

# Precoding for the AWGN Channel with Discrete Interference

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

**Abstract**—For a state-dependent DMC with input alphabet  $\mathcal{X}$  and state alphabet  $\mathcal{S}$  where the i.i.d. state sequence is known causally at the transmitter, it is shown that by using at most  $|\mathcal{X}||\mathcal{S}| - |\mathcal{S}| + 1$  out of  $|\mathcal{X}|^{|\mathcal{S}|}$  input symbols of the Shannon's associated channel, the capacity is achievable. As an example of state-dependent channels with side information at the transmitter,  $M$ -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. For the special case where the Gaussian noise power is zero, a sufficient condition, which is independent of interference, is given for the capacity to be  $\log_2 M$  bits per channel use. The problem of maximization of the transmission rate under the constraint that the channel input given any current interference symbol is uniformly distributed over the channel input alphabet is investigated. 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 received a great deal of attention. 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 AWGN channel [1]. Therefore, the interference does not incur any loss in the capacity. This result was extended to arbitrary interference (random or deterministic) by Erez *et al.* [2]. 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 [3] in causal knowledge setting and by Gel'fand and Pinsker [4] 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  $\mathcal{X} = \{x_1, \dots, x_{|\mathcal{X}|}\}$ , a finite output alphabet  $\mathcal{Y}$ , and transition probabilities  $p(y|x, s)$ , where the state  $s$  takes on values in a finite alphabet  $\mathcal{S} = \{1, \dots, |\mathcal{S}|\}$ .

<sup>1</sup>This work was supported by Nortel, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Ontario Centres of Excellence (OCE).

Shannon [3] 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 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  $|\cdot|$  denotes the cardinality of a set. Any of the functions can be represented by a  $|\mathcal{S}|$ -tuple  $(x_{i_1}, x_{i_2}, \dots, x_{i_{|\mathcal{S}|}})$  composed of elements of  $\mathcal{X}$ , implying that the value of the function at state  $s$  is  $x_{i_s}, s = 1, 2, \dots, |\mathcal{S}|$ .

The capacity is given by [3]

$$C = \max_{p(t)} I(T; Y), \quad (1)$$

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

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

This paper is organized as follows. In section II, we derive an upper bound on the cardinality of the Shannon's associated channel input alphabet to achieve the capacity. In section III, we introduce our channel model. In section IV, we investigate the capacity of the channel in the absence of noise. In section V, we consider maximizing the transmission rate under the constraint that the channel input given any current interference symbol is uniformly distributed over the channel input alphabet. We present the general structure of a communication system for the channel with causally-known discrete interference in section VI. We conclude this paper in section VII.

## II. A BOUND ON THE CARDINALITY OF THE SHANNON'S ASSOCIATED CHANNEL INPUT ALPHABET

We can obtain the pmf of the channel output  $Y$  as

$$\begin{aligned} p_Y(y) &= \sum_{s \in \mathcal{S}} p_S(s) p_{Y|S}(y|s) \\ &= \sum_{s \in \mathcal{S}} p_S(s) \sum_{x \in \mathcal{X}} p_{X|S}(x|s) p_{Y|X,S}(y|x, s) \\ &= \sum_{s \in \mathcal{S}} p_S(s) \sum_{x \in \mathcal{X}} p_{X_s}(x) p_{Y|X,S}(y|x, s). \end{aligned} \quad (2)$$

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

$$C = \max_{p_{i_1 i_2 \cdots i_{|\mathcal{S}|}}} I(X_1 X_2 \cdots X_{|\mathcal{S}|}; Y). \quad (3)$$

The mutual information between  $T$  and  $Y$  is the difference between the entropies  $H(Y)$  and  $H(Y|T)$ . It can be seen from (2) that  $p_Y(y)$ , and hence  $H(Y)$ , are uniquely determined by the marginal pmfs  $\{p_{X_s}(x_i)\}_{i=1}^{|\mathcal{X}|}$ ,  $s = 1, \dots, |\mathcal{S}|$ . The conditional entropy  $H(Y|T)$  is given by

$$\begin{aligned} H(Y|T) &= H(Y|X_1 X_2 \cdots X_{|\mathcal{S}|}) \\ &= \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} h_{i_1 \cdots i_{|\mathcal{S}|}} p_{i_1 \cdots i_{|\mathcal{S}|}}, \end{aligned} \quad (4)$$

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

There are  $|\mathcal{X}|^{|\mathcal{S}|}$  variables involved in the maximization problem (3). 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 (3).

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

*Proof:* Denote by  $\{\hat{p}_i^{(s)}\}_{i=1}^{|\mathcal{X}|} = \{\hat{p}_{X_s}(x_i)\}_{i=1}^{|\mathcal{X}|}$  the pmf of  $X_s$ ,  $s = 1, 2, \dots, |\mathcal{S}|$ , induced by a capacity-achieving joint pmf  $\{\hat{p}_{i_1 \cdots i_{|\mathcal{S}|}}\}_{i_1, \dots, i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$ . We limit the search for a capacity-achieving joint pmf to those joint pmfs that yield the same marginal pmfs as  $\{\hat{p}_{i_1 \cdots i_{|\mathcal{S}|}}\}_{i_1, \dots, i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$ . By limiting the search to this smaller set, the maximum of  $I(X_1 \cdots X_{|\mathcal{S}|}; Y)$  remains unchanged since the capacity-achieving joint pmf  $\{\hat{p}_{i_1 \cdots i_{|\mathcal{S}|}}\}_{i_1, \dots, i_{|\mathcal{S}|}=1}^{|\mathcal{X}|}$  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  $X_1, \dots, X_{|\mathcal{S}|}$ . Therefore, the maximization problem in (3) reduces to the linear minimization problem

$$\begin{aligned} \min_{p_{i_1 \cdots i_{|\mathcal{S}|}}} & \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} h_{i_1 \cdots i_{|\mathcal{S}|}} p_{i_1 \cdots i_{|\mathcal{S}|}} \\ \text{s. t.} & \sum_{i_2=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|}=1}^{|\mathcal{X}|} p_{i_1 \cdots i_{|\mathcal{S}|}} = \hat{p}_{i_1}^{(1)}, \quad i_1 = 1, \dots, |\mathcal{X}|, \\ & \vdots \\ & \sum_{i_1=1}^{|\mathcal{X}|} \cdots \sum_{i_{|\mathcal{S}|-1}=1}^{|\mathcal{X}|} p_{i_1 \cdots i_{|\mathcal{S}|}} = \hat{p}_{i_{|\mathcal{S}|}}^{(|\mathcal{S}|)}, \quad i_{|\mathcal{S}|} = 1, \dots, |\mathcal{X}|, \\ & p_{i_1 \cdots i_{|\mathcal{S}|}} \geq 0, \quad i_1, \dots, i_{|\mathcal{S}|} = 1, 2, \dots, |\mathcal{X}|. \end{aligned} \quad (5)$$

There are  $|\mathcal{X}|^{|\mathcal{S}|}$  equality constraints in (5) out of which  $|\mathcal{X}|^{|\mathcal{S}|} - |\mathcal{S}| + 1$  are linearly independent. From the theory of linear programming, the minimum of (5), and hence the

maximum of  $I(X_1 \cdots X_{|\mathcal{S}|}; Y)$ , is achieved by a feasible solution with at most  $|\mathcal{X}|^{|\mathcal{S}|} - |\mathcal{S}| + 1$  nonzero variables. ■ Theorem 1 states that at most  $|\mathcal{X}|^{|\mathcal{S}|} - |\mathcal{S}| + 1$  out of  $|\mathcal{X}|^{|\mathcal{S}|}$  input symbols of the associated channel are needed to be used with positive probability to achieve the capacity. However, in general one does not know which of the inputs must be used to achieve the capacity. If we knew the marginal pmfs for  $X_1, \dots, X_{|\mathcal{S}|}$  induced by a capacity-achieving joint pmf, we could obtain the capacity-achieving joint pmf itself by solving the linear program (5).

### III. THE CHANNEL MODEL

We consider data transmission over the channel

$$Y = X + S + N, \quad (6)$$

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

$$\mathcal{X} = \{x_1, x_2, \dots, x_M\}, \quad (7)$$

$Y$  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

$$S = \{s_1, s_2, \dots, s_Q\} \quad (8)$$

with probabilities  $r_1, r_2, \dots, r_Q$ , 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  $(X_1, X_2, \dots, X_Q)$ , where  $X_j$  is the random variable that represents the channel input when the current interference symbol is  $s_j$ ,  $j = 1, \dots, Q$ .

The likelihood function for the associated channel is given by

$$\begin{aligned} f_{Y|T}(y|t) &= \sum_{j=1}^Q r_j f_{Y|X,S}(y|x_{i_j}, s_j) \\ &= \sum_{j=1}^Q r_j f_N(y - x_{i_j} - s_j), \end{aligned} \quad (9)$$

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

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

### IV. THE NOISE-FREE CHANNEL

We consider a special case where the noise power is zero in (6). 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<sup>1</sup>: The decoder just needs to partition the 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. We do this by considering a one-shot coding scheme which uses only  $M$  (out of  $M^Q$ ) inputs of the associated channel.

In a one-shot coding scheme, a message is encoded to a single input of the associated channel. Any input of the associated channel can be represented by a  $Q$ -tuple composed of elements of  $\mathcal{X}$ . Given that the current interference symbol is  $s_j$ , the  $j$ th element of the  $Q$ -tuple is sent through the channel. Therefore, one single message can result in (up to)  $Q$  symbols at the output. For convenience, we consider the output symbols corresponding to a single message as a multi-set<sup>2</sup> of size (exactly)  $Q$ . If the  $M$  multi-sets at the output corresponding to  $M$  different messages are mutually disjoint, reliable transmission through the channel is possible.

Unfortunately, we cannot always find  $M$  inputs of the associated channel such that the corresponding multi-sets are mutually disjoint. For example, consider a channel with the input alphabet  $\mathcal{X} = \{0, 1, 2, 4\}$  and the interference alphabet  $\mathcal{S} = \{0, 1, 3\}$ . It is easy to check that for this channel we cannot find four triples composed of elements of  $\mathcal{X}$  such that the corresponding multi-sets are mutually disjoint. 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 arithmetic progression. Then the capacity of the noise-free channel

$$Y = X + S, \quad (10)$$

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

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

$$\mathcal{Y}^{(q)} = \{x_1 + s_q, x_2 + s_q, \dots, x_M + s_q\}, \quad q = 1, \dots, Q. \quad (11)$$

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

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

<sup>1</sup>This is true even if the interference sequence is unknown to the encoder.

<sup>2</sup>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.

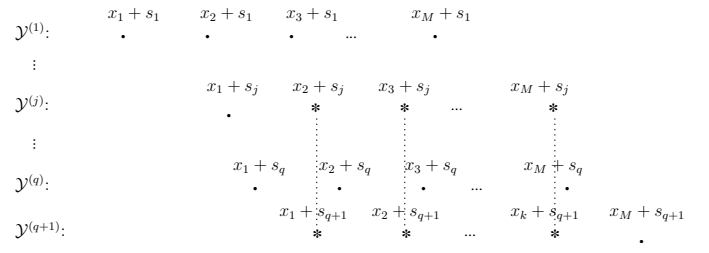


Fig. 1. The elements of  $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(q+1)}$  shown as shifted version of each other. The elements of  $\mathcal{Y}^{(q+1)}$  up to  $x_k + s_{q+1}$  appear in  $\mathcal{Y}^{(j)}$ .

We prove by induction on  $Q$  that there exist  $M$  mutually-disjoint (multi-)sets of size  $Q$  composed of the elements of  $\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \dots, \mathcal{Y}^{(Q)}$  (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.

For  $Q = 1$ , the statement of the theorem is true since we can take  $\{x_1 + s_1\}, \{x_2 + s_1\}, \dots, \{x_M + s_1\}$  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  $\mathcal{Y}^{(q+1)} = \{x_1 + s_{q+1}, x_2 + s_{q+1}, \dots, x_M + s_{q+1}\}$ . 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  $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(q)}$  (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  $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(q)}$  and  $s_{q+1} > s_q > \dots > s_1$ , exactly one of the sets  $\mathcal{Y}^{(1)}, \dots, \mathcal{Y}^{(q)}$ , 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. 1. Since any of the disjoint multi-sets of size  $Q$  contain just one element of  $\mathcal{Y}^{(j)}$ , the elements of  $\mathcal{Y}^{(q+1)}$  up to  $x_k + s_{q+1}$  appear in different multi-sets of size  $Q = q$ . We can form the disjoint multi-sets of size  $q+1$  by including these common elements in the corresponding multi-sets and including the elements of  $\{x_{k+1} + s_{q+1}, \dots, x_M + s_{q+1}\}$  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  $Q = 2$ . Because in the second iteration,

we do not need the arithmetic progression condition to form  $M$  mutually-disjoint multi-sets of size two.

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  $q$ th 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  $X_1, \dots, X_Q$ .

## V. UNIFORM TRANSMISSION

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

$$p_i^{(1)} = p_i^{(2)} = \dots = p_i^{(Q)} = \frac{1}{M}, \quad i = 1, 2, \dots, M, \quad (12)$$

for which we show how to obtain the optimal input probability assignment. We call a transmission scheme that induces uniform distribution on  $X_1, \dots, X_Q$  as *uniform transmission*. The uniform distribution for  $X_1, \dots, X_Q$  implies uniform distribution for  $X$ , the input to the state-dependent channel defined in (6).

In the previous section, we established that the capacity achieving pmf for the asymptotic case of noise-free channel induces uniform distributions on  $X_1, \dots, X_Q$  (provided that we can find  $M$   $Q$ -tuples such that the corresponding multi-sets are mutually disjoint).

Considering the constraints in (12), the maximization of  $I(X_1 \cdots X_Q; Y)$  is reduced to the linear minimization problem

$$\begin{aligned} \min_{p_{i_1 \dots i_Q}} \quad & \sum_{i_1=1}^M \cdots \sum_{i_Q=1}^M h_{i_1 \dots i_Q} p_{i_1 \dots i_Q} \\ \text{s. t.} \quad & \sum_{i_2=1}^M \cdots \sum_{i_Q=1}^M p_{i_1 \dots i_Q} = \frac{1}{M}, \quad i_1 = 1, \dots, M, \\ & \vdots \\ & \sum_{i_1=1}^M \cdots \sum_{i_{Q-1}=1}^M p_{i_1 \dots i_Q} = \frac{1}{M}, \quad i_Q = 1, \dots, M, \\ & p_{i_1 \dots i_Q} \geq 0, \quad i_1, \dots, i_Q = 1, 2, \dots, M. \end{aligned} \quad (13)$$

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  $\{p_{i_1 \dots i_Q}\}_{i_1, \dots, i_Q=1}^M$  that induce uniform marginal distributions on  $X_1, X_2, \dots, X_Q$  is achieved by a joint pmf with at most  $MQ - Q + 1$  nonzero elements.

This result is independent of the coefficients  $\{h_{i_1 \dots i_Q}\}$ . However, which probability assignment with at most  $MQ -$

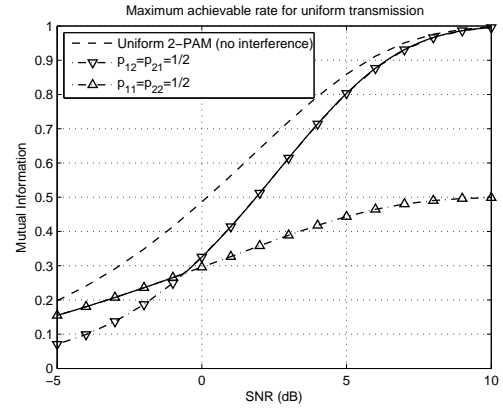


Fig. 2. Maximum mutual information vs. SNR for the channel with  $\mathcal{X} = \{-1, +1\}$  and  $r_1 = r_2 = \frac{1}{2}$ .

$Q + 1$  nonzero elements is optimal depends on the coefficients  $\{h_{i_1 \dots i_Q}\}$ . The coefficient  $h_{i_1 \dots i_Q}$  is determined by the interference levels  $s_1, \dots, s_Q$ , the probability of interference levels  $r_1, \dots, r_Q$ , the noise power  $P_N$ , and the signal points  $x_1, x_2, \dots, x_M$ . The optimal probability assignment is obtained by solving the linear programming problem (13) using the simplex method [6].

### A. Two-Level Interference

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

*Theorem 3:* The maximum of  $I(X_1 X_2; Y)$  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  $\frac{1}{M}$ .

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

$$\mathbf{A}\mathbf{p} = \mathbf{1}, \quad (14)$$

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

For  $Q = 2$ , it is easy to check that  $\mathbf{A}$  is the vertex-edge incidence matrix of  $K_{M, M}$ , the complete bipartite graph with  $M$  vertices at each part. Therefore,  $\mathbf{A}$  is a totally unimodular matrix<sup>3</sup> [5]. 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. ■

Fig. 2 depicts the maximum mutual information (for the uniform transmission scenario) vs. SNR for the channel with  $\mathcal{X} = \{-1, +1\}$  and equiprobable interference symbols. The mutual information vs. SNR curve for the interference-free AWGN channel with equiprobable input alphabet  $\{-1, +1\}$

<sup>3</sup>A totally unimodular matrix is a matrix for which every square submatrix has determinant 0, 1, or  $-1$ .

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.

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* [5], which can be solved using low-complexity algorithms such as the *Hungarian method* [6].

### B. Integrality Constraint for the $Q$ -Level Interference

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

For  $Q \neq 2$ ,  $\mathbf{A}$  is not a totally unimodular matrix. Therefore, not all extreme points of the feasible region defined by  $\mathbf{A}\mathbf{p} = \mathbf{1}$ ,  $\mathbf{p} \geq \mathbf{0}$ , are integer vectors. However, at the expense of possible loss in rate, we may add the integrality constraint in this case. The resulting optimization problem is called the *multi-dimensional assignment problem* [7]. The optimal solution of (13) with the integrality constraint, will be a vector with exactly  $M$  nonzero elements with the value  $\frac{1}{M}$ . Therefore, any encoding scheme just needs to use  $M$  symbols of the associated channel with equal probabilities, simplifying the encoding operation.

## VI. OPTIMAL PRECODING

The general structure of a communication system for the channel defined in (6) is shown in fig. 3. Any encoding and decoding scheme for the associated channel can be translated to an encoding and decoding scheme for the original channel defined in (6). A message  $w$  is encoded into a block of length  $n$  composed of input symbols of the associated channel  $t \sim (x_{i_1}, x_{i_2}, \dots, x_{i_Q})$ . There are  $M^Q$  input symbols. However, we showed that the maximum rate with uniformity and integrality constraints can be achieved by using just  $M$  input symbols of the associated channel with equal probabilities. The optimal  $M$  input symbols of the associated channel are obtained by solving the linear programming problem (13) with the integrality constraint. Those  $M$  input symbols of the associated channel define the optimal precoding operation: For any  $t$  that belongs to the set of  $M$  optimal input symbols, the precoder sends the  $q$ th component of  $t$  if the current interference symbol is  $s_q$ ,  $q = 1, \dots, Q$ . Based on the received sequence, the receiver decodes  $\hat{w}$  as the transmitted message.

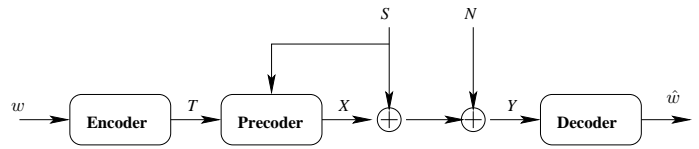


Fig. 3. General structure of the communication system for channels with causally-known discrete interference.

## VII. CONCLUSION

In this paper, we proved that the capacity of an SD-DMC with finite input alphabet  $\mathcal{X}$  and finite state alphabet  $\mathcal{S}$  and with causally known i.i.d. state sequence at the encoder can be achieved by using at most  $|\mathcal{X}||\mathcal{S}| - |\mathcal{S}| + 1$  out of  $|\mathcal{X}|^{|\mathcal{S}|}$  input symbols of the associated channel. As an example of state-dependent channels with side information at the encoder, 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.

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

We considered the transmission schemes with uniform pmfs for  $X_1, \dots, X_Q$ . For this so called uniform transmission, the optimal input probability assignment with at most  $MQ - Q + 1$  nonzero elements can be obtained by solving the linear optimization problem (13). The optimal solution to (13) with the integrality constraint has exactly  $M$  nonzero elements. For the case  $Q = 2$ , we showed that the integrality constraint does not reduce the maximum achievable rate. The loss in rate (if there is any) by imposing the integrality constraint for the general case is a problem to be explored.

## ACKNOWLEDGMENT

The authors would like to thank Gerhard Kramer and Syed Ali Jafar for pointing out that the proof of theorem 1, which was originally given for the channel model considered in this paper, works for any SD-DMC with causal side information at the encoder.

## REFERENCES

- [1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439-441, May 1983.
- [2] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3820-3833, Nov. 2005.
- [3] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289-293, Oct. 1958.
- [4] 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.
- [5] G. Nemhauser and L. Wolsey, *Integer and combinatorial optimization*, John Wiley & Sons, 1988.
- [6] B. Kerkó, *Linear Programming*, Translated by J. H. L. Ahrens and C. M. Safe. Sir Isaac Pitman & Sons Ltd., 1968.
- [7] W. P. Pierskalla, "The multidimensional assignment problem," *Operations Research* 16, 1968, p. 422-431.