Precoding for Channels with Discrete Interference

Hamid Farmanbar and Amir K. Khandani

Coding and Signal Transmission Laboratory
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, N2L 3G1

Technical Report UW-ECE #2006-24
November 24, 2006

Precoding for Channels with Discrete Interference

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

Abstract

M-ary signal transmission over AWGN channel with additive Q-ary interference where the sequence of i.i.d. interference symbols is known causally at the transmitter is considered. Shannon's theorem for channels with side information at the transmitter is used to formulate the capacity of the channel. It is shown that by using at most $MQ - Q + 1$ out of M^Q inputs of the *associated* channel, the capacity is achievable. We consider the maximization of the transmission rate over input probability assignments for the associated channel that induce uniform distribution on the input to the actual channel. For this setting, the general structure of a communication system with optimal precoding is proposed.

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 main result on such channels was obtained by Costa who showed that the capacity of 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 AWGN channel [3]. Therefore, the interference does not incur any loss in the capacity. This result was extended to non-Gaussian interference in [4]. The result obtained by Costa does not hold for the case that the sequence of interference symbols is known causally 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 [5] in causal knowledge setting and by Gel'fand and Pinsker [6] in non-causal knowledge setting.

This paper is organized as follows. Section II, provides some background on channels with causally-known interference at the encoder. In section III, we introduce the channel model. In section IV, we investigate the capacity of the channel introduced in section III. In section V, we consider maximizing the transmission rate when the channel input is uniform. The general structure of a communication system for the channel with causally-known discrete interference is given in section VI. We extend the uniform transmission scheme to continuous-input alphabet in section VII. We conclude this paper in section VIII.

II. Preliminaries

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.

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

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

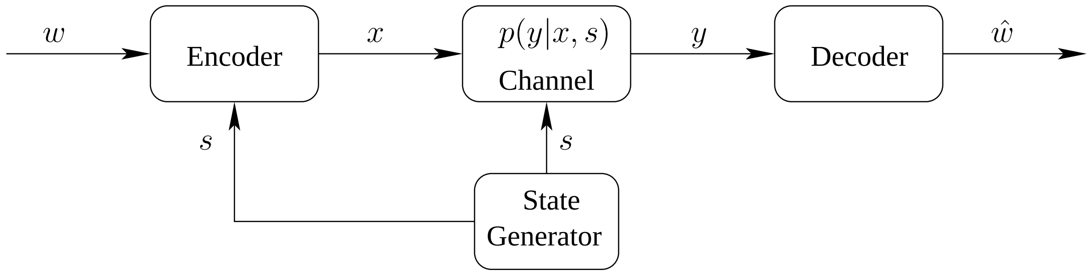

Fig. 1. SD-DMC with state information at the encoder.

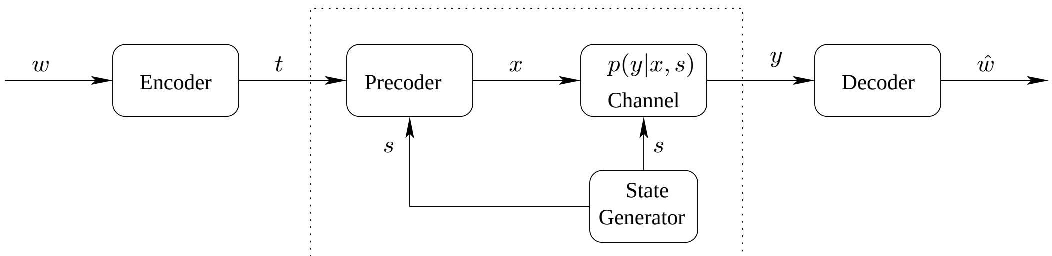

Fig. 2. The associated regular DMC.

Shannon [5] showed that the capacity of an SD-DMC where the i.i.d. state sequence is known causally at the encoder is equal to the capacity of an *associated* regular (without state) DMC with an extended input alphabet $\mathcal{T}$ and the same output alphabet $\mathcal{Y}$. The input alphabet of the associated channel is the set of indices of 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 $(x_{i_1}, x_{i_2}, \ldots, x_{i_{|\mathcal{S}|}})$ of elements of $\mathcal{X}$, implying that the value of the function at state s is $x_{i_s}, s = 1, 2, \ldots, |\mathcal{S}|$.

The transition probabilities for the associated channel are given by [5]

$$p(y|t) = \sum_{s=1}^{|\mathcal{S}|} p(s) p(y|x_{i_s}, s), \tag{2}$$

where t denotes the index of the function represented by $(x_{i_1}, x_{i_2}, \ldots, x_{i_{|\mathcal{S}|}})$. Also,

$$p(y(1) \cdots y(n)|t(1) \cdots t(n)) = \prod_{k=1}^{n} p(y(k)|t(k)). \tag{3}$$

The capacity is given by [5]

$$C = \max_{p(t)} I(T;Y), \tag{4}$$

where the maximization is taken over the probability mass function (pmf) of the random variable T.

Any encoding and decoding scheme for the associated channel can be translated into an encoding and decoding scheme for the original state-dependent channel with the same probability of error [5]. An encoder for the associated channel encodes a message w to $(t(1), \ldots, t(n))$. The translated encoding scheme for the original state-dependent channel is to map the message w to $(x(1), x(2), \ldots, x(n))$, where $x(k) = s$th component of $t(k)$ if the state at time k is s, $s = 1, 2, \ldots, |\mathcal{S}|$, and $k = 1, 2, \ldots, n$. The block diagram of the associated regular DMC is shown in fig. 2.

In the capacity formula (4), we can alternatively replace T with $(X_1, \ldots, X_{|\mathcal{S}|})$, where X_s is the input to the state-dependent channel when the state is $s, s = 1, \ldots, |\mathcal{S}|$.

III. The Channel Model

We consider data transmission over the channel

$$Y = X + S + N, \tag{5}$$

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

$$\mathcal{X} = \{x_1, x_2, \ldots, x_M\}. \tag{6}$$

We may assume that $x_1 < x_2 < \cdots < x_M$. 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

$$\mathcal{S} = \{s_1, s_2, \ldots, s_Q\} \tag{7}$$

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_q is the random variable that represents the channel input when the current interference symbol is s_q, $q = 1, \ldots, Q$.

The likelihood function for the associated channel is given by

$$\begin{aligned} f_{Y|T}(y|t) &= \sum_{q=1}^{Q} r_q f_{Y|X,S}(y|x_{i_q}, s_q) \\ &= \sum_{q=1}^{Q} r_q f_N(y - x_{i_q} - s_q), \end{aligned} \tag{8}$$

where f_N denotes the pdf of the noise N, and t is the index of $(x_{i_1}, x_{i_2}, \ldots, x_{i_Q})$. The pdf of Y is then given by

$$\begin{aligned} f_Y(y) &= \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} p_{i_1 i_2 \cdots i_Q} \left(\sum_{q=1}^{Q} r_q f_N(y - x_{i_q} - s_q) \right) \\ &= \sum_{q=1}^{Q} r_q \sum_{i=1}^{M} p_i^{(q)} f_N(y - x_i - s_q), \end{aligned} \tag{9}$$

where $p_{i_1 i_2 \cdots i_Q} = \Pr\{X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q}\}$, $p_i^{(q)} = \Pr\{X_q = x_i\}$.

IV. The capacity

The capacity of the associated channel, which is the same as the capacity of the original channel defined in section III, is the maximum of $I(T;Y) = I(X_1 X_2 \cdots X_Q; Y)$ over the joint pmf values $p_{i_1 i_2 \cdots i_Q}$, i.e.,

$$C = \max_{p_{i_1 i_2 \cdots i_Q}} I(X_1 X_2 \cdots X_Q; Y). \tag{10}$$

The mutual information between T and Y is the difference between differential entropies $h(Y)$ and $h(Y|T)$. It can be seen from (9) that $f_Y(y)$, and hence $h(Y)$, are uniquely determined by the marginal pmfs $\{p_i^{(q)}\}_{i=1}^{M}$, $q = 1, \ldots, Q$. The conditional entropy $h(Y|T)$

is given by

$$\begin{aligned} h(Y|T) &= h(Y|X_1X_2\cdots X_Q) \\ &= \sum_{i_1=1}^{M}\cdots\sum_{i_Q=1}^{M} p_{i_1\cdots i_Q} h(Y|X_1=x_{i_1},\ldots,X_Q=x_{i_Q}) \\ &= \sum_{i_1=1}^{M}\cdots\sum_{i_Q=1}^{M} p_{i_1\cdots i_Q} h_{i_1\cdots i_Q}, \end{aligned} \tag{11}$$

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

There are M^Q variables involved in the maximization problem (10). 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 (10).

Theorem 1: The capacity of the associated regular channel is achieved by using at most $MQ-Q+1$ out of M^Q inputs with nonzero probabilities.

Proof: Denote by $\{\hat{p}_i^{(q)}\}_{i=1}^{M}$ the pmf of X_q, $q=1,2,\ldots,Q$, induced by a capacity-achieving joint pmf $\{\hat{p}_{i_1\cdots i_Q}\}_{i_1,\ldots,i_Q=1}^{M}$. We limit the search for a capacity-achieving joint pmf to those joint pmfs that yield the same marginal pmfs as $\{\hat{p}_{i_1\cdots i_Q}\}_{i_1,\ldots,i_Q=1}^{M}$. By limiting the search to this smaller set, the maximum of $I(X_1\cdots X_Q;Y)$ remains unchanged since the capacity-achieving joint pmf $\{\hat{p}_{i_1\cdots i_Q}\}_{i_1,\ldots,i_Q=1}^{M}$ is in the smaller set. But all joint pmfs in the smaller set yield the same $h(Y)$ since they induce the same marginal pmfs on $X_1,\ldots,X_Q$. Therefore, the maximization problem in (10) reduces to the linear minimization problem

$$
\begin{aligned}
\min_{p_{i_1\cdots i_Q}} \quad & \sum_{i_1=1}^{M}\cdots\sum_{i_Q=1}^{M} h_{i_1\cdots i_Q} p_{i_1\cdots i_Q} \\
\text{subject to} \quad & \\
& \sum_{i_2=1}^{M}\cdots\sum_{i_Q=1}^{M} p_{i_1\cdots i_Q} = \hat{p}_{i_1}^{(1)}, \qquad i_1 = 1,2,\ldots,M, \\
& \qquad \vdots \qquad\qquad\qquad\qquad\qquad \vdots \\
& \sum_{i_1=1}^{M}\cdots\sum_{i_{Q-1}=1}^{M} p_{i_1\cdots i_Q} = \hat{p}_{i_Q}^{(Q)}, \qquad i_Q = 1,2,\ldots,M, \\
& p_{i_1\cdots i_Q} \geq 0, \qquad i_1,\ldots,i_Q = 1,2,\ldots,M.
\end{aligned} \tag{12}
$$

There are MQ equality constraints in (12) out of which $MQ - Q + 1$ are linearly independent. From the theory of linear programming, the minimum of (12), and hence the maximum of $I(X_1 \cdots X_Q; Y)$, is achieved by a feasible solution with at most $MQ-Q+1$ nonzero variables. ■

Theorem 1 states that at most $MQ - Q + 1$ out of M^Q inputs 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 inputs must be used to achieve the capacity. If we knew the marginal pmfs for $X_1, \ldots, X_Q$ induced by a capacity-achieving joint pmf, we could obtain the capacity-achieving joint pmf itself by solving the linear program (12).

A. *The Noise-Free Channel*

We consider a special case where the noise power is zero in (5). 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

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

set of all possible channel output values into M subsets of size Q corresponding to M possible inputs, 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.

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.,

$$\mathcal{Y}^{(q)} = \{x_1 + s_q, x_2 + s_q, \ldots, x_M + s_q\}, \quad q = 1, \ldots, Q. \tag{13}$$

The union of $\mathcal{Y}^{(q)}$s is the set of all possible outputs of the noise-free channel. Our purpose is to find M mutually-disjoint (multi-)sets [2] 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 inputs 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.

Unfortunately, we cannot always find M mutually disjoint (multi-)sets of size Q. 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. In fact, by entropy calculations we can show that the capacity of the channel in this example is less than 2 bits. However, if we put 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

[2] A multi-set differs from a set in that each member has a multiplicity. For example, $\{1, 2, 2, 3\}$ is a multi-set of size four.

arithmetic progression. Then the capacity of the noise-free channel

$$Y = X + S, \tag{14}$$

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

Proof: 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). 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. 3. 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

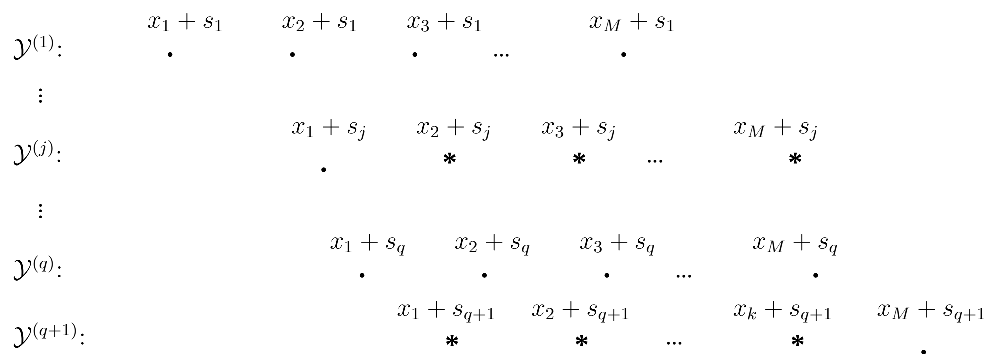

Fig. 3. The elements of $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(q+1)}$ shown as shifted version of each other. The elements of $\mathcal{Y}^{(q+1)}$ up to $x_k + s_{q+1}$ appear in $\mathcal{Y}^{(j)}$. In the figure, $k = M - 1$.

$x_k + s_{q+1}$ 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 $\{x_{k+1} + s_{q+1}, \dots, x_M + s_{q+1}\}$ 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 of $Q = 2$ [9].

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 elements of $\mathcal{S}$.

The proof of theorem 2 is actually a constructive algorithm for finding M (out of M^Q) inputs 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 $\mathcal{X}$, $q = 1, \dots, Q$. This is due to the fact that each (multi-)set contains one element from each $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(Q)}$. Therefore, a uniform distribution on the M Q-tuples induces uniform distribution on $X_1, \dots, X_Q$.

As an example, consider a noise-free channel with the input alphabet $\mathcal{X} = \{-3, -1,$

$+1, +3\}$ and the interference alphabet $\mathcal{S} = \{-2, 0, +1\}$. We have

$$\mathcal{Y}^{(1)} = \{-5, -3, -1, +1\}, \tag{15}$$

$$\mathcal{Y}^{(2)} = \{-3, -1, +1, +3\}, \tag{16}$$

$$\mathcal{Y}^{(3)} = \{-2, 0, +2, +4\}. \tag{17}$$

At the first iteration, we start with disjoint sets

$$\{-5\}, \{-3\}, \{-1\}, \{+1\}. \tag{18}$$

At the second iteration, we include the elements of $\mathcal{Y}^{(2)}$ in the current disjoint sets to obtain the disjoint (multi-)sets

$$\{-5, +3\}, \{-3, -3\}, \{-1, -1\}, \{+1, +1\}. \tag{19}$$

Since the elements of $\mathcal{Y}^{(3)}$ have not appeared in the current disjoint (multi-)sets, we can include the elements of $\mathcal{Y}^{(3)}$ in the current disjoint (multi-)sets arbitrarily. For example, we obtain the disjoint (multi-)sets

$$\{-5, +3, +4\}, \{-3, -3, -2\}, \{-1, -1, 0\}, \{+1, +1, +2\}. \tag{20}$$

The corresponding four triples are $(-3, +3, +3)$,$(-1, -3, -3)$,$(+1, -1, -1)$, and $(+3, +1, +1)$. We use these triples to send any of four different messages A, B, C, D, respectively. Message A can produce the outputs -5 or $+3$ or $+4$. Message B can produce -3 or -2. Message C produces -1 or 0 and message D produces $+1$ or $+2$. So, the output alphabet is partitioned into four mutually disjoint sets corresponding to four different messages. Therefore, error-free transmission is possible in one use of the channel.

V. Uniform Transmission

In the sequel, we study the maximization of the transmission 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.,

$$p_i^{(1)} = p_i^{(2)} = \cdots = p_i^{(Q)} = \frac{1}{M}, \qquad i = 1, 2, \ldots, M, \tag{21}$$

for which we show how to obtain the optimal input probability assignment. We call a transmission scheme that induces uniform distribution on $X_1, \ldots, X_Q$ as uniform transmission. Uniform distribution for $X_1, \ldots, X_Q$ implies uniform distribution for X, the input to the state-dependent channel defined in (5).

The optimality of uniform pmfs for $X_1, \ldots, X_Q$ was established for the asymptotic case of noise-free channel in the previous section (provided that we can find M Q-tuples such that the corresponding (multi)-sets are mutually disjoint).

Considering the constraints in (21), the maximization of $I(X_1 \cdots X_Q; Y)$ is reduced to the linear minimization problem

$$
\begin{aligned}
\min_{p_{i_1 \cdots i_Q}} \quad & \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} h_{i_1 \cdots i_Q} p_{i_1 \cdots i_Q} \\
\text{subject to} \quad & \\
& \sum_{i_2=1}^{M} \cdots \sum_{i_Q=1}^{M} p_{i_1 \cdots i_Q} = \frac{1}{M}, \qquad i_1 = 1, 2, \ldots, M, \\
& \quad \vdots \qquad\qquad\qquad\qquad\qquad \vdots \\
& \sum_{i_1=1}^{M} \cdots \sum_{i_{Q-1}=1}^{M} p_{i_1 \cdots i_Q} = \frac{1}{M}, \qquad i_Q = 1, 2, \ldots, M, \\
& p_{i_1 \cdots i_Q} \geq 0, \qquad i_1, \ldots, i_Q = 1, 2, \ldots, M.
\end{aligned} \tag{22}
$$

The same argument used in the last part of the proof of theorem 1 can be used to show that the maximum is achieved by using at most $MQ - Q + 1$ inputs 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

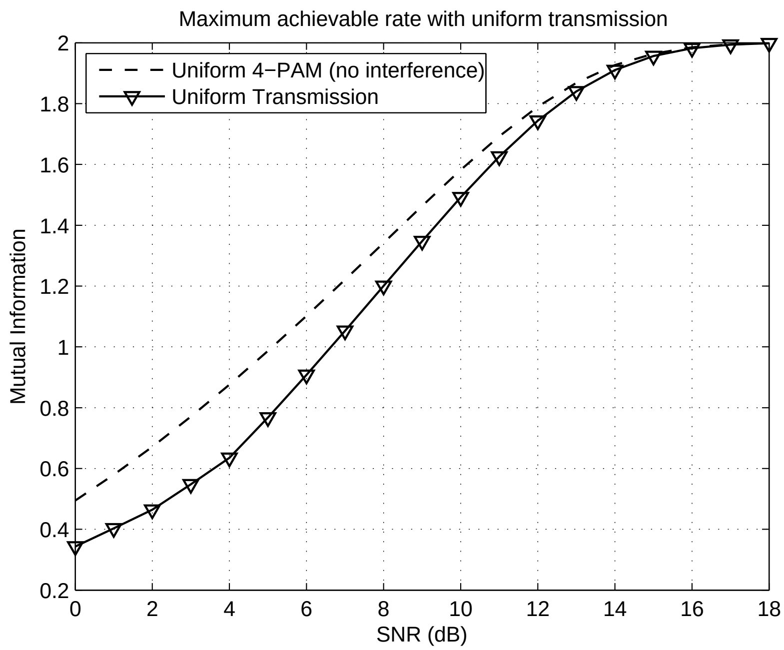

Fig. 4. Maximum mutual information vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-3, -1, +1, +3\}$ and $r_1 = r_2 = r_3 = r_4 = \frac{1}{4}$.

signal points $x_1, x_2, \ldots, x_M$. The optimal probability assignment is obtained by solving the linear programming problem (22) using the simplex method [8].

Fig. 4 depicts the maximum mutual information (for uniform transmission) 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

$$Y = X + N, \tag{23}$$

with equiprobable input alphabet $\{-3, -1, +1, +3\}$ is plotted for comparison purposes. It can be observed that the achievable rate approaches $\log_2 4 = 2$ bit per channel use as SNR increases complying with the fact that we established in section IV for the noise-free channel.

The equality constraints of (22) can be written in matrix form as

$$\mathbf{A}\mathbf{p} = \mathbf{1}, \tag{24}$$

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.

If the number of interference levels is two, i.e., $Q = 2$, we can make a stronger statement than corollary 1. In [9], it is shown that the maximum of $I(X_1 X_2; Y)$ 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 inputs of the associated channel with probability $\frac{1}{M}$. So, for the case $Q = 2$, if we add the integrality constraint to the set of constraints in (24), i.e., $\mathbf{p}$ is integer, we still obtain the same optimal solution. The resulting integer linear optimization problem is called the *assignment problem* [7], which can be solved using low-complexity algorithms such as the *Hungarian method* [8].

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 $\frac{1}{M}$.

We may add the integrality constraint for arbitrary Q. The resulting optimization problem is called the *multi-dimensional assignment problem* [10]. As explained earlier, this simplifies the encoding operation, however there may be some loss in rate unlike the case of $Q = 2$.

It is worth mentioning that with the integrality constraint, the optimal solution of (22) is a joint pmf of $X_1, \cdots, X_Q$ for which $X_2, \cdots, X_Q$ can be presented as a function of X_1.

In the sequel, we further investigate the optimal solution of (22). 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.,

$$h_{i_1 \cdots i_Q} = g(x_{i_1} - x_{i_2}, x_{i_1} - x_{i_3}, \ldots, x_{i_1} - x_{i_Q}), \tag{25}$$

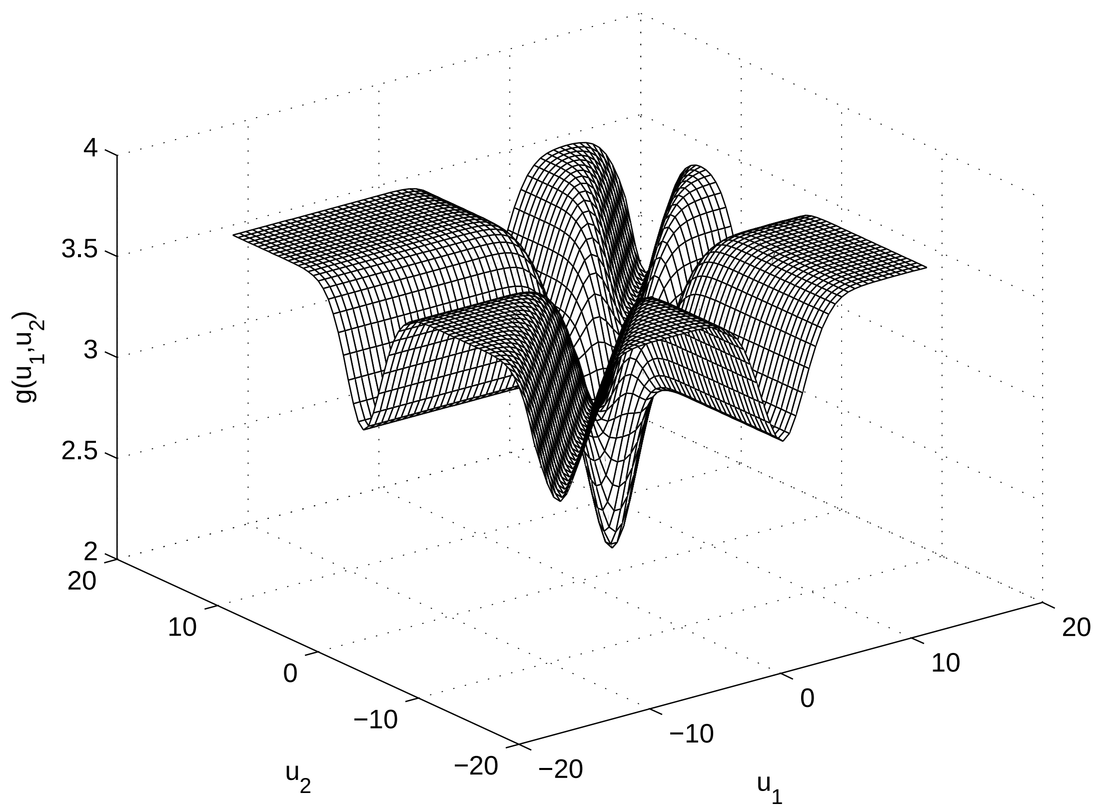

Fig. 5. The plot of $g(u_1, u_2)$ with parameters $r_1 = r_2 = r_3 = \frac{1}{3}, s_1 = -2, s_2 = 0, s_3 = +2, P_N = 1$.

where g is a given by

$$g(u_1, \ldots, u_{Q-1}) = -\int_{-\infty}^{+\infty} \left(r_1 f_N(z) + \sum_{q=2}^{Q} r_q f_N(z + u_{q-1} + s_1 - s_q) \right) \times \log_2 \left(r_1 f_N(z) + \sum_{q=2}^{Q} r_q f_N(z + u_{q-1} + s_1 - s_q) \right) dz. \quad (26)$$

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. 5. It can also be shown that g is lower-bounded by the differential entropy of the noise, $h(N)$, and upper-bounded by $h(N) + H(S)$, where $H(S)$ is the entropy of the discrete interference.

Theorem 3: If g is convex in the $(Q-1)$-cube $\{(u_1, \ldots, u_{Q-1}) :$

$x_1 - x_M \le u_i \le x_M - x_1, i = 1, 2, \ldots, Q-1\}$, then the optimal solution to (22) is

$$\tilde{p}_{i_1 \cdots i_Q} = \begin{cases} \frac{1}{M}, & \text{if} \quad i_1 = \cdots = i_Q \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

Proof: Define random variables $U_i = X_1 - X_{i+1}$, $i = 1, \ldots, Q-1$. The objective function in (22) can be written as

$$\begin{aligned} \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} p_{i_1 \cdots i_Q} h_{i_1 \cdots i_Q} &= \sum_{i_1=1}^{M} \cdots \sum_{i_Q=1}^{M} \text{Pr}\left\{X_1 = x_{i_1}, \ldots, X_Q = x_{i_Q}\right\} \times \\ &\qquad g(x_{i_1} - x_{i_2}, \ldots, x_{i_1} - x_{i_Q}) \\ &= \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \text{Pr}\{X_1 = x_{i_1}, X_2 = x_{i_1} - u_{j_1}, \ldots, \\ &\qquad X_Q = x_{i_1} - u_{j_{Q-1}}\} \times g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\ &= \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \text{Pr}\{X_1 = x_{i_1}, X_1 - X_2 = u_{j_1}, \ldots, \\ &\qquad X_1 - X_Q = u_{j_{Q-1}}\} \times g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\ &= \sum_{j_1} \cdots \sum_{j_{Q-1}} \sum_{i_1=1}^{M} \text{Pr}\{X_1 = x_{i_1}, U_1 = u_{j_1}, \ldots, \\ &\qquad U_{Q-1} = u_{j_{Q-1}}\} \times g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\ &= \sum_{j_1} \cdots \sum_{j_{Q-1}} \text{Pr}\left\{U_1 = u_{j_1}, \ldots, U_{Q-1} = u_{j_{Q-1}}\right\} \times \\ &\qquad g(u_{j_1}, \ldots, u_{j_{Q-1}}) \\ &= \text{E}[g(U_1, \ldots, U_{Q-1})], \end{aligned} \tag{28}$$

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

$$\begin{aligned} \text{E}[g(U_1, \ldots, U_{Q-1})] &\ge g\left(\text{E}[U_1, \ldots, U_{Q-1}]\right) \\ &= g(0, \ldots, 0). \end{aligned} \tag{29}$$

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

$$X_1 = X_2 = \cdots = X_Q. \tag{30}$$

The joint pmf in (27) satisfies both the constraints in (22) and (30), 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 [9]

$$x_M - x_1 \leq s_1 - s_2 + u^* \sqrt{P_N}, \tag{31}$$

where $u^* \approx 1.636$. In general ($Q \geq 2$), when the power of the noise P_N is sufficiently large, the function g will be convex in the $(Q-1)$-cube.

Theorem 3 has an interesting interpretation: Given the condition of the theorem 3 satisfied, the optimal precoder sends the same signal in the channel regardless of the current interference symbol. In other words, the optimal precoder for uniform transmission ignores the interference. In fact, 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 3 holds. Note that this might not hold true for a capacity achieving coding scheme without any constraints on the marginal pmfs of $X_1, \ldots, X_Q$.

VI. Optimal Precoding

The general structure of a communication system for the channel (5) is shown in fig. 6. In fact, fig. 6 is the same as fig. 2 for the special case of the state-dependent channel defined in (5). Any encoding and decoding scheme for the associated channel can be translated to an encoding and decoding scheme for the original channel (5). A message w is encoded to a block of length n of indices $t \sim (x_{i_1}, x_{i_2}, \ldots, x_{i_Q})$. According to Corollary 1, at most $MQ - Q + 1$ out of M^Q indices are needed to achieve the maximum rate for uniform transmission. Those indices are obtained by solving the linear programming problem (22). For each t, the precoder sends the qth component of t if the current interference symbol is s_q, $q = 1, \ldots, Q$. Based on the received signal Y, the receiver decodes $\hat{w}$ as the transmitted message.

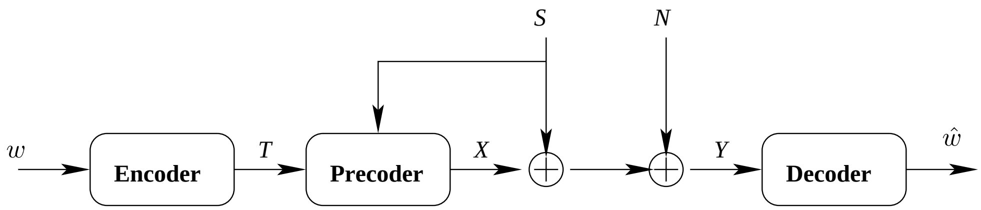

Fig. 6. General structure of the communication system for channels with causally-known discrete interference.

VII. Extension to Continuous Input Alphabet

We can extend the uniform transmission scheme introduced in section V to the case that 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

$$\begin{aligned}
\min_{f_{X_1 \cdots X_Q}} \quad & \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} h(x_1, \ldots, x_Q) f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_1 \cdots dx_Q \\
\text{subject to} \quad & \\
& \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_2 \cdots dx_Q = \frac{1}{\Delta}, \qquad x_1 \in A_\Delta, \\
& \quad \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad \vdots \\
& \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \cdots \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) dx_1 \cdots dx_{Q-1} = \frac{1}{\Delta}, \qquad x_Q \in A_\Delta, \\
& f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) \geq 0, \qquad x_1, \ldots, x_Q \in A_\Delta,
\end{aligned} \tag{32}$$

where $h(x_1, \ldots, x_Q) = h(Y | X_1 = x_1, \ldots, X_Q = x_Q)$. We are interested in solutions to (32) that are of the form

$$f_{X_1 \cdots X_Q}(x_1, \ldots, x_Q) = \frac{1}{\Delta} \delta\left(|x_2 - \xi_1(x_1)| + |x_3 - \xi_2(x_1)| + \cdots + |x_Q - \xi_{Q-1}(x_1)|\right), \tag{33}$$

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 (33) describes random variables $X_1, \ldots, X_Q$, $Q-1$ of which are functions of the other random variable. It is easy to check that (33) satisfies the constraints in (32). The objective value corresponding to the joint pdf (33) is

$$\frac{1}{\Delta}\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} h\left(x_1, \xi_1(x_1), \ldots, \xi_{Q-1}(x_1)\right) dx_1, \tag{34}$$

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

A. Comparison to Modulo Operation

The modulo operation was proposed in [4] as a precoding method for channels with known (discrete or continuous) interference at the transmitter. Based on the input symbol of the associated channel V and the current interference symbol S, the precoder sends [4]

$$X = [V - \alpha S] \mod \Delta, \tag{35}$$

where α is a constant. It is assumed that V is distributed uniformly in A_Δ.

In our setting where the interference is discrete with Q levels, we have

$$\begin{aligned} X_1 &= [V - \alpha s_1] \mod \Delta, \\ X_2 &= [V - \alpha s_2] \mod \Delta, \\ &\vdots \\ X_Q &= [V - \alpha s_Q] \mod \Delta, \end{aligned} \tag{36}$$

where X_q is the random variable that represents the channel input when the current interference symbol is s_q, $q = 1, \ldots, s_Q$. Since V is uniformly distributed in A_Δ, $X_1, \ldots, X_Q$ will be uniformly distributed in A_Δ. Therefore, modulo operation 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 the following

$$\begin{aligned} X_2 &= [X_1 + \alpha(s_1 - s_2)] \mod \Delta, \\ X_3 &= [X_1 + \alpha(s_1 - s_3)] \mod \Delta, \\ &\vdots \\ X_Q &= [X_1 + \alpha(s_1 - s_Q)] \mod \Delta. \end{aligned} \tag{37}$$

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 operation scheme fits in the category of joint pdfs in (33). The bijective functions corresponding to the modulo operation scheme are given by (37). These functions are circular shifts of each other. The modulo operation gives a feasible solution to (32) which is not (necessarily) an optimal solution.

VIII. Conclusion

In this paper, we investigated M-ary signal transmission over AWGN channel with additive Q-level interference, where the sequence of interference symbols is known causally at the transmitter. According to Shannon's theorem for channels with side information at the transmitter, the associated channel will be a channel with M^Q inputs. We proved that by choosing at most $MQ - Q + 1$ inputs with positive probability the capacity is achievable. Then we focused on transmission schemes that induce uniform marginal pmfs on $X_1, \cdots, X_Q$. 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 (22). In the special case where the function g is convex in a specific region, the optimal solution is given by theorem 3. The optimal solution determines the optimal precoding to be used in the general structure of the communication system for the channel with causally-known discrete interference.

APPENDIX

A.

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

$$\tilde{Y} = \tilde{S} + N. \tag{38}$$

Since

$$0 \leq I(\tilde{Y};\tilde{S}) \leq H(\tilde{S}), \tag{39}$$

we have

$$0 \leq h(\tilde{Y}) - h(\tilde{Y}|\tilde{S}) \leq H(\tilde{S}), \tag{40}$$

or equivalently,

$$\begin{aligned} h(N) \leq h(\tilde{Y}) &\leq h(N) + H(\tilde{S}) \\ &= h(N) + H(S). \end{aligned} \tag{41}$$

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

[6] 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.

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

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

[9] H. Farmanbar and A. K. Khandani, “Precoding for channels with two-level interference,” Technical Report UW-E&CE #2006-23.

[10] W. P. Pierskalla, “The multidimensional assignment problem,” *Operations Research* 16, 1968, p. 422-431