Communication over Channels with Causal Side Information at the Transmitter

by

Hamidreza Farmanbar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This work deals with communication over the AWGN channel with additive discrete interference, where the sequence of interference symbols is known causally at the transmitter. We use Shannon's treatment for channels with side information at the transmitter as a framework to derive "optimal precoding" and "channel code design criterion" for the channel with known interference at the transmitter.

Communication over Shannon's state-dependent discrete memoryless channel where the state sequence is known causally at the transmitter requires encoding over the so-called *associated* channel which has exponential input alphabet cardinality with respect to the number of states. We show that by using at most linearly many input symbols of the *associated* channel, the capacity is achievable.

In particular, we consider M-ary signal transmission over the AWGN channel with additive Q-ary interference where the sequence of i.i.d. interference symbols is known causally at the transmitter. We investigate the problem of maximization of the transmission rate under the uniformity constraint, where the channel input given any current interference symbol is uniformly distributed over the channel input alphabet. For this setting, we propose the general structure of a communication system with optimal precoding. We also investigate the extension of the proposed precoding scheme to continuous channel input alphabet.

We also consider the problem of channel code design with causal side information at the encoder. We derive the code design criterion at high SNR by defining a new distance measure between the input symbols of the Shannon's *associated* channel. For the case of the binary-input channel, i.e., $M = 2$, we show that it is sufficient

to use only two (out of 2^Q) input symbols of the *associated* channel in encoding as far as the distance spectrum of code is concerned. This reduces the problem of channel code design for the binary-input AWGN channel with known interference at the encoder to design of binary codes for the binary symmetric channel where the Hamming distance among codewords is the major factor in the performance of the code.

Acknowledgements

This work would not have been possible without the help and support of many people. It is a pleasure to convey my gratitude to them all.

I would like to acknowledge and thank my supervisor, Professor Amir K. Khandani for his valuable guidance and support throughout the course of my graduate studies at the University of Waterloo.

I would also like to thank the members of my thesis committee, Professors W. Yu, P. Fieguth, M. O. Damen, and M. Uysal for taking the time to carefully read my thesis and providing me with insightful comments and suggestions.

I wish to thank everyone with whom I have worked with in the Coding and Signal Transmission Laboratory for many useful discussions.

I wish to give my sincere thanks to my family, my parents and my wife, for their constant encouragement and love throughout these years.

Dedication

To my wife,

Roghayeh,

for her support.

Contents

List of Figures

Chapter 1

Introduction

Information transmission over channels with known interference at the transmitter has been a major focus of research due to its application in various communication problems. A remarkable result on such channels was obtained by Costa who showed that the capacity of the additive white Gaussian noise (AWGN) channel with additive Gaussian i.i.d. interference where the sequence of interference symbols is known non-causally at the transmitter is the same as the capacity of the AWGN channel [1]. Therefore, the interference does not incur any loss in the capacity. This result was extended to arbitrary interference (random or deterministic) by Erez *et. al.* [2]. Following Costa's "Writing on Dirty Paper" famous title [1], a coding strategy for the channel with non-causally known interference at the transmitter is referred to as "dirty paper coding" (DPC). By analogy, a coding strategy for the channel with causally-known interference at the transmitter is sometimes referred to as "dirty tape coding" (DTC).

The result obtained by Costa does not hold for the case that the sequence of

interference symbols is known causally at the transmitter. In fact, the capacity is unknown in this case and unlike the non-causal knowledge setting, the capacity depends on the interference. The only definitive result in this case is due to Erez *et. al.* [2], who showed that for the worst-case interference, at the limit of high SNR, the loss in capacity due to not having the future samples of the interference at the transmitter is exactly the ultimate shaping gain $\frac{1}{2}\log\left(\frac{2\pi e}{12}\right) \approx 0.254$ bit.

Recently, dirty paper coding (DPC) has emerged as a building block in multiuser communication. In particular, there has been considerable research studying the application of dirty paper coding to broadcast over multiple-input multiple-output (MIMO) channels. In such systems, for a given user, the signals sent to other users are considered as interference. Since all signals are known to the transmitter, successive "dirty paper" cancelation can be used in transmission after some linear preprocessing [4]. It was shown that DPC in fact achieves the sum capacity of the MIMO broadcast channel [5, 6, 7]. Most recently, it has been shown that the same is true for the entire capacity region of the MIMO broadcast channel [8, 9, 10]. Another important application of DPC is information embedding or watermarking [11, 12, 13, 14, 15, 16], where a host signal is modeled as interference onto which a watermark signal is embedded.

These developments motivate finding realizable dirty paper coding techniques. Building upon [2], Erez and ten Brink [50] proposed a practical code design based on vector quantization via trellis shaping and using powerful channel codes. Due to the complexity of implementation, their scheme uses the knowledge of interference up to six future symbols rather than the whole interference sequence. Bennatan *et al.* [51] gave another design based on superposition coding and successive cancelation

decoding. Their design uses a trellis coded quantizer with memory length nine and a low density parity check (LDPC) code as channel code. Yu *et al.* [52] gave a design based on convolutional shaping and channel codes.

The schemes that use the interference sequence up to the current symbol can be used as low-complexity solutions for the dirty paper problem. For example, in [11], scalar lattice quantization is proposed for data-hiding even though in that context, the host signal is clearly known non-causally.

In this work, we consider the AWGN channel with additive discrete interference where the sequence of i.i.d. interference symbols is known causally at the transmitter. The discrete interference model is more appropriate for many practical applications. For example, in the MIMO broadcast channel where the transmitter uses a finite constellation, the interference caused by the other users is discrete rather than continuous. We are interested in the capacity of the channel, optimal precoding scheme, and channel code design for the channel.

1.1 Channels with Side Information at the Transmitter

Channels with known interference at the transmitter are special cases of channels with side information at the transmitter which were considered for the first time by Shannon [17]. Shannon considered a discrete memoryless channel (DMC) whose transition matrix depends on the channel state. A state-dependent discrete memoryless channel (SD-DMC) is defined by a finite input alphabet $\mathcal{X}$, a finite output

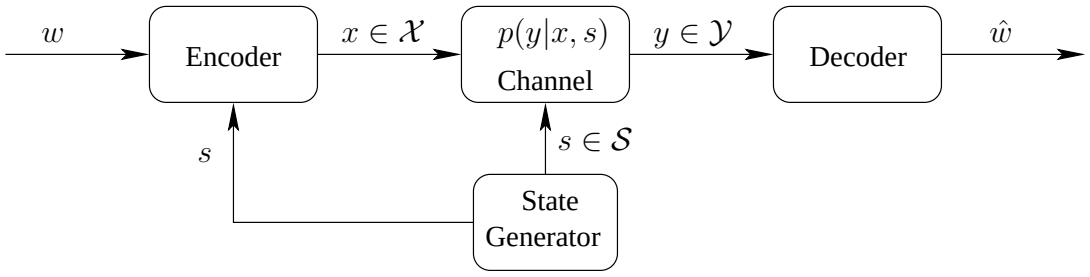

Figure 1.1: SD-DMC with state information at the encoder.

alphabet $\mathcal{Y}$, and transition probabilities $p(y|x,s)$, where the state s takes on values in a finite alphabet $\mathcal{S}$. The block diagram of a state-dependent channel with state information at the encoder is shown in fig. 1.1.

We may consider two different settings for the knowledge of state sequence at the encoder: causal or non-causal. In the causal knowledge setting, the encoder maps a message w into $\mathcal{X}^n$ as

$$x_i = f_i\left(w, s_1, \ldots, s_i\right), \quad i = 1, \ldots, n, \tag{1.1}$$

whereas in the non-causal knowledge setting, the encoder observes the entire state sequence to generate every symbol of the code sequence; i.e.,

$$x_i = f_i\left(w, s_1, \ldots, s_n\right), \quad i = 1, \ldots, n. \tag{1.2}$$

In either case (causal or non-causal), the receiver decodes the message from the whole received sequence as $\hat{w} = g(y_1^n)$.

Shannon considered the case where the i.i.d. state sequence is known causally at the encoder and obtained the capacity formula [17]. The case where the i.i.d. state sequence is known non-causally at the encoder was considered by Kuznetsov and Tsybakov in the context of coding for memories with defective cells [18]. Gel'fand and Pinsker obtained the capacity formula for this case [19].

Shannon's capacity formula was generalized by Salehi [20] for the case that a noisy version of the state sequence is available at both encoder and decoder. Caire and Shamai [21] investigated the case that the state sequence is not memoryless. The capacity results with non-causal side information at the encoder were generalized to the case were rate-limited side information is available at both encoder and decoder [22, 23].

Shannon showed that it is sufficient to consider the coding schemes that use only the current state symbol in the encoding process to achieve the capacity of an SD-DMC with i.i.d. state sequence which is known causally at the encoder [17]. The SD-DMC can be used in the way shown in fig. 1.2 to transmit information. A precoder is added in front of the SD-DMC. A message w is mapped into $\mathcal{T}^n$, where $\mathcal{T}$ is a new alphabet. The output of the precoder ranges over $\mathcal{X}$ and depends on the current interference symbol. The regular (without state) channel from T to Y is defined by the transition probabilities

$$p(y|t) = \sum_{s \in \mathcal{S}} p(s) p(y|x = t(s), s), \tag{1.3}$$

where $p(s)$ is the probability of the state s. The channel from T to Y is a discrete memoryless channel, i.e.,

$$p(y_1^n|t_1^n) = \prod_{i=1}^{n} p(y_i|t_i) \tag{1.4}$$

The DMC defined in (1.3) is called the *associated* channel. The codes for the *associated* channel describe the codes for the SD-DMC that use only the current state symbols in the encoding operation. In order to describe all coding schemes for the SD-DMC that use only the current state symbol in the encoding process, $\mathcal{T}$ must include all functions from the state alphabet to the input alphabet of the

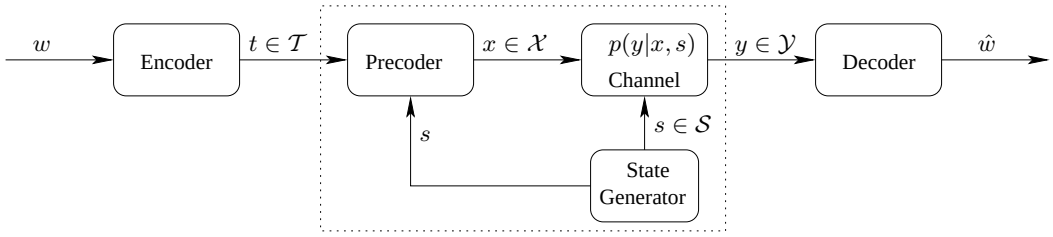

Figure 1.2: The *associated* regular DMC.

state-dependent channel. There are a total of $|\mathcal{X}|^{|\mathcal{S}|}$ of such functions, where $|.|$ denotes the cardinality of a set. Any of the functions can be represented by a $|\mathcal{S}|$-tuple $(x_1, x_2, \ldots, x_{|\mathcal{S}|})$ composed of elements of $\mathcal{X}$, implying that the value of the function at state s is $x_s, s = 1, 2, \ldots, |\mathcal{S}|$.

The capacity of the *associated* channel, which is the same as the capacity of the state-dependent channel with state sequence known causally at the transmitter, is given by [17]

$$C_c = \max_{p(t)} I(T;Y), \tag{1.5}$$

where the maximization is taken over the probability mass function (pmf) of the random variable T. In the capacity formula (1.5), we can alternatively replace the random variable T with $(X_1, \ldots, X_{|\mathcal{S}|})$, where X_s is the random variable that represents the input to the state-dependent channel when the state is $s, s = 1, \ldots, |\mathcal{S}|$.

The capacity with non-causal side information at the transmitter is given by [19]

$$C_{nc} = \max_{p(t,x|s)} \left\{ I(T;Y) - I(T;S) \right\}. \tag{1.6}$$

The capacity achieving coding scheme for the channel with known non-causal side information at the transmitter is based on random binning [24]. The following rough argument illustrates the achievability of (1.6). Generate $2^{nI(T;Y)}$ typical sequences

according to $p(t)$ and distribute them randomly into 2^{nR} bins. Then each bin contains $2^{n(I(T;Y)-R)}$ typical sequences. Only a fraction $2^{-nI(T;S)}$ of the sequences in each bin are jointly typical with the state sequence. Reliable communication is possible if with high probability every bin contains a sequence jointly typical with the state sequence. Therefore, all rates $R < I(T;Y) - I(T;S)$ are achievable.

1.2 Precoding for Canceling Known Interference

Consider the Costa channel model shown in fig. 1.3. In this model, S is additive white Gaussian interference, N is AWGN with power P_N, and Channel input X is constrained to have power P_X. Costa used the Gelfand-Pinsker capacity formula (1.6) as follows to obtain the capacity of the AWGN channel with additive Gaussian interference which is known non-causally at the transmitter. Pick a joint distribution on S, X, and T such that X is Gaussian with power P_X and independent of S, and $T = X + \alpha S$ where $\alpha = \frac{P_X}{P_X+P_N}$. With this choice of α, $X - \alpha(X+N)$ is independent of $X+N$ and $Y = X+S+N$ since they are jointly Gaussian and uncorrelated. Then compute conditional differential entropies

$$\begin{aligned} h(T|S) &= h(X+\alpha S|S) \\ &= h(X|S) \\ &= h(X), \end{aligned} \tag{1.7}$$

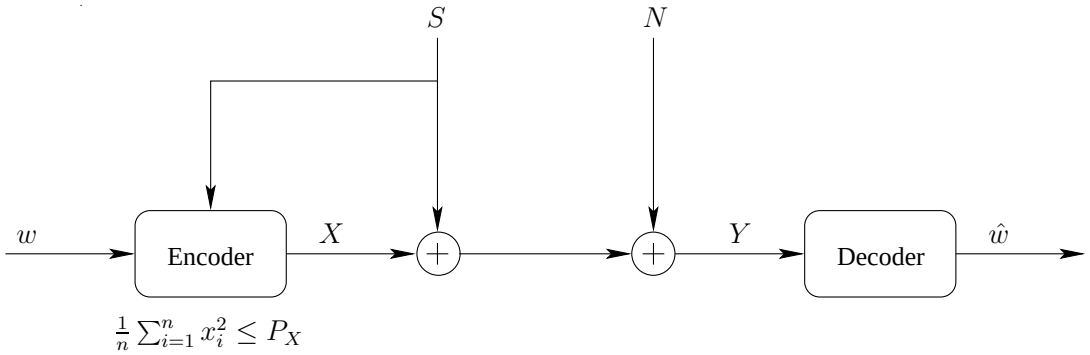

Figure 1.3: The Costa channel model.

and

$$\begin{aligned} h(T|Y) &= h(X+\alpha S|Y) \\ &= h(X+\alpha(S-Y)|Y) \\ &= h(X-\alpha(X+N)|Y) \\ &= h(X-\alpha(X+N)) \\ &= h(X-\alpha(X+N)|X+N) \\ &= h(X|X+N). \end{aligned} \tag{1.8}$$

Then the achievable rate is given by

$$\begin{aligned} I(T;Y)-I(T;S) &= h(T|S)-h(T|Y) \\ &= h(X)-h(X|X+N) \\ &= I(X;X+N) \\ &= \frac{1}{2}\log\left(1+\frac{P_X}{P_N}\right). \end{aligned} \tag{1.9}$$

The capacity cannot be larger than the capacity of the AWGN channel. Therefore, the capacity is given by (1.9). This result has since been generalized to the case

where the interference has any power-limited distribution [12] and to the case of arbitrary interference sequence (random or deterministic) [2], and to the case of colored Gaussian interference and noise [25].

As mentioned earlier, the capacity of the Gelfand-Pinsker channel, and in particular, the capacity of the dirty paper channel is obtained by the random binning method. This method typically produces unstructured codes, which are not suitable for practical applications. Zamir *et. al.* suggested a structured coding approach based on nested lattices [26]. Let Λ be an n-dimensional lattice with fundamental Voronoi region Λ_0 and second moment P_X. Also let $\mathbf{U}$ be a vector random variable uniformly distributed over Λ_0. The vector random variable $\mathbf{U}$ is called as *dither* and is assumed to be available at both transmitter and receiver. The transmission scheme is as follows. For any $\mathbf{v} \in \Lambda_0$, the transmitter sends

$$\mathbf{x} = [\mathbf{v} - \alpha\mathbf{s} - \mathbf{u}] \mod \Lambda_0, \tag{1.10}$$

and the receiver computes

$$\mathbf{y}' = [\alpha\mathbf{y} + \mathbf{u}] \mod \Lambda_0. \tag{1.11}$$

Due to using dither which is uniformly distributed in Λ_0, $\mathbf{X}$ is uniformly distributed in Λ_0 and is independent of $\mathbf{V}$ and $\mathbf{S}$ and has power P_X. The resulting channel is a modulo-Λ_0 additive noise channel described by the following equivalent channel model [3]

$$\mathbf{Y}^{'} = \mathbf{V} + \mathbf{N}^{'} \mod \Lambda_0, \tag{1.12}$$

with

$$\mathbf{N}^{'} = [-(1-\alpha)\mathbf{X} + \alpha\mathbf{N}] \mod \Lambda_0. \tag{1.13}$$

The mutual information of the channel is maximized by a uniform input, giving

$$\begin{aligned}\frac{1}{n}I(\mathbf{V};\mathbf{Y}') &= \frac{1}{n}h(\mathbf{Y}') - \frac{1}{n}h(\mathbf{N}') && (1.14)\\ &= \frac{1}{2}\log\frac{P_X}{G(\Lambda)} - \frac{1}{n}h(\mathbf{N}'), && (1.15)\end{aligned}$$

where $G(\Lambda) = \frac{1}{n}\frac{\int_{\Lambda_0}\|\mathbf{x}\|^2 d\mathbf{x}}{|\Lambda_0|^{1+2/n}}$ is the normalized second moment of Λ and $|\Lambda_0|$ is the volume of Λ_0. Taking $\alpha = \frac{P_X}{P_X + P_N}$ we have

$$\begin{aligned}\frac{1}{n}\mathrm{var}(-(1-\alpha)\mathbf{X} + \alpha\mathbf{N}) &= (1-\alpha)^2\frac{1}{n}\mathrm{var}(\mathbf{X}) + \alpha^2\frac{1}{n}\mathrm{var}(\mathbf{N}) && (1.16)\\ &= \alpha P_N. && (1.17)\end{aligned}$$

Since for a given second moment, a Gaussian random vector has the largest entropy, it follows that

$$\begin{aligned}\frac{1}{n}h(\mathbf{N}') &\leq \frac{1}{n}h((1-\alpha)\mathbf{X} + \alpha\mathbf{N}) && (1.18)\\ &\leq \frac{1}{2}\log(2\pi e\alpha P_N), && (1.19)\end{aligned}$$

where the first inequality follows because the modulo operation can only decrease the entropy. Substituting (1.19) in (1.15), we obtain the following lower bound on the achievable rate as a function of $G(\Lambda)$,

$$\frac{1}{n}I(\mathbf{V};\mathbf{Y}') \geq \frac{1}{2}\log\left(1 + \frac{P_X}{P_N}\right) - \frac{1}{2}\log 2\pi e G(\Lambda). \quad (1.20)$$

Thus, in principle, for a given lattice Λ, the gap to the capacity may be made smaller than $\log 2\pi e G(\Lambda)$. For optimal lattices for quantization, we have $G(\Lambda) \to \frac{1}{2\pi e}$; and the gap goes to zero.

The capacity of the AWGN channel with additive i.i.d. interference which is known causally at the transmitter is not known. In fact, there is some loss due to

not having the future interference symbols at the transmitter. The scalar version of precoding scheme defined by (1.10) and (1.11) can be used for the case that the sequence of interference symbols is known causally at the transmitter as follows. Based on the input symbol V and the current interference symbol S, the transmitter sends

$$X = [V - \alpha S - U] \mod \Delta, \tag{1.21}$$

where the dither U is uniformly distributed in $A_\Delta = \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ and $\alpha = \frac{P_X}{P_X + P_N}$. The receiver computes

$$Y' = [\alpha Y + U] \mod \Delta, \tag{1.22}$$

where $Y = X + S + N$ is the channel output. The receiver output Y' is related to the input V as [3]

$$Y^{'} = [V + N^{'}] \mod \Delta, \tag{1.23}$$

where $N^{'}$ is independent of V and is given by

$$N^{'} = [-(1-\alpha)X + \alpha N] \mod \Delta. \tag{1.24}$$

The optimality of the scalar modulo precoding scheme defined in (1.21) and (1.22) is only proved for the worst-case interference and at the limit of high SNR [2]. The mutual information between V and Y' from (1.23) is

$$I(V; Y') = h(Y') - h(N') \tag{1.25}$$

A uniform distribution for V in A_Δ will make Y' uniformly distributed, and hence, maximizes the above mutual information

$$I(V; Y') = \log \Delta - h(N'). \tag{1.26}$$

The above rate has been depicted vs. SNR in Fig. 1.4. The loss in capacity at high SNR is shown to be the ultimate shaping gain 1.53 dB. At low SNR, the loss in capacity is even more.

Due to using common randomness (dither at both transmitter and receiver), the distribution of interference does not affect the achievable rate for the modulo precoding scheme defined in (1.21) and (1.22). In fact, the rate (1.26) is achievable for the *worst-case* interference, which is equivalent to "strong and smooth" interference [2]. Removing common randomness from the modulo precoding scheme can increase the rate. For example, consider the case that the interference has small power. Then it is more reasonable to consider the interference as part of the noise instead of using the dithered modulo precoding, which results in higher transmit power. The increase in rate and improvement in error probability by removing the common randomness from the dithered modulo precoding scheme has been investigated in [27].

The modulo precoding scheme requires modulo operation both at the transmitted and receiver sides. According to data processing inequality [29, 30], modulo operation at the receiver may introduces loss in rate. The gain in rate obtained by removing the modulo operation at the receiver is investigated in [28]. For the case of strong interference, however, it has been shown that modulo operation at the receiver does not reduce the rate [28].

The first causal precoding scheme was proposed by Willems [31, 32]. It is easy to show that his proposed scheme is the same as the modulo precoding scheme with $\alpha = 1$.

Recently, dirty paper/tape coding arguments have been generalized for mul-

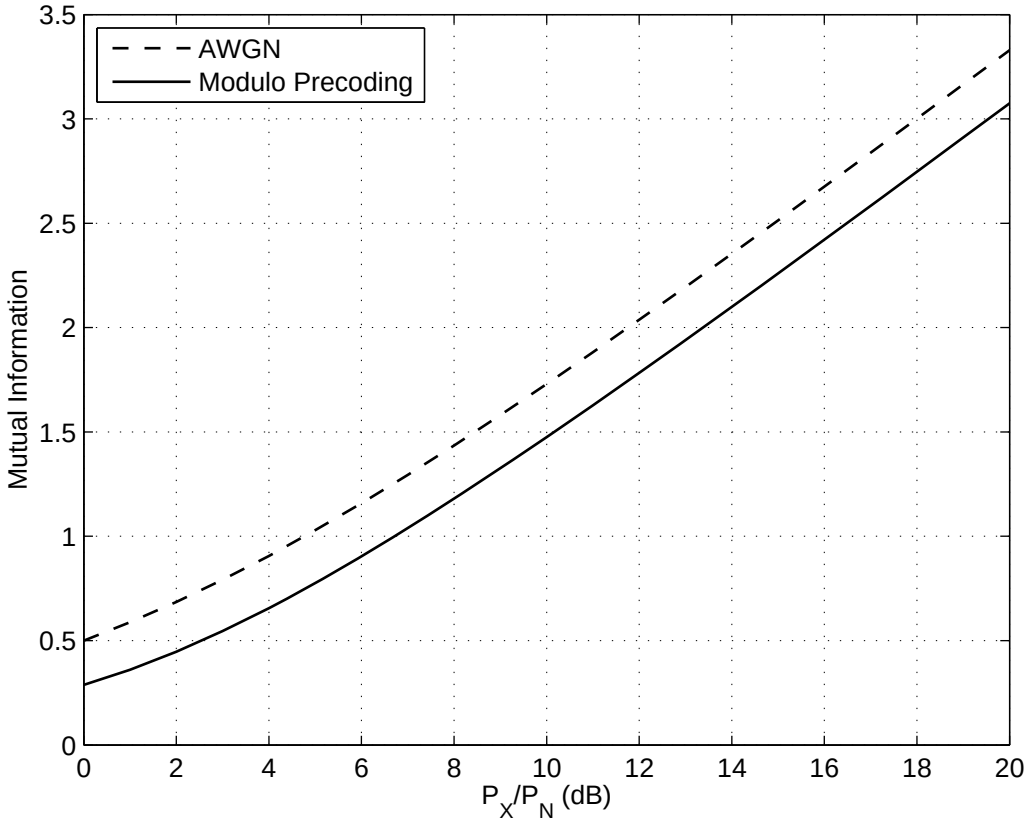

Figure 1.4: Achievable rate with modulo precoding.

tiuser channels. In particular, it has been shown that Costa's result is valid for the Gaussian broadcast channel, the Gaussian multiple-access channel, and the physically degraded Gaussian relay channel [33]. Extensions of the Shannon's result to the physically degraded broadcast and relay channels are given in [34], [35]. The capacity region of the multiple-access channel with rate-limited noncausal side information at the transmitters has been investigated in [36]. The sum-capacity of the multiple-access channel with non-causal independent side information is obtained in [37].

1.3 Thesis Organization

In chapter 2, we investigate the problem of precoding for the AWGN channel with additive discrete interference where the sequence of i.i.d. interference symbols is known causally at the transmitter. We consider both discrete and continuous channel input alphabets. We begin the chapter by introducing an upper bound on the cardinality of a capacity achieving distribution for the general state-dependent channel model considered by Shannon. The linear-programming-based argument for obtaining the upper bound provides some insight on a capacity achieving distribution and also serves as a guideline to obtain the optimal precoding scheme. The results presented in this chapter has been published in part in [38], [39], [40].

In chapter 3, we consider the problem of channel code design for the same channel. Our design does not rely on the suboptimal (in terms of capacity) scheme of modulo precoding for the causally known interference [2], [53]. Instead, we consider code design for the Shannon's *associated* channel over all possible input

symbols. Another distinction between our work and the related research in the field is that we consider a finite channel input alphabet rather than a continuous one. The results presented in this chapter has been published in part in [41, 42].

Chapter 4 contains conclusions and directions for future research.

Chapter 2

Precoding for the AWGN Channel with Discrete Interference

The design of precoding schemes to maximize the information transmission rate is the subject of this chapter. In section 2.1, we derive an upper bound on the cardinality of a capacity achieving input distribution for the Shannon's *associated* channel. In section 2.2, we introduce our channel model. In section 2.3, we investigate the capacity of the channel in the absence of noise. In section 2.4, we consider maximizing the transmission rate under the uniformity constraint, where the channel input given any current interference symbol is uniformly distributed over the channel input alphabet. We summarize the optimal precoding scheme with uniformity and integrality constraints in section 2.5. We extend the uniform transmission scheme to continuous-input alphabet in section 2.6.

2.1 An Upper Bound

Consider the SD-DMC shown in fig. 1.1. We may explicitly define the channel input and state alphabets as

$$\mathcal{X} = \{x_1, x_2, \ldots, x_{|\mathcal{X}|}\}, \tag{2.1}$$

$$\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}, \tag{2.2}$$

respectively. We can obtain the pmf of the channel output Y as

$$\begin{aligned} p_Y(y) &= \sum_{s=1}^{|\mathcal{S}|} p_S(s) p_{Y|S}(y|s) \\ &= \sum_{s=1}^{|\mathcal{S}|} p_S(s) \sum_{i=1}^{|\mathcal{X}|} p_{X|S}(x_i|s) p_{Y|X,S}(y|x_i, s) \\ &= \sum_{s=1}^{|\mathcal{S}|} p_S(s) \sum_{i=1}^{|\mathcal{X}|} p_{X_s}(x_i) p_{Y|X,S}(y|x_i, s), \end{aligned} \tag{2.3}$$

where X_s is the channel input given the current state is s. The capacity of the *associated* channel, which is the same as the capacity of the original state-dependent channel, is the maximum of $I(T;Y) = I(X_1 X_2 \cdots X_{|\mathcal{S}|}; Y)$ over the joint pmf values $p_{i_1 i_2 \cdots i_{|\mathcal{S}|}} = \Pr\{X_1 = x_{i_1}, \ldots, X_{|\mathcal{S}|} = x_{i_{|\mathcal{S}|}}\}$, i.e.,

$$C = \max_{p_{i_1 i_2 \cdots i_{|\mathcal{S}|}}} I(X_1 X_2 \cdots X_{|\mathcal{S}|}; Y). \tag{2.4}$$

The mutual information between T and Y is the difference between the entropies $H(Y)$ and $H(Y|T)$. It can be seen from (2.3) that $p_Y(y)$, and hence $H(Y)$, are uniquely determined by the marginal pmfs $\{p_{X_s}(x_i)\}_{i=1}^{|\mathcal{X}|}$, $s = 1, \ldots, |\mathcal{S}|$. The conditional entropy $H(Y|T)$ is given by

$$\begin{aligned} H(Y|T) &= H(Y|X_1 X_2 \cdots X_{|\mathcal{S}|}) \\ &= \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} h_{i_1 \cdots i_{|\mathcal{S}|}} p_{i_1 \cdots i_{|\mathcal{S}|}}, \end{aligned} \tag{2.5}$$

where $h_{i_1\cdots i_{|\mathcal{S}|}} = H(Y|X_1 = x_{i_1}, \ldots, X_{|\mathcal{S}|} = x_{i_{|\mathcal{S}|}})$.

There are $|\mathcal{X}|^{|\mathcal{S}|}$ variables involved in the maximization problem (2.4). Each variable represents the probability of an input symbol of the *associated* channel. The following theorem regards the number of nonzero variables required to achieve the maximum in (2.4).

Theorem 1. *The capacity of the* associated *channel is achieved by using at most* $|\mathcal{X}||\mathcal{S}| - |\mathcal{S}| + 1$ *out of* $|\mathcal{X}|^{|\mathcal{S}|}$ *input symbols with nonzero probabilities.*

Proof. Denote by $\{\hat{p}_i^{(s)}\}_{i=1}^{|\mathcal{X}|} = \{\hat{p}_{X_s}(x_i)\}_{i=1}^{|\mathcal{X}|}$ the pmf of X_s, $s = 1, 2, \ldots, |\mathcal{S}|$, induced by a capacity-achieving joint pmf $\{\hat{p}_{i_1\cdots i_{|\mathcal{S}|}}\}_{i_1,\ldots,i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$. We limit the search for a capacity-achieving joint pmf to the set of joint pmfs that yield the same marginal pmfs as $\{\hat{p}_{i_1\cdots i_{|\mathcal{S}|}}\}_{i_1,\ldots,i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$. By limiting the search to this new set, the maximum of $I(X_1 \cdots X_{|\mathcal{S}|}; Y)$ remains unchanged since the capacity-achieving joint pmf $\{\hat{p}_{i_1\cdots i_{|\mathcal{S}|}}\}_{i_1,\ldots,i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$ is in the new set. But all joint pmfs in the new set yield the same $H(Y)$ since they induce the same marginal pmfs on $X_1, \ldots, X_{|\mathcal{S}|}$. Therefore, the maximization problem in (2.4) reduces to the linear minimization problem

$$
\begin{aligned}
&\min_{p_{i_1\cdots i_{|\mathcal{S}|}}} \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} h_{i_1\cdots i_{|\mathcal{S}|}} p_{i_1\cdots i_{|\mathcal{S}|}} \\
&\text{s. t.} \\
&\qquad \sum_{i_2=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} p_{i_1\cdots i_{|\mathcal{S}|}} = \hat{p}_{i_1}^{(1)}, \qquad\qquad i_1 = 1, \ldots, |\mathcal{X}|, \\
&\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots \\
&\qquad \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|-1}=1}^{|\mathcal{X}|} p_{i_1\cdots i_{|\mathcal{S}|}} = \hat{p}_{i_{|\mathcal{S}|}}^{(|\mathcal{S}|)}, \qquad i_{|\mathcal{S}|} = 1, \ldots, |\mathcal{X}|, \\
&\qquad p_{i_1\cdots i_{|\mathcal{S}|}} \geq 0, \qquad\qquad i_1, \ldots, i_{|\mathcal{S}|} = 1, 2, \ldots, |\mathcal{X}|. \qquad (2.6)
\end{aligned}
$$

There are $|\mathcal{X}||\mathcal{S}|$ equality constraints in (2.6) out of which $|\mathcal{X}||\mathcal{S}|-|\mathcal{S}|+1$ are linearly independent. From the theory of linear programming, the minimum of (2.6), and hence the maximum of $I(X_1 \cdots X_{|\mathcal{S}|}; Y)$, is achieved by a feasible solution with at most $|\mathcal{X}||\mathcal{S}| - |\mathcal{S}| + 1$ nonzero variables. □

Theorem 1 states that at most $|\mathcal{X}||\mathcal{S}| - |\mathcal{S}| + 1$ out of $|\mathcal{X}|^{|\mathcal{S}|}$ input symbols of the *associated* channel are needed to be used with positive probability to achieve the capacity. However, in general one does not know which of the input symbols must be used to achieve the capacity. If we knew the marginal pmfs for $X_1, \ldots, X_{|\mathcal{S}|}$ induced by a capacity-achieving joint pmf, we could obtain the capacity-achieving joint pmf itself by solving the linear program (2.6).

2.2 The Channel Model

We consider data transmission over the channel

$$Y = X + S + N, \tag{2.7}$$

where X is the channel input, which takes on values in a fixed real constellation

$$\mathcal{X} = \{x_1, x_2, \ldots, x_M\}, \tag{2.8}$$

Y is the channel output, N is additive white Gaussian noise with power P_N, and the interference S is a discrete random variable that takes on values in

$$\mathcal{S} = \{s_1, s_2, \ldots, s_Q\} \tag{2.9}$$

with probabilities $r_1, r_2, \ldots, r_Q$, respectively. The sequence of i.i.d. interference symbols is known causally at the encoder.

The above channel can be considered as a special case of state-dependent channels considered by Shannon with one exception, that the channel output alphabet is continuous. In our case, the likelihood function $f_{Y|X,S}(y|x,s)$ is used instead of the transition probabilities. We denote the input to the *associated* channel by T, which can also be represented as $(X_1, X_2, \ldots, X_Q)$, where X_j is the random variable that represents the channel input when the current interference symbol is s_j, $j = 1, \ldots, Q$.

The likelihood function for the *associated* channel is given by

$$\begin{aligned} f_{Y|T}(y|t) &= \sum_{j=1}^{Q} r_j f_{Y|X,S}(y|x_{i_j}, s_j) \\ &= \sum_{j=1}^{Q} r_j f_N(y - x_{i_j} - s_j), \end{aligned} \tag{2.10}$$

where f_N denotes the pdf of the Gaussian noise N, and t is the input symbol of the *associated* channel represented by $(x_{i_1}, x_{i_2}, \ldots, x_{i_Q})$.

According to theorem 1, the capacity of our channel is obtained by using at most $MQ - Q + 1$ out of M^Q input symbols of the *associated* channel.

2.3 The Noise-Free Channel

We consider a special case where the noise power is zero in (2.7). We call this channel as the noise-free channel. The following argument shows that $\log_2 M$ is an upper bound on the capacity of the noise-free channel: If the interference sequence is also made available at the decoder, then the capacity will be the average of capacities of different realizations of the noise-free channel over different interference

symbols, which is equal to $\log_2 M$. But making the interference known to the decoder does not reduce the capacity. Thus, $\log_2 M$ is a upper bound on the capacity.

In the absence of noise, the channel output Y takes on at most MQ different values since different X and S pairs may yield the same sum. If Y takes on exactly MQ different values, then it is easy to see that the capacity is $\log_2 M$ bits [1]: The decoder just needs to partition the set of all possible channel output values into M subsets of size Q corresponding to M possible input symbols, and decide that which subset the current received symbol belongs to.

In general, where the cardinality of the channel output symbols can be less than MQ, we will show that under some condition on the channel input alphabet, there exists a coding scheme that achieves the rate $\log_2 M$ in one use of the channel. We do this by considering a one-shot coding scheme which uses only M (out of M^Q) input symbols of the *associated* channel.

In a one-shot coding scheme, a message is encoded to a single input of the *associated* channel. Any input of the *associated* channel can be represented by a Q-tuple composed of elements of $\mathcal{X}$. Given that the current interference symbol is s_j, the jth element of the Q-tuple is sent through the channel. Therefore, one single message can result in (up to) Q symbols at the output. For convenience, we consider the output symbols corresponding to a single message as a multi-set[2] of size (exactly) Q. If the M multi-sets at the output corresponding to M

[1] This is true even if the interference sequence is unknown to the encoder.

[2] A multi-set differs from a set in that each member may have a multiplicity greater than one. For example, $\{1, 3, 3, 7\}$ is a multi-set of size four where 3 has multiplicity two.

different messages are mutually disjoint, reliable transmission through the channel is possible.

Unfortunately, we cannot always find M input symbols of the *associated* channel such that the corresponding multi-sets are mutually disjoint. For example, consider a channel with the input alphabet $\mathcal{X} = \{0, 1, 2, 4\}$ and the interference alphabet $\mathcal{S} = \{0, 1, 3\}$. It is easy to check that for this channel we cannot find four triples composed of elements of $\mathcal{X}$ such that the corresponding multi-sets are mutually disjoint. To see this, consider the sets $\{0, 1, 2, 4\}$, $\{1, 2, 3, 5\}$, and $\{3, 4, 5, 7\}$, which in fact are the sets of the channel output symbols when the interference symbol is 0, 1, and 3, respectively. We are looking for four mutually-disjoint multi-sets of size three composed of the elements of the above sets (one element from each). In order to have mutually-disjoint multi-sets, the repeated elements $1, 1$ must be in the same multi-set. The same is true for the other repeated elements $2, 2$, $3, 3$, $4, 4$, and $5, 5$. But we only have four multi-sets of size three, which makes it impossible. In fact, we can obtain the capacity of the corresponding *associated* channel (, which is the same as the capacity of the channel in the example) using the Arimoto-Blahut algorithm [43, 44] to see that the capacity of the channel in this example is 1.8869 bits.

However, if we impose some constraint on the channel input alphabet, the rate $\log_2 M$ is achievable.

Theorem 2. *Suppose that the elements of the channel input alphabet $\mathcal{X}$ form an arithmetic progression. Then the capacity of the noise-free channel*

$$Y = X + S, \tag{2.11}$$

where the sequence of interference symbols is known causally at the encoder equals $\log_2 M$ *bits.*

Proof. Let $\mathcal{Y}^{(q)}$ be the set of all possible outputs of the noise-free channel when the interference symbol is s_q, i.e.,

$$\mathcal{Y}^{(q)} = \{x_1 + s_q, x_2 + s_q, \ldots, x_M + s_q\}, \quad q = 1, \ldots, Q. \tag{2.12}$$

The union of $\mathcal{Y}^{(q)}$s is the set of all possible outputs of the noise-free channel.

Without loss of generality, we can assume that $s_1 < s_2 < \cdots < s_Q$. The elements of $\mathcal{Y}^{(q)}$ form an arithmetic progression, $q = 1, \ldots, Q$. Furthermore, these Q arithmetic progressions are shifted versions of each other.

We prove by induction on Q that there exist M mutually-disjoint multi-sets of size Q composed of the elements of $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \ldots, \mathcal{Y}^{(Q)}$ (one element from each). If we can find such M multi-sets of size Q, then we can obtain the corresponding M Q-tuples of elements of $\mathcal{X}$ by subtracting the corresponding interference terms from the elements of the multi-sets. These M Q-tuples can serve as the input symbols of the *associated* channel to be used for sending any of M distinct messages through the channel without error in one use of the channel, hence achieving the rate $\log_2 M$ bits per channel use.

For $Q = 1$, the statement of the theorem is true since we can take $\{x_1+s_1\}, \{x_2+s_1\}, \ldots, \{x_M + s_1\}$ as mutually-disjoint sets of size one.

Assume that there exist M mutually-disjoint multi-sets of size $Q = q$. For $Q = q + 1$, we will have the new set of channel outputs $\mathcal{Y}^{(q+1)} = \{x_1 + s_{q+1}, x_2 + s_{q+1}, \ldots, x_M + s_{q+1}\}$. We consider two possible cases:

Case 1: None of the elements of $\mathcal{Y}^{(q+1)}$ appear in any of the multi-sets of size $Q = q$.

In this case, we include the elements of $\mathcal{Y}^{(q+1)}$ in the M multi-sets arbitrarily (one element is included in each multi-set). It is obvious that the resulting multi-sets of size $Q = q + 1$ are mutually disjoint.

Case 2: Some of the elements of $\mathcal{Y}^{(q+1)}$ appear in some of the multi-sets of size $Q = q$.

Suppose that the largest element of $\mathcal{Y}^{(q+1)}$ which appears in any of the sets $\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(q)}$ (or equivalently, in any of the multi-sets of size $Q = q$) is $x_k + s_{q+1}$ for some $1 \leq k \leq M - 1$. Then since $\mathcal{Y}^{(q+1)}$ is shifted version of each $\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(q)}$ and $s_{q+1} > s_q > \cdots > s_1$, exactly one of the sets $\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(q)}$, say $\mathcal{Y}^{(j)}$ for some $1 \leq j \leq q$, contains all elements of $\mathcal{Y}^{(q+1)}$ up to $x_k + s_{q+1}$. See fig. 2.1. Since any of the disjoint multi-sets of size Q contain just one element of $\mathcal{Y}^{(j)}$, the elements of $\mathcal{Y}^{(q+1)}$ up to $x_k + s_{q+1}$ appear in different multi-sets of size $Q = q$. We can form the disjoint multi-sets of size $q+1$ by including these common elements in the corresponding multi-sets and including the elements of $\{x_{k+1} + s_{q+1}, \ldots, x_M + s_{q+1}\}$ in the remaining multi-sets arbitrarily. □

The condition on the channel input alphabet in the statement of theorem 2 is a sufficient condition for the channel capacity to be $\log_2 M$. However, it is not a necessary condition. For example, the statement of theorem 2 without that condition is true for the case $Q = 2$. Because in the second iteration, we do not need the arithmetic progression condition to form M mutually-disjoint multi-sets of size two.

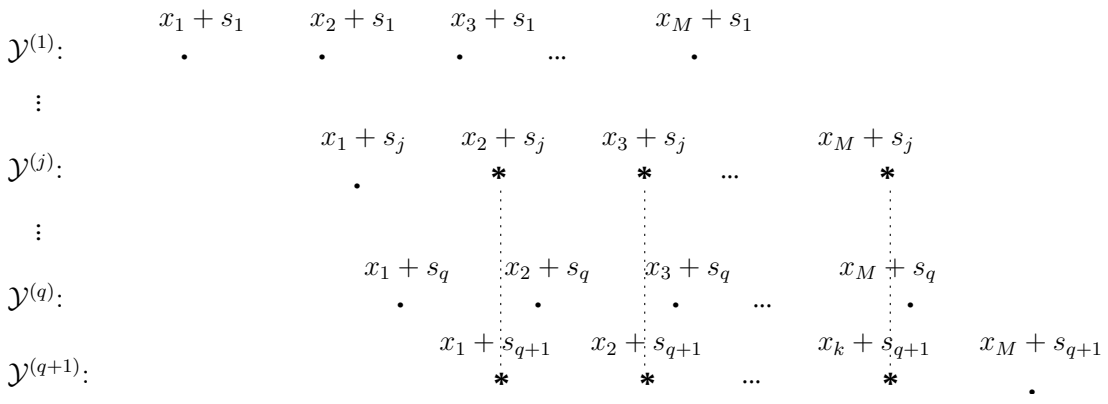

Figure 2.1: The elements of $\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(q+1)}$ shown as shifted version of each other. The elements of $\mathcal{Y}^{(q+1)}$ up to $x_k + s_{q+1}$ appear in $\mathcal{Y}^{(j)}$.

It is worth mentioning that in the proof of theorem 2, we did not use the assumption that the interference sequence is i.i.d.. In fact, the interference sequence could be any arbitrary varying sequence of the elements of $\mathcal{S}$.

The proof of theorem 2 is actually a constructive algorithm for finding M (out of M^Q) input symbols of the *associated* channel to be used with probability $\frac{1}{M}$ to achieve the rate $\log_2 M$ bits.

It is interesting to see that the set containing the qth elements of the M Q-tuples obtained by the constructive algorithm is $\mathcal{X}$, $q = 1, \ldots, Q$. This is due to the fact that each multi-set contains one element from each $\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(Q)}$. Therefore, a uniform distribution on the M Q-tuples induces uniform distributions on $X_1, \ldots, X_Q$.

2.4 Uniform Transmission

In the sequel, we study the maximization of the rate $I(X_1 \cdots X_Q; Y)$ over joint pmfs $\{p_{i_1 \cdots i_Q}\}_{i_1,\ldots,i_Q=1}^{M}$ that induce uniform marginal distributions on $X_1, \ldots, X_Q$, i.e.,

$$p_i^{(1)} = p_i^{(2)} = \cdots = p_i^{(Q)} = \frac{1}{M}, \qquad i = 1, 2, \ldots, M, \tag{2.13}$$

for which we show how to obtain the optimal input probability assignment. We call a transmission scheme that induces uniform distributions on $X_1, \ldots, X_Q$ as *uniform transmission*. Uniform distributions for $X_1, \ldots, X_Q$ implies uniform distribution for X, the input to the state-dependent channel defined in (2.7).

In the previous section, we established that the capacity achieving pmf for the asymptotic case of noise-free channel induces uniform distributions on $X_1, \ldots, X_Q$ (provided that we can find M Q-tuples such that the corresponding multi-sets are mutually disjoint). Therefore, imposing the uniformity constraint given in (2.13) does not reduce the transmission rate in the asymptotic case of noise-free channel. However, in the general case where the noise power is not zero there will be some loss in rate due to imposing the uniformity constraint.

Imposing the uniformity constraint along with the integrality constraint (which will be explained later on in this section), however, simplifies the encoding operation for the *associated* channel as will be shown in this section. Furthermore, we will show in section 2.6 that our precoding scheme with both uniformity and integrality constraints provides higher rates than the existing modulo precoding scheme of [2].

Considering the uniformity constraints given in (2.13), the maximization of

$I(X_1 \cdots X_Q; Y)$ is reduced to the linear minimization problem

$$\begin{aligned}
\min_{p_{i_1 \cdots i_Q}} \quad & \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} h_{i_1 \cdots i_Q} p_{i_1 \cdots i_Q} \\
\text{subject to} \quad & \\
& \sum_{i_2=1}^{M} \cdots \sum_{i_Q=1}^{M} p_{i_1 \cdots i_Q} = \frac{1}{M}, \qquad i_1 = 1, 2, \ldots, M, \\
& \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\
& \sum_{i_1=1}^{M} \cdots \sum_{i_{Q-1}=1}^{M} p_{i_1 \cdots i_Q} = \frac{1}{M}, \qquad i_Q = 1, 2, \ldots, M, \\
& p_{i_1 \cdots i_Q} \geq 0, \qquad\qquad\qquad\qquad i_1, \ldots, i_Q = 1, 2, \ldots, M. \quad (2.14)
\end{aligned}$$

The equality constraints of (2.14) can be interpreted as the following. We assign $p_{i_1 \cdots i_Q}$ to the element $(i_1, \ldots, i_Q)$ of an M by $M \cdots$ by M (Q times) array. For $Q = 2$, the equality constraints of (2.14) mean that every row and every column of the array adds up to $\frac{1}{M}$. For $Q > 2$, the equality constraints can be interpreted accordingly.

The same argument used in the last part of the proof of theorem 1 can be used to show that the maximum rate with uniformity constraint is achieved by using at most $MQ - Q + 1$ input symbols of the *associated* channel with positive probabilities. This is restated in the following corollary.

Corollary 1. *The maximum of $I(X_1 \cdots X_Q; Y)$ over joint pmfs $\{p_{i_1 \cdots i_Q}\}_{i_1, \ldots, i_Q = 1}^{M}$ that induce uniform marginal distributions on $X_1, X_2, \ldots, X_Q$ is achieved by a joint pmf with at most $MQ - Q + 1$ nonzero elements.*

This result is independent of the coefficients $\{h_{i_1 \cdots i_Q}\}$. However, which probability assignment with at most $MQ-Q+1$ nonzero elements is optimal depends on the

coefficients $\{h_{i_1 \cdots i_Q}\}$. The coefficient $h_{i_1 \cdots i_Q}$ is determined by the interference levels $s_1, \ldots, s_Q$, the probability of interference levels $r_1, \ldots, r_Q$, the noise power P_N, and the signal points $x_1, x_2, \ldots, x_M$. The optimal probability assignment is obtained by solving the linear programming problem (2.14) using the simplex method [46].

2.4.1 The Two-Level Interference

If the number of interference levels is two, i.e., $Q = 2$, we can make a stronger statement than corollary 1.

Theorem 3. *The maximum of $I(X_1X_2;Y)$ over $\{p_{i_1 i_2}\}_{i_1,i_2=1}^{M}$ with uniform marginal pmfs for X_1 and X_2 is achieved by using exactly M out of M^2 input symbols of the* associated *channel with probability $\frac{1}{M}$.*

Proof. The equality constraints of (2.14) can be written in matrix form as

$$\mathbf{Ap} = \mathbf{1}, \tag{2.15}$$

where $\mathbf{A}$ is a zero-one $MQ \times M^Q$ matrix, $\mathbf{p}$ is M times the vector containing all $p_{i_1 \cdots i_Q}$s in lexicographical order, and $\mathbf{1}$ is the all-one $MQ \times 1$ vector.

For $Q = 2$, it is easy to check that $\mathbf{A}$ is the vertex-edge incidence matrix of $K_{M,M}$, the complete bipartite graph with M vertices at each part. Therefore, $\mathbf{A}$ is a totally unimodular matrix[3] [45]. Hence, the extreme points of the feasible region $F = \{\mathbf{p} : \mathbf{Ap} = \mathbf{1}, \mathbf{p} \geq \mathbf{0}\}$ are integer vectors. Since the optimal value of a linear optimization problem is attained at one of the extreme points of its feasible region,

[3]A totally unimodular matrix is a matrix for which every square submatrix has determinant 0, 1, or -1.

the minimum in (2.14) is achieved at an all-integer vector $\mathbf{p}^*$. Considering that $\mathbf{p}^*$ satisfies (2.15), it can only be a zero-one vector with exactly M ones. □

As an example, the optimal solution for a channel with $\mathcal{X} = \{-3, -1, +1, +3\}$ and $\mathcal{S} = \{-2, 2\}$ with equiprobable interference symbols is illustrated in fig. 2.2. The points circled in the array correspond to the input symbols to the *associated* channel that must be chosen with probability $\frac{1}{4}$ in order to achieve the maximum rate in the uniform transmission scenario.

Fig. 2.3 depicts the maximum mutual information (for the uniform transmission scenario) vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$ and equiprobable interference symbols. The mutual information vs. SNR curve for the interference-free AWGN channel with equiprobable input alphabet $\{-1, +1\}$ is plotted for comparison purposes. As it can be seen, for low SNR, the input probability assignment $p_{11} = p_{22} = \frac{1}{2}$ is optimal, whereas at high SNR, the input probability assignment $p_{12} = p_{21} = \frac{1}{2}$ is optimal. The maximum achievable rate for uniform transmission is the upper envelope of the two curves corresponding to different input probability assignments. Also, it can be observed that the achievable rate approaches $\log_2 2 = 1$ bit per channel use as SNR increases complying with the fact that we established in theorem 2 for the noise-free channel.

It turns out from the proof of theorem 3 that the optimum solution of the linear optimization problem, $\mathbf{p}^*$, is a zero-one vector. So, if we add the integrality constraint to the set of constraints in (2.15), we still obtain the same optimal solution. The resulting integer linear optimization problem is called the *assignment problem* [45], which can be solved using low-complexity algorithms such as the

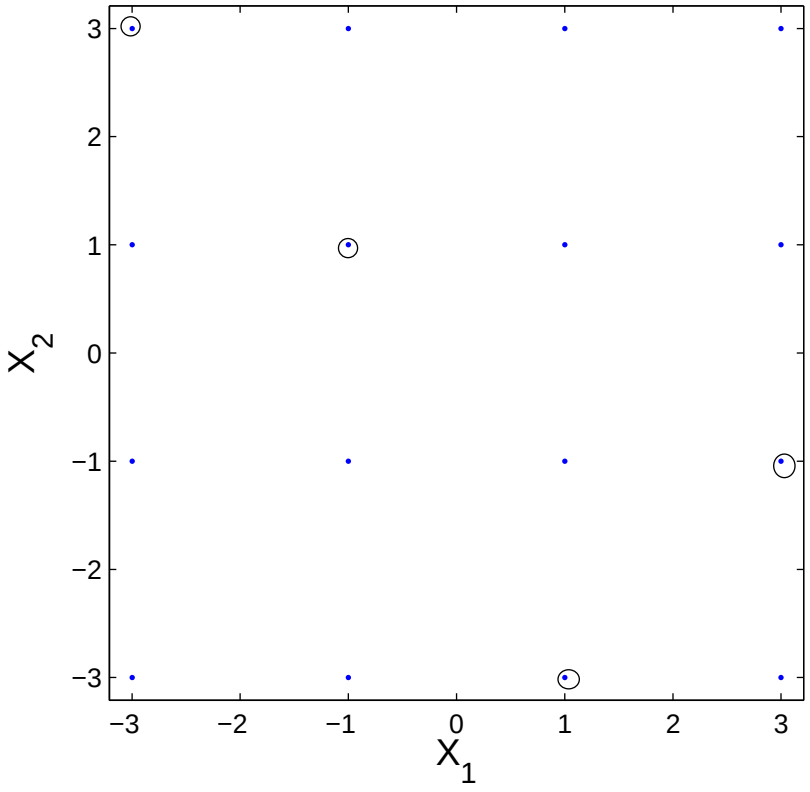

Figure 2.2: Optimal solution for 4-PAM input with parameters $r_1 = r_2 = \frac{1}{2}, s_1 = -2, s_2 = +2, P_N = 1$.

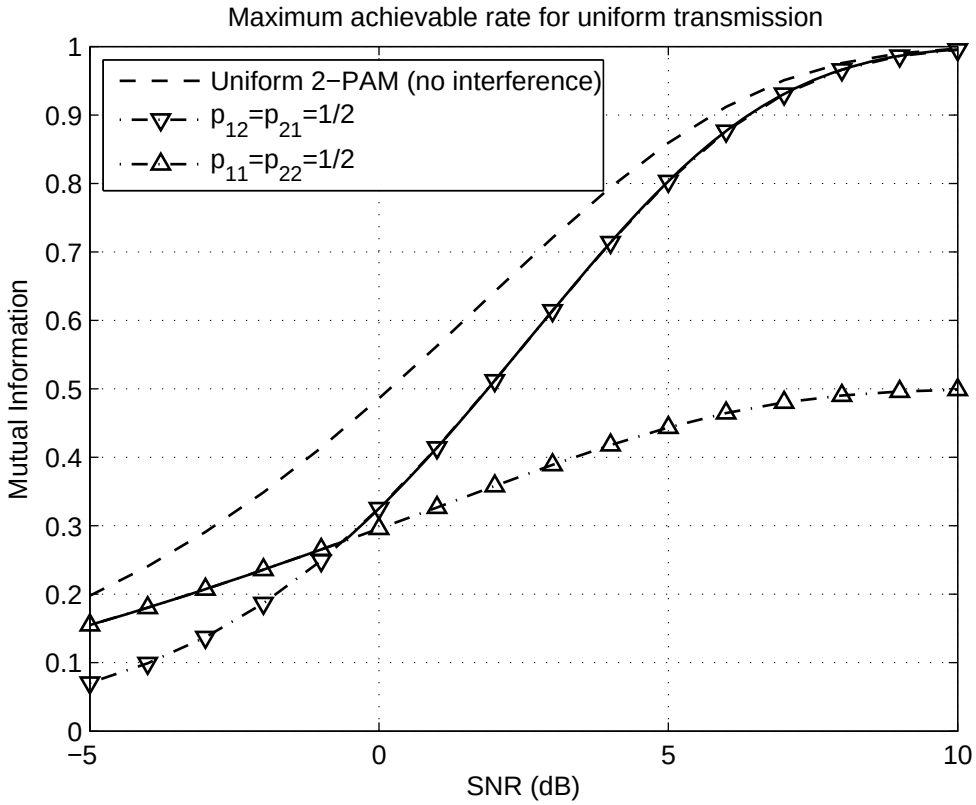

Figure 2.3: Maximum mutual information vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$ and $r_1 = r_2 = \frac{1}{2}$.

Hungarian method [46].

2.4.2 Integrality Constraint for the Q-Level Interference

The fact that for the case $Q = 2$, there exists an optimal $\mathbf{p}$ which is a zero-one vector with exactly M ones simplifies the encoding operation. Because any encoding scheme just needs to work on a subset of size M of the *associated* channel input alphabet with equal probabilities $\frac{1}{M}$.

For $Q \neq 2$, $\mathbf{A}$ is not a totally unimodular matrix. Therefore, not all extreme points of the feasible region defined by $\mathbf{Ap} = \mathbf{1}, \mathbf{p} \geq \mathbf{0}$, are integer vectors. However, at the expense of possible loss in rate, we may add the integrality constraint (i.e., $\mathbf{p}$ integer) in this case. The resulting optimization problem is called the *multi-dimensional assignment problem* [47]. The optimal solution of (2.14) with the integrality constraint, will be a vector with exactly M nonzero elements with the value $\frac{1}{M}$. Therefore, any encoding scheme just needs to use M symbols of the *associated* channel with equal probabilities, simplifying the encoding operation.

Fig. 2.4 depicts the maximum mutual information for uniform transmission with the integrality constraint vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-3, -1, +1, +3\}$ and with equiprobable interference symbols. The mutual information vs. SNR curve for the interference-free AWGN channel with equiprobable input alphabet $\{-3, -1, +1, +3\}$ is plotted for comparison purposes. It is interesting to mention that we obtained the exact same curves as in fig. 2.4 without imposing the integrality constraints.

It is worth mentioning that, with the integrality constraint, the optimal solution

of (2.14) is a joint pmf of $X_1, \ldots, X_Q$ for which $X_2, \ldots, X_Q$ can be presented as a function of X_1.

2.4.3 Explicit Optimal Solutions

In the sequel, we further investigate the optimal solution of (2.14). It can be shown that the coefficient $h_{i_1 \cdots i_Q} = h(Y|X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q})$ is a function of $x_{i_1} - x_{i_2}, x_{i_1} - x_{i_3}, \ldots, x_{i_1} - x_{i_Q}$, i.e.,

$$h_{i_1 \cdots i_Q} = g(x_{i_1} - x_{i_2}, x_{i_1} - x_{i_3}, \ldots, x_{i_1} - x_{i_Q}), \tag{2.16}$$

where g is a given by

$$\begin{aligned} g(u_1, \ldots, u_{Q-1}) \; &= -\int_{-\infty}^{+\infty} \left(r_1 f_N(z) + \textstyle\sum_{q=2}^{Q} r_q f_N(z + u_{q-1} + s_1 - s_q) \right) \times \\ &\qquad \log_2 \left(r_1 f_N(z) + \textstyle\sum_{q=2}^{Q} r_q f_N(z + u_{q-1} + s_1 - s_q) \right) dz. \end{aligned} \tag{2.17}$$

The plot of $g(.)$ for $Q = 2$ with parameters $r_1 = \frac{1}{2}, r_2 = \frac{1}{2}, s_1 = -2, s_2 = +2, P_N = 1$ is shown in fig. 2.5. The plot of $g(.)$ for $Q = 3$ with parameters $r_1 = r_2 = r_3 = \frac{1}{3}, s_1 = -2, s_2 = 0, s_3 = +2, P_N = 1$ is shown in fig. 2.6. In appendix A, it has been shown that g is lower bounded by the differential entropy of the noise, $h(N)$, and is upper-bounded by $h(N) + H(S)$, where $H(S)$ is the entropy of the discrete interference.

We may assume that x_1 and x_M are the smallest and the largest elements of the input alphabet $\mathcal{X}$, respectively. Then the following theorem gives an explicit solution to (2.14) under some circumstances.

Theorem 4. *If g is convex in the $(Q-1)$-cube $\{(u_1, \ldots, u_{Q-1}) :$ $x_1 - x_M \leq u_i \leq x_M - x_1, i = 1, 2, \ldots, Q-1\}$, then the optimal solution to (2.14)*

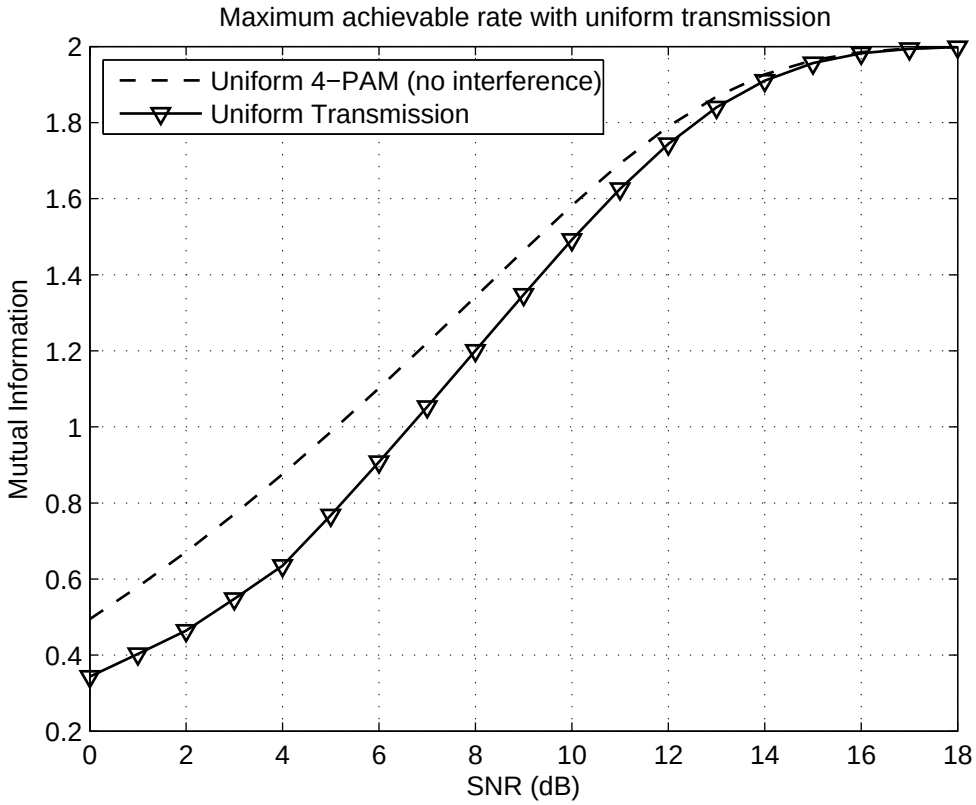

Figure 2.4: Maximum mutual information vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-3, -1, +1, +3\}$ and $r_1 = r_2 = r_3 = r_4 = \frac{1}{4}$.

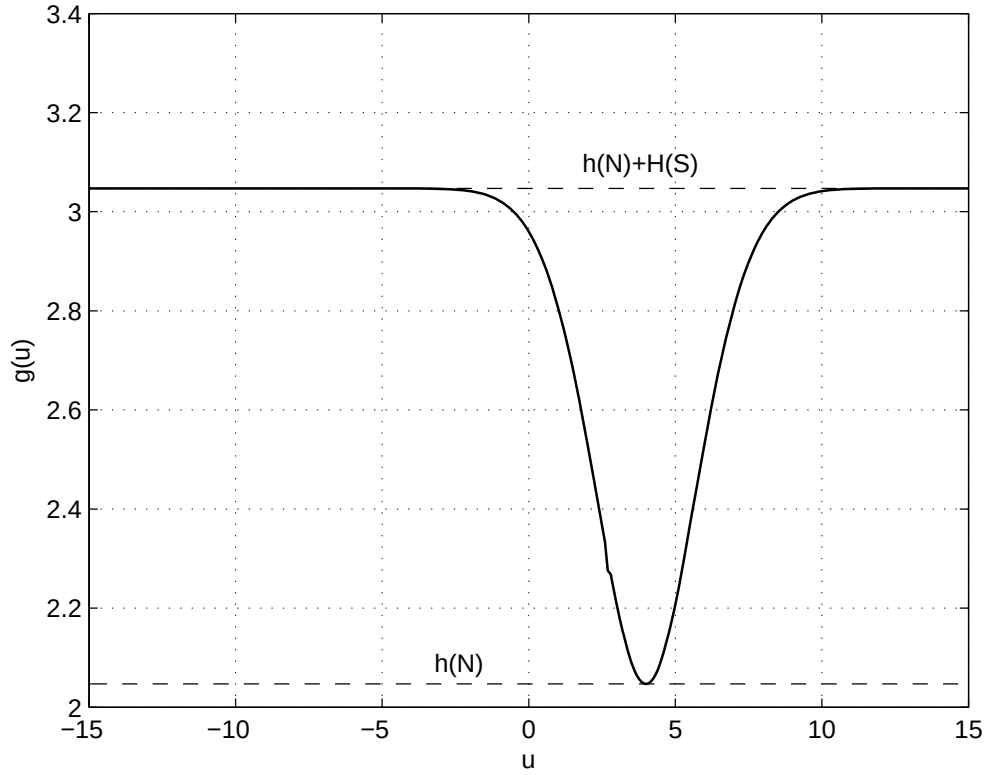

Figure 2.5: The plot of $g(u)$ for $r_1 = \frac{1}{2}, r_2 = \frac{1}{2}, s_1 = -2, s_2 = +2, P_N = 1$.

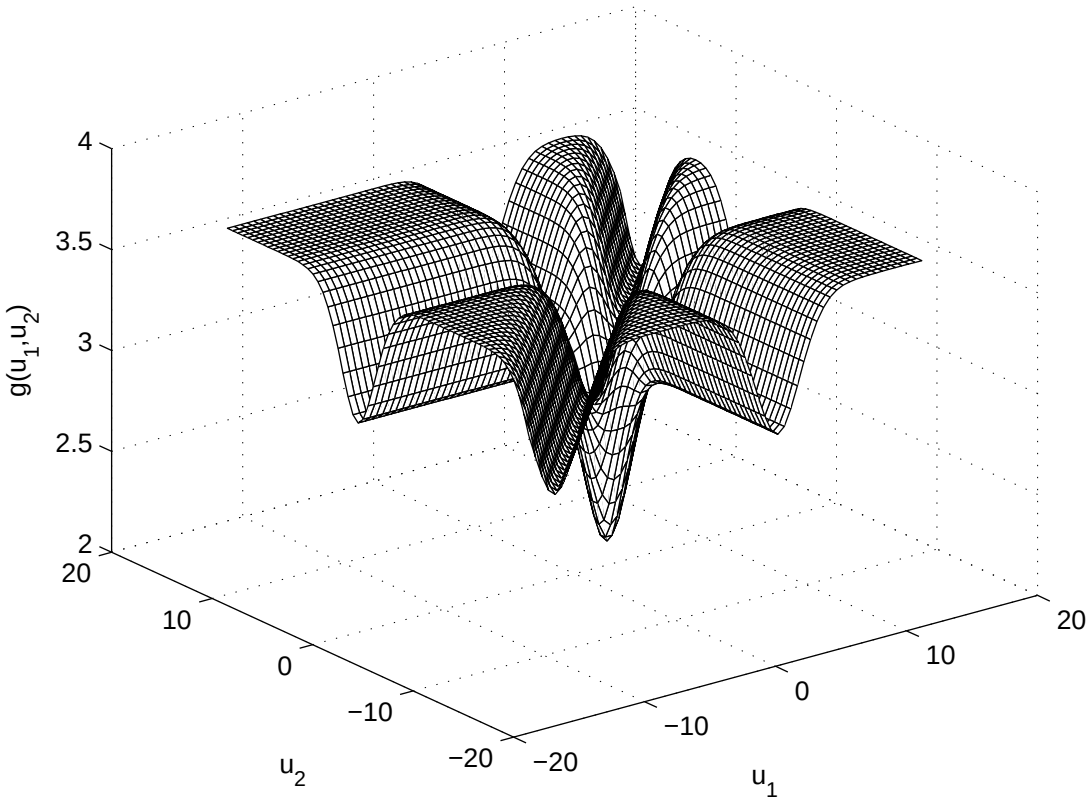

Figure 2.6: The plot of $g(u_1, u_2)$ with parameters $r_1 = r_2 = r_3 = \frac{1}{3}, s_1 = -2, s_2 = 0, s_3 = +2, P_N = 1$.

is

$$\tilde{p}_{i_1\cdots i_Q} = \begin{cases} \frac{1}{M}, & \textit{if} \quad i_1 = \cdots = i_Q \\ 0, & \textit{otherwise.} \end{cases} \tag{2.18}$$

Proof. Define random variables $U_i = X_1 - X_{i+1}$, $i = 1, \ldots, Q-1$. The objective function in (2.14) can be written as

$$\begin{aligned}
& \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} \Pr\left\{X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q}\right\} g(x_{i_1} - x_{i_2}, \ldots, x_{i_1} - x_{i_Q}) \\
= & \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \Pr\left\{X_1 = x_{i_1}, X_2 = x_{i_1} - u_{j_1}, \ldots, X_Q = x_{i_1} - u_{j_{Q-1}}\right\} \times \\
& \qquad\qquad g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\
= & \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \Pr\left\{X_1 = x_{i_1}, X_1 - X_2 = u_{j_1}, \ldots, X_1 - X_Q = u_{j_{Q-1}}\right\} \times \\
& \qquad\qquad g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\
= & \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \Pr\left\{X_1 = x_{i_1}, U_1 = u_{j_1}, \ldots, U_{Q-1} = u_{j_{Q-1}}\right\} g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\
= & \sum_{j_1} \cdots \sum_{j_{Q-1}} \Pr\left\{U_1 = u_{j_1}, \ldots, U_{Q-1} = u_{j_{Q-1}}\right\} g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\
= & \ \mathrm{E}[g(U_1, \ldots, U_{Q-1})],
\end{aligned} \tag{2.19}$$

where E[.] denotes the expectation operator. Now, considering the convexity of g, apply the Jensen's Inequality

$$\begin{aligned}
\mathrm{E}[g(U_1, \ldots, U_{Q-1})] & \geq g\left(\mathrm{E}[U_1, \ldots, U_{Q-1}]\right) \\
& = g(0, \ldots, 0).
\end{aligned} \tag{2.20}$$

Equality holds when the random variables $U_1, \ldots, U_{Q-1}$ take the value zero with probability one, or equivalently,

$$X_1 = X_2 = \cdots = X_Q. \tag{2.21}$$

The joint pmf in (2.18) satisfies both the constraints in (2.14) and (2.21), so it is the optimal solution. □

For $Q = 2$, the convexity of g in the interval $[x_1 - x_M, x_M - x_1]$ is equivalent to

$$x_M - x_1 \leq s_1 - s_2 + u^* \sqrt{P_N}, \tag{2.22}$$

where $u^* \approx 1.636$ and $s_1 < s_2$. The proof can be found in appendix B. In general ($Q \geq 2$), when the power of the noise P_N is sufficiently large, g will be convex in the $(Q-1)$-cube.

Theorem 4 has an interesting interpretation: Given the condition of theorem 4 satisfied, the optimal precoder sends the same symbol in the channel regardless of the current interference symbol. In other words, the optimal precoder for uniform transmission ignores the interference. In fact, as it can be seen from (2.20), any transmission scheme that forces $X_1, \ldots, X_Q$ to have the same statistical average does not benefit from the causal knowledge of interference symbols at the transmitter if the condition of theorem 4 is satisfied. Note that this might not hold true for a capacity achieving coding scheme without any constraints on the marginal pmfs of $X_1, \ldots, X_Q$.

The following theorem holds for the case $Q = 2$ and when the input alphabet $\mathcal{X}$ is symmetric w.r.t. the origin, i.e.,

$$x_i = -x_{M+1-i}, \quad i = 1, \ldots, M. \tag{2.23}$$

For example, a regular PAM constellation satisfies (2.23).

Theorem 5. *If the input alphabet $\mathcal{X}$ is symmetric w.r.t. the origin, and if g is concave in the interval $[x_1 - x_M, x_M - x_1]$, then*

$$\tilde{p}_{ij} = \begin{cases} \frac{1}{M}, & \text{if} \quad i + j = M + 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.24}$$

is an optimal solution to (2.14).

Proof. We rewrite (2.14) for the case $Q = 2$ as

$$\begin{aligned} \min_{p_{ij}} \quad & \sum_{i=1}^{M} \sum_{j=1}^{M} h_{ij} p_{ij} \\ \text{subject to} \quad & \\ & \sum_{j=1}^{M} p_{ij} = \frac{1}{M}, \quad i = 1, 2, \ldots, M, \\ & \sum_{i=1}^{M} p_{ij} = \frac{1}{M}, \quad j = 1, 2, \ldots, M, \\ & p_{ij} \geq 0, \qquad i, j = 1, 2, \ldots, M. \end{aligned} \tag{2.25}$$

We assign p_{ij} to the element (i, j) of an M by M array (See fig. 2.2). The equality constraints of (2.25) mean that every row and every column of the array adds up to $\frac{1}{M}$. We make the observation that if $\{p_{ij}\}_{i,j=1,2,\ldots,M}$ is a feasible solution of (2.25), then $\{q_{ij}\}_{i,j=1,2,\ldots,M}$, where $q_{ij} = p_{(M+1-j)(M+1-i)}$, will also be a feasible solution of (2.25). Furthermore, due to (2.23) and the fact that $h_{ij} = g(x_i - x_j)$, $\{p_{ij}\}$ and $\{q_{ij}\}$ yield the same objective value. Therefore, if $\{p_{ij}\}$ is an optimal solution of (2.25), $\{q_{ij}\}$ will be an optimal solution too. The convex combination of the two optimal solutions $\{\theta_{ij} = \frac{1}{2} p_{ij} + \frac{1}{2} q_{ij}\}$ is also an optimal solution with the following symmetry property

$$\theta_{ij} = \theta_{(M+1-j)(M+1-i)}. \tag{2.26}$$

In fact, (2.26) describes a solution which is symmetric w.r.t. the main diagonal of the array. So far, we have established the existence of an optimal solution to (2.25) with the symmetry property (2.26). Now, suppose that a symmetric optimal solution to (2.25) has nonzero entries

$$p_{ij} = p_{(M+1-j)(M+1-i)} = p, \tag{2.27}$$

where $i + j \neq M + 1$. Now, if we add p to the main diagonal entries $p_{(M+1-j)j}$ and $p_{i(M+1-i)}$ and turn p_{ij} and $p_{(M+1-j)(M+1-i)}$ to zero, the constraints of (2.25) are not violated. However, the change in the objective function will be proportional to

$$\begin{aligned} &h(Y|X_1 = x_i, X_2 = x_{M+1-i}) + h(Y|X_1 = x_{M+1-j}, X_2 = x_j) \\ &-h(Y|X_1 = x_i, X_2 = x_j) - h(Y|X_1 = x_{M+1-j}, X_2 = x_{M+1-i}), \end{aligned}$$

which is equal to $g(2x_i) + g(-2x_j) - 2g(x_i - x_j)$ which is non-positive by concavity of g. Hence, we have not increased the objective value by the process described above. We can repeat the process until all nonzero entries lie on the main diagonal without increasing the objective value. Therefore, (2.24) is an optimal solution of (2.25). □

It can be shown that g is concave in the interval $[x_1 - x_M, x_M - x_1]$ if and only if

$$x_M - x_1 \leq s_2 - s_1 - u_0\sqrt{P_N}. \tag{2.28}$$

See appendix B for the proof.

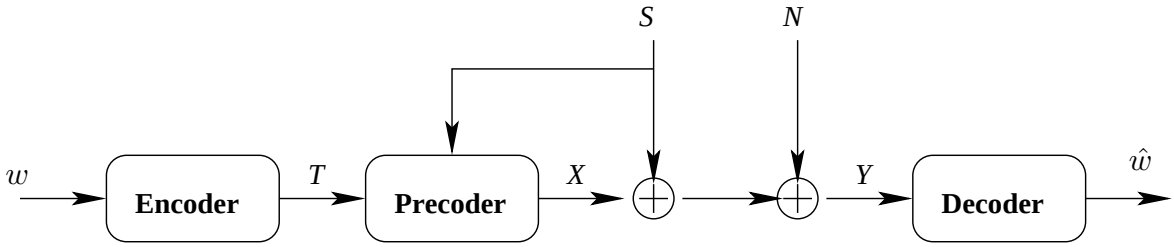

Figure 2.7: General structure of the communication system for channels with causally-known discrete interference.

2.5 Optimal Precoding

The general structure of a communication system for the channel defined in (2.7) is shown in fig. 2.7. In fact, fig. 2.7 is the same as fig. 1.2 for the special case of the state-dependent channel defined in (2.7). Any encoding and decoding scheme for the *associated* channel can be translated to an encoding and decoding scheme for the original channel defined in (2.7). A message w is encoded to a block of length n composed of input symbols of the *associated* channel $t \sim (x_{i_1}, x_{i_2}, \ldots, x_{i_Q})$. There are M^Q input symbols. However, we showed that the maximum rate with uniformity and integrality constraints can be achieved by using just M input symbols of the *associated* channel with equal probabilities. The optimal M input symbols of the *associated* channel are obtained by solving the linear programming problem (2.14) with the integrality constraint. Those M input symbols of the *associated* channel define the optimal precoding operation: For any t that belongs to the set of M optimal input symbols, the precoder sends the qth component of t if the current interference symbol is s_q, $q = 1, \ldots, Q$. Based on the received sequence, the receiver decodes $\hat{w}$ as the transmitted message.

2.6 Extension to Continuous Input Alphabet

We can extend the uniform transmission scheme introduced in section 2.4 to the case where the channel input alphabet $\mathcal{X}$ is continuous. For the continuous input alphabet case, we consider the maximization of the transmission rate $I(X_1 \cdots X_Q; Y)$ over joint pdfs $f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q)$ that induce uniform marginal distributions on $X_1, \ldots, X_Q$ in the interval $A_\Delta = \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$.

Since $h(Y)$ is the same for all joint pdfs $f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q)$ that induce uniform marginal pdfs on $X_1, \ldots, X_Q$, the maximization of the transmission rate reduces to the linear minimization problem

$$
\begin{aligned}
\min_{f_{X_1 \cdots X_Q}} \quad & \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} h(x_1, \ldots, x_Q) f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_1 \cdots dx_Q \\
\text{subject to} \quad & \\
& \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_2 \cdots dx_Q = \frac{1}{\Delta}, \qquad x_1 \in A_\Delta, \\
& \quad \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \qquad\quad \vdots \\
& \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_1 \cdots dx_{Q-1} = \frac{1}{\Delta}, \qquad x_Q \in A_\Delta, \\
& f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) \geq 0, \qquad\qquad\qquad\qquad x_1, \ldots, x_Q \in A_\Delta,
\end{aligned}
\tag{2.29}
$$

where $h(x_1, \ldots, x_Q) = h(Y | X_1 = x_1, \ldots, X_Q = x_Q)$. We are interested in solutions to (2.29) that are of the form

$$
f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) = \frac{1}{\Delta} \delta \left(|x_2 - \xi_1(x_1)| + |x_3 - \xi_2(x_1)| + \cdots + |x_Q - \xi_{Q-1}(x_1)| \right),
\tag{2.30}
$$

where $\delta(.)$ is the Dirac's delta function, $|.|$ denote absolute value, and $\xi_1, \xi_2, \ldots, \xi_{Q-1}$ are bijective functions from A_Δ to A_Δ.

The joint pdf in (2.30) describes random variables $X_1, \ldots, X_Q$, $Q-1$ of which are functions of the other random variable. Solutions of the form (2.30) can be considered as the continuous extension of solutions to (2.14) with the integrality constraint for the discrete input alphabet case. It is easy to check that (2.30), with the given condition that $\xi_1, \xi_2, \ldots, \xi_{Q-1}$ are bijective function from A_Δ to A_Δ, satisfies the constraints in (2.29). The objective value corresponding to the joint pdf (2.30) is

$$\frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} h\left(x_1, \xi_1(x_1), \ldots, \xi_{Q-1}(x_1)\right) dx_1, \tag{2.31}$$

which is to be minimized over bijective functions $\xi_1, \xi_2, \ldots, \xi_{Q-1}$.

2.6.1 Comparison to Modulo Precoding

The modulo precoding was originally proposed by Tomlinson and Harashima [48, 49] for the ISI channel. Then it was extended in [2] as a precoding method for channels with known (discrete or continuous) interference at the transmitter. The main idea is as follows. Based on the input symbol of the *associated* channel V and the current interference symbol S, the precoder sends [2]

$$X = [V - \alpha S] \mod \Delta, \tag{2.32}$$

where $\alpha = \frac{P_X}{P_X + P_N}$ (P_X is the power of X) and V is distributed uniformly in A_Δ.

In our setting where the interference is discrete with Q levels, (2.32) results in

$$X_q = [V - \alpha s_q] \mod \Delta, \qquad q = 1, \ldots, Q, \tag{2.33}$$

where X_q is the random variable that represents the channel input when the current interference symbol is s_q, $q = 1, \ldots, Q$. Since V is uniformly distributed in A_Δ,

$X_1, \ldots, X_Q$ will be uniformly distributed in A_Δ. Therefore, modulo precoding is indeed a uniform transmission scheme. We can remove V from the above equations and express $X_2, \ldots, X_Q$ in terms of X_1 as

$$X_q = [X_1 + \alpha(s_1 - s_q)] \mod \Delta, \qquad q = 2, \ldots, Q. \tag{2.34}$$

Since $X_2, \ldots, X_Q$ are functions of X_1, the joint pdf $f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q)$ corresponding to the modulo precoding fits in the category of joint pdfs in (2.30). The bijective functions corresponding to the modulo precoding are given by (2.34). These functions are circular shifts of each other.

The modulo precoding corresponds to a feasible solution to (2.29) which is not an optimal solution. For example, we may follow the line of proof of theorem 4 to show that for large P_N, where g becomes convex in the hyper-cube $\{(u_1, \ldots, u_{Q-1}) : -\Delta \leq u_i \leq \Delta, i = 1, \ldots, Q-1\}$, the optimal bijective functions are given by $\xi_1(x) = \cdots = \xi_{Q-1}(x) = x$, which are different from the functions given in (2.34).

To make the example more specific, consider a channel with $\mathcal{X} = A_\Delta = [-1, +1]$ and $\mathcal{S} = \{-\frac{1}{2}, +\frac{1}{2}\}$. According to (2.22), $g(u)$ will be convex if we choose $P_N = 3.363$. Then we will have $\alpha = \frac{P_X}{P_X + P_N} = \frac{0.333}{0.333 + 3.363} \approx 0.09$. Therefore, the bijective function corresponding to modulo precoding is given by

$$X_2 = [X_1 - 0.09] \mod 2, \tag{2.35}$$

while the optimal precoding corresponds to $X_2 = X_1$ in this example.

Chapter 3

Channel Code Design with Causal Side Information at the Encoder

In this chapter, we consider the problem of channel code design for the M-ary input AWGN channel with additive Q-ary interference where the i.i.d. sequence of interference symbols is known causally at the transmitter. In section 3.1, we introduce the channel model. In section 3.2, we derive the code design criterion at high SNR. In section 3.3, we consider channels with binary input for which we show that the design criterion derived in section 3.2 reduces to maximizing the Hamming distance. In section 3.4, we consider a special case for which the result for the binary channel also holds for the M-ary channel. In section 3.5, we consider a more general channel model for which the main results of this chapter hold.

3.1 The Channel Model

We consider data transmission over the channel

$$Y = X + S + N, \tag{3.1}$$

where X is the channel input, which takes on values in a real finite set $\mathcal{X}$, Y is the channel output, N is additive white Gaussian noise with power $P_N = \sigma^2$, and the interference S is a discrete random variable that takes on values in a real finite set $\mathcal{S}$. The sequence of i.i.d. interference symbols is known causally at the encoder.

The above channel can be considered as a special case of the state-dependent channel considered by Shannon with one exception, that the channel output alphabet is continuous. In our case, the likelihood function $f_{Y|X,S}(y|x,s)$ is used instead of the transition probabilities. We denote the input to the *associated* channel by T, which can be considered as a function from $\mathcal{S}$ to $\mathcal{X}$. We denote the cardinality of $\mathcal{X}$ and $\mathcal{S}$ by M and Q, respectively. Then the cardinality of $\mathcal{T}$ will be M^Q, which is the number of all functions from $\mathcal{S}$ to $\mathcal{X}$.

The likelihood function for the *associated* channel is given by

$$\begin{aligned} f_{Y|T}(y|t) &= \sum_{s\in\mathcal{S}} p(s) f_{Y|X,S}(y|t(s),s) \\ &= \sum_{s\in\mathcal{S}} p(s) f_N(y - t(s) - s), \end{aligned} \tag{3.2}$$

where $p(s)$ is the probability of the interference symbol s and f_N denotes the pdf of the Gaussian noise N.

Although in this work, we consider a fixed channel input alphabet $\mathcal{X}$, the transmitted power is not fixed in general. In fact, for probability distribution $p(s)$ on $\mathcal{S}$

and for a given coding scheme for the *associated* channel which induces probability distribution $p(t)$ on the symbols of $\mathcal{T}$, the transmitted power is given by

$$\begin{aligned} E[X^2] &= \sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}} p(t)p(s)E[X^2|t,s] \\ &= \sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}} p(t)p(s)t^2(s). \end{aligned} \tag{3.3}$$

Thus, in general, the transmitted power depends on the probability distribution on the interference alphabet. The binary-input channel with $\mathcal{X} = \{-x, x\}$ is an exception, however, for which we have $t^2(s) = x^2$ for all $s \in \mathcal{S}$. Therefore, for any coding scheme and any probability distribution on the interference alphabet, the transmitted power is equal to x^2.

In this work, we do not impose any constraint on the power of the transmitted signal. However, in the performance comparisons given in sections 3.3 and 3.4 for different scenarios, we will make sure that the transmitted power is the same in all scenarios.

3.2 The Code Design Criterion

Any coding scheme for the *associated* channel defined by (3.2) translates to a coding scheme for the actual channel defined by $f_{Y|X,S}(y|x,s)$. We use the pairwise error probability (PEP) approach to derive the code design criterion at high SNR. Since in this work, we consider fixed channel input and interference alphabets, the high SNR scenario is realized by making the noise power σ^2 sufficiently small. This is equivalent to scale up the transmitted signal and the interference by the same factor for a given fixed noise power.

Suppose that the messages w_1 and w_2 are encoded into $t_1^n \equiv t_1 t_2 \dots t_n$ and $r_1^n \equiv r_1 r_2 \dots r_n$, respectively, where t_i and r_i belong to the alphabet $\mathcal{T}$, $i = 1, \dots, n$. Using maximum likelihood decoding, the probability of the event that message w_2 is decoded given message w_1 was sent is given by

$$\begin{aligned}
\Pr\{w_1 \to w_2 | w_1\} &= \sum_{s_1^n} p(s_1^n) \Pr\{w_1 \to w_2 | w_1, s_1^n\} \\
&= \sum_{s_1^n} p(s_1^n) \Pr\left\{ f_{Y|T}(y_1^n | t_1^n) \le f_{Y|T}(y_1^n | r_1^n) | w_1, s_1^n \right\} \\
&= \sum_{s_1^n} p(s_1^n) \Pr\left\{ \prod_{i=1}^{n} f_{Y|T}(y_i | t_i) \le \prod_{i=1}^{n} f_{Y|T}(y_i | r_i) | w_1, s_1^n \right\} \\
&= \sum_{s_1^n} p(s_1^n) \Pr\left\{ \prod_{i=1}^{n} \sum_{s \in \mathcal{S}} p(s) f_N(y_i - t_i(s) - s) \le \right. \\
&\qquad \left. \prod_{i=1}^{n} \sum_{s \in \mathcal{S}} p(s) f_N(y_i - r_i(s) - s) | w_1, s_1^n \right\}. \quad (3.4)
\end{aligned}$$

where $s_1^n \equiv s_1 \cdots s_n \in \mathcal{S}^n$ represents the interference sequence during the transmission. In appendix C, we have shown that the above error probability at high SNR is given by

$$\Pr\{w_1 \to w_2 | w_1\} = O\left(Q\left(\frac{\sqrt{\sum_{i=1}^{n} d_{\mathsf{SI}}^2(t_i, r_i)}}{2\sigma} \right) \right), \quad (3.5)$$

where

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2} \right) dy, \quad (3.6)$$

and $d_{\mathsf{SI}}(t, r)$ (SI stands for side information), the distance between two input symbols of the *associated* channel t and r, is defined as

$$d_{\mathsf{SI}}(t, r) = \min_{s_1, s_2 \in \mathcal{S}} |t(s_1) + s_1 - r(s_2) - s_2|. \quad (3.7)$$

According to (3.5), at high SNR, the code design criterion is to maximize the minimum distance between the codewords with the distance measure defined in

(3.7).

In order to see how the knowledge of interference at the encoder can result in larger distances between codewords, consider the channel model introduced in section 3.1 with the exception that the interference sequence is not known at the encoder. In this case, the discrete interference is considered as noise. In order to obtain the PEP for this channel, suppose that messages v_1 and v_2 are encoded into $x_1^n \equiv x_1 \cdots x_n \in \mathcal{X}^n$ and $z_1^n \equiv z_1 \cdots z_n \in \mathcal{X}^n$, respectively. Similarly, it can be shown that the PEP at high SNR is given by

$$\Pr\{v_1 \to v_2 | v_1\} = O\left(Q\left(\frac{\sqrt{\sum_{i=1}^{n} d^2(x_i, z_i)}}{2\sigma}\right)\right), \tag{3.8}$$

where $d(x, z)$, the distance between two symbols x and z of $\mathcal{X}$ is defined as

$$d(x, z) = \min_{s_1, s_2 \in \mathcal{S}} |x + s_1 - z - s_2|. \tag{3.9}$$

Comparing (3.7) and (3.9), it becomes clear that larger distances among codewords are possible for the channel with side information at the encoder. In fact, the distance $d(x, z)$ is equal to $d_{\mathsf{SI}}(t, r)$ for $t = (x, \ldots, x)$ and $r = (z, \ldots, z)$. However, $\mathcal{T}$ has many other symbols, which may yield larger distances. For example, consider the channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$. For the case without side information at the encoder, we can compute the distances between symbols of $\mathcal{X}$ according to (3.9) as $d(1, 1) = d(-1, -1) = d(1, -1) = 0$. Hence, according to (3.8), it is impossible to transmit data over this channel with low error probability even at high SNR. For the case with side information at the encoder, the four symbols of the *associated* channel can be represented as $u_1 = (-1, +1), u_2 = (+1, -1), u_3 = (+1, +1), u_4 = (-1, -1)$. Using (3.7), it is easy to check that the distances between all pairs of the symbols are zero except for $d_{\mathsf{SI}}(u_1, u_2)$ which is 2. As will be seen in section 3.3, u_1 and u_2

can be used in the encoding to achieve arbitrarily low error probabilities as SNR increases.

It is worth mentioning that the distance measures defined in (3.7) or (3.9) do not satisfy the triangle inequality. For example, again consider the channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$. The distances between all pairs of the input symbols of the *associated* channel are zero except for $d_{\mathsf{SI}}(u_1, u_2)$ which is 2. Therefore, the triangle inequality does not hold for $d_{\mathsf{SI}}(u_1, u_3)$, $d_{\mathsf{SI}}(u_3, u_2)$, and $d_{\mathsf{SI}}(u_1, u_2)$.

3.3 The Binary Channel

We call the channel introduced in (3.1) a *binary channel* when the channel accepts binary input, i.e., $M = 2$. There is no constraints on the cardinality of the interference alphabet. For the binary channel, the size of $\mathcal{T}$ is 2^Q. However, we may not need to use all the symbols of the alphabet in the encoding. In this section, we show that it is sufficient to use only two symbols of $\mathcal{T}$ in the encoding as far as the distance spectrum of the code is concerned. We begin with the following lemma for the binary channel.

Lemma 1. *For the binary channel, there exist at least two symbols in $\mathcal{T}$ with nonzero distance.*

Proof. We may explicitly denote the channel input and interference alphabets by $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{S} = \{s_1, \ldots, s_Q\}$, where $x_1 < x_2$ and $s_1 < s_2 < \cdots < s_Q$. From the definition of distance in (3.7), it is sufficient to show that there exist two elements t and r in $\mathcal{T}$ such that the corresponding multi-sets (of size Q) $\{t(s_1) +$

$s_1, \ldots, t(s_Q) + s_Q\}$ and $\{r(s_1) + s_1, \ldots, r(s_Q) + s_Q\}$ are disjoint. We prove this by induction on Q.

The statement of the lemma holds for $Q = 1$ since we may take $t = (x_1)$ and $r = (x_2)$. Then the sets $\{x_1 + s_1\}$ and $\{x_2 + s_1\}$ are disjoint. Now suppose that the statement of the lemma is true for some Q. Therefore, there exist two Q-tuples composed of elements of $\mathcal{X}$ (two input symbols of the *associated* channel) such that the corresponding multi-sets are disjoint. We prove that the statement of the lemma hold for $Q + 1$.

The element $x_2 + s_{Q+1}$ is larger than any element of the two multi-sets (of size Q). Hence, it does not belong to any of the multi-sets. If $x_1 + s_{Q+1}$ does not belong to any of the multi-sets too, then we can include the new elements $x_1 + s_{Q+1}$ and $x_2 + s_{Q+1}$ in the multi-sets of size Q arbitrarily (one elements in each multi-set). The resulting multi-sets of size $Q + 1$ will be disjoint. If $x_1 + s_{Q+1}$ belongs to one of the multi-set of size Q, we include it in that multi-set and include $x_2 + s_{Q+1}$ in the other multi-set to form the new disjoint multi-sets of size $Q + 1$. The two $(Q + 1)$-tuples (the two input symbols of the *associated* channel) are then obtained from the two multi-sets of size $Q + 1$ by subtracting the interference symbols from their elements. □

Lemma 1 is in fact a special case of theorem 2, which was stated in the context of capacity.

Let u_1 and u_2 be two input symbols of the *associated* channel with the maximum distance among all pairs of input symbols of the *associated* channel. Since $d_{\mathsf{SI}}(u_1, u_2) > 0$ (according to Lemma 1), we have $u_1(s) \neq u_2(s), \forall s \in \mathcal{S}$, otherwise,

from (3.7), $d_{\mathsf{SI}}(u_1, u_2) = 0$. We choose an arbitrary interference symbol $s \in \mathcal{S}$ to partition $\mathcal{T}$ as follows. We put $t \in \mathcal{T}$ in $\mathcal{T}_1$ if $t(s) = u_1(s)$, otherwise (i.e., $t(s) = u_2(s)$) we put t in $\mathcal{T}_2$. Note that the distance between any two symbols in $\mathcal{T}_j$ is zero, $j = 1, 2$.

Suppose that a codebook is designed for the binary channel with codewords composed of elements of $\mathcal{T}$. We construct a new codebook from the original one by replacing the elements of the codewords that belong to $\mathcal{T}_1$ by u_1 and replacing the elements of the codewords that belong to $\mathcal{T}_2$ by u_2. Since the codewords of the new codebook are composed of just two elements, we may call the new code a binary code.

Theorem 6. *The distance spectrum of the binary code constructed by the procedure described above is at least as good as the distance spectrum of the original code.*

Proof. Consider any two codewords $(t_1, \ldots, t_n)$ and $(r_1, \ldots, r_n)$ from the original codebook, where $t_i, r_i \in \mathcal{T}$. The squared distance between the two codewords is equal to $\sum_{i=1}^{n} d^2_{\mathsf{SI}}(t_i, r_i)$. For any $i \in \{1, 2, \ldots, n\}$, we consider two cases:

Case 1: t_i and r_i belong to the same partition. Then $d_{\mathsf{SI}}(t_i, r_i) = 0$, so the replacement will not change the distance.

Case 2: t_i and r_i belong to different partitions. Then since $d_{\mathsf{SI}}(t_i, r_i) \leq d_{\mathsf{SI}}(u_1, u_2)$, the replacement will not decrease the distance. □

According to theorem 6, as far as the distance spectrum of the code in concerned, it is sufficient to use two symbols of $\mathcal{T}$ with the maximum distance, namely u_1 and u_2, in the encoding for a binary channel. Since $\mathcal{T}$ has size 2^Q for the binary channel,

a brute-force search for finding two symbols in $\mathcal{T}$ with the maximum distance will have exponential complexity with respect to Q. We have proposed an algorithm with polynomial complexity for finding two symbols with the maximum distance in appendix D.

Since it is sufficient to use u_1 and u_2 in the encoding for the binary channel, we can define the Hamming distance between any two codewords, which is the number of positions at which the two codewords are different. Consider two codewords $c_1 = (t_1, \ldots, t_n)$ and $c_2 = (r_1, \ldots, r_n)$ with elements from the binary set $\{u_1, u_2\}$. The squared distance between these codewords is given by

$$\sum_{i=1}^{n} d^2_{\mathsf{SI}}(t_i, r_i) = d^2_{\mathsf{SI}}(u_1, u_2) d_H(c_1, c_2), \tag{3.10}$$

where $d_H(c_1, c_2)$ is the Hamming distance between c_1 and c_2. Therefore, the problem of designing codes for the binary channel where the interference sequence is known causally at the encoder reduces to the design of codes for the binary symmetric channel. The only difference is that the coding is over the set $\{u_1, u_2\}$ rather than $\{0, 1\}$.

3.3.1 Comparison with the Interference-Free Channel

If we were to use a binary code for the interference-free binary channel with the input alphabet $\mathcal{X} = \{x_1, x_2\}$, then the Euclidean distance between any two codewords c_1 and c_2 of length n for the interference-free channel would be

$$d^2_E(c_1, c_2) = (x_1 - x_2)^2 d_H(c_1, c_2), \tag{3.11}$$

where d_E denotes the Euclidean distance.

Using (3.10) and (3.11), we can compare the performance of a zero-one binary code for the binary channel with causal side information at the encoder with the same zero-one binary code for the interference-free binary channel. In the case of channel with side information, zero and one are mapped to u_1 and u_2, and in the case of the interference-free channel, zero and one are mapped to x_1 and x_2, respectively. Note that u_1 and u_2 are functions from the interference alphabet $\mathcal{S}$ to the channel input alphabet $\mathcal{X} = \{x_1, x_2\}$.

It is clear from (3.7) that

$$d_{\mathsf{SI}}(u_1, u_2) \leq |x_1 - x_2|. \tag{3.12}$$

Therefore, using (3.10) and (3.11), the distance spectrum of the code for the interference-free channel is at least as good as the distance-spectrum of the code for the channel with known interference at the encoder. Of course, this is not surprising. However, it is interesting to search for the conditions that (3.12) is satisfied with equality.

If (3.12) is satisfied with equality, the distance spectrum of the two codes will be the same. In other words, if (3.12) is satisfied with equality, the knowledge of interference at the encoder enables us to achieve the same performance (in terms of order of probability of error) as the interference-free case at high SNR.

We may explicitly denote the interference alphabet by $\mathcal{S} = \{s_1, \ldots, s_Q\}$, where $s_1 < s_2 < \cdots < s_Q$. Then the following theorem holds.

Theorem 7. *$d_{\mathsf{SI}}(u_1, u_2) = |x_1 - x_2|$ if and only if*

$$\min_{i \neq j} |s_i - s_j| \geq |x_1 - x_2|.$$

Proof. If $\min|s_i - s_j| \geq |x_1 - x_2|$, we may take $u_1 = (x_1, x_2, x_1, \ldots)$ and $u_2 = (x_2, x_1, x_2, \ldots)$. Then we have

$$\begin{aligned} d_{\mathsf{SI}}(u_1, u_2) &= \min_{i,j} |u_1(s_i) + s_i - u_2(s_j) - s_j| \\ &= \min\{|x_1 + s_k - x_2 - s_k|, |x_1 + s_{2k_1+1} - x_2 - s_{2k_2+1}|_{k_1 \neq k_2} \\ &\qquad |x_1 + s_{2k_1+1} - x_1 - s_{2k_2}|_{k_1,k_2}, |x_2 + s_{2k_1} - x_2 - s_{2k_2+1}|_{k_1,k_2}\} \\ &= \min\{|x_1 - x_2|, |x_1 + s_{2k_1+1} - x_2 - s_{2k_2+1}|_{k_1 \neq k_2}, |s_{2k_1+1} - s_{2k_2}|_{k_1,k_2}\}. \end{aligned} \tag{3.13}$$

We also have

$$\begin{aligned} |x_1 + s_{2k_1+1} - x_2 - s_{2k_2+1}| &\geq |s_{2k_1+1} - s_{2k_2+1}| - |x_1 - x_2| \\ &\geq 2\min|s_i - s_j| - |x_1 - x_2| \qquad \text{for } k_1 \neq k_2 \\ &\geq |x_1 - x_2| \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} |s_{2k_1+1} - s_{2k_2}| &\geq \min|s_i - s_j| \quad \forall\, k_1, k_2 \\ &\geq |x_1 - x_2|. \end{aligned} \tag{3.15}$$

Therefore, $d_{\mathsf{SI}}(u_1, u_2) = |x_1 - x_2|$.

For the other direction, suppose that $\min|s_i - s_j| < |x_1 - x_2|$. We will show that $d_{\mathsf{SI}}(u_1, u_2) < |x_1 - x_2|$. Suppose that $s_k, s_{k+1} \in \mathcal{S}$ achieve the minimum of $|s_i - s_j|$ and t_1 and t_2 are arbitrary elements of $\mathcal{T}$. We consider two non-trivial cases:

Case 1: $t_1(s_k) = t_1(s_{k+1}) = x_1$ and $t_2(s_k) = t_2(s_{k+1}) = x_2$. Then $d_{\mathsf{SI}}(t_1, t_2) \leq |t_1(s_{k+1}) + s_{k+1} - t_2(s_k) - s_k| < |x_1 - x_2|$.

Case 2: $t_1(s_k) = x_1, t_1(s_{k+1}) = x_2$ and $t_2(s_k) = x_2, t_2(s_{k+1}) = x_1$. Then $d_{\mathsf{SI}}(t_1, t_2) \leq |t_1(s_k) + s_k - t_2(s_{k+1}) - s_{k+1}| < |x_1 - x_2|$. □

As an example, consider a binary channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$ and equiprobable interference symbols. The two symbols with the maximum distance in the input alphabet of the *associated* channel are $u_1 = (-1, +1), u_2 = (+1, -1)$. We have simulated the error probability performance of the above uncoded system with maximum likelihood decoding. The error probability vs. SNR $\left(= \frac{1}{\sigma^2}\right)$ for the above channel is plotted in fig. 3.1. The error probability curve for the interference-free channel with $\mathcal{X} = \{-1, +1\}$ is plotted for comparison. For the interference-free channel, $P_e = Q(\frac{1}{\sigma})$. It is easy to check that in this example, $d_{\mathsf{SI}}(u_1, u_2) = |x_1 - x_2| = 2$. As it can be seen, the error probability curves decay at the same rate with increasing SNR as expected. The error probability curve for the scenario that the interference is not known at the encoder, is plotted for comparison. In this scenario, the error probability curve reaches an error floor of $\frac{1}{4}$.

Another example is illustrated in fig. 3.2. For this example, $\mathcal{X} = \{-1, +1\}, \mathcal{S} = \{-1, 0, +1\}$. We can find by inspection two symbols of the *associated* channel input alphabet with the maximum distance as $u_1 = (-1, -1, +1), u_2 = (+1, +1, -1)$. Here, we have $d_{\mathsf{SI}}(u_1, u_2) = 1 < |x_1 - x_2| = 2$. Therefore, the error probability curve for the channel with known interference at the encoder does not decay as fast as the error probability curve for the interference-free channel. For the scenario that the interference is not known at the encoder, the error probability curve reaches an error floor of $\frac{1}{6}$.

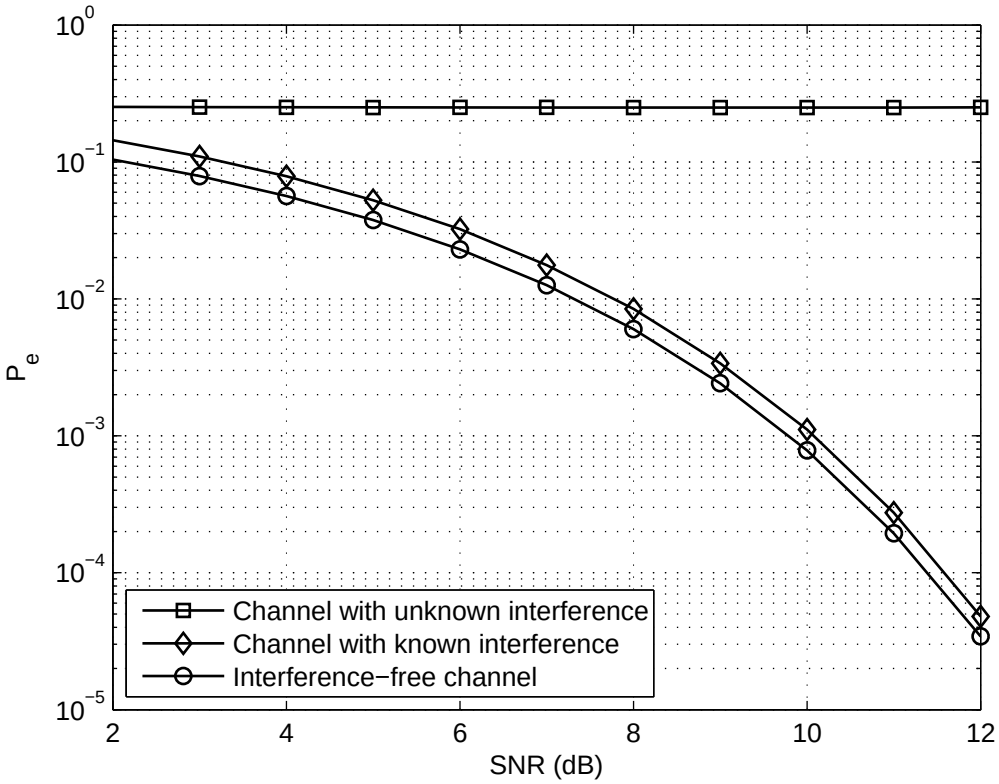

Figure 3.1: Error probability vs. SNR for the binary input AWGN channel with/without known/unknown interference. $\mathcal{X} = \mathcal{S} = \{-1, +1\}$.

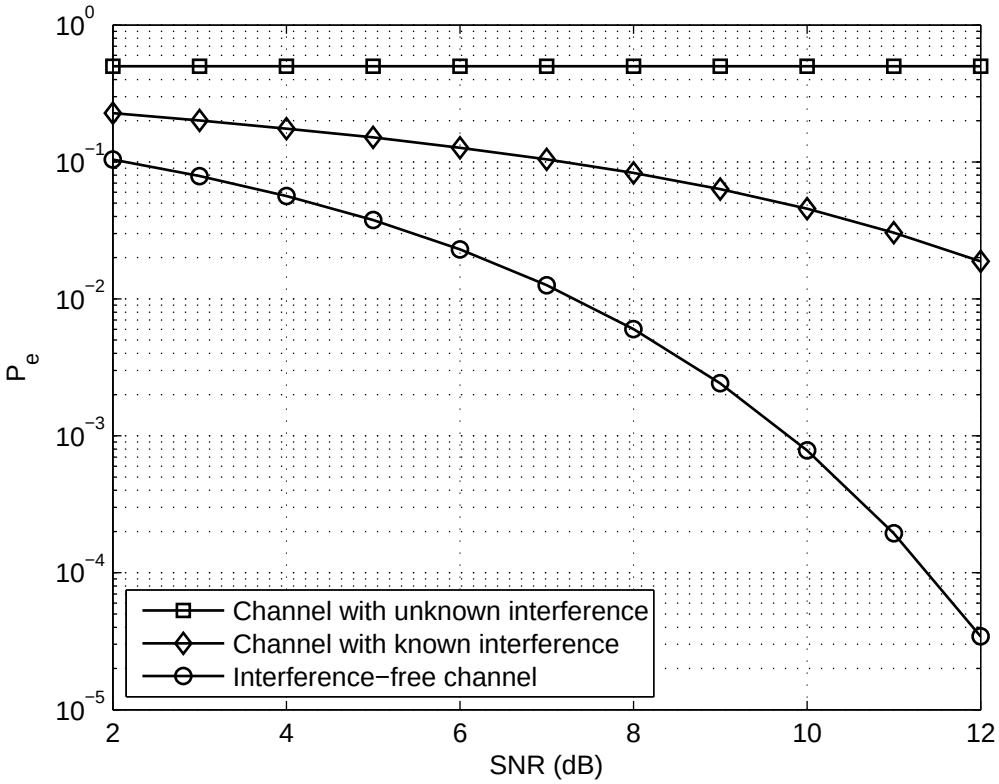

Figure 3.2: Error probability vs. SNR for the binary input AWGN channel with/without known/unknown interference. $\mathcal{X} = \{-1, +1\}$, $\mathcal{S} = \{-1, 0, +1\}$.

3.4 The M-ary Channel

In general, the statement of theorem 6 is not extendable to the case with $M > 2$ channel input symbols. In fact, by using more than M input symbols of the *associated* channel, we can obtain a better codebook in terms of distance spectrum than any other codebook composed of just M input symbols of the *associated* channel. An example showing this is given in appendix E. However, under some condition on the channel input and interference alphabets, the statement of theorem 6 can be generalized to the case with $M > 2$.

Theorem 8. *As far as the distance spectrum of code is concerned, it is sufficient to use M (out of M^Q) input symbols of the* associated *channel in the encoding if*

$$\min_{s_i, s_j \in \mathcal{S}} |s_i - s_j| \geq 2 \max_{x_i, x_j \in \mathcal{X}} |x_i - x_j|.$$

Proof. Consider the M input symbols of the *associated* channel $u_1 = (x_1, \ldots, x_1)$, $u_2 = (x_2, \ldots, x_2)$, ..., $u_M = (x_M, \ldots, x_M)$. We use these symbols to partition the *associated* channel input alphabet $\mathcal{T}$ as follows. Put $t \in \mathcal{T}$ in $\mathcal{T}_i$ if the first element of t is x_i, $i = 1, 2, \ldots, M$. Note that $\mathcal{T}_i$ has size M^{Q-1} and the distance between any two symbols in $\mathcal{T}_i$ is zero, $i = 1, 2, \ldots, M$. For any $p, q = 1, \ldots, M$, we have

$$\begin{aligned} d_{\mathsf{SI}}(u_p, u_q) &= \min_{k_1, k_2} |x_p + s_{k_1} - x_q - s_{k_2}| \\ &= \min \left\{ |x_p - x_q|, |x_p + s_{k_1} - x_q - s_{k_2}|_{k_1 \neq k_2} \right\}. \end{aligned} \tag{3.16}$$

We also have

$$\begin{aligned} |x_p + s_{k_1} - x_q - s_{k_2}| &\geq |s_{k_1} - s_{k_2}| - |x_p - x_q| \\ &\geq 2 \max |x_i - x_j| - |x_p - x_q| \quad \text{for } k_1 \neq k_2 \\ &\geq |x_p - x_q|, \end{aligned} \tag{3.17}$$

Therefore, $d_{\mathsf{SI}}(u_p, u_q) = |x_p - x_q|$. Note that the distance between any two symbols from $\mathcal{T}_p$ and $\mathcal{T}_q$ is at most $|x_p - x_q| = d_{\mathsf{SI}}(u_p, u_q)$.

Suppose that a codebook is designed with codewords composed of possibly all elements of $\mathcal{T}$. We construct a new codebook from the original one by replacing the elements of the codewords that belong to $\mathcal{T}_i$ by u_i, $i = 1, 2, \ldots, M$. It is easy to check that the distance spectrum of the new code is at least as good as the distance spectrum of the original code. □

According to theorem 8, it is sufficient to use only the symbols $u_1, \ldots, u_M$ in the encoding. But any of these symbols is a constant function from $\mathcal{S}$ to $\mathcal{X}$. Therefore, the same symbol enters the channel regardless of the current interference symbol. This suggests that the knowledge of interference symbols at the encoder is not helpful in terms of distance spectrum improvement provided that the condition of theorem 8 is satisfied. In fact, with the condition of theorem 8, we have

$$d_{\mathsf{SI}}(u_i, u_j) = d(x_i, x_j) = d_E(x_i, x_j), \qquad i, j = 1, \ldots, M. \tag{3.18}$$

where $d(.,.)$, defined in (3.9), is the distance measure when the interference is not known at the encoder and $d_E(.,.)$ is the Euclidean distance measure. Therefore, the error probability performance of a code for the channel with known/unknown interference at the encoder will be the same as the performance of the same code for the interference-free channel at high SNR.

It is worth mentioning that for the above-mentioned three scenarios the codes for the interference-free channel, the channel with known interference at the encoder, and the channel with unknown interference use the same transmitted power.

3.5 A More General Channel Model

Although we have considered the AWGN channel with additive interference so far, our treatment applies to more general channels characterized by

$$Y = f(X, S) + N, \tag{3.19}$$

where f is an arbitrary function of two variables, S is the channel state which is known causally at the encoder, X is the channel input, and N is white Gaussian noise. Another special case of this more general channel is the fast fading channel

$$Y = SX + N, \tag{3.20}$$

where S is the fading coefficient. For the general channel model (3.19), the distance between two symbols t and r of $\mathcal{T}$ is defined as

$$d_{\mathsf{SI}}(t, r) = \min_{s_1, s_2 \in \mathcal{S}} |f(t(s_1), s_1) - f(t(s_2), s_2)|. \tag{3.21}$$

Theorem 6 on the binary channel also holds for the general channel model. However, the maximum distance among pairs of symbols of $\mathcal{T}$ may be zero; i.e., lemma 1 does not hold true in general. Theorems 7 and 8 do not hold for the more general channel model in (3.19) and are specific to the AWGN with additive interference channel model.

Chapter 4

Conclusion

In chapter 2, we investigated M-ary signal transmission over AWGN channel with additive Q-level interference, where the sequence of i.i.d. interference symbols is known causally at the transmitter. According to Shannon's theorem for channels with side information at the transmitter, the capacity of our channel is the same as the capacity of an associated regular (without state) channel with M^Q input symbols. We proved that by using at most $MQ-Q+1$ (out of M^Q) input symbols the capacity is achievable.

For the noise-free channel, provided that the signal points are equally spaced, we proposed a one-shot coding scheme that uses M input symbols of the *associated* channel to achieves the capacity $\log_2 M$ bits regardless of the interference.

We considered the maximization of the transmission rate with the constraint that $X_1, \ldots, X_Q$ are uniformly distributed over the channel input alphabet. For this so called uniform transmission, the optimal input probability assignment (again with at most $MQ-Q+1$ nonzero elements) can be obtained by solving the linear

optimization problem (2.14). The optimal solution to (2.14) with the integrality constraint has exactly M nonzero elements. For the case $Q = 2$, we showed that the integrality constraint does not reduce the maximum achievable rate. The loss in rate (if there is any) by imposing the integrality constraint for the general case is a problem to be explored.

We extended the uniform transmission scheme to continuous channel input alphabet. We showed that modulo precoding is a uniform transmission scheme, which enabled us to compare the performances of our precoding scheme and modulo precoding.

In chapter 3, we derived the code design criterion at high SNR for our channel model. The code design is over an input alphabet $\mathcal{T}$ of size M^Q. We defined a new distance between the elements of $\mathcal{T}$. The performance of a code for our channel at high SNR is governed by the minimum distance between the codewords with elements from $\mathcal{T}$. We may not need to use all symbols of $\mathcal{T}$ in the encoding. In particular, we showed that for the case $M = 2$, as far as the distance spectrum of the code is concerned, we just need to use two symbols of $\mathcal{T}$ with the maximum distance among all pairs of symbols. This reduces the code design problem for our channel to code design for binary symmetric channel which has been well researched in the literature.

We showed that for the binary channel it is possible to obtain the same performance (in terms of the order of error probability) as the interference-free channel at high SNR if the minimum spacing between interference symbols is not less than the spacing of the channel input alphabet symbols.

For the general M-ary channel, we proved that if the minimum spacing between

interference symbols is larger than twice the maximum spacing between channel input symbols, then it is sufficient to use M (out of M^Q) symbols of $\mathcal{T}$ in the encoding as far as the distance spectrum of code is concerned. The problem without that constraint on the channel input and interference alphabets can be investigated in the future.

Appendix A

Bounds for

$h(Y|X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q})$

Denote by $\tilde{S}$ the random variable that takes on $x_{i_1} + s_1, x_{i_2} + s_2, \ldots, x_{i_Q} + s_Q$ with probabilities $r_1, r_2, \ldots, r_Q$, respectively. Also, denote by $\tilde{Y}$ the random variable $Y|X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q}$. Then

$$\tilde{Y} = \tilde{S} + N. \tag{A.1}$$

Since

$$0 \leq I(\tilde{Y}; \tilde{S}) \leq H(\tilde{S}), \tag{A.2}$$

we have

$$0 \leq h(\tilde{Y}) - h(\tilde{Y}|\tilde{S}) \leq H(\tilde{S}), \tag{A.3}$$

or equivalently,

$$\begin{aligned} h(N) \leq h(\tilde{Y}) &\leq h(N) + H(\tilde{S}) \\ &= h(N) + H(S). \end{aligned} \tag{A.4}$$

Appendix B

Necessary and Sufficient Conditions for the Convexity/Concavity of g

The function g given in (2.17) for the case $Q = 2$ can be considered as a function of u and parameters s_1, s_2, P_N as

$$\begin{aligned} g(u) &= g(u, s_1, s_2, P_N) \\ &= g(u + s_1 - s_2, 0, 0, P_N) \\ &= g\left(\frac{u + s_1 - s_2}{\sqrt{P_N}}, 0, 0, 1\right) + \log_2 \sqrt{P_N}. \end{aligned} \tag{B.1}$$

Denote by u_0 and $-u_0$ the inflection points of $g(u, 0, 0, 1)$. We can obtain u_0 numerically as $u_0 \approx 1.636$. Then the inflection points of $g(u)$ are

$$\alpha_1 = s_2 - s_1 - u_0\sqrt{P_N}, \tag{B.2}$$

$$\alpha_2 = s_2 - s_1 + u_0\sqrt{P_N}, \tag{B.3}$$

The function g is convex in the interval $[\alpha_1, \alpha_2]$ and is concave anywhere else.

The function g is convex in the interval $[x_1 - x_M, x_M - x_1]$ if and only if $[x_1 - x_M, x_M - x_1] \subseteq [\alpha_1, \alpha_2]$. This gives (2.22).

The function g is concave in the interval $[x_1 - x_M, x_M - x_1]$ if and only if $[x_1 - x_M, x_M - x_1] \subseteq (-\infty, \alpha_1]$ or $[x_1 - x_M, x_M - x_1] \subseteq [\alpha_2, \infty)$. This gives (2.28).

Appendix C

Derivation of Code Design Criterion at High SNR

Define

$$\mathcal{A}_i = \{t_i(s) + s : s \in \mathcal{S}\}, \qquad i = 1, \ldots, n, \tag{C.1}$$

$$\mathcal{B}_i = \{r_i(s) + s : s \in \mathcal{S}\}, \qquad i = 1, \ldots, n. \tag{C.2}$$

It is worth mentioning that the cardinality of $\mathcal{A}_i$ (or $\mathcal{B}_i$) can be less than Q, $i = 1, \ldots, n$, since different interference symbols may yield the same element in $\mathcal{A}_i$ (or $\mathcal{B}_i$). For any $i = 1, \ldots, n$, we have

$$\sum_{s \in \mathcal{S}} p(s) f_N(y - t_i(s) - s) = \sum_{a \in \mathcal{A}_i} p(a) f_N(y - a), \tag{C.3}$$

$$\sum_{s \in \mathcal{S}} p(s) f_N(y - r_i(s) - s) = \sum_{b \in \mathcal{B}_i} p(b) f_N(y - b), \tag{C.4}$$

where $p(a)$ and $p(b)$ are obtained from $p(s)$ according to

$$p(a) = \sum_{s\in\mathcal{S}:t_i(s)+s=a} p(s), \tag{C.5}$$

$$p(b) = \sum_{s\in\mathcal{S}:r_i(s)+s=b} p(s). \tag{C.6}$$

For any sequence $a_1^n \equiv a_1\cdots a_n \in \mathcal{A}_1\times\cdots\times\mathcal{A}_n$ and $b_1^n \equiv b_1\cdots b_n \in \mathcal{B}_1\times\cdots\times\mathcal{B}_n$, we define the events

$$E_1(a_1^n) = \bigcap_{i=1}^{n}\left(a_i = \arg\min_{a\in\mathcal{A}_i}|y_i-a|\right), \tag{C.7}$$

$$E_2(b_1^n) = \bigcap_{i=1}^{n}\left(b_i = \arg\min_{b\in\mathcal{B}_i}|y_i-b|\right), \tag{C.8}$$

given that w_1 has been sent and the interference sequence s_1^n has occurred. The event $E_1(a_1^n)$ simply means that a_i is the closest point to the received signal y_i (given w_1 has been sent and the interference sequence s_1^n has occurred) among all points of $\mathcal{A}_i$ for all $i = 1,\ldots,n$.

Any term in the error probability in (3.4) can be written as

$$\begin{aligned}
&\Pr\left\{\prod_{i=1}^{n}\sum_{a\in\mathcal{A}_i}p(a)f_N(y_i-a) \le \prod_{i=1}^{n}\sum_{b\in\mathcal{B}_i}p(b)f_N(y_i-b)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\prod_{i=1}^{n}\sum_{a\in\mathcal{A}_i}p(a)f_N(y_i-a) \le \prod_{i=1}^{n}\sum_{b\in\mathcal{B}_i}p(b)f_N(y_i-b), E_1(a_1^n), E_2(b_1^n)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\prod_{i=1}^{n}f_N(y_i-a_i)\left(p(a_i)+\sum_{\substack{a\in\mathcal{A}_i\\ a\neq a_i}}p(a)\frac{f_N(y_i-a)}{f_N(y_i-a_i)}\right)\right.\\
&\qquad\left.\le \prod_{i=1}^{n}f_N(y_i-b_i)\left(p(b_i)+\sum_{\substack{b\in\mathcal{B}_i\\ b\neq b_i}}p(b)\frac{f_N(y_i-b)}{f_N(y_i-b_i)}\right), E_1(a_1^n), E_2(b_1^n)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2 \ge \sum_{i=1}^{n}(y_i-b_i)^2 + K\sigma^2, E_1(a_1^n), E_2(b_1^n)|w_1,s_1^n\right\},
\end{aligned} \tag{C.9}$$

where $K = K(y_1^n, a_1^n, b_1^n)$ is given by

$$K(y_1^n, a_1^n, b_1^n) = 2\sum_{i=1}^{n} \log \frac{p(a_i) + \sum_{\substack{a\in\mathcal{A}_i \\ a\neq a_i}} p(a)\frac{f_N(y_i-a)}{f_N(y_i-a_i)}}{p(b_i) + \sum_{\substack{b\in\mathcal{B}_i \\ b\neq b_i}} p(b)\frac{f_N(y_i-b)}{f_N(y_i-b_i)}}. \tag{C.10}$$

Given the events $E_1(a_1^n)$ and $E_2(b_1^n)$, it is easy to check that $K(y_1^n, a_1^n, b_1^n)$ is bounded as

$$K_1(a_1^n) = 2\sum_{i=1}^{n} \log p(a_i) < K(y_1^n, a_1^n, b_1^n) < K_2(b_1^n) = 2\sum_{i=1}^{n} \log \frac{1}{p(b_i)}. \tag{C.11}$$

As we consider the high SNR regime, we may assume that the noise power is sufficiently small so that the error probability (3.4) can be well approximated by

$$\sum_{s_1^n} p(s_1^n) \sum_{a_1^n} \sum_{b_1^n} \Pr\left\{ \sum_{i=1}^{n} (y_i - a_i)^2 \geq \sum_{i=1}^{n} (y_i - b_i)^2, E_1(a_1^n), E_2(b_1^n) | w_1, s_1^n \right\}. \tag{C.12}$$

Any term in the summation (C.12) can be upper bounded as

$$\begin{aligned}
& \Pr\left\{ \sum_{i=1}^{n} (y_i - a_i)^2 \geq \sum_{i=1}^{n} (y_i - b_i)^2, E_1(a_1^n), E_2(b_1^n) | w_1, s_1^n \right\} \\
\leq\ & \Pr\left\{ \sum_{i=1}^{n} (y_i - c_i)^2 \geq \sum_{i=1}^{n} (y_i - b_i)^2, E_1(a_1^n), E_2(b_1^n) | w_1, s_1^n \right\} \\
\leq\ & \Pr\left\{ \sum_{i=1}^{n} (y_i - c_i)^2 \geq \sum_{i=1}^{n} (y_i - b_i)^2 | w_1, s_1^n \right\} \\
=\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n} |c_i - b_i|^2}}{2\sigma} \right) \\
\leq\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n} d_{\mathsf{SI}}^2(t_i, r_i)}}{2\sigma} \right),
\end{aligned} \tag{C.13}$$

where

$$c_i = t_i(s_i) + s_i, \quad i = 1, \ldots, n. \tag{C.14}$$

The first inequality is due to the fact that given $E_1(a_1^n)$, we have $|y_i - a_i| \leq |y_i - c_i|, i = 1, \ldots, n$.

In the following, we show that the upper bound (C.13) is tight for the term(s) in the summation (C.12) satisfying

$$\{a_i, b_i\} = \arg \min_{\substack{a \in \mathcal{A}_i \\ b \in \mathcal{B}_i}} |a - b|, \quad i = 1, \ldots, n, \tag{C.15}$$

and

$$a_i = c_i, \quad i = 1, \ldots, n. \tag{C.16}$$

Any term in (C.12) equals the integral of the joint probability distribution of $y_1^n \equiv y_1 \cdots y_n$ (given w_1, s_1^n) over the region in the n-dimensional Euclidean space defined by

$$\left\{ y_1^n : \sum_{i=1}^{n} (y_i - a_i)^2 \geq \sum_{i=1}^{n} (y_i - b_i)^2, E_1(a_1^n), E_2(b_1^n) \right\}. \tag{C.17}$$

This region is illustrated by the shaded area ABCD in fig. C.1 for $n = 2$. The horizontal and vertical boundaries of ABCD correspond to the events $E_1(a_1^2)$ and $E_2(b_1^2)$. The elements of $\mathcal{A}_i$ and $\mathcal{B}_i$ are shown by $\circ$ and $\times$, respectively. The other boundary of ABCD which corresponds to $\sum_{i=1}^{2}(y_i - a_i)^2 \geq \sum_{i=1}^{2}(y_i - b_i)^2$ is the perpendicular bisector of the line segment connecting a_1^2 to b_1^2. We may consider an n-cube inside this region with sides equal to some $\delta > 0$ as shown in fig. C.1 and perform the integration over this smaller region to obtain a lower bound for the term(s) in the summation (C.12) satisfying (C.15) and (C.16).

In summary, for the terms in (C.12) which satisfy (C.15) and (C.16), we have

$$\begin{aligned}
&\Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2 \geq \sum_{i=1}^{n}(y_i-b_i)^2, E_1(a_1^n), E_2(b_1^n)|w_1, s_1^n\right\} \\
\geq\ & \left[1-Q\left(\frac{\delta}{2\sigma}\right)\right]^{n-1}\left[Q\left(\frac{\|b_1^n-a_1^n\|}{2\sigma}\right)-Q\left(\frac{\|b_1^n-a_1^n\|+\delta}{2\sigma}\right)\right] \\
\simeq\ & Q\left(\frac{\|b_1^n-a_1^n\|}{2\sigma}\right) \qquad \text{as}\,\sigma\to 0 \\
=\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n} d_{\mathsf{SI}}^2(t_i,r_i)}}{2\sigma}\right), \qquad \text{(C.18)}
\end{aligned}$$

where the right hand side of the inequality in (C.18) equals the integral of the joint probability distribution of $y_1^n \equiv y_1 \cdots y_n$ (given w_1, s_1^n) over the smaller region, which is obtained by using the fact that y_1^n is Gaussian centered at $c_1^n = a_1^n$ and by applying the necessary rotation.

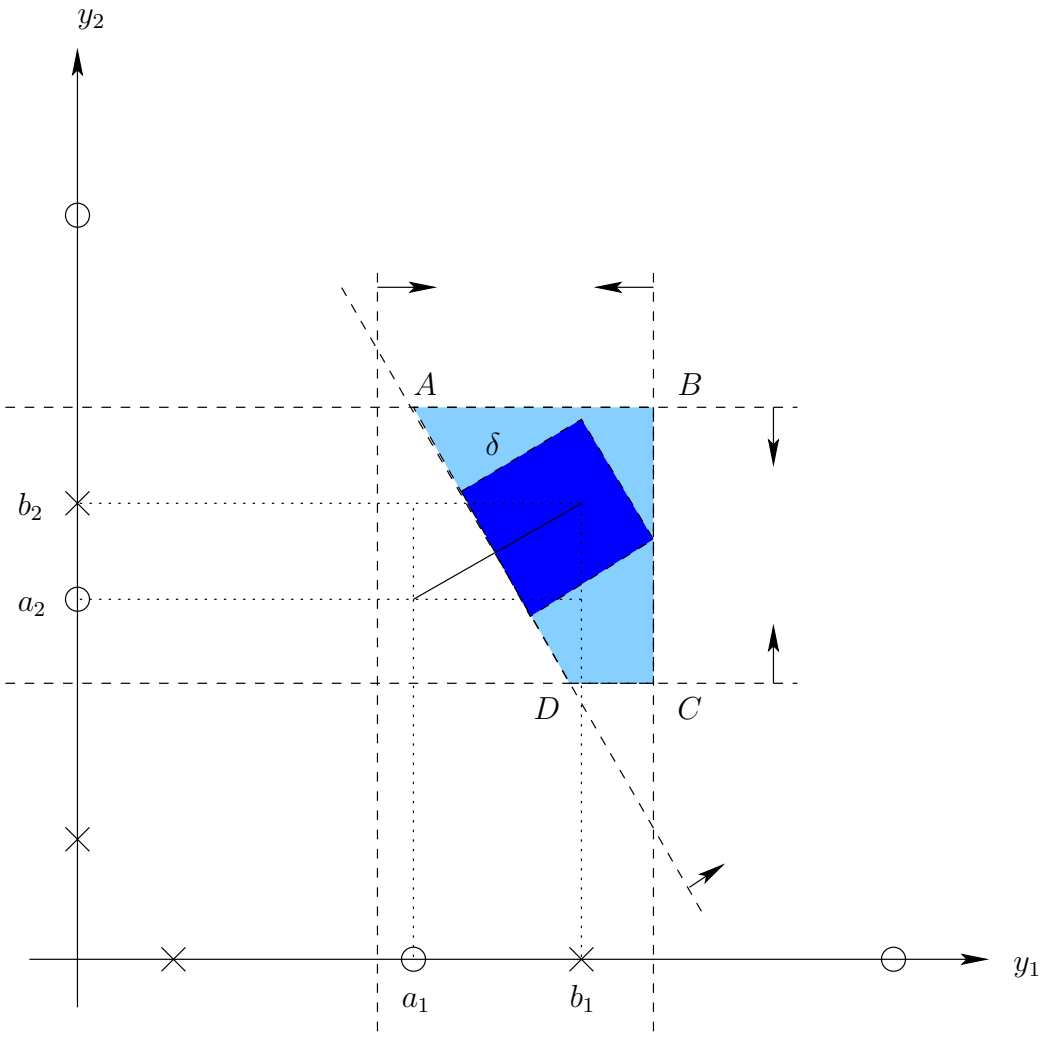

Figure C.1: Illustrating the regions of integration for dimension $n = 2$.

Appendix D

A Polynomial Complexity Algorithm for Finding Two Symbols of $\mathcal{T}$ with the Maximum Distance

We propose an algorithm for finding two symbols of $\mathcal{T}$ with distance greater than or equal to some $d_0 > 0$. Then we explain how to find two symbols in $\mathcal{T}$ with the maximum distance. Consider the bipartite graph $G(U, V, E)$ shown in fig. D.1 with $2Q$ vertices at each part. Each of the non-intersecting sets $U_1, \cdots, U_Q$ contains two vertices of the upper part U and each of the nonintersecting sets $V_1, \cdots, V_Q$ contains two vertices of the lower part V. The vertices of the sets $U_i = \{u_{i1}, u_{i2}\}$ and $V_i = \{v_{i1}, v_{i2}\}$ are labeled by the elements of the set $\mathcal{X} + s_i = \{x_1 + s_i, x_2 + s_i\}$, $i = 1, \ldots, Q$. A vertex in U_i is connected to a vertex in V_j if the absolute value of

the difference of their labels is greater than or equal to d_0, $i, j = 1, \ldots, Q$.

From the definition of distance in (3.7), there exist two symbols in $\mathcal{T}$ with distance $d \geq d_0$ if and only if G has a complete bipartite subgraph $K_{Q,Q}$ with exactly one vertex in each U_i and each V_j. If such a subgraph exists, we label the edges of the subgraph by 1 and we label the rest of the edges of G by 0. We denote the label of edge e by $y_e \in \{0, 1\}$. Such a labeling satisfies the following set of constraints

$$\sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i = 1, \ldots, Q, \tag{D.1}$$

$$\sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i = 1, \ldots, Q, \tag{D.2}$$

$$y_e \in \{0, 1\}. \tag{D.3}$$

Note that by definition, an edge of a graph is a set of two vertices. Therefore, the notation $e \cap U_i$ in (D.1) is meaningful. The equations (D.1) and (D.2) state that the sum of the labels of the edges going out of any U_i and V_i is Q.

We devise an objective function for the constraints (D.1), (D.2), and (D.3) such that the objective function takes a *given* maximum value only for a labeling with label 1 for the edges of the subgraph $K_{Q,Q}$ and label 0 for the rest of the edges.

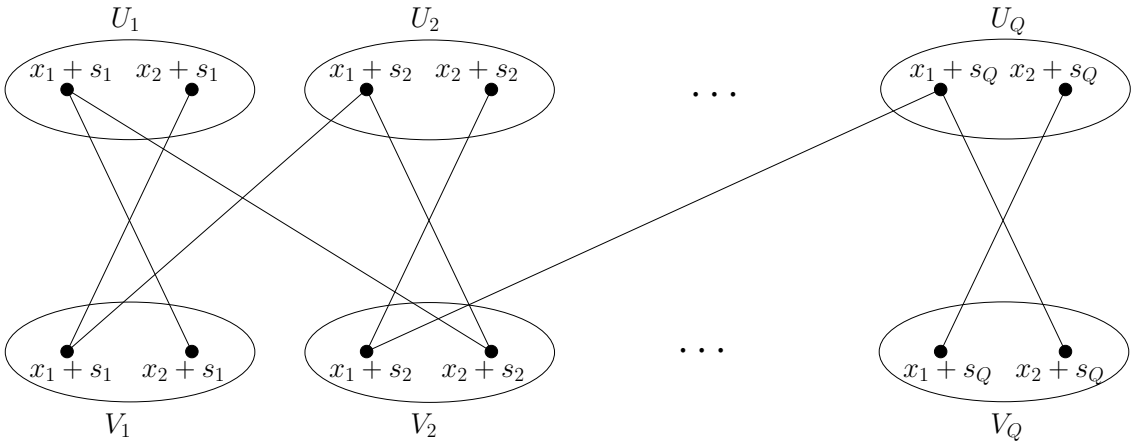

Figure D.1: Graph representation for the problem of finding two symbols of $\mathcal{T}$ with the maximum distance.

Consider the following optimization problem

$$\max_{y_e} \quad \sum_{i=1}^{Q}\sum_{j=1}^{2}\left(\sum_{e:u_{ij}\in e} y_e\right)^2 + \sum_{i=1}^{Q}\sum_{j=1}^{2}\left(\sum_{e:v_{ij}\in e} y_e\right)^2$$

subject to

$$\begin{aligned} &\sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q, \\ &\sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q, \\ &y_e \in \{0,1\}. \end{aligned} \tag{D.4}$$

In the following, we find the maximum of the above optimization problem for the foregoing labeling. Given the constraints of (D.4), we have

$$\sum_{j=1}^{2}\left(\sum_{e:u_{ij}\in e} y_e\right) = \sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q, \tag{D.5}$$

$$\sum_{j=1}^{2}\left(\sum_{e:v_{ij}\in e} y_e\right) = \sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q. \tag{D.6}$$

If the sum of two nonnegative variables is constant, then the sum of their squares

takes its maximum if one of the variables is zero. Therefore, for any $i = 1, \ldots, Q$, the maximum of

$$\sum_{j=1}^{2} \left(\sum_{e:u_{ij} \in e} y_e \right)^2$$

and

$$\sum_{j=1}^{2} \left(\sum_{e:v_{ij} \in e} y_e \right)^2$$

will be Q^2 and this maximum occurs if and only if one vertex in any of $U_1, \ldots, U_Q$ and $V_1, \ldots, V_Q$ is connected to Q edges with label 1 and the other vertex in any of $U_1, \ldots, U_Q$ and $V_1, \ldots, V_Q$ is not connected to any edge with label 1. This is equivalent to the existence of a subgraph $K_{Q,Q}$. Then the maximum of the objective function in (D.4) will be $Q \times Q^2 + Q \times Q^2 = 2Q^3$.

We may relax the integrality constraint (D.3) and change equality signs in (D.1) and (D.2) to inequality signs to obtain the following optimization program

$$\begin{aligned} \max_{y_e} \quad & \sum_{i=1}^{Q} \sum_{j=1}^{2} \left(\sum_{e:u_{ij} \in e} y_e \right)^2 + \sum_{i=1}^{Q} \sum_{j=1}^{2} \left(\sum_{e:v_{ij} \in e} y_e \right)^2 \\ \text{subject to} \quad & \\ & \sum_{e:e \cap U_i \neq \phi} y_e \leq Q, \qquad i = 1, \ldots, Q, \\ & \sum_{e:e \cap V_i \neq \phi} y_e \leq Q, \qquad i = 1, \ldots, Q, \\ & 0 \leq y_e \leq 1. \end{aligned} \tag{D.7}$$

Using the same argument as in the previous paragraph, the value $2Q^3$ is also achievable for the above maximization problem if and only if a subgraph $K_{Q,Q}$ of the graph G exists. The above optimization problem is a *quadratic programming* problem [54] with convex objective function and can be solved in polynomial time [55] in terms of the number of edges of G, which is at most $4Q^2$.

In summary, we turned the problem of finding two symbols in $\mathcal{T}$ with distance at least $d_0 > 0$ into the quadratic programming problem (D.7). If the maximum value of (D.7) is $2Q^3$, then two such symbols are obtained from the optimal solution of (D.7). Otherwise, two such symbols do not exist.

To find two symbols in $\mathcal{T}$ with the maximum distance, we need to run the described algorithm for a few values for d_0. We can obtain an upper bound on the number of possible distances between symbols of $\mathcal{T}$. From the definition of distance in (3.7), a loose upper bound is $M^2Q^2 = 4Q^2$. By using the binary search algorithm [56], the search over possible distances can be done with logarithmic complexity with respect to the number of possible distances.

It is worth mentioning that our proposed algorithm can be extended to find $K \geq 2$ symbols of $\mathcal{T}$ with the maximum minimum distance among K symbols for the general case $M \geq 2$.

Appendix E

Using More Than M Symbols of $\mathcal{T}$ $(M > 2)$

Consider the channel with $\mathcal{X} = \{1, 4, 5, 7\}$ and $\mathcal{S} = \{0, 4\}$. Consider the following codebook with six codewords of length two that uses seven symbols of the *associated* channel.

$$\begin{aligned}
&\text{Codeword 1}: \quad ((4,1),(5,1)) \\
&\text{Codeword 2}: \quad ((4,1),(1,5)) \\
&\text{Codeword 3}: \quad ((5,4),(5,4)) \\
&\text{Codeword 4}: \quad ((5,4),(4,5)) \\
&\text{Codeword 5}: \quad ((1,5),(4,1)) \\
&\text{Codeword 6}: \quad ((1,5),(1,4))
\end{aligned} \tag{E.1}$$

The minimum distance of the above code is 3. However, it can be verified by a computer program that any code for this channel with codebook size six and length

two that uses any four symbols of the *associated* channel yields a minimum distance less than 3.

Bibliography

[1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439-441, May 1983. 1

[2] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3820-3833, Nov. 2005. 1, 2, 9, 11, 12, 14, 26, 43

[3] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," in *Proc. IEEE Int. Symp.Inform. Theory and Its Applications (ISITA)*, Nov. 2000, pp. 681-684. 9, 11

[4] G. Caire and S. Shamai,"On achievable throughput of a multiple antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691-1706, Jul. 2003. 2

[5] W. Yu and J. M. Cioffi,"Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875-1892, Sep. 2004. 2

[6] S. Viswanath, N. Jindal, and A. Goldsmith,"Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcsat channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003. 2

[7] P. Viswanath and D. Tse,"Sum capacity of the multiple-antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1912-1921, Jul. 2003. 2

[8] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, A. Goldsmith, "Outer bounds for Gaussian multi-antenna broadcast channels", in *Proc. of DIMACS workshop on Signal processing for wireless Transmission*, Oct. 2002, Rutgers, NJ, USA. 2

[9] P. Viswanath and D. Tse, "On the capacity of the multiple antenna broadcast channel," in *Proc. DIMACS Workshop on Signal Processing for Wireless Transmissions*, Oct. 2002, Rutgers, NJ, USA. 2

[10] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of Gaussian multiple-input multiple-output channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936-3964, Sept. 2006. 2

[11] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001. 2, 3

[12] A. Cohen and A. Lapidoth,"The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639-1667, Jun. 2002. 2, 9

[13] P. Moulin and J. A. O'Sullivan,"Information-Theoretic Analysis of Information Hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563-593, Mar. 2003. 2

[14] I. J. Cox, M. L. Mileer, and A. L. McKellips, "Watermarking as communications with side information," in *Proc. IEEE (Special Issue on Identification and Protection of Multimedia Information)*, vol. 87, no. 7, pp. 1127-1141, Jul. 1999. 2

[15] R. F. H. Fischer, R. Tzschoppe, and R. Bauml, "Lattice Costa scheme using subspace projection for digital watermarking," *Europ. Trans. Telecommunication*, vol. 15, no. 4, pp. 351-361, Jul/Aug. 2004. 2

[16] J. J. Eggers, J. K. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," in *Proc. IEE Colloquium: Secure Images and Image Authentication*, London, U.K., pp. 4/1-4/6, Apr. 2000. 2

[17] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289-293, Oct. 1958. 3, 4, 5, 6

[18] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, no. 2, pp. 52-60, Apr.-June 1974. 4

[19] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19-31, Jan. 1980. 4, 6

[20] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at the encoder and decoder," *Proc. Inst. Elec. Eng.-Pt. I*, vol. 139, no. 2, pp. 113-117, Apr. 1992. 5

[21] G. Caire and S. Shamai,"On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007-2019, Sep. 1999. 5

[22] C. Heegard and A. El Gamal,"On the capacity of computer memories with defects," *IEEE Trans. Inform. Theory*, vol. 29, no. 5, pp. 731-739, Sep. 1983. 5

[23] A. Rosenzweig, Y. Steinberg, and S. Shamai,"On channels with partial state information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1817-1830, May 2005. 5

[24] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471-480, July 1973. 6

[25] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on Colored Paper," *IEEE International Symposium on Information Theory (ISIT)*, p.302, 2001. 9

[26] R. Zamir, S. Shamai, and U. Erez,"Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250-1276, June 2002. 9

[27] H. Farmanbar and A. K. Khnadani, "On precoding for channels with known interference at the transmitter," in *Proc. 2005 Conference on Information sciences and Systems (CISS 2005)*, Baltimore, MD, Mar. 16-18, 2005. 12

[28] H. Farmanbar, M. Rashidpour, and A. K. Khandani, "A dirty paper coding approach without modulo operation at the receiver," in *Proc. 2005 Canadian*

Workshop on Information Theory (CWIT 2005), Montreal, Quebec, pp. 179-182. 12

[29] T. M. Cover and J. A. Thomas, Elements of Information Theory, Wiley, New York, 1991. 12

[30] R. G. Gallager, Information Theory and Reliable Communication, John Wiley & Sons Inc., 1968. 12

[31] F. M. J. Willems, "On Gaussian Channels with side information at the transmitter," in *Proc. 9th Sym. Information Theory in the Benelux*, Enschede, The Netherland, May 1988, pp. 129-135. 12

[32] F. M. J. Willems, "Signalling for the Gaussian Channel with side information at the transmitter," in *Proc. Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 348. 12

[33] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proceedings of IEEE International Symposium on Information Theory*, Chicago, Illinois, p. 534, June/July 2004. 14

[34] S. Sigurjonsson and Y.-H. Kim, "On multiple user channels with causal state information at the transmitters," in *Proceedings of IEEE International Symposium on Information Theory*, pp. 72–76, Adelaide, Australia, September 2005. 14

[35] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information, *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2867-2877, Aug. 2005. 14

[36] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3992-4003, Nov. 2005. 14

[37] S. A. Jafar, "Capacity with causal and non-causal side information - a unified view", *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5468-5475, Dec. 2006. 14

[38] H. Farmanbar and A. K. Khandani, "Transmission over channels with known two-level interference at the transmitter," in *Proc. IEEE Biennial Symposium on Communications (QBSC 2006)*, Kingston, Ontario, pp. 279-282, May 29-June 1, 2006. 14

[39] H. Farmanbar and A. K. Khandani, "Precoding for the AWGN channel with discrete interference," in *Proc. 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, pp. 2186-2190, June 24-29, 2007. 14

[40] H. Farmanbar and A. K. Khandani, "Precoding for the AWGN channel with discrete interference," *Submitted to IEEE Transactions on Information Theory*, Mar. 2007. 14

[41] H. Farmanbar, S. Oveis Gharan, and A. K. Khandani, "Channel code design with causal side information at the encoder," in *Proc. 2007 IEEE Canadian Workshop on Information Theory (CWIT 2007)*, Edmonton, Alberta, pp. 144-147, June 6-8, 2007. 15

[42] H. Farmanbar, S. Oveis Gharan, and A. K. Khandani, "Channel code design with causal side information at the encoder," *Submitted to IEEE Transactions on Information Theory*, Oct. 2007. 15

[43] S. Aromoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 14-20, Jan. 1972. 22

[44] R. E. Blahut, "Computation of channel capacity and ratedistortion functions," *IEEE Transactions Information on Theory*, vol. 18, pp. 460473, July 1972. 22

[45] G. Nemhauser and L. Wolsey, *Integer and combinatorial optimization*, John Wiley & Sons, 1988. 28, 29

[46] B. Krekó, *Linear Programming*, Translated by J. H. L. Ahrens and C. M. Safe. Sir Isaac Pitman & Sons Ltd., 1968. 28, 32

[47] W. P. Pierskalla, "The multidimensional assignment problem," *Operations Research* 16, p. 422-431, 1968. 32

[48] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138-139, Mar. 1971. 43

[49] M. Miyakawa and H. Harashima, "A method of code conversion for a digital communication channel with intersymbol interference," *Trans. Inst. Electron. Commun. Eng. Japan*, vol. 52-A, pp. 272-273, Jun. 1969. 43

[50] U. Erez, and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3417-3432, Oct. 2005. 2

[51] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1872-1889, May 2006. 2

[52] W. Yu, D. P. Varodayan, and J. M. Cioffi "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. Commun*, vol. 53, no. 7, pp. 1220-1230, July 2005. 3

[53] G. Caire and S. Shamai,"Writing on dirty tape with LDPC codes," in *Proc. DIMACS Workshop on Signal Processing for Wireless Transmission*, Piscataway, NJ, Oct. 7-9, 2002. 14

[54] R. Fletcher, *Practical Methods of Optimization*, 2nd edition, John Wiley & Sons, Inc., New York, 1987. 77

[55] M. K. Kozlov, S. P. Tarasov, and L. G. Khachiyan, "Polynomial solvability of convex quadratic programming," *in Sov. Math.*, Dokl. 20, pp. 1108-1111, 1979. 77

[56] D. E. Knuth, *The Art of Computer Programming*, Volume 3: *Sorting and Searching*, 3rd edition, Addison-Wesley, 1997. 78