Transmission over Channels with Known Two-Level Interference at the Transmitter

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— **Uniform M-PAM signal transmission over AWGN channel with two-level interference where the sequence of 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 exactly M out of M^2 inputs of the *associated* channel the capacity is achievable. Based on this, the general structure of a communication system with optimum precoding for the channel under consideration 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 AWGN channel with 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 any type of 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. 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 $\mathcal{S}$. 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 $\mathcal{T}$ that indexes all functions from the state alphabet to the input alphabet of the state-dependent channel and with the same output alphabet as 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
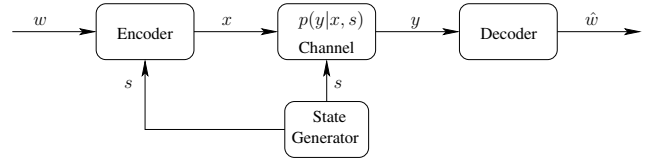

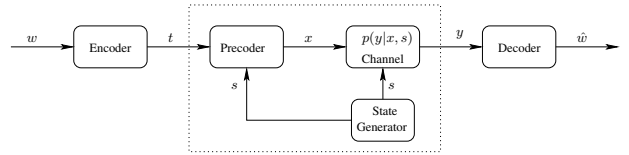Fig. 1. SD-DMC with state information at the encoder.

Fig. 2. The associated regular DMC.

by a $|\mathcal{S}|$-tuple $(x_1, x_2, \ldots, x_{|\mathcal{S}|})$, implying that the value of the function at state s is $x_s, s = 1, 2, \ldots, |\mathcal{S}|$.

The transition probabilities for the associated channel are given by

$$p(y|t) = \sum_{s=1}^{|\mathcal{S}|} p(s)p(y|x_s, s), \qquad (1)$$

where t denotes the index of the function represented by $(x_1, x_2, \ldots, x_{|\mathcal{S}|})$. The SD-DMC and its associated regular DMC are shown in figs. 1 and 2, respectively.

In the next section, we introduce the channel model. In section III, we investigate the capacity of the channel introduced in section II. The general structure of a communication system for the channel with known discrete interference is given in section IV. We conclude this paper in section V.

II. THE CHANNEL MODEL

We consider M-ary Pulse Amplitude Modulation (M-PAM) signaling over the channel

$$Y = X + S + N, \qquad (2)$$

where X is the channel input, which takes on values in $\{x_1 = -(M - 1), x_2 = -(M - 3), \ldots, x_M = M - 1\}$, 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 $\{s, -s\}$ with equal probabilities. 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 (X_1, X_2), where X_1 and X_2 take value in M-PAM constellation.

The likelihood function for the associated channel is given by

$$
\begin{aligned}
f_{Y|T}(y|t) &= \frac{1}{2} f_{Y|T,S}(y|t, s) + \frac{1}{2} f_{Y|T,S}(y|t, -s) \\
&= \frac{1}{2} f_{Y|X,S}(y|x_i, s) + \frac{1}{2} f_{Y|X,S}(y|x_j, -s) \\
&= \frac{1}{2} f_N(y - x_i - s) + \frac{1}{2} f_N(y - x_j + s), \quad (3)
\end{aligned}
$$

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

$$
\begin{aligned}
f_Y(y) &= \sum_{i=1}^{M} \sum_{j=1}^{M} p_{ij} \left(\frac{1}{2} f_N(y - x_i - s) \right) + \\
&\quad \sum_{i=1}^{M} \sum_{j=1}^{M} p_{ij} \left(\frac{1}{2} f_N(y - x_j + s) \right) \\
&= \frac{1}{2} \sum_{i=1}^{M} p_i^{(1)} f_N(y - x_i - s) + \\
&\quad \frac{1}{2} \sum_{j=1}^{M} p_j^{(2)} f_N(y - x_j + s), \quad (4)
\end{aligned}
$$

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

III. THE CAPACITY

The capacity of the associated channel, which is the same as the capacity of the original channel defined in section II, is the maximum of $I(T; Y) = I(X_1 X_2; Y)$ over the joint pmf values $p_{ij} = \Pr\{X_1 = x_i, X_2 = x_j\}$, where x_i and x_j belong to M-PAM constellation, i.e.,

$$
C = \max_{p_{ij}} I(X_1 X_2; Y). \quad (5)
$$

The mutual information between T and Y is the difference between differential entropies $h(Y)$ and $h(Y|T)$. It can be seen from (4) 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

$$
\begin{aligned}
h(Y|T) &= h(Y|X_1 X_2) \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} p_{ij} h(Y|X_1 = x_i, X_2 = x_j) \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} p_{ij} h_{ij}, \quad (6)
\end{aligned}
$$

where $h_{ij} = h(Y|X_1 = x_i, X_2 = x_j)$. Assuming that the marginal pmfs for X_1 and X_2 are given, the maximization problem in (5) reduces to the linear minimization problem

$$
\min_{p_{ij}} \quad \sum_{i=1}^{M} \sum_{j=1}^{M} h_{ij} p_{ij}
$$

subject to

$$
\begin{aligned}
\sum_{j=1}^{M} p_{ij} &= p_i^{(1)}, \quad i = 1, 2, \ldots, M, \\
\sum_{i=1}^{M} p_{ij} &= p_j^{(2)}, \quad j = 1, 2, \ldots, M, \\
p_{ij} &\geq 0, \quad\quad i, j = 1, 2, \ldots, M. \quad (7)
\end{aligned}
$$

In this paper, we are not going to find the actual capacity C since it is a difficult problem. However, we will study the optimal probability distribution $\{p_{ij}\}_{i,j=1,2,\ldots,M}$ under the constraint that X_1 and X_2 are uniformly distributed, i.e.,

$$
p_i^{(1)} = p_i^{(2)} = \frac{1}{M}, \quad i = 1, 2, \ldots, M. \quad (8)
$$

This constraint makes the input to our channel defined in (2) uniform.

There are M^2 variables involved in the linear minimization problem (7). 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 optimum value of (7).

Theorem 1: Denote by C^U the maximum of $I(X_1 X_2; Y)$ over p_{ij}s with uniform marginal pmfs for X_1 and X_2. Then C^U can be achieved by using only M inputs of the associated channel with nonzero probabilities.

The maximization of C^U is equivalent to the minimization of (7) with marginal pmfs given in (8). Considering the constraints in (7), theorem 1 implies that the capacity can be achieved by some set of p_{ij}s, exactly M of them being $\frac{1}{M}$ and the others being zero.

Proof: The equality constraints of (7) can be written in matrix form as

$$
\mathbf{A}\mathbf{p} = \mathbf{1}, \quad (9)
$$

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 $\mathbf{p}_i = \left[p_{i1} p_{i2} \ldots p_{iM} \right]^T, i = 1, 2, \ldots, M$, 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 (7) is achieved at an all-integer vector $\mathbf{p}^*$. Considering that $\mathbf{p}^*$ satisfies (9), it can only be a zero-one vector with exactly M ones. ∎

It turns out from the proof of theorem 1 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
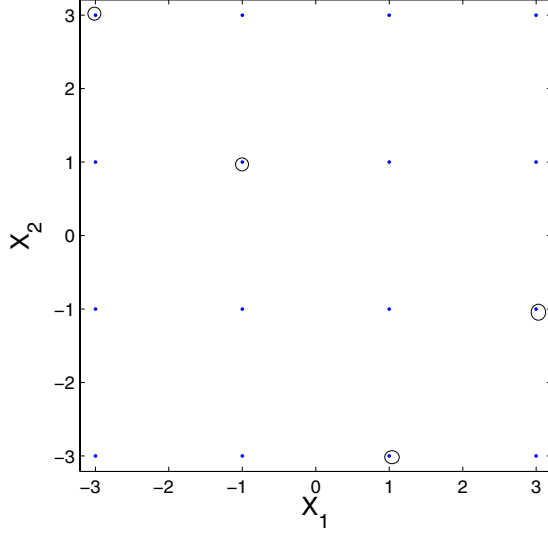
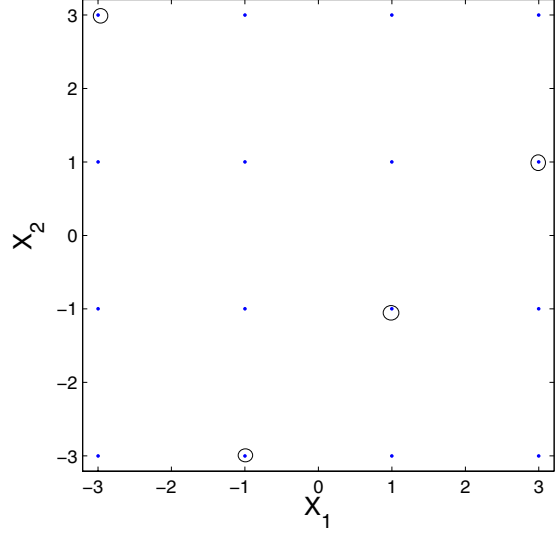Fig. 3. Optimal solution for 4-PAM input with parameters $s = 2, P_N = 1, P_X = 5$.

Fig. 4. Optimal solution for 4-PAM input with parameters $s = 2.5, P_N = 9, P_X = 5$.

constraints in (9), 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].

According to theorem 1, to achieve C^U, we need to use only M out of M^2 inputs of the associated channel with equal 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 parameter s, the noise power P_N, and the signal points x_i, x_j.

As example, the optimal solutions for two different scenarios with 4-PAM constellation ($X_1, X_2 \in \{-3, -1, +1, +3\}$) are illustrated in figs. 3 and 4. The points circled in the array correspond to the inputs to the associated channel that should be chosen with probability $\frac{1}{4}$.

In the sequel, we further investigate the optimal solution of (7). 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.,

$$h_{ij} = g(x_i - x_j), \tag{10}$$

where $g(.)$ is a given function. The minimum of $g(u)$ is the differential entropy of the noise $h(N)$, which occurs at $u = 2s$ and its supremum is $1 + h(N)$, which is approached when $u \to \pm\infty$. The plot of $g(.)$ for $s = 2, P_N = 1$ is shown in fig. 5.

Using (10), it can be shown that if $\{p_{ij}\}_{i,j=1,2,...,M}$ is an optimum solution to (7), then $\{q_{ij}\}_{i,j=1,2,...,M}$ with $q_{ij} = \frac{1}{2}\left(p_{ij} + p_{(M+1-j)(M+1-i)}\right)$ is also an optimum solution of (7) and has the following symmetry property
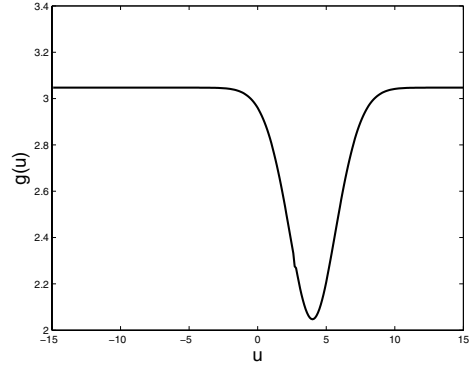
$$q_{ij} = q_{(M+1-j)(M+1-i)}. \tag{11}$$

Fig. 5. The plot of $g(u)$ for $s = 2, P_N = 1$.

The symmetry property (11) translates to the symmetry of the array in figs. 3 and 4 with respect to the main diagonal. We use the properties given in (10) and (11) to prove the following theorem.

Theorem 2: If the function g is strictly concave in the interval $[2x_1, 2x_M]$, then the optimal solution to (7) corresponds to an array with probabilities $\frac{1}{M}$ on the main diagonal and zero elsewhere.

Proof: Suppose that the optimum solution has a nonzero entry p at position (x_i, x_j) which is not on the main diagonal, i.e., $i \neq j$. Then by the symmetry property (11), there must be another entry at location $(-x_j, -x_i)$ with the same value p. Now, if we add p to the main diagonal entries located at $(x_i, -x_i)$ and $(-x_j, x_j)$ and turn the probabilities of the two symmetric entries to zero, the constraints of (7) are not violated. However, the change in the objective function will

be proportional to

$$h(Y|X_1 = x_i, X_2 = -x_i) + h(Y|X_1 = -x_j, X_2 = x_j)$$
$$-h(Y|X_1 = x_i, X_2 = x_j) - h(Y|X_1 = -x_j, X_2 = -x_i),$$

which by property (10) is equal to $g(2x_i) + g(-2x_j) - 2g(x_i - x_j)$ which is negative by concavity of g. This contradicts the optimality of the solution with nonzero entries that are not located on the main diagonal. ∎

It can be shown that the function g is strictly concave in the interval $[2x_1, 2x_M]$ if and only if

$$\sqrt{1 + 3P_X} \le 1 + s - \frac{1}{2}u^*\sqrt{P_N}, \quad (12)$$

where $u^* \cong 1.636$ and $P_X = \frac{M^2 - 1}{3}$ is the transmitted signal power.

Theorem 3: If the function g is convex in the interval $[2x_1, 2x_M]$, then the optimal solution to (7) corresponds to an array with probabilities $\frac{1}{M}$ on the main skew diagonal and zero elsewhere.

Proof: Define random variable $U = X_1 - X_2$. We denote the pmf of U by p_U. The objective function in (7) can be written as

$$\sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij}h_{ij} = \sum_{i=1}^{M}\sum_{j=1}^{M} p_{ij}g(x_i - x_j)$$
$$= \sum_{k}\sum_{j=1}^{M} p_{(k+j)j}g(x_{k+j} - x_j)$$
$$= \sum_{k}\sum_{j=1}^{M} p_{(k+j)j}g(u_k)$$
$$= \sum_{k} p_U(u_k)g(u_k)$$
$$= \mathrm{E}[g(U)], \quad (13)$$

where $u_k = x_{k+j} - x_j$ (u_k does not depend on j) and E[.] denotes the expectation operator. Now, considering the convexity of g, apply the Jensen's Inequality

$$\mathrm{E}[g(U)] \ge g(\mathrm{E}[U])$$
$$= g(0). \quad (14)$$

Equality holds when $U \equiv 0$, or equivalently, $X_1 \equiv X_2$. Therefore, for minimization of the objective function, only the entries of the main skew diagonal must be nonzero. ∎

It can be shown that the function g is convex in the interval $[2x_1, 2x_M]$ if and only if

$$\sqrt{1 + 3P_X} \le 1 - s + \frac{1}{2}u^*\sqrt{P_N}. \quad (15)$$

IV. OPTIMAL PRECODING

Any encoding and decoding scheme for the associated channel can be translated to an encoding and decoding scheme for the original channel (2) with the same probability of error [5]. The general structure of a communication system for the channel (2) is shown in fig. 6. Message w is encoded to a block of length n of indices $t \sim (x_i, x_j)$. According to
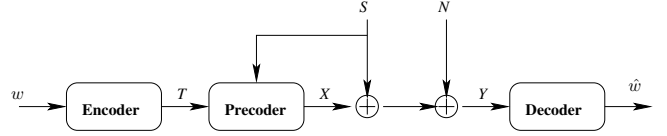

Fig. 6. General structure of the communication system for channels with causally-known discrete interference.

theorem 1, only M of the indices are needed to achieve the capacity. Those M indices are obtained by solving the linear programming problem (7). For each t, the precoder sends either x_i (if $S = s$) or x_j (if $S = -s$). Based on the received signal Y, the receiver decodes $\hat{w}$ as the transmitted message.

V. CONCLUSION

In this paper, we have investigated uniform M-PAM signal transmission over AWGN channel with two-level interference which is known causally at the transmitter. According to the Shannon's theorem for channels with side information at the transmitter, the associated channel will be a channel with M^2 inputs. We have proved that by choosing only M inputs with probability $\frac{1}{M}$ the capacity (with uniform marginal pmfs for X_1 and X_2) is achievable. The optimal M inputs can be obtained by solving the linear optimization problem (7). In some special cases where the function $g(u)$ is concave or convex in the interval $[2x_1, 2x_M]$, the optimal solution is given by theorems 2 and 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 two-level interference.

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.