Channel Code Design with Causal Side Information at the Encoder

Hamid Farmanbar, Shahab Oveis Gharan, and Amir K. Khandani

Department of Electrical and Computer Engineering

University of Waterloo

Waterloo, Ontario, N2L 3G1

Technical Report UW-ECE #2007-30

September 27, 2007

Channel Code Design with Causal Side Information at the Encoder

Hamid Farmanbar, Shahab Oveis Gharan, and Amir K. Khandani

Coding and Signal Transmission Laboratory
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, N2L 3G1
Email: {hamid,shahab,khandani}@cst.uwaterloo.ca

Abstract

The problem of channel code design for the M-ary input AWGN channel with additive Q-ary interference where the sequence of i.i.d. interference symbols is known causally at the encoder is considered. The code design criterion at high SNR is derived by defining a new distance measure between the input symbols of the Shannon's *associated* channel. For the case of binary-input channel, i.e., $M = 2$, it is shown that it is sufficient to use only two (out of 2^Q) input symbols of the *associated* channel in the encoding as long as the distance spectrum of the code is concerned. This reduces the problem of channel code design for the binary-input AWGN channel with known interference at the encoder to the 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.

Index Terms

Causal side information, Shannon's associated channel, channel coding, pairwise error probability.

I. INTRODUCTION

Information transmission over channels with known interference at the transmitter has recently found applications in various communication problems such as digital watermarking [1] and broadcast schemes [2]. 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 [3]. Therefore, the interference does not incur any loss in the capacity. This result was extended to arbitrary (random or deterministic) interference in [4] by using a precoding scheme based on multi-dimensional lattice quantization. The result obtained by Costa does not hold for the case that the sequence of interference symbols is known causally at the transmitter.

Following Costa's "Writing on Dirty Paper" famous title [3], when the interference is known non-causally at the transmitter, the channel is referred to as "dirty paper" channel.

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 [2]. 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].

These developments motivate finding realizable dirty paper coding techniques. Building upon [4], Erez and ten Brink [9] 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. Wei Yu *et al.* [11] gave a design based on convolutional shaping and channel codes. Bennatan *et al.* [10] 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.

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 [1], scalar lattice quantization is proposed for data-hiding even though in that context, the host signal in clearly known non-causally.

In this paper, we consider the problem of channel code design for the M-ary input AWGN channel with additive causally-known discrete interference. The discrete model for interference is more appropriate for many practical applications. For example, in the MIMO broadcast channel, the interference caused by other users is discrete rather than continuous.

Our design does not rely on the suboptimal (in terms of capacity) scheme of scalar lattice quantization for the causally known interference [4], [12]. 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.

This paper is organized as follows. In the next section, we summarize Shannon's work on channels with causal side information at the transmitter. In section III, we introduce the channel model. In section IV, we derive the code design criterion for AWGN channel with causally-known discrete interference at the encoder. In section V, we consider channels with binary input for which we show that the design criterion derived in section IV reduces to maximizing the Hamming distance. In section VI, we consider a special case for which the result for the binary channel also holds for the M-ary channel. We conclude this paper in section VII.

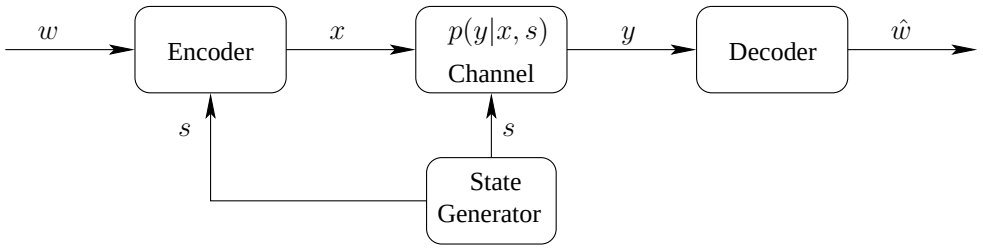

Fig. 1. SD-DMC with state information at the encoder.

II. Channels with Side Information at the Transmitter

Channels with known interference at the transmitter are special case of channels with side information at the transmitter which were considered by Shannon [13] in the causal knowledge setting and by Gel'fand and Pinsker [14] in the non-causal knowledge setting.

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 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.

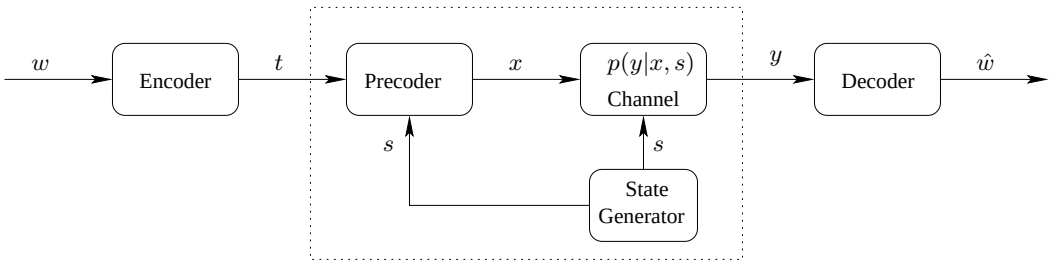

Fig. 2. The associated regular DMC.

In the causal knowledge setting, the encoder maps a message w into $\mathcal{X}^n$ as

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

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 known causally at the encoder [13].

The SD-DMC can be used in the way shown in fig. 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

$$q(y|t) = \sum_{s \in \mathcal{S}} p(s) p(y|x = t(s), s), \tag{2}$$

where $p(s)$ is the probability of the state s. The DMC defined in (2) 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 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 $x_s, s = 1, 2, \ldots, |\mathcal{S}|$.

III. The Channel Model

We consider data transmission over the channel

$$Y = X + S + N, \tag{3}$$

where X is the channel input, which takes on values in a fixed real constellation $\mathcal{X}$, Y is the channel output, N is additive white Gaussian noise with power σ^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 all functions from $\mathcal{S}$ to $\mathcal{X}$.

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

$$\begin{aligned} f_{Y|T}(y|t) &= \sum_{s\in\mathcal{S}} p(s) f_{Y|X,S}(y|t(s),s) \\ &= \sum_{s\in\mathcal{S}} p(s) f_N(y-t(s)-s), \end{aligned} \tag{4}$$

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

IV. The Code Design Criterion

Any coding scheme for the associated channel defined by (4) 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. Suppose that the messages w_1 and w_2 are encoded into $t_1^n \equiv t_1 t_2 \ldots t_n$ and $r_1^n \equiv r_1 r_2 \ldots r_n$, respectively, where t_i's and r_i's belong to the alphabet $\mathcal{T}$. Using maximum likelihood decoding, the probability of the event that message w_2 is decoded given message w_1 was

sent is given by

$$\begin{aligned}
\Pr\{w_1 \to w_2|w_1\} &= \sum_{s_1^n} p(s_1^n)\Pr\{w_1 \to w_2|w_1, s_1^n\} \\
&= \sum_{s_1^n} p(s_1^n)\Pr\left\{f_{Y|T}(y_1^n|t_1^n) < f_{Y|T}(y_1^n|r_1^n)|w_1, s_1^n\right\} \\
&= \sum_{s_1^n} p(s_1^n)\Pr\left\{\prod_{i=1}^{n} f_{Y|T}(y_i|t_i) < \prod_{i=1}^{n} f_{Y|T}(y_i|r_i)|w_1, s_1^n\right\} \\
&= \sum_{s_1^n} p(s_1^n)\Pr\left\{\prod_{i=1}^{n}\sum_{s\in\mathcal{S}} p(s) f_N(y_i - t_i(s) - s) < \right. \\
&\qquad \left. \prod_{i=1}^{n}\sum_{s\in\mathcal{S}} p(s) f_N(y_i - r_i(s) - s)|w_1, s_1^n\right\}. \qquad (5)
\end{aligned}$$

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

$$\Pr\{w_1 \to w_2|w_1\} \propto Q\left(\frac{\sqrt{\sum_{i=1}^{n} d_{SI}^2(t_i, r_i)}}{2\sigma}\right), \qquad (6)$$

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

$$d_{SI}(t, r) = \min_{s_1, s_2 \in \mathcal{S}} |t(s_1) + s_1 - r(s_2) - s_2|. \qquad (7)$$

According to (6), at high SNR, the code design criterion is to maximize the minimum distance between the codewords with the distance measure defined in (7).

In order to see how the knowledge of interference at the encoder can result in increased distances between codewords, consider the channel model introduced in section III 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

$$\Pr\{v_1 \to v_2|v_1\} \propto Q\left(\frac{\sqrt{\sum_{i=1}^{n} d^2(x_i, z_i)}}{2\sigma}\right), \qquad (8)$$

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

$$d(x, z) = \min_{s_1, s_2 \in \mathcal{S}} |x + s_1 - z - s_2|. \tag{9}$$

It becomes clear by comparing (7) and (9) that higher distances between codewords are possible for the channel with side information at the encoder.

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 (9) as $d(1, 1) = d(-1, -1) = d(1, -1) = 0$. Hence, according to (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 (7), it is easy to check that the distances between all pairs of the symbols are zero except for $d_{SI}(u_1, u_2)$ which is 2. As will be seen in section V, 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 (7) or (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 $d_{SI}(u_1, u_2)$ which is 2. Therefore, the triangle inequality does not hold for $d_{SI}(u_1, u_3)$, $d_{SI}(u_3, u_2)$, and $d_{SI}(u_1, u_2)$.

V. The Binary Channel

Any code designed for the regular associated channel translates to a code for the actual channel with known interference at the encoder. The alphabet size of the associated channel is M^Q. However, we might not need to use all the symbols of the alphabet in the encoding scheme as long as the distance spectrum of the code is concerned.

For example, we consider the case where $M = 2$, i. e., when the channel accepts binary input. However, there is no constraints on the cardinality of interference alphabet.

We call this channel the *binary channel*. For the binary channel, the size of $\mathcal{T}$ is 2^Q. The following lemma holds 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 (7), it is sufficient to show that there exist two elements t and r in $\mathcal{T}$ such that the corresponding multi-sets [1] (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, the 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 in [15] which is stated in the context of capacity.

Let u_1 and u_2 be two input symbols of the *associated* channel with the maximum

[1] 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.

distance among all pairs of input symbols of the associated channel. Since $d_{SI}(u_1, u_2) > 0$, we have $u_1(s) \neq u_2(s), \forall s \in \mathcal{S}$, otherwise, from (7), $d_{SI}(u_1, u_2) = 0$. 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., $t(s) = u_2(s)$) 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 current 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 1: The distance spectrum of the binary code constructed by the procedure described above is at least as good as the distance spectrum of the old code.

Proof: Consider any two codewords $(t_1, \ldots, t_n)$ and $(r_1, \ldots, r_n)$ from the old codebook, where $t_i, r_i \in \mathcal{T}$. The squared distance between the two codewords is equal to $\sum_{i=1}^{n} d_{SI}^2(t_i, r_i)$. 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 $d_{SI}(t_i, r_i) = 0$, so the replacement will not change the distance.

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

According to theorem 1, as long as the distance spectrum of the code in concerned, it is sufficient to use two symbols of $\mathcal{T}$ with maximum distance, namely u_1 and u_2, in the encoding for a binary channel. Since for the binary channel $\mathcal{T}$ has size 2^Q, 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 II.

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 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

$$\sum_{i=1}^{n} d_{SI}^2(t_i, r_i) = d_{SI}^2(u_1, u_2) d_H(c_1, c_2), \tag{10}$$

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 $\{0, 1\}$.

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

$$d_E^2(c_1, c_2) = (x_1 - x_2)^2 d_H(c_1, c_2), \tag{11}$$

where d_E denotes the Euclidean distance.

Using (10) and (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 (7) that

$$d_{SI}(u_1, u_2) \leq |x_1 - x_2|. \tag{12}$$

Therefore, using (10) and (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 (12) is satisfied with equality.

If (12) is satisfied with equality, the distance spectrum of the two codes will be the same. In particular, the slope of error probability curves at high SNR (which corresponds to the minimum distance of the codebook) with maximum likelihood decoding will be the same for the two cases. In other words, if (12) is satisfied with equality, the knowledge of interference at the encoder enables us to achieve the same performance as the interference-free case at high SNR.

Theorem 2: $d_{SI}(u_1, u_2) = |x_1 - x_2|$ if and only if

$$\min_{i \neq j} |s_i - s_j| \geq |x_1 - x_2|. \tag{13}$$

Proof: If $\min |s_i - s_j| \geq |x_1 - x_2|$, then we may take $u = (x_1, x_2, x_1, \ldots)$ and $v = (x_2, x_1, x_2, \ldots)$. Then it is easy to check that $d_{SI}(u_1, u_2) = |x_1 - x_2|$.

For the other direction, suppose that $\min |s_i - s_j| < |x_1 - x_2|$. We will show that $d_{SI}(u_1, u_2) < |x_1 - x_2|$. Suppose that $s_k, s_l \in \mathcal{S}$ achieve the minimum of $|s_i - s_j|$ and t_1 and t_2 arbitrary elements of $\mathcal{T}$. Then, we consider two possibilities:

Case 1: $t_1(s_k) = t_1(s_l) = x_1$ and $t_2(s_k) = t_2(s_l) = x_2$. Then $|t_1(s_l) + s_l - t_2(s_k) - s_k| < |x_1 - x_2|$.

Case 2: $t_1(s_k) = x_1, t_1(s_l) = x_2$ and $t_2(s_k) = x_2, t_2(s_l) = x_1$. Then $|t_1(s_k) + s_k - t_2(s_l) - s_l| < |x_1 - x_2|$. ■

As an example, consider a binary channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$ and with 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 channel without error control coding and with maximum likelihood decoding. The error probability vs. SNR (not in dB) for the above channel is plotted in fig. 3.

The error probability curve for the interference-free binary 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 for this example, $d_{SI}(u_1, u_2) = |x_1 - x_2| = 2$. As it can be seen, the curves have the same slopes as expected at high SNR. Note that if the interference

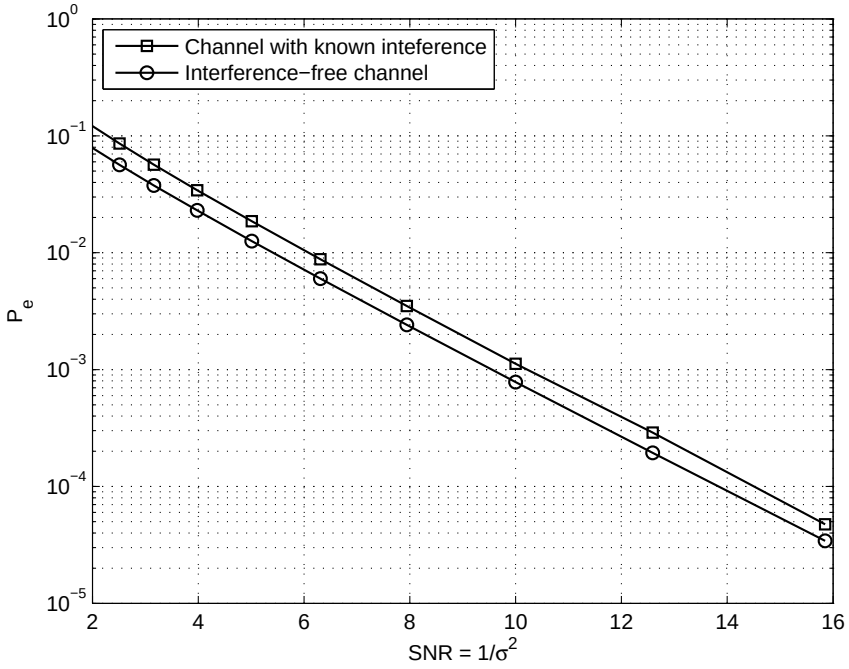

Fig. 3. Error probability vs. SNR for the binary input AWGN channel with/without (known) interference. $\mathcal{X} = \mathcal{S} = \{-1, +1\}$.

were not known at the encoder, the error probability curve would reach an error floor of $\frac{1}{4}$.

Another example is illustrated in fig. 4. 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 $d_{SI}(u_1, u_2) = 1 < |x_1 - x_2| = 2$. 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.

VI. The M-ary Channel

In general, the statement of theorem 1 is not extendable to the case with $M > 2$ channel input symbols. In fact, by using more than M input symbols of the associated

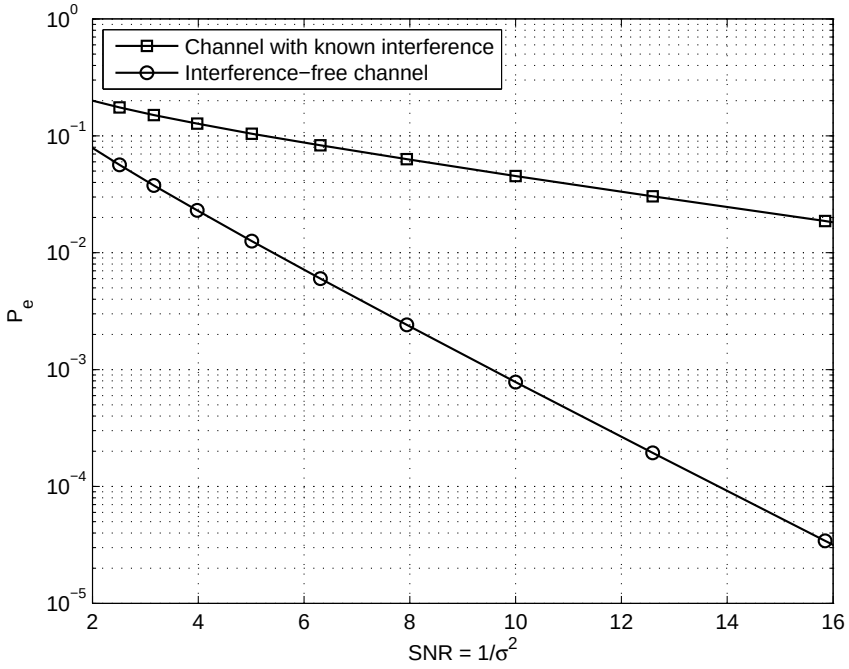

Fig. 4. Error probability vs. SNR for the binary input AWGN channel with/without (known) interference. $\mathcal{X} = \{-1, +1\}, \mathcal{S} = \{-1, 0, +1\}$.

channel, we can obtain a better codebook in terms of distance spectrum than any other codebook composed of M symbols of the associated channel. As an example, 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.

$$((4,1),(5,1))$$

$$((4,1),(1,5))$$

$$((5,4),(5,4))$$

$$((5,4),(4,5))$$

$$((1,5),(4,1))$$

$$((1,5),(1,4))$$

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.

Under some condition on the channel input and interference alphabets, the statement of theorem 1 can be generalized to the case with $M > 2$.

Theorem 3: As long 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

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

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, $i = 1, 2, \ldots, M$.

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

In theorem 3, we showed that given the the condition of the theorem satisfied, it is sufficient to use the associated channel input 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. In fact, given the condition of theorem 3 satisfied, the knowledge of interference symbols at the encoder is not helpful (in terms of distance spectrum improvement).

VII. Conclusion

In this paper, we derived the code design criterion at high SNR for the M-ary input AWGN channel with additive Q-level interference, where the sequence of interference

symbols is known causally at the encoder. Although we considered the AWGN channel with additive interference, our treatment applies to more general channels characterized by

$$Y = f(X, S) + N, \tag{14}$$

where f is a function of two variables and S is the channel state which is known causally at the encoder. Another special case of this more general channel is the fast fading channel

$$Y = SX + N, \tag{15}$$

where S is the fading coefficient.

The code design is over an input alphabet $\mathcal{T}$ of size M^Q. For the general channel model (14), the distance between two symbols t and r of $\mathcal{T}$ is defined as

$$d_{SI}(t, r) = \min_{s_1, s_2 \in \mathcal{S}} |f(t(s_1), s_1) - f(t(s_2), s_2)|. \tag{16}$$

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 long 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 past fifty years.

It is worth mentioning that lemma 1 and theorems 2 and 3 do not hold for the more general channel model in (14) and are specific to the AWGN channel with additive interference.

APPENDIX I
DERIVATION OF CODE DESIGN CRITERION AT HIGH SNR

Define

$$\mathcal{A}_i = \{t_i(s) + s : s \in \mathcal{S}\}, \qquad i = 1, \ldots, n, \tag{17}$$

$$\mathcal{B}_i = \{r_i(s) + s : s \in \mathcal{S}\}, \qquad i = 1, \ldots, n. \tag{18}$$

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

$$\sum_s p(s) f_N(y - t_i(s) - s) = \sum_{a \in \mathcal{A}_i} p(a) f_N(y - a), \tag{19}$$

$$\sum_s p(s) f_N(y - r_i(s) - s) = \sum_{b \in \mathcal{B}_i} p(b) f_N(y - b), \tag{20}$$

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

$$p(a) = \sum_{s \in \mathcal{S}: t_i(s) + s = a} p(s), \tag{21}$$

$$p(b) = \sum_{s \in \mathcal{S}: r_i(s) + s = b} p(s). \tag{22}$$

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

$$E_1(a_1^n) : a_i = \arg\min_{a \in \mathcal{A}_i} |y_i - a|, \quad i = 1, \ldots, n, \tag{23}$$

$$E_2(b_1^n) : b_i = \arg\min_{b \in \mathcal{B}_i} |y_i - b|, \quad i = 1, \ldots, n,$$

given that w_1 has been sent and the interference sequence s_1^n has occurred. Any term in

the error probability in (5) can be written as

$$\begin{aligned}
&\Pr\left\{\prod_{i=1}^{n}\sum_{a\in\mathcal{A}_i}p(a)f_N(y_i-a)<\prod_{i=1}^{n}\sum_{b\in\mathcal{B}_i}p(b)f_N(y_i-b)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\prod_{i=1}^{n}\sum_{a\in\mathcal{A}_i}p(a)f_N(y_i-a)<\prod_{i=1}^{n}\sum_{b\in\mathcal{B}_i}p(b)f_N(y_i-b),E_1(a_1^n),E_2(b_1^n)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\prod_{i=1}^{n}f_N(y_i-a_i)\left(p(a_i)+\sum_{\substack{a\in\mathcal{A}_i\\a\neq a_i}}p(a)\frac{f_N(y_i-a)}{f_N(y_i-a_i)}\right)\right.\\
&\qquad\left.<\prod_{i=1}^{n}f_N(y_i-b_i)\left(p(b_i)+\sum_{\substack{b\in\mathcal{B}_i\\b\neq b_i}}p(b)\frac{f_N(y_i-b)}{f_N(y_i-b_i)}\right),E_1(a_1^n),E_2(b_1^n)|w_1,s_1^n\right\}\\
&=\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2>\sum_{i=1}^{n}(y_i-b_i)^2+K\sigma^2,E_1(a_1^n),E_2(b_1^n)|w_1,s_1^n\right\},
\end{aligned}\tag{24}$$

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

$$K(y_1^n,a_1^n,b_1^n)=2\sum_{i=1}^{n}\log\frac{p(a_i)+\sum_{\substack{a\in\mathcal{A}_i\\a\neq a_i}}p(a)\frac{f_N(y_i-a)}{f_N(y_i-a_i)}}{p(b_i)+\sum_{\substack{b\in\mathcal{B}_i\\b\neq b_i}}p(b)\frac{f_N(y_i-b)}{f_N(y_i-b_i)}}.\tag{25}$$

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

$$K_1(a_1^n)=2\sum_{i=1}^{n}\log p(a_i)<K(y_1^n,a_1^n,b_1^n)<K_2(b_1^n)=2\sum_{i=1}^{n}\log\frac{1}{p(b_i)}.\tag{26}$$

As we are considering the high SNR regime, we may assume that the noise power is sufficiently small so that the error probability (5) can be well approximated by

$$\sum_{s_1^n}p(s_1^n)\sum_{a_1^n}\sum_{b_1^n}\Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2>\sum_{i=1}^{n}(y_i-b_i)^2,E_1(a_1^n),E_2(b_1^n)|w_1,s_1^n\right\}.\tag{27}$$

Any term in the summation (27) can be upper bounded as

$$\begin{aligned}
& \Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2 > \sum_{i=1}^{n}(y_i-b_i)^2, E_1(a_1^n), E_2(b_1^n)|w_1, s_1^n\right\} \\
\leq\ & \Pr\left\{\sum_{i=1}^{n}(y_i-c_i)^2 > \sum_{i=1}^{n}(y_i-b_i)^2, E_1(a_1^n), E_2(b_1^n)|w_1, s_1^n\right\} \\
\leq\ & \Pr\left\{\sum_{i=1}^{n}(y_i-c_i)^2 > \sum_{i=1}^{n}(y_i-b_i)^2|w_1, s_1^n\right\} \\
=\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n}|c_i-b_i|^2}}{2\sigma}\right) \\
\leq\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n}d_{SI}^2(t_i,r_i)}}{2\sigma}\right), \qquad (28)
\end{aligned}$$

where

$$c_i = t_i(s_i) + s_i, \quad i = 1, \ldots, n. \qquad (29)$$

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

Now, we show that the upper bound (28) is tight for the term(s) in the summation (27) satisfying

$$\{a_i, b_i\} = \arg\min_{\substack{a\in\mathcal{A}_i \\ b\in\mathcal{B}_i}} |a-b|, \quad i = 1, \ldots, n, \qquad (30)$$

and

$$a_i = c_i, \quad i = 1, \ldots, n. \qquad (31)$$

Any term in (27) 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

$$\left\{y_1^n : \sum_{i=1}^{n}(y_i-a_i)^2 > \sum_{i=1}^{n}(y_i-b_i)^2, E_1(a_1^n), E_2(b_1^n)\right\}. \qquad (32)$$

This region is illustrated by the shaded area ABCD in fig. 5 for $n = 2$. For the terms in (27) which satisfy (30) and (31) the region defined in (32) includes the region

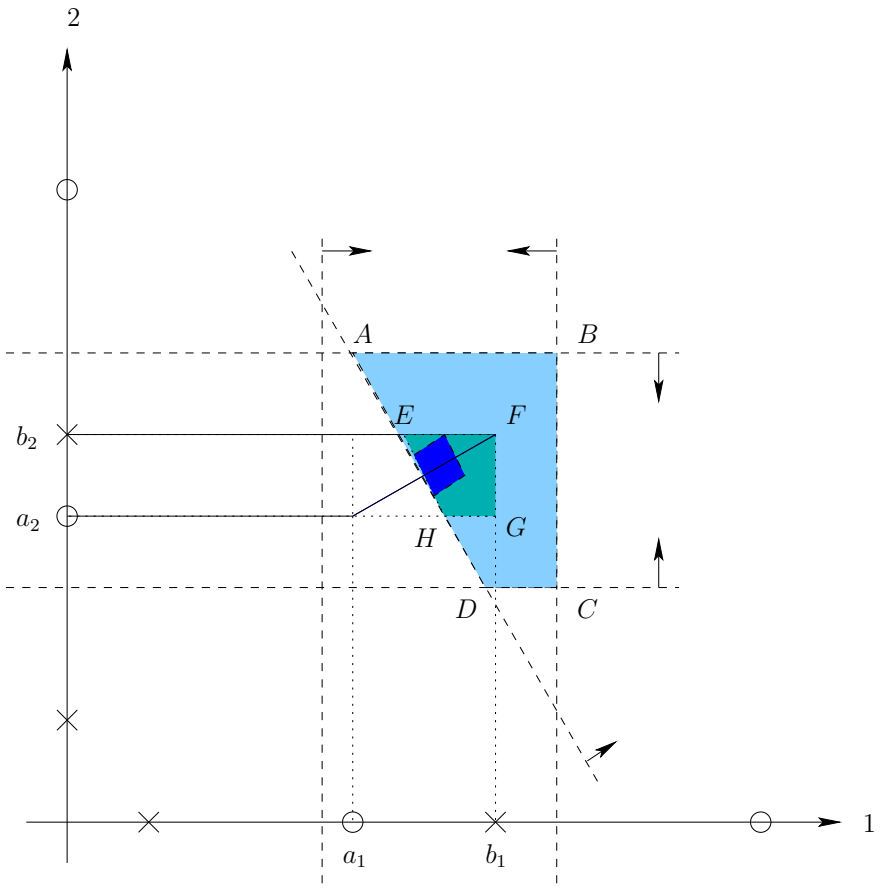

Fig. 5. Illustrating the regions of integration for dimension $n = 2$.

defined by

$$\left\{ \sum_{i=1}^{n} (y_i - a_i)^2 > \sum_{i=1}^{n} (y_i - b_i)^2, \min(a_i, b_i) < y_i < \max(a_i, b_i), i = 1, \ldots, n \right\}. \quad (33)$$

The above region is illustrated by the shaded area EFGH in fig. 5. This region is not empty since we may assume that $a_i \neq b_i$, $i = 1, \ldots, n$. We may consider an n-cube inside this region with sides equal to $\delta > 0$ as shown in fig. 5 and perform the integration over this smaller region.

In summary,

$$\begin{aligned}
& \Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2>\sum_{i=1}^{n}(y_i-b_i)^2, E_1(a_1^n), E_2(b_1^n)|w_1,s_1^n\right\} \\
\geq\ & \Pr\left\{\sum_{i=1}^{n}(y_i-a_i)^2>\sum_{i=1}^{n}(y_i-b_i)^2, \min(a_i,b_i)<y_i<\max(a_i,b_i), i=1,\ldots,n|w_1,s_1^n\right\} \\
\geq\ & \left[1-Q\left(\frac{\delta}{2\sigma}\right)\right]^{n-1}\left[Q\left(\frac{\|b_1^n-a_1^n\|}{2\sigma}\right)-Q\left(\frac{\|b_1^n-a_1^n\|+\delta}{2\sigma}\right)\right] \\
\approx\ & Q\left(\frac{\|b_1^n-a_1^n\|}{2\sigma}\right) \qquad \text{as}\,\sigma\to 0 \\
=\ & Q\left(\frac{\sqrt{\sum_{i=1}^{n}d_{SI}^2(t_i,r_i)}}{2\sigma}\right). \qquad (34)
\end{aligned}$$

APPENDIX II

A POLYNOMIAL COMPLEXITY ALGORITHM FOR FINDING TWO INPUT SYMBOLS OF $\mathcal{T}$ WITH THE MAXIMUM DISTANCE

Consider the bipartite graph G shown in fig. 6 with $2Q$ vertices at each part. Each of the non-intersecting sets $U_1,\cdots,U_Q$ contains two vertices of the upper part and each of the nonintersecting sets $V_1,\cdots,V_Q$ contains two vertices of the lower part. 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 some $d_0\geq 0$, $i,j=1,\ldots,Q$.

From the definition of distance in (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

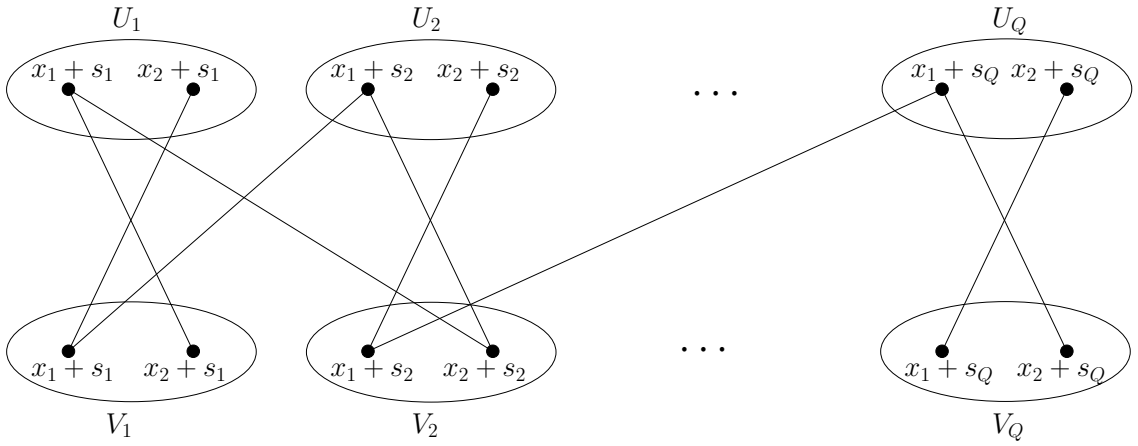

Fig. 6. A bipartite graph.

$y_e \in \{0,1\}$. Such a labeling satisfies the following set of constraints

$$\sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i=1,\ldots,Q, \tag{35}$$

$$\sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i=1,\ldots,Q, \tag{36}$$

$$y_e \in \{0,1\}. \tag{37}$$

Note that by definition an edge of a graph is a set of two vertices. Therefore, the notation $e\cap U_i$ in (35) is meaningful. The equations (35) and (36) 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 (35), (36), and (37) such that the objective function takes a *given* maximum value only for the labeling with label 1 for the edges of the subgraph $K_{Q,Q}$ and label 0 for the rest of the edges. 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

$$\begin{aligned}
&\sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q,\\
&\sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q,\\
&y_e \in \{0,1\}.
\end{aligned} \tag{38}$$

We have

$$\sum_{j=1}^{2}\left(\sum_{e:u_{ij}\in e} y_e\right) = \sum_{e:e\cap U_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q, \tag{39}$$

$$\sum_{j=1}^{2}\left(\sum_{e:v_{ij}\in e} y_e\right) = \sum_{e:e\cap V_i\neq\phi} y_e = Q, \qquad i = 1,\ldots,Q. \tag{40}$$

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 the subgraph $K_{Q,Q}$. Then the maximum of the objective function in (38) will be $Q\times Q^2 + Q\times Q^2 = 2Q^3$.

We may relax the integrality constraint (37) and change equality signs in (35) and (36) to inequality signs to obtain the following optimization program

$$\begin{aligned}
\max_{y_e} \quad & \sum_{i=1}^{Q}\sum_{j=1}^{2}\left(\sum_{e:u_{ij}\in e} y_e\right)^2 + \sum_{i=1}^{Q}\sum_{j=1}^{2}\left(\sum_{e:v_{ij}\in e} y_e\right)^2 \\
\text{subject to} \quad & \\
& \sum_{e:e\cap U_i\neq\phi} y_e \leq Q, \qquad i=1,\ldots,Q, \\
& \sum_{e:e\cap V_i\neq\phi} y_e \leq Q, \qquad i=1,\ldots,Q, \\
& 0 \leq y_e \leq 1. \qquad (41)
\end{aligned}$$

It is easy to check that the value $2Q^3$ is achievable for the above maximization problem too if and only if a subgraph $K_{Q,Q}$ of the graph G exists. The above optimization problem is a *quadratic programming* problem [16] and can be solved in polynomial time [17] 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 (41). If the maximum value of (41) is $2Q^3$, then two such symbols are obtained from the optimal solution of (41). 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 (7), a loose upper bound is $M^2Q^2 = 4Q^2$. By using the binary search algorithm [18], the search over possible distances can be done with logarithmic complexity with respect to the number of possible distances.

REFERENCES

[1] 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] 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.

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

[4] 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.

[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.

[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.

[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.

[8] H. Weingarten, Yosef 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

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

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

[17] 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.

[18] D. Knut, *The Art of Computer Programming*, Volume 3: *Sorting and Searching*, 3rd edition, Addison-Wesley, 1997.