the combined the spectrum of the received signals at all the receive antennas.

Specifically, the entire received sequence on the kth antenna is chopped up into T segments of the same length. Then fast Fourier transform (FFT) is performed for each segment and the absolute values of the spectrum of the received signal on each antenna are combined, resulting in a sequence of sample observations $\{\sum_{k=1}^{n_r} \mathbf{s}_{k,1}, \cdots, \sum_{k=1}^{n_r} \mathbf{s}_{k,T}\}$ where $\mathbf{s}_{k,t}$ denotes the absolute value of the spectrum of the tth segment of the received signal on the kth antenna, for $k = 1, \cdots n_r$ and $t = 1, \cdots, T$. Denote $\mathbf{s}_t = \sum_{k=1}^{n_r} \mathbf{s}_{k,t}$ for $t = 1, \cdots, T$.

Parallel to the discussions in the previous section, the presence of targets can then be detected based on

- estimation of change points in the distribution of peak values on the combined spectrum;
- distance between adjacent sample observations, i.e.,

$$D(\tau) = |\mathbf{s}_\tau - \mathbf{s}_{\tau-1}|,$$

for $\tau = 2, \cdots, T$.

5.5 Simulation Results

In this section, we present numerical results to evaluate the performance of the proposed schemes. We assume that

- without target, there are two paths - LOS and a reflected path by the transmitting airplane;
- a target introduces up to 3 additional paths.

The distance between two airplanes is set between 300m and 30000m. Attenuation is assumed to be proportional to the inverse of the path length with reflection coefficient 0.5. For simulation involving ULAs, transmit and receive angles are both randomly chosen between $-\pi$ and π. Carrier frequency is set at 120MHz and quadrature phase shift keying (QPSK) is adopted. The parameter settings are summarized in Table 5.1.

Table 5.1: Parameter settings for ULA simulation.

	LOS Path	Reflected Path
transmit angle ϕ_{ti}	$U(-\pi, \pi)$	$U(-\pi, \pi)$
receive angle ϕ_{ri}	$U(-\pi, \pi)$	$U(-\pi, \pi)$
path length r_i	$U(300, 30000)$	Calculated from geometry
reflection factor α_i	1	0.5

For nonlinear arrays, antenna element locations are set according to [71, Table 1] which are repeated in Table 5.2. The distance between antenna pairs are calculated based on the geometry of antenna elements. For example, denote the longitudinal and lateral coordinates of the kth antenna at the transmitting airplane and the lth antenna at the receiving airplane are (x_k, y_k) and (x_l, y_l), respectively. Then the LOS path length between two elements is

$$d_{kl,LOS} = \sqrt{(d - x_k - x_l)^2 + (w - y_k - y_l)^2 + h^2}$$

where d is the longitudinal separation of the two airplanes, travelling oppositely in the longitudinal direction. The lateral and height separation of the two airplanes are w and h respectively. Without loss of generality, we set the position of the transmitting airplane to be the origin. Clearly, the kth transmit antenna element is placed at $(x_k, y_k, 0)$ and the position of lth receive antenna is $(d - x_l, w - y_l, h)$. Then, the length of the path reflected by target at (x_i, y_i, z_i) can be obtained as

$$d_{kli} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + z_i^2} + \sqrt{(d - x_i - x_l)^2 + (w - y_i - y_l)^2 + (z_i - h)^2}.$$

Table 5.2: Location of antenna elements on the F-35C (m)

element number	1	2	3	4	5	6
longitudinal position x	-3.6	-3.18	-2.57	-1.96	-1.36	1.57
lateral position y	-6.55	-5.5	-4.45	-3.40	-2.36	-1.31
element number	7	8	9	10	11	12
longitudinal position x	1.57	-1.36	-1.96	-2.57	-3.18	-3.60
lateral position y	1.31	2.36	3.40	4.45	5.5	6.55

For the simulation of angular domain target detection, a total number of 100 channel matrices are generated for each Monte Carlo run, where a target may appear any time in the sequence or does not show up at all. Standard least squares (LS) training-based technique is applied for MIMO channel estimation with orthogonal pilot signals [85, Section II]. The window size w is empirically chosen to be 3.

For the simulation of Doppler domain target detection, Doppler frequency associated with any physical path is randomly chosen to be between $-\pi$ and π. In each Monte Carlo run, 100 received sequences of length 128 are collected on each antenna and a target may or may not appear.

Fig. 5.2 illustrates the empirical divergence measure $\hat{Q}$. In this simulation, 100 channel matrices are generated for the uniform linear array case where a target is present starting at the 47th sample. The empirical divergence measure $\hat{Q}$ is computed based on the proposed angular domain peak detection method and plotted for each candidate change point with α set at 1.0 and 1.5. Clearly, $\hat{Q}$ is maximized at the true change point in both cases. In the following simulations, we set α to be 1.0.

Fig. 5.3 compares the true change points and estimated ones for the uniform linear array case using the nonparametric change detector in Section 5.2. The result contains 20 Monte Carlo runs, and in each run, 100 sample sequences are collected and a change point is randomly set between 10 and 90, so that sufficient data is accumulated both before and

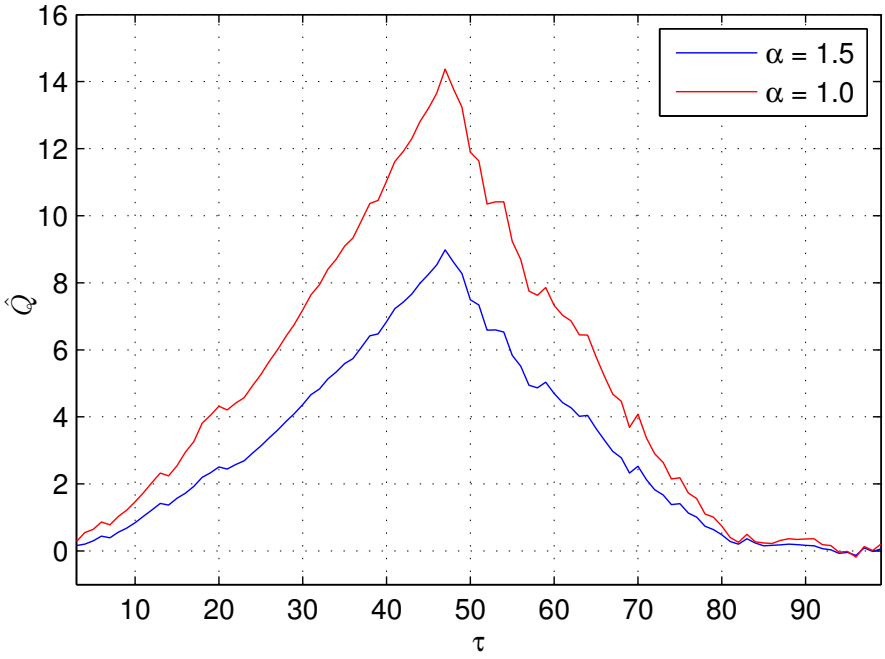


Fig. 5.2: Empirical divergence measure $\hat{Q}$ vs candidate change point τ.

after change. As shown in the figure, the estimated change points based on the peak values in the angular domain matrix (blue curve with circle markers) are very close to the true change points (black curve with star markers). The estimated change points obtained by directly feeding the channel matrix to the change detector are also plotted as a benchmark (red curve with square markers). Clearly, the performance of change point estimation based on raw data is not as good as that of the proposed method, which is also illustrated using the ROC curves in Fig. 5.5.

Fig. 5.4 shows the estimation of change points for a nonlinear array using Doppler information. When signal-to-noise ratio (SNR) is set at 20dB, the estimated change points (blue curve with circle markers) are very close to the true change points with occasional instances with large estimation errors. When the SNR drops to 10dB, the estimation is less accurate.

Fig. 5.5 compares the ROC curves of different target detection methods for uniform linear arrays in the noisefree case. In the angular domain, the peak-detection-based de-

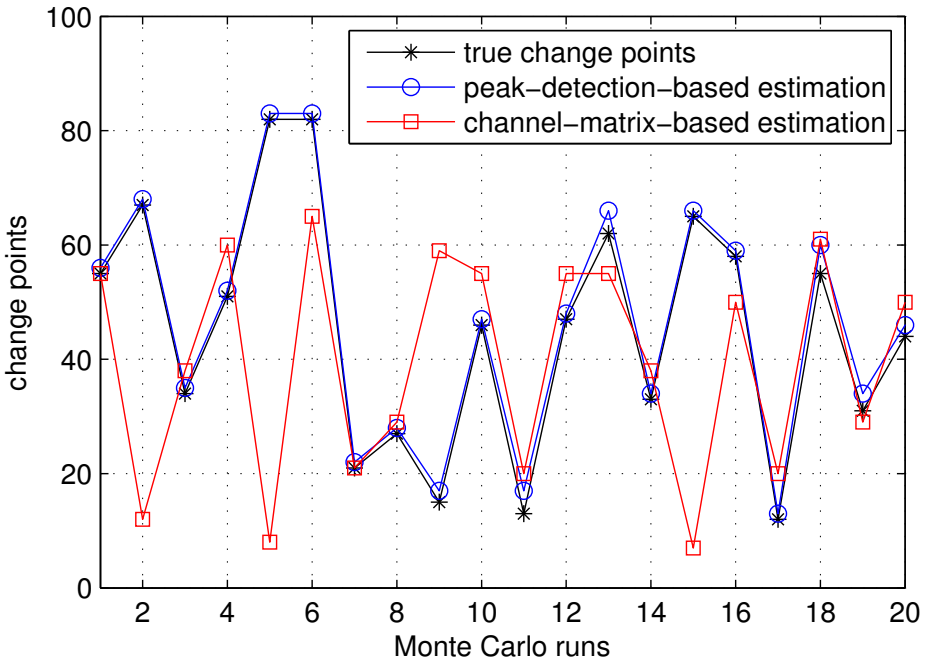


Fig. 5.3: Simulation of 20 runs of change detector for a uniform linear array.

tector performs much better than the channel-matrix-based detector which directly inputs the entire sequence of estimated channel matrices to the nonparametric distribution change detector. The performance of the Doppler-frequency-based detector is close to that of the angular domain peak-detection-based detector, as they both depend on peak detection but in different transform domains.

Fig. 5.6 shows the ROC curves of the proposed Euclidean-distance-based angular domain target detection methods under varying signal-to-noise ratio. The solid curves correspond to the cases where both transmit and receive antennas are uniform linear arrays with 12 elements. The dashed curves indicate ULA only on the receiver side. As expected, the performance increases with the increase of SNR. The performance of the cases with ULAs on both sides is better than that of ULA on one side, due to the additional dimension of angular resolution.

Fig. 5.7 compares the ROC curves of the Doppler domain target detection methods where both transmit and receive antenna arrays are nonlinear. The performance improves

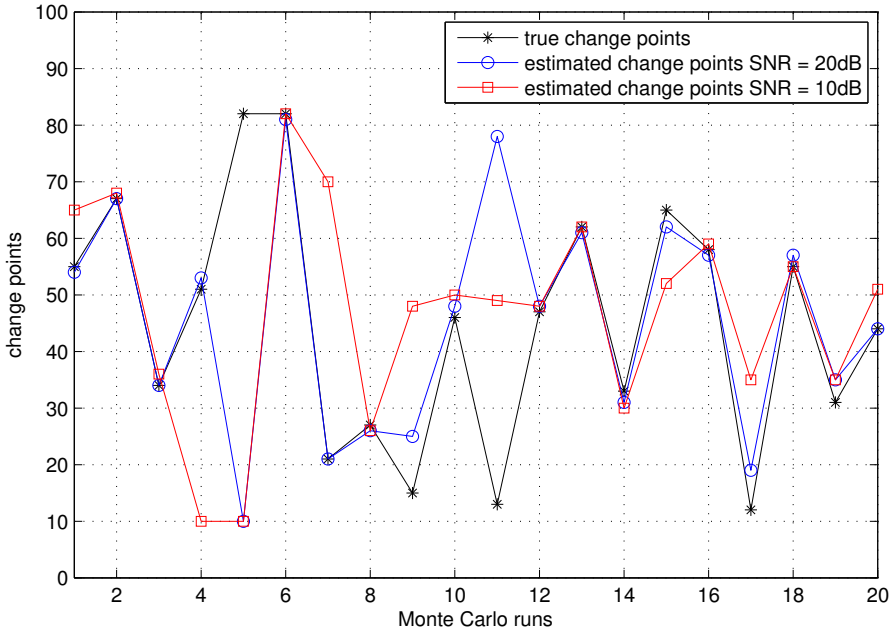


Fig. 5.4: Simulation of 20 runs of change detector for a non-linear array.

with the increase of SNR. Euclidean-distance-based detectors generally outperform peak-detection-based ones, because the former preserve more information regarding the targets.

5.6 Summary

This chapter investigated the dual-use of airborne MIMO. In addition to communications, received signals are also used for target detection that exploits the scarcity of scatterers in an airborne environment. Nonparametric target detection methods were proposed based on peak detection as well as Euclidean distance between sample observations. For the case of uniform linear arrays, informative statistics were obtained through estimation of angular domain channel matrix. For the case of nonlinear arrays, parameters were selected based on Doppler frequency analysis on the received signal.

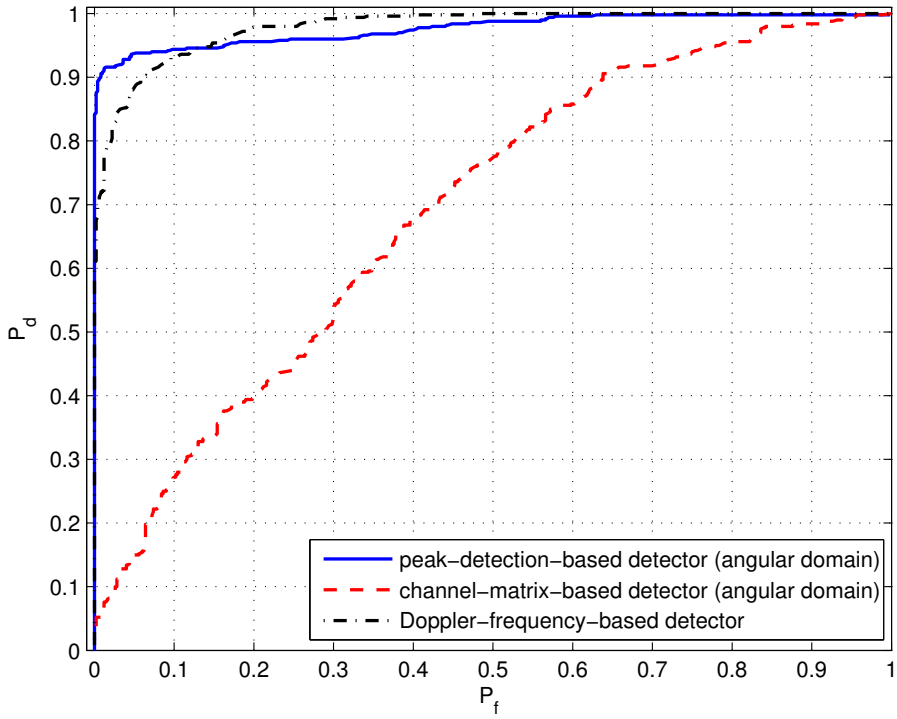


Fig. 5.5: ROC curves of detection for a uniform linear array in the noisefree case.

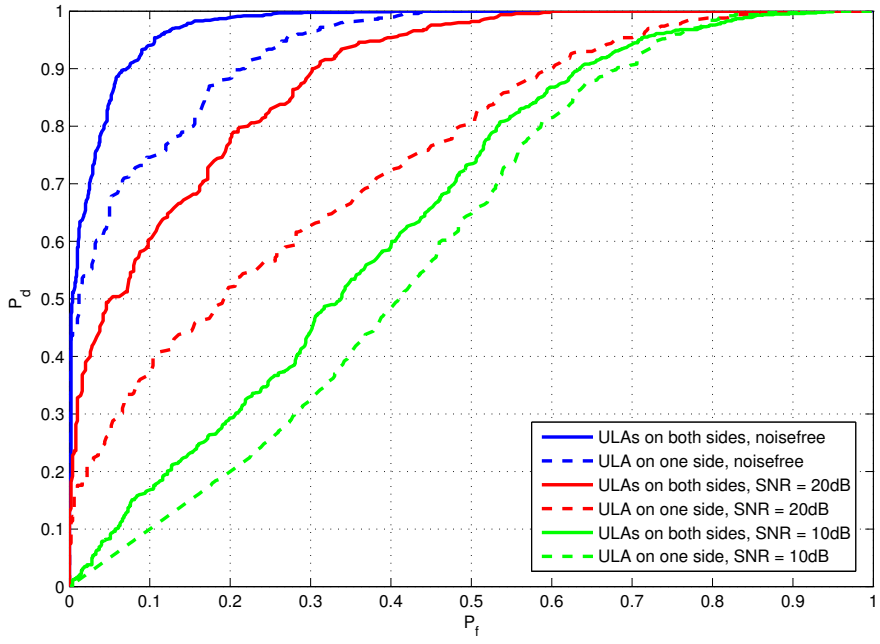


Fig. 5.6: ROC curves of the Euclidean-distance-based angular domain target detection methods under varying channel condition.

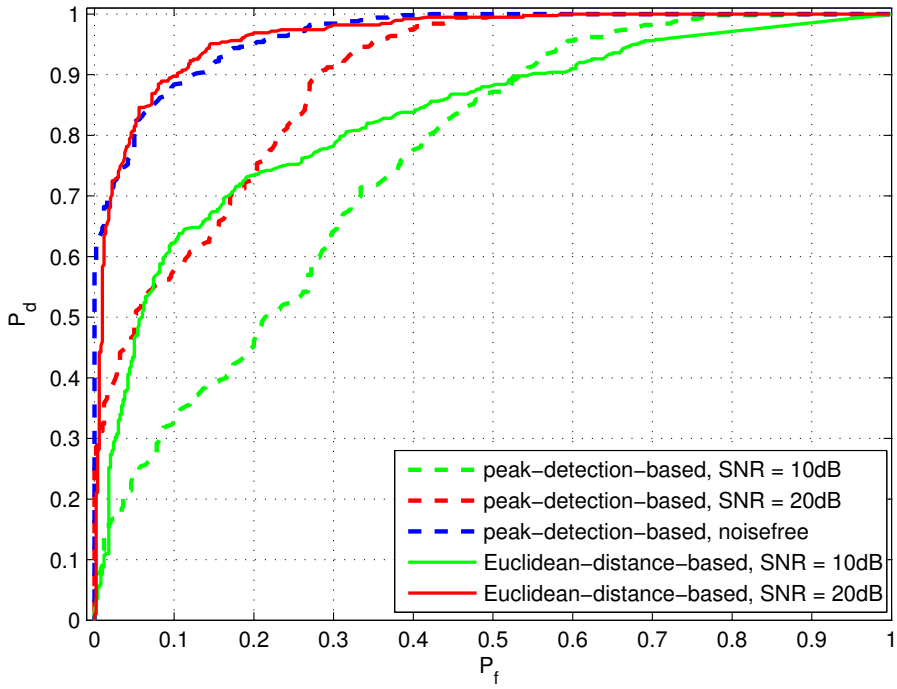


Fig. 5.7: ROC curves of Doppler domain target detection under varying channel condition. Both transmit and receive antennas are nonlinear arrays.

CHAPTER 6
CONCLUSIONS

In this thesis, we investigated some of the research challenges in multi-mission wireless systems. We first considered theoretical limits on communication rate when interference is present. Specifically, a simple coding scheme was developed for a class of Gaussian interference channels with mixed interference. The achievable rate region of such a scheme was then shown to be equivalent to Costa's *noiseberg* region for the weak one-sided Gaussian interference channel. This allows for an indirect proof that this simple coding scheme results in an achievable rate region that is equivalent to the Han-Kobayashi region with Gaussian input, which is by far the largest achievable rate region for this class of Gaussian interference channel with mixed interference.

The power allocation policies for an estimation system was then studied, where a remote sensor observes a Gaussian sequence and sends it through a Gaussian channel. The remote sensor is powered by energy harvested from the environment, so the energy available for data transmission typically fluctuates over time. With the aim of minimizing the mean squared error at the receiver, we developed optimal power management strategies under various conditions using convex optimization techniques and dynamic programming. The structure of the optimal power allocation schemes was analyzed, which provides insight on the design of heuristic algorithms that are able to simplify the computation and

extend the operation horizon to infinity.

Related to the remote estimation problem, we gave an information theoretic formulation of a multi-functioning radio where communication and sensing are both required. Specifically, we considered a remote estimation system where the sensor simultaneously transmits a source sequence and an independent message. The objective was to study the optimal coding trade-off between the message rate and sequence estimation. For point-to-point channels, we showed that a simple separation scheme is optimal, where the total channel capacity is split into two components, one for message communication and the other for sequence transmission. The same conclusion, however, does not hold for the multi-user case - we showed that such separation-based schemes lead to strictly suboptimal rate-distortion trade-off for the Gaussian broadcast channels. In fact, future work can be focused on exploring coding schemes for simultaneous transmission of messages and source sequences under other multi-user scenarios such as Gaussian multiple access channels and discrete channels.

Finally we investigated the dual-use of airborne MIMO communication systems. Specifically, the lack of scatterers in a typical airborne environment was leveraged for the detection of airborne targets. With uniform linear arrays, angular domain decomposition of channel matrices was utilized and nonparametric methods were applied to detect the change in the resolvable paths in the angular domain. Doppler frequency analysis was utilized for both linear and nonlinear arrays and target detection was accomplished by detection of change in the frequency components. In fact, the use of airborne RF systems can be extended beyond target detection. For example, one can explore methods for target localization and tracking, as location and velocity information can be extracted from angular domain channel decomposition and Doppler frequency analysis for airborne MIMO communication systems.

APPENDIX A
PROOF OF EQUIVALENCE BETWEEN $\mathcal{R}_{\mathcal{O}}$ AND $\mathcal{R}_{\mathcal{N}}$

In this appendix, we simplify $\mathcal{R}_{\mathcal{O}}$ and prove that it is equivalent with $\mathcal{R}_{\mathcal{N}}$.

Comparing this region with $\mathcal{R}_{\mathcal{N}}$, we see that (2.17)(2.18) and (2.3)(2.4) are identical. Then it remains to show that (2.16)(2.19)-(2.21) are redundant given (2.17)(2.18).

Eq. (2.16) is redundant since

$$\begin{aligned}
&\frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right)\leq\frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}\right)\\
\Leftrightarrow\quad & 1+\frac{P_{1A}}{\bar{\lambda}}+\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}+\frac{P_{1A}}{\bar{\lambda}}\cdot\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}\leq 1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}} \qquad \text{(A.1)}\\
\Leftrightarrow\quad & \frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\leq\frac{P_{1C}}{\bar{\lambda}}\\
\Leftrightarrow\quad & a^2\bar{\lambda}+a^2P_{1A}\leq\bar{\lambda}+a^2P_{1A}+P_2\\
\Leftrightarrow\quad & (1-a^2)\bar{\lambda}+P_2\geq 0,
\end{aligned}$$

which is trivially true.

Eq. (2.19) is redundant, since it is superseded by (2.17) and (2.18), i.e., we will show

$$\frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right)+\frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right)$$
$$\leq\frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}\right),$$

or, equivalently,

$$\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{P_2}{\bar{\lambda}+a^2P_{1A}}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)$$
$$+\frac{P_2}{\bar{\lambda}+a^2P_{1A}}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}\leq\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2P_2}{\bar{\lambda}}. \tag{A.2}$$

In order to prove (A.2), it suffices to show

$$\frac{P_2}{\bar{\lambda}+a^2P_{1A}}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\leq\frac{b^2P_2}{\bar{\lambda}} \tag{A.3}$$

and

$$\frac{P_2}{\bar{\lambda}+a^2P_{1A}}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}+\frac{a^2P_{1C}}{\bar{\lambda}+a^2P_{1A}+P_2}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\leq\frac{P_{1C}}{\bar{\lambda}}. \tag{A.4}$$

Eq. (A.3) is equivalent to

$$\frac{b^2}{\bar{\lambda}}\geq\frac{1+\frac{P_{1A}}{\bar{\lambda}}}{\bar{\lambda}+a^2P_{1A}}$$
$$\Leftrightarrow\quad(b^2-1)\bar{\lambda}+(a^2b^2-1)P_{1A}\geq 0$$

which is obviously true. Eq. (A.4) is equivalent to

$$\begin{aligned} & \frac{P_{1C}}{\bar{\lambda}} \geq \frac{a^2 P_{1C}}{\bar{\lambda}+a^2 P_{1A}+P_2}\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)\left(1+\frac{P_2}{\bar{\lambda}+a^2 P_{1A}}\right) \qquad \text{(A.5)} \\ \Leftrightarrow \quad & \bar{\lambda}+a^2 P_{1A}+P_2 \geq a^2(\bar{\lambda}+P_{1A})\left(1+\frac{P_2}{\bar{\lambda}+a^2 P_{1A}}\right) \\ \Leftrightarrow \quad & \bar{\lambda}(1-a^2)(\bar{\lambda}+a^2 P_{1A}+P_2) \geq 0. \end{aligned}$$

Thus (A.3) and (A.4) are true and (2.19) is redundant.

Comparing (2.17)(2.18) and (2.20), we see that (2.20) being redundant is equivalent to

$$\begin{aligned} & \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right) \leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}}\right) \qquad \text{(A.6)} \\ \Leftrightarrow \quad & 1+\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}+a^2 P_{1A}}+\frac{P_{1A}}{\bar{\lambda}}\frac{P_2}{\bar{\lambda}+a^2 P_{1A}} \leq 1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}} \end{aligned}$$

which is equivalent to (A.3). Thus (2.20) is redundant.

Comparing (2.17)(2.18) and (2.21), we see that (2.21) being redundant is equivalent to

$$\begin{aligned} & 2*\frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right)+\frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right) \\ & \leq \frac{1}{2}\log\left(1+\frac{P_{1A}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}}\right)+\frac{1}{2}\log\left(1+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}}\right). \end{aligned}$$

It suffices to show (A.6) and

$$\frac{1}{2}\log\left(1+\frac{a^2\frac{P_{1C}}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}+\frac{P_2}{\bar{\lambda}}}\right)+\frac{1}{2}\log\left(1+\frac{\frac{P_2}{\bar{\lambda}}}{1+a^2\frac{P_{1A}}{\bar{\lambda}}}\right) \leq \frac{1}{2}\log\left(1+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}}\right)$$

which is equivalent to

$$1+\frac{P_2}{\bar{\lambda}+a^2 P_{1A}}+\frac{P_2}{\bar{\lambda}+a^2 P_{1A}}\frac{a^2 P_{1C}}{\bar{\lambda}+a^2 P_{1A}+P_2}+\frac{a^2 P_{1C}}{\bar{\lambda}+a^2 P_{1A}+P_2} \leq 1+\frac{P_{1C}}{\bar{\lambda}}+\frac{b^2 P_2}{\bar{\lambda}}.$$

Then it is sufficient to show

$$\frac{P_2}{\bar{\lambda} + a^2 P_{1A}} \leq \frac{b^2 P_2}{\bar{\lambda}}$$

and

$$\left(1 + \frac{P_2}{\bar{\lambda} + a^2 P_{1A}}\right) \frac{a^2 P_{1C}}{\bar{\lambda} + a^2 P_{1A} + P_2} \leq \frac{P_{1C}}{\bar{\lambda}},$$

which are trivially true from (A.3) and (A.5). Thus (2.21) is redundant given (2.17) and (2.18).

APPENDIX B
PROOF OF LEMMA 3.2.1

For the converse, consider the case where all harvested energy $\sum_{j=0}^{K-1} H^j$ is available at the first block. Denote $h(\cdot)$ as the MSE contributed from a single block as a function of the transmit power, e.g., $h(\cdot) = h_1(\cdot)$ in (3.4) if the transmit power is known to the receiver. Since $h(\cdot)$ is convex, the optimal strategy is to allocate energy evenly to all K blocks and the corresponding MSE is

$$D(K) = \frac{1}{K} \sum_{j=1}^{K} h\left(\frac{1}{K} \sum_{j=0}^{K-1} H^j\right). \tag{B.1}$$

As $K \to \infty$, $D(K)$ converges to $h\left(\mathbb{E}[H]\right)$, i.e.,

$$\lim_{K\to\infty} D(K) = h\left(\mathbb{E}[H]\right).$$

The MSE achievable for the general case is lower bounded by $D(K)$ in (B.1), i.e.,

$$\mathbb{E}\left[D\right] \geq D(K).$$

For achievability, by the law of large numbers, for any $\epsilon > 0$, there exists $K_\epsilon > 0$ such that

$$\mathcal{P}\left(\frac{1}{K}\sum_{j=0}^{K-1} H^j > \mathbb{E}\left[H^j\right] - \epsilon\right) \geq \mathcal{P}\left(\left|\frac{1}{K}\sum_{j=0}^{K-1} H^j - \mathbb{E}\left[H^j\right]\right| < \epsilon\right) > 1 - \epsilon,$$

for all $K > K_\epsilon$. Consider the following power allocation scheme

$$P^j = \begin{cases} \mathbb{E}\left[H^j\right] - \epsilon, & j > K_\epsilon, \\ 0, & j \leq K_\epsilon, \end{cases}$$

with MSE

$$\mathbb{E}\left[D\right] = \frac{1}{K}\left(K_\epsilon \sigma_S^2 + \sum_{j=K_\epsilon+1}^{K} h\left(\mathbb{E}\left[H^j\right] - \epsilon\right)\right).$$

By the continuity of $h(\cdot)$ function, we have

$$\lim_{K\to\infty} \mathbb{E}\left[D\right] \leq h\left(\mathbb{E}[H]\right) + \delta_\epsilon, \tag{B.2}$$

where $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

APPENDIX C
DERIVATION OF EQUATION (3.12)

From (3.3), the MMSE estimator with $\hat{P}^j$ is

$$\hat{S}_i^j = \frac{\sqrt{\hat{P}^j}\sigma_S Y_i^j}{\sigma_Z^2 + \hat{P}^j} = \frac{\sqrt{\hat{P}^j}\sigma_S}{\sigma_Z^2 + \hat{P}^j}\left(\frac{\sqrt{P^j}}{\sigma_S}S_i^j + Z_i^j\right).$$

Thus the corresponding MSE is

$$\begin{aligned}
\tilde{D} &= \frac{1}{KN}\sum_{j=1}^{K}\sum_{i=1}^{N}\mathbb{E}\left[\hat{S}_i^j - S_i^j\right]^2 \\
&= \frac{1}{KN}\sum_{j=1}^{K}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\sqrt{P^j\hat{P}^j} - \hat{P}^j - \sigma_Z^2}{\sigma_Z^2 + \hat{P}^j}S_i^j + \frac{\sqrt{\hat{P}^j}\sigma_S}{\sigma_Z^2 + \hat{P}^j}Z_i^j\right]^2 \\
&= \frac{1}{K}\sum_{j=1}^{K}\frac{1}{\left(\sigma_Z^2 + \hat{P}^j\right)^2}\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[\left(\hat{P}^j + \sigma_Z^2 - \sqrt{P^j\hat{P}^j}\right)^2\left(S_i^j\right)^2 + \hat{P}^j\sigma_S^2\left(Z_i^j\right)^2\right] \\
&= \frac{1}{K}\sum_{j=1}^{K}\frac{\left(\hat{P}^j\sigma_Z^2 + \left(\sigma_Z^2 + \hat{P}^j - \sqrt{\hat{P}^j P^j}\right)^2\right)\sigma_S^2}{(\sigma_Z^2 + \hat{P}^j)^2}.
\end{aligned}$$

REFERENCES

[1] I. F. Akiyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, is. 4, pp. 393-422, Jan. 2002.

[2] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part IâĂŤFundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54-63, Jan. 1997.

[3] Y. Zhang, N. Meratnia and P. Havinga, "Outlier detection techniques for wireless sensor networks: a survey," *Proc. IEEE Communications Surveys and Tutorials*, vol. 12, no. 2, pp. 159-170, May 2010.

[4] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 6-28, Dec. 2004.

[5] G. K. Ottman, H. F. Hofmann, A. C. Bhatt, and G. A. Lesieutre, "Adaptive piezoelectric energy harvesting circuit for wireless remote power supply," in *IEEE Trans. Power Electronics*, vol. 17, issue 5, pp. 669-676, Sep. 2002.

[6] S. P. Beeby, M. J. Tudor, and N. M. White, "Energy harvesting vibration sources for microsystems applications," in *Meas. Sci. Technol.*, vol. 17, issue 12, pp. R175-R195, Dec. 2006.

[7] V. Raghunathan, A. Kansal, J. Hsu, J. Friedman, and M. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems," in *Proc. 4th Int. Symp. Information Processing in Sensor Networks*, Los Angeles, CA, pp. 463-468, Apr. 2005.

[8] H. D. Griffiths, "From a different perspective: principles, practice and potential of bistatic radar," *Proc. Int. Conf. on Radar*, Adelaide, Australia, pp. 1-7, Sep. 2003.

[9] D. K. P. Tan, H. Sun, Y. Lu, M. Lesturgie, and H. L. Chan, "Passive radar using global system for mobile communication signal: theory, implementation and measurements," *Proc. Inst. Elect. Eng.-Radar, Sonar, Navigat.*, vol. 152, pp. 116-123, Jun. 2005.

[10] Q. Wang, C. Hou, and Y. Lu, "An experimental study of wimax-based passive radar," *IEEE Trans. Microwave Theory & Tech*, vol. 58, pp. 3502-3510, Dec. 2010.

[11] S. Xu, B. Chen, and P. Zhang, "Radar-communication integration based on DSSS techniques," *Proc. Int. Conf. on Signal Processing*, Beijing, China, pp. 1-7, Nov. 2006.

[12] P. Barrenechea, F. Elferink, and J. Janssen, "FMCW radar with broadband communication capability," *Proc. the 4th European Radar Conf.*, Munich, Germany, pp. 130-133, Oct. 2007.

[13] M. Jamil, H.-J. Zepernick, and M. I. Pettersson, "On integrated radar and communication systems using oppermann sequences," *Proc. IEEE Military Commun. Conf.*, San Diego, CA, pp. 1-6, Nov. 2008.

[14] M. Braun, C. Sturm, A. Niethammer, and F. K. Jondral, "Parametrization of joint OFDM-based radar and communication systems for vehicular applications," *Proc. Symp. Personal Indoor and Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, pp. 3020-3024, Sep. 2009.

[15] M. K. Hoang and R. Haeb-Umbach, "Parameter estimation and classification of censored Gaussian data with application to WiFi indoor positioning," *Proc. IEEE IntâĂŹl Conf. on Acoust., Speech and Signal Process (ICASSP)*, Vancouver, BC, pp. 3721-3725, May. 2013.

[16] H. Zou, H. Jiang, X. Lu and L. Xie "An online sequential extreme learning machine approach to WiFi based indoor positioning," *Proc. Internet of Things (WF-IoT), 2014 IEEE World Forum* , Seoul, South Korea, pp. 111-116, Mar. 2014.

[17] C. Yang and H.Shao, "WiFi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, pp. 150-157, Mar. 2015.

[18] W. Su, J. D. Matyjas, M. J. Gans and S. Batalama, "On the capacity of airborne MIMO communications", in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012, pp. 4629-4634.

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

[20] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 569-570, Sep. 1975.

[21] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786-788, Nov. 1981.

[22] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49-60, Jan. 1981.

[23] H. Chong, M. Motani, H. K. Garg and H. E. Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188-3195, July 2008.

[24] X. Shang and B. Chen, "A new computable achievable rate region for the Gaussian interference channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.

[25] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 620–643, Feb. 2009.

[26] M. H. M. Costa, "Noisebergs in Z gaussian interference channels," in *ITA Workshop*, San Diego, CA, Feb. 2011.

[27] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, pp. 607–615, Sep. 1985.

[28] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *Proc. IEEE INFOCOM*, Rome, Italy, pp. 85-96, July 2001.

[29] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in Ad Hoc wireless networks," in *Proc. ACM MobiCom*, Anchorage, AK, pp. 1548-1557, Apr. 2001.

[30] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: A short preamble MAC protocol for duty-cycled wireless sensor networks," in *Proc. ACM SenSys*, Boulder, CO, pp. 307-320, Oct./Nov. 2006.

[31] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Communications*, vol. 60, pp. 220-230, Jan. 2012.

[32] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1732-1743, Sep. 2011.

[33] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Processing*, vol. 60, pp. 4808-4818, Sep. 2012.

[34] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 1180-1189, Mar. 2012.

[35] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1326-1336, Apr. 2010.

[36] C. Huang, R. Zhang, and S. Cui, "Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraints," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1074-1087, Feb. 2014.

[37] O. Orhan, D. Gunduz, and E. Erkip, "Delay-constrained distortion minimization for energy harvesting transmission over a fading channel," *Proc. IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, pp. 1794-1798, July 2013.

[38] P. Castiglione, O. Simeone, E. Erkip, and T. Zemen, "Energy management policies for energy-neutral source-channel coding," *IEEE Trans. Communications*, vol. 60, pp. 2668-2678, Sep. 2012.

[39] A. Limmanee, S. Dey and A. Ahlen, "Distortion minimization via multiple sensors under energy harvesting constraints," in *Proc. IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, pp. 225-229, June 2013.

[40] C. Huang, Y. Zhou, T. Jiang, P. Zhang, and S. Cui, "Power allocation for joint estimation with energy harvesting constraints," in *Proc. IEEE Int'l Conf. on Acoust., Speech and Signal Process (ICASSP)*, Vancouver, BC, pp. 4804-4808, May 2013.

[41] A. Nayyar, T. Basar, D. Teneketzis, and V. V. Veeravalli, "Communication scheduling and remote estimation with energy harvesting sensor," in *Proc. IEEE 51st Annual Conference on Decision and Control (CDC)*, Maui, HI, pp. 843-848, Dec. 2012.

[42] D. P. Bertsekas, *Dynamic Programming and Optimal Control Vol. 1.* Belmont, MA: Athena Scientific, 1995.

[43] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, pp. 558-567, Oct. 1965.

[44] B. Rimoldi, "Beyond the separation principle: a broader approach to source-channel codings," in *Proc. 4th Int. ITG Conf. Source and Channel Coding*, Berlin, Germany, Jan. 2002.

[45] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1147-1158, May 2003.

[46] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," in *Proc. Information Theory and Applications (ITA) Workshop*, La Jolla, CA, pp. 177-182, Jan. 2007.

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

[48] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, issue. 9, pp. 1249-1274, Sep. 2000.

[49] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1486-1495, Apr. 2005.

[50] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1850-1859, May 2008.

[51] W. Liu and B. Chen, "Message transmission and state estimation over Gaussian broadcast channels," *Proc. Information Sciences and Systems, 43rd Annual Conference on (CISS)*, Baltimore, MD, pp. 147-151, Mar. 2009.

[52] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. Inf. Theory*, vol. 59, pp. 3709-3719, Jun. 2013.

[53] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439-441, May 1983.

[54] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1147-1158, May 2003.

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

[56] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 and 623-656, Jul. and Oct. 1948.

[57] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. Control*, vol. 2, pp. 101-112, Jun. 1959.

[58] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, pp. 269-275, Mar. 1991.

[59] C. Tian, S. Diggavi, and S. Shamai, "The achievable distortion region of sending a bivariate Gaussian source on the Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6419-6427, Oct. 2011.

[60] J. Li and P. Stoica, *MIMO Radar Signal Processing.* Wiley, 2008.

[61] J. Li and P. Stoica, "MIMO radar with colocated antennas: review of some recent work," in *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106-114, Sep. 2007.

[62] A. M. Haimovich, R. S. Blum and L. Cimini, "MIMO radar with widely separated antennas," in *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 116-129, Jan. 2008.

[63] H. Chen, X. Li, W. Jiang and Z. Zhuang, “MIMO radar sensitivity analysis of antenna position for direction finding," in *IEEE Trans. Signal Processing*, vol. 60, no. 10, pp. 5201-5216, Oct. 2012.

[64] E. Fishler, A. Haimovich, R. S. Blum, L. J. Cimini, D. Chizhik and R. A. Valenzuela, “Spatial diversity in radars-models and detection performance", in *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 823-838, Mar. 2006.

[65] T. Naghibi and F. Behnia, “MIMO radar waveform design in the presence of clutter", in *IEEE Trans. Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 770-781, Apr. 2011.

[66] P. Wang, H. Li and B. Himed, “Moving target detection using distributed MIMO radar in clutter with nonhomogeneous power", in *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4809-4820, Sep. 2011.

[67] G. Cui, J. Liu, H. Li and B. Himed, “Target detection for passive radar with noisy reference channel", in *IEEE Proc. Radar Conference*, Cincinnati, OH, May 2014, pp. 144-148.

[68] J. Brown, K. Woodbridge, A. Stove and S. Watts, “Air target detection using airborne passive bistatic radar", in *Electronics Letters*, vol. 46, no. 20, pp. 1396-1397, Sep. 2010.

[69] I. Suberviola, I. Mayordomo and J. Mendizabal, “Experimental results of air target detection with a GPS forward-scattering radar", in *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 1, pp. 47-51, Jan. 2012.

[70] M. Cherniakov, D. Nezlin and K. Kubik, “Air target detection via bistatic radar based on LEOs communication signals", in *IEE Proc. Radar Sonar and Navigation*, vol. 149, no. 1, pp. 33-38, Feb. 2002.

[71] M. J. Gans, “Aircraft free-space MIMO communications,” in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2009, pp. 663-666.

[72] M. H. Hayes, *Statistical Digital Signal Processing And Modeling.* John Wiley & Sons, Inc., 1996.

[73] D. S. Matteson and N.A. James, “A nonparametric approach for multiple change point analysis of multivariate data", in *journal of American Statistical Association*, vol. 109, is. 505, pp. 334-345, Nov. 2013.

[74] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application.* Prentice Hall, 1993.

[75] J. Zhang, “Powerful goodness-of-fit tests based on likelihood ratio", in *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 64, pp. 281-294, 2002.

[76] C. Zou and F. Tsung, “Likelihood ratio-based distribution-free EWMA control charts", in *Journal of Quality Technology*, vol. 42, pp. 174-196, 2010.

[77] J. R. Jian and P. T. Helo, “Optimization design of a CUSUM control chart based on taguchi’s loss function", in *The International Journal of Advanced Manufacturing Technology*, vol. 35, pp. 1234-1243, Oct. 2006.

[78] B. P. Carlin, A. E. Gelfand and A. F. Smith, “Hierarchical Bayesian analysis of changepoint problems", in *Applied Statistics*, vol. 41, no. 2, pp. 389-405, 1992.

[79] M. Lavielle and G. Teyssiere, “Detection of multiple change-points in multivariate time series", in *Lithuanian Mathematical Journal*, vol. 46, no. 3, pp. 287-306, 2006.

[80] S. Chakraborti and M. A. van de Wiel, “A nonparametric control chart based on the Mann-Whitney statistic", in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pp. 156-172, 2008.

[81] D. M. Hawkins and Q. Deng, “A nonparametric change-point control chart", in *Journal of Quality Technology*, vol. 42, pp. 165-173, 2010.

[82] G. J. Ross and N. M. Adams, “Two nonparametric control charts for detecting arbitrary distribution changes", in *Journal of Quality Technology*, vol. 44, no. 2, pp. 102-116, Apr. 2012.

[83] G. J. Szekely and M. L. Rizzo, “Hierarchical clustering via joint between-within distances: extending ward’s minimum variance method", in *Journal of Classification*, vol. 22, no. 2, pp. 151-183, 2005.

[84] W. Hoeffding, “The strong law of large numbers for U-statistics", in *Technical Report 302*, North Carolina State University, Dept. of Statistics, 1961.

[85] M. Biguesh and A. B. Gershman, “Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals", in *IEEE Trans. Signal Processing*, vol. 54, is. 3, pp. 884-893, Mar. 2006.

VITA

NAME OF AUTHOR: Yu Zhao

EDUCATION:

M.S. May 2014 Syracuse Univeristy, NY, US

B.S. July 2009 Beijing Institute of Technology, Beijing, China

PUBLICATIONS:

Y. Zhao, B. Chen, and R. Zhang, "Optimal power management for remote estimation with an energy harvesting sensor," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 6471-6480, Nov. 2015.

Y. Zhao and B. Chen, "Capacity theorems for multi-functioning radios," *Proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, pp. 2406-2410, July 2014.

K. Borle, F. Zhu, Y. Zhao, and B. Chen, "A software radio design for communications in uncoordinated networks," *Proc. IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Toronto, ON, pp. 254-258, June 2014.

Y. Zhao, B. Chen, and R. Zhang, "Optimal power allocation for an energy harvesting estimation system," *Proc. IEEE Int'l Conf. on Acoust., Speech and Signal Process (ICASSP)*, Vancouver, BC, pp. 4549 - 4553, May 2013.

Y. Zhao, F. Zhu, and B. Chen, "The Han-Kobayashi region for a class of Gaussian interference channels with mixed interference," *Proc. IEEE International Symposium*

on Information Theory (ISIT), Boston, MA, pp. 2791 - 2795, July 2012.