University of Waterloo

Precoding for Channels with Two-Level 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 #2006-23
October 18, 2006

Precoding for Channels with Two-Level 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 two-level 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 $2M-1$ out of M^2 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 which we show that by using exactly M out of M^2 inputs of the associated channel the maximum is achievable. Based on this, 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 conclude this paper in section VII.

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

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

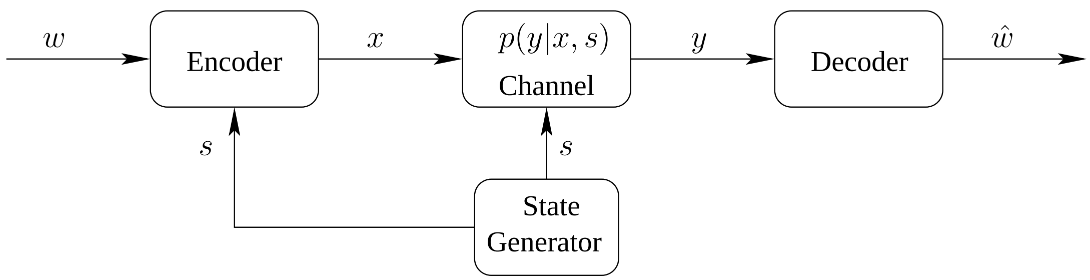

Fig. 1. SD-DMC with state information at the encoder.

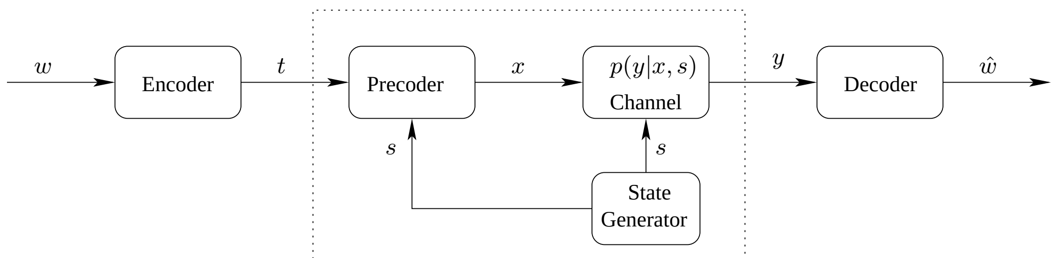

Fig. 2. The associated regular DMC.

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, $s = 1, 2, \ldots, |S|$, 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 $(X_1, \ldots, X_{|S|})$, where X_s is the input to the state-dependent channel when the state is $s, s = 1, \ldots, |S|$.

III. The Channel Model

We consider M-ary signaling over channel

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

where X is the channel input, which takes on values in the real set $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$, where $x_1 < x_2 < \cdots < x_M$, 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 $\mathcal{S} = \{s_1, s_2\}$, where $s_1 < s_2$, with probabilities r_1, r_2, 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 $f(y|x, s)$ is used instead of the transition probabilities. We denote the input to the associated channel by T, which can also be represented as the pair (X_1, X_2), where X_1 and X_2 take values in $\mathcal{X}$.

The likelihood function for the associated channel is given by

$$\begin{aligned} f_{Y|T}(y|t) &= r_1 f_{Y|T,S}(y|t,s_1) + r_2 f_{Y|T,S}(y|t,s_2) \\ &= r_1 f_{Y|X,S}(y|x_i,s_1) + r_2 f_{Y|X,S}(y|x_j,s_2) \\ &= r_1 f_N(y-x_i-s_1) + r_2 f_N(y-x_j-s_2), \end{aligned} \quad (6)$$

where f_N denotes the pdf of noise N, and t represents (x_i, x_j). The pdf of Y is then given by

$$\begin{aligned} f_Y(y) &= \sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij} \left(r_1 f_N(y-x_i-s_1) + r_2 f_N(y-x_j-s_2)\right) \\ &= r_1 \sum_{i=1}^{M} p_i^{(1)} f_N(y-x_i-s_1) + r_2 \sum_{j=1}^{M} p_j^{(2)} f_N(y-x_j-s_2), \end{aligned} \quad (7)$$

where $p_{ij} = \Pr\{X_1 = x_i, X_2 = x_j\}$, $p_i^{(1)} = \Pr\{X_1 = x_i\}$, and $p_j^{(2)} = \Pr\{X_2 = x_j\}, i, j = 1, 2, \ldots, M$.

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 $I(T;Y) = I(X_1X_2;Y)$ over the joint pmf values $p_{ij} = \Pr\{X_1 = x_i, X_2 = x_j\}$, where x_i and x_j belong to $\mathcal{X}$, i.e.,

$$C = \max_{p_{ij}} \; I(X_1X_2;Y). \quad (8)$$

The mutual information between T and Y is the difference between differential entropies $h(Y)$ and $h(Y|T)$. It can be seen from (7) that $f_Y(y)$, and hence $h(Y)$, are uniquely determined by the marginal pmfs $\{p_i^{(1)}\}_{i=1}^M$ and $\{p_j^{(2)}\}_{j=1}^M$. The conditional entropy $h(Y|T)$ is given by

$$\begin{aligned} h(Y|T) &= h(Y|X_1X_2) \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij} h(Y|X_1 = x_i, X_2 = x_j) \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij} h_{ij}, \end{aligned} \quad (9)$$

where $h_{ij} = h(Y|X_1 = x_i, X_2 = x_j)$.

There are M^2 variables involved in the maximization problem (8). 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 optimal value of (8).

Theorem 1: The capacity of the associated regular channel is achieved by using at most $2M-1$ out of M^2 inputs with nonzero probabilities.

Proof: Denote by $\{\hat{p}_i^{(1)}\}_{i=1}^M$ and $\{\hat{p}_i^{(2)}\}_{i=1}^M$ the pmf of X_1, and X_2, induced by a capacity-achieving joint pmf $\{\hat{p}_{ij}\}_{i,j=1}^M$. We limit the search for a capacity-achieving joint pmf to those joint pmfs that yield the same marginal pmfs as $\{\hat{p}_{ij}\}_{i,j=1}^M$. By limiting the search to this smaller set, the maximum of $I(X_1X_2;Y)$ remains unchanged since the capacity-achieving joint pmf $\{\hat{p}_{ij}\}_{i,j=1}^M$ 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, X_2. Therefore, the maximization problem in (8) reduces to the linear minimization problem

$$\begin{aligned} \min_{p_{ij}} \quad & \sum_{i=1}^M \sum_{j=1}^M h_{ij} p_{ij} \\ \text{subject to} \quad & \\ & \sum_{j=1}^M p_{ij} = \hat{p}_i^{(1)}, \quad i = 1, 2, \ldots, M, \\ & \sum_{i=1}^M p_{ij} = \hat{p}_j^{(2)}, \quad j = 1, 2, \ldots, M, \\ & p_{ij} \geq 0, \qquad i, j = 1, 2, \ldots, M. \end{aligned} \tag{10}$$

There are $2M$ equality constraints in (10) out of which $2M-1$ are linearly independent. From the theory of linear programming, the minimum of (10), and hence the maximum of $I(X_1X_2;Y)$, is achieved by a feasible solution with at most $2M-1$ nonzero variables. ■

Theorem 1 states that at most $2M-1$ out of M^2 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 and X_2 induced by a capacity-achieving joint pmf, we could obtain the capacity-achieving joint pmf itself by solving the linear program (10).

We consider a special case where the noise power is zero in (5), for which we will obtain the capacity. In the absence of noise, the channel output Y can take on at most $2M$ different values. If Y takes on exactly $2M$ different values, then it is easy to see that the capacity is $\log_2 M$ bits even if the interference sequence is unknown to the encoder: The decoder just needs to partition the set of all possible channel outputs into M subsets of size 2 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 alphabet can be less than $2M$, the same result holds:

Theorem 2: The capacity of the noise-free channel

$$Y = X + S, \tag{11}$$

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

Proof: We show that there exists a coding scheme that uses just M inputs of the associated channel with probability $\frac{1}{M}$ and achieves the rate $\log_2 M$ bits in one use of the channel. It is sufficient to show that there exist M pairs (x_i, x_j) of elements of $\mathcal{X}$, such that the corresponding (multi-)sets, $\{x_i + s_1, x_j + s_2\}$, are mutually disjoint. Denote by $Y^{(i)}$ the set of the possible sums when the interference symbol is s_i, $i = 1, 2$. We must match the elements of $Y^{(1)}$ to the elements of $Y^{(2)}$ such that the resulting M (multi-)sets of size 2 are mutually disjoint. This can be done by matching those elements of $Y^{(1)}$ and $Y^{(2)}$ that are equal and matching the other elements of the two sets arbitrarily. Once the mutually disjoint (multi-)sets are found, we can obtain the corresponding pairs. ■

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 an arbitrary varying sequence of s_1s and s_2s.

As an example of the noise-free case, consider a channel with input alphabet $\mathcal{X} = \{-3, -1, +1, +3\}$ and interference alphabet $\mathcal{S} = \{-2, +2\}$. The output alphabet will be $\mathcal{Y} = \{-5, -3, -1, +1, +3, +5\}$. Based on the algorithm given in the proof of theorem 2, we may choose 4 pairs $(-3, +1)$, $(-1, +3)$, $(+1, -3)$, $(+3, -1)$. We use these pairs to send any of four different messages A, B, C, D, respectively. Message A can produce the outputs -5 or $+3$. Message B can produce -3 or $+5$. Message C produces -1, and message D produces $+1$. 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.

In the noise-free case, to achieve the capacity $\log_2 M$, the M pairs must be chosen with equal probabilities. This induces uniform distribution on X_1 and X_2.

V. Uniform Transmission

In the sequel, we study the maximization of the transmission rate $I(X_1 X_2; Y)$ over joint pmfs $\{p_{ij}\}_{i,j=1}^{M}$ that induce uniform marginal distributions on X_1 and X_2, i.e.,

$$p_i^{(1)} = p_i^{(2)} = \frac{1}{M}, \qquad i = 1, 2, \ldots, M, \tag{12}$$

for which we show how to obtain the optimal input probability assignment. We call a transmission scheme that induces uniform distribution on X_1 and X_2 as uniform transmission. Uniform distribution for X_1 and X_2 implies uniform distribution for X, the input to the state-dependent channel defined in (5).

The optimality of uniform pmfs for X_1 and X_2 was established for the asymptotic case of noise-free channel in the previous section. Furthermore, we will show that (12) results in simplification of the encoding scheme (mapping the message to the set of indices $\mathcal{T}$) for the associated channel. In fact, we will show that the rate-maximizing joint pmf with the constraints in (12) is uniform on a subset of size M of $\mathcal{T}$, which simplifies the encoding operation.

Theorem 3: The maximum of $I(X_1X_2;Y)$ over $\{p_{ij}\}_{i,j=1}^{M}$ 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 $\frac{1}{M}$.

Proof: Considering the constraints in (12), the maximization of $I(X_1X_2;Y)$ is reduced to the linear minimization problem

$$\begin{aligned} \min_{p_{ij}} \quad & \sum_{i=1}^{M}\sum_{j=1}^{M} h_{ij}p_{ij} \\ \text{subject to} \quad & \\ & \sum_{j=1}^{M} p_{ij} = \frac{1}{M}, \quad i = 1,2,\ldots,M, \\ & \sum_{i=1}^{M} p_{ij} = \frac{1}{M}, \quad j = 1,2,\ldots,M, \\ & p_{ij} \geq 0, \qquad\quad i,j = 1,2,\ldots,M. \end{aligned} \tag{13}$$

The equality constraints of (13) can be written in matrix form as

$$\mathbf{A}\mathbf{p} = \mathbf{1}, \tag{14}$$

where $\mathbf{A}$ is a zero-one $2M \times M^2$ matrix, $\mathbf{p} = M\left[\mathbf{p}_1^T\mathbf{p}_2^T \ldots \mathbf{p}_M^T\right]^T$, where $\mathbf{p}_m = [p_{m1}p_{m2}\ldots p_{mM}]^T$, $m = 1,2,\ldots,M$, and $\mathbf{1}$ is the all-one $2M \times 1$ vector. It is easy to check that $\mathbf{A}$ is the vertex-edge incidence matrix of $K_{M,M}$, the complete bipartite graph with $2M$ vertices. Therefore, $\mathbf{A}$ is a totally unimodular matrix [7]. Hence, the extreme points of the feasible region $F = \{\mathbf{p} : \mathbf{A}\mathbf{p} = \mathbf{1}, \mathbf{p} \geq \mathbf{0}\}$ are integer vectors. Since the optimal value of a linear optimization problem is attained at one of the extreme points of its feasible region, the minimum in (13) is achieved at an all-integer vector $\mathbf{p}^*$. Considering that $\mathbf{p}^*$ satisfies (14), it can only be a zero-one vector with exactly M ones. ■

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 (14), 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].

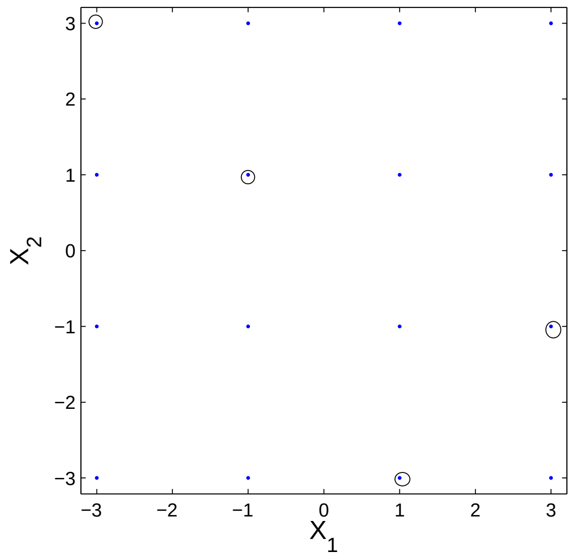

Fig. 3. Optimal solution for 4-PAM input with parameters $r_1 = r_2 = \frac{1}{2}s_1 = -2, s_2 = +2, P_N = 1$.

According to theorem 3, to achieve the maximum of $I(X_1X_2;Y)$ over $\{p_{ij}\}_{i,j=1}^{M}$ with uniform marginal pmfs for X_1 and X_2, we need to use only M out of M^2 inputs of the associated channel with probabilities $\frac{1}{M}$. This result is independent of the value of the coefficients $\{h_{ij}\}$. However, which probability assignment with M nonzero elements is optimum depends on the coefficients $\{h_{ij}\}$. The coefficient h_{ij} is determined by the interference levels s_1, s_2, the probability of interference levels r_1, r_2, the noise power P_N, and the signal points $x_1, x_2, \ldots, x_M$.

As example, the optimal solutions for two different scenarios with 4-PAM constellation ($\mathcal{X} = \{-3, -1, +1, +3\}$) are illustrated in figs. 3 and 4. The points circled in the array correspond to the inputs to the associated channel that must be chosen with probability $\frac{1}{4}$.

Fig. 5 depicts the maximum mutual information (for uniform transmission) 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

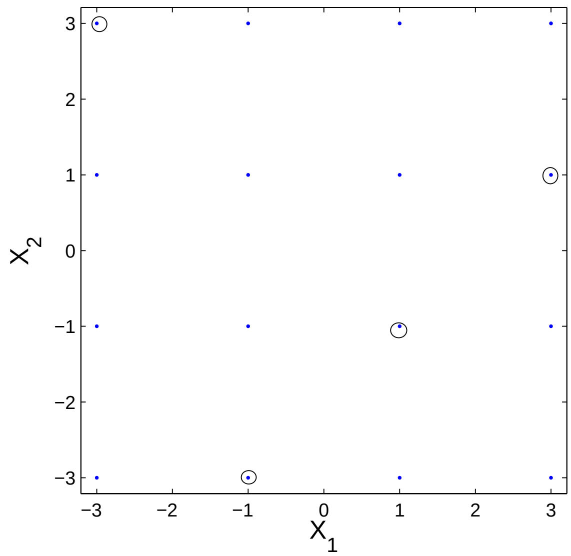

Fig. 4. Optimal solution for 4-PAM input with parameters $r_1 = r_2 = \frac{1}{2}, s_1 = -2.5, s_2 = +2.5, P_N = 9$.

equiprobable input alphabet $\{-1, +1\}$ is plotted for comparison purposes. As it can be seen, for low SNRs, the input probability assignment $p_{11} = p_{22} = \frac{1}{2}$ is optimal, whereas at high SNRs, 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 section IV for the noise-free channel.

Fig. 6 depicts the maximum mutual information vs. SNR for the channel with $\mathcal{X} = \{-3, -1, +1, +3\}$, and equiprobable interference symbols from the alphabet $\mathcal{S} = \{-2, +2\}$.

In the sequel, we further investigate the optimal solution of (13). It can be easily shown that the conditional entropy $h_{ij} = h(Y|X_1 = x_i, X_2 = x_j)$ is a function of $x_i - x_j$, i.e.,

$$h_{ij} = g(x_i - x_j), \tag{15}$$

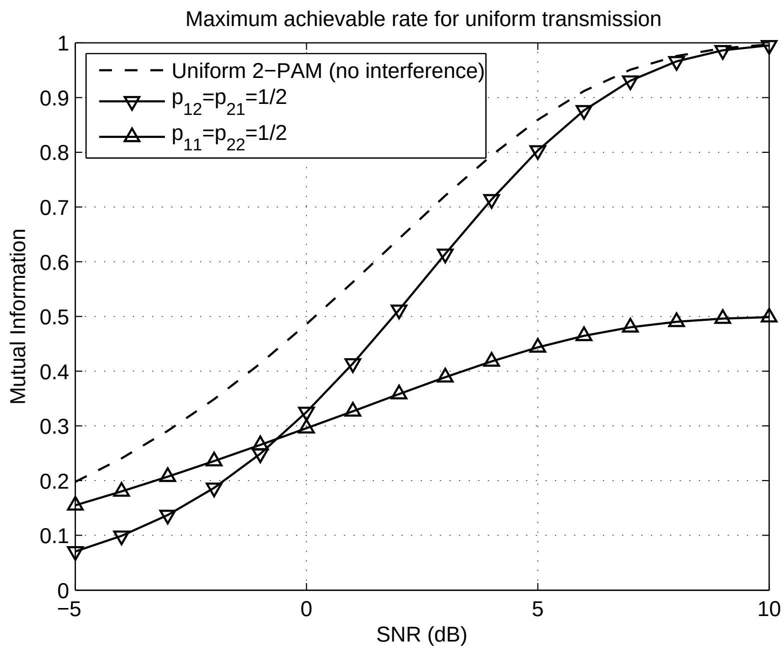

Fig. 5. Maximum mutual information vs. SNR for the channel with $\mathcal{X} = \mathcal{S} = \{-1, +1\}$ and $r_1 = r_2 = \frac{1}{2}$.

where $g(.)$ is given by

$$g(u) \quad = -\int_{-\infty}^{+\infty} \left[\frac{r_1}{\sqrt{2\pi P_N}} \exp\left(-\frac{(z-u+s_2-s_1)^2}{2P_N}\right) + \frac{r_2}{\sqrt{2\pi P_N}} \exp\left(-\frac{z^2}{2P_N}\right)\right] \times \log_2 \left[\frac{r_1}{\sqrt{2\pi P_N}} \exp\left(-\frac{(z-u+s_2-s_1)^2}{2P_N}\right) + \frac{r_2}{\sqrt{2\pi P_N}} \exp\left(-\frac{z^2}{2P_N}\right)\right] dz. \quad (16)$$

The plot of $g(.)$ for $r_1 = \frac{1}{2}, r_2 = \frac{1}{2}, s_1 = -2, s_2 = +2, P_N = 1$ is shown in fig. 7. In part A of the Appendix, 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 two-level interference.

Theorem 4: If g is convex in the interval $[x_1 - x_M, x_M - x_1]$, then

$$\tilde{p}_{ij} = \begin{cases} \frac{1}{M}, & \text{if} \quad i = j \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

is the optimal solution to (13).

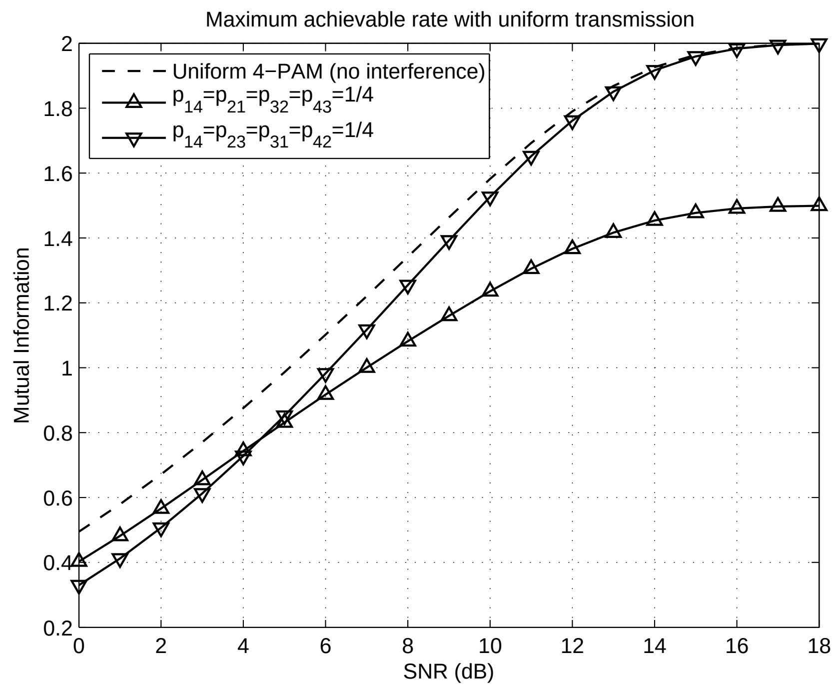

Fig. 6. maximum mutual information vs. SNR for the channel with $\mathcal{X} = \{-3, -1, +1, +3\}$, $\mathcal{S} = \{-2, +2\}$ and $r_1 = r_2 = \frac{1}{2}$.

Proof: Define random variable $U = X_1 - X_2$. The objective function in (13) can

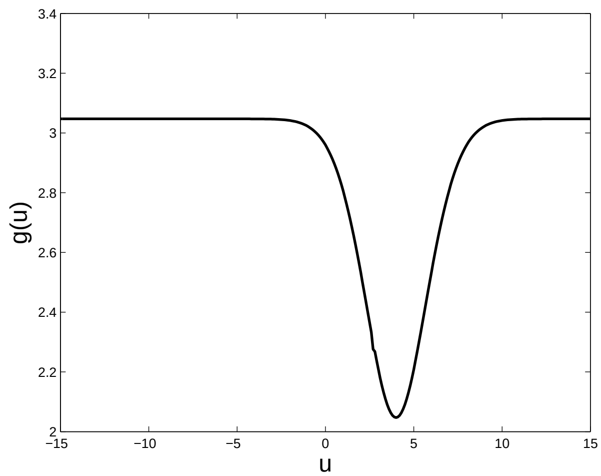

Fig. 7. The plot of $g(u)$ for $r_1 = \frac{1}{2}, r_2 = \frac{1}{2}, s_1 = -2, s_2 = +2, P_N = 1$.

be written as

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij}h_{ij} &= \sum_{i=1}^{M}\sum_{j=1}^{M} \Pr\{X_1 = x_i, X_2 = x_j\} g(x_i - x_j) \\ &= \sum_{k}\sum_{i=1}^{M} \Pr\{X_1 = x_i, X_2 = x_i - u_k\} g(u_k) \\ &= \sum_{k}\sum_{i=1}^{M} \Pr\{X_1 = x_i, X_1 - X_2 = u_k\} g(u_k) \\ &= \sum_{k}\sum_{i=1}^{M} \Pr\{X_1 = x_i, U = u_k\} g(u_k) \\ &= \sum_{k} \Pr\{U = u_k\} g(u_k) \\ &= \mathrm{E}[g(U)],\end{aligned}$$

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

$$\begin{aligned}\mathrm{E}[g(U)] &\geq g\left(\mathrm{E}[U]\right) \\ &= g(0). \end{aligned} \tag{18}$$

Equality holds when $U \equiv 0$, or equivalently,

$$X_1 = X_2. \tag{19}$$

The joint pmf in (17) satisfies both (19) and the constraints of (13). Therefore, it is the optimal solution. ■

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

$$x_M - x_1 \leq s_1 - s_2 + u_0\sqrt{P_N}, \tag{20}$$

where $u_0 \approxeq 1.636$. See the Appendix, part B.

Theorem 4 has an interesting interpretation. Given the condition of theorem 4 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 and X_2 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 holds. Note that this might not hold true for a capacity achieving coding scheme without any constraints on the marginal pmfs of X_1 and X_2.

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

$$x_i = -x_{M+1-i}, \quad i = 1, \ldots, M. \tag{21}$$

For example, a regular PAM constellation satisfies (21).

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

$$\tilde{p}_{ij} = \begin{cases} \frac{1}{M}, & \text{if} \quad i + j = M + 1 \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

is an optimal solution to (13).

Proof: We first make the observation that if $\{p_{ij}\}_{i,j=1,2,\ldots,M}$ is a feasible solution of (13), 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 (13). Furthermore, due to (21), $\{p_{ij}\}$ and $\{q_{ij}\}$ yield the same objective value. Therefore, if $\{p_{ij}\}$ is an optimal solution of (13), $\{q_{ij}\}$ will be an optimal solution too. The convex combination of these two optimal solutions, $\{\theta_{ij} = \frac{1}{2}p_{ij} + \frac{1}{2}q_{ij}\}$, is also an optimal solution which has the following symmetry property

$$\theta_{ij} = \theta_{(M+1-j)(M+1-i)}. \tag{23}$$

Suppose that a symmetric optimal solution to (13) has nonzero entries

$$p_{ij} = p_{(M+1-j)(M+1-i)} = p, \tag{24}$$

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 (13) 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. Therefore, (22) is an optimal solution of (13). ■

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

$$x_M - x_1 \leq s_2 - s_1 - u_0\sqrt{P_N}. \tag{25}$$

See the Appendix, part B.

VI. Optimal Precoding

The general structure of a communication system for the channel (5) is shown in fig. 8. In fact, fig. 8 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 $t \sim (x_i, x_j)$. According to theorem 3,

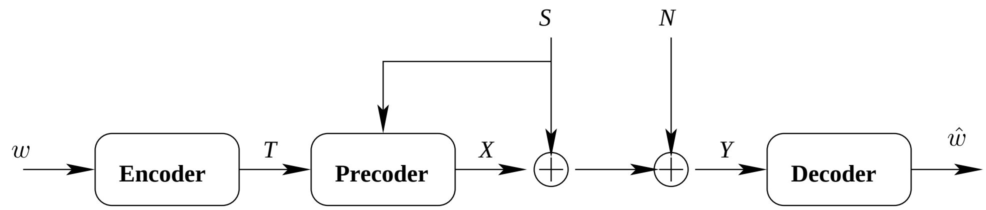

Fig. 8. General structure of a communication system for channels with causally-known discrete interference.

only M of the indices are needed to achieve the maximum rate for uniform transmission. Those M indices are obtained by solving the linear programming problem (13). For each t, the precoder sends either x_i (if $S = s_1$) or x_j (if $S = s_2$). Based on the received signal Y, the receiver decodes $\hat{w}$ as the transmitted message.

VII. Conclusion

In this paper, we investigated M-ary signal transmission over AWGN channel with additive two-level interference, where the sequence of i.i.d. interference symbols is known causally at the transmitter. According to the Shannon's theorem for channels with side information at the transmitter, the associated channel has an input alphabet of size M^2. We proved that by using at most $2M - 1$ inputs the capacity is achievable. Then we focused on transmission schemes that induce uniform marginal pmfs on X_1 and X_2. For this so called uniform transmission, the maximum rate is obtained by using only M inputs of the associated channel with probability $\frac{1}{M}$. The optimal M inputs can be obtained by solving the linear optimization problem (13). In some special cases where the function $g(u)$ is concave or convex in the interval $[x_1 - x_M, x_M - x_1]$, the optimal solution is given by theorems 4 and 5. The optimal solution determines the optimal precoding to be used in the general structure of a communication system for the channel with causally-known two-level interference.

APPENDIX

A.

Denote by $\tilde{S}$ the random variable that takes on x_i+s_1 and x_j+s_2 with probabilities r_1 and r_2, respectively. Also, denote by $\tilde{Y}$ the random variable $Y|X_1 = x_i, X_2 = x_j$. Then

$$\tilde{Y} = \tilde{S} + N. \tag{26}$$

Since

$$0 \le I(\tilde{Y};\tilde{S}) \le H(\tilde{S}), \tag{27}$$

we have

$$0 \le h(\tilde{Y}) - h(\tilde{Y}|\tilde{S}) \le H(\tilde{S}), \tag{28}$$

or equivalently,

$$\begin{aligned} h(N) \le h(\tilde{Y}) &\le h(N) + H(\tilde{S}) \\ &= h(N) + H(S). \end{aligned} \tag{29}$$

B.

The function $g(u)$ given in (16) can be considered as a function of u and parameters s_1, s_2, P_N. We have

$$\begin{aligned} g(u) &= g(u, s_1, s_2, P_N) \\ &= g(u + s_1 - s_2, 0, 0, P_N) \\ &= g\left(\frac{u + s_1 - s_2}{\sqrt{P_N}}, 0, 0, 1\right) + \log_2 \sqrt{P_N}. \end{aligned} \tag{30}$$

We can obtain the inflection points of $g(u, 0, 0, 1)$ numerically as u_0 and $-u_0$, where $u_0 \approxeq 1.636$. Therefore, the inflection points of $g(u)$ are

$$\alpha_1 = s_2 - s_1 - u_0\sqrt{P_N}, \tag{31}$$

$$\alpha_2 = s_2 - s_1 + u_0\sqrt{P_N}, \tag{32}$$

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 (20).

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 (25).

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.