

# Universal A Posteriori Metrics Game

Emmanuel Abbe  
 LCM, EPFL  
 Lausanne, 1015, Switzerland  
 emmanuel.abbe@epfl.ch

Rethnakaran Pulikoonattu  
 Broadcom Inc,  
 Irvine, CA, USA, 92604  
 rethna@broadcom.com

**Abstract**—Over binary input channels, uniform distribution is a universal prior, in the sense that it allows to maximize the worst case mutual information over all binary input channels, ensuring at least 94.2% of the capacity. In this paper, we address a similar question, but with respect to a universal generalized linear decoder. We look for the best collection of finitely many a posteriori metrics, to maximize the worst case mismatched mutual information achieved by decoding with these metrics (instead of an optimal decoder such as the Maximum Likelihood (ML) tuned to the true channel). It is shown that for binary input and output channels, two metrics suffice to actually achieve the same performance as an optimal decoder. In particular, this implies that there exist a decoder which is generalized linear and achieves at least 94.2% of the compound capacity on any compound set, without the knowledge of the underlying set.

## I. INTRODUCTION

We consider the problem where a communication system is to be designed without explicit knowledge of the channel. Here, neither the transmitter nor the receiver have access to the exact channel law. The goal is to devise a single coding strategy which will ensure reliable communication over the unknown channel picked for transmission. We assume i.i.d. realizations of the channel at each channel use, i.e., we are interested in a universal coding framework for communicating over discrete memoryless channels (DMC). In this paper, we present results for DMC's with binary input and output alphabets, which we refer to as binary memoryless channels (BMC). Our goal is to design decoders which have a linear structure and which entails reliable communication at the largest possible rate in this setting. In the next section, we revise in more details the notion of universality and linearity for DMC's, and their attribute. We will then formulate our problem as a game where the decoder has to pick the decoding metrics, i.e., a generalized linear decoder, before nature select a channel for communication.

### A. Universality

If the channel over which communication takes place is unknown at both the transmitter and the receiver but belongs to a set of DMC's  $\mathcal{S}$ , then we are in the setting of *compound channels*. Let us denote by  $\mathcal{X}$  the input alphabet and  $\mathcal{Y}$  the output alphabet of any DMC in  $\mathcal{S}$ . The objective is to design a code (i.e., an encoder and decoder pair) which will provide a mean for reliable communication, independently of which  $W \in \mathcal{S}$  is picked up (by nature) for transmission. The

compound channel problem has been extensively studied in the literature such as [5],[7],[9],[10],[12],[16] and [17].

The highest achievable rate, known as the *compound capacity*  $C(\mathcal{S})$  of a set  $\mathcal{S}$  of channels is established in [5] and is given by:

$$C(\mathcal{S}) \triangleq \max_P \inf_{W \in \mathcal{S}} I(P, W) \quad (1)$$

where the maximization is over all possible probability distributions  $P$  on  $\mathcal{X}$ , and the infimum is performed over all the channels  $W$  in the compound set  $\mathcal{S}$ .

We call a decoder to be universal if the compound capacity can be achieved by the employed decoding rule. In [5], a decoder that maximizes a uniform mixture of likelihoods, over a dense set of possible channels is proposed as a universal decoder. In the literature, a decoder which allows us to achieve the same random coding exponent (without the knowledge of true channel) as the maximum likelihood (ML) decoder tuned to the true channel is called a *universal decoder*. The maximum mutual information (MMI) decoder introduced in [13] is a universal decoder. The MMI decoder computes the empirical mutual information (EMI) between a given received output and each codewords in the codebook, and declares the element with the highest EMI score as the sent codeword. There has been a number of other universal decoders proposed in the literature, such as Lempel-Ziv (LZ) based algorithm [23], and the merged likelihood decoder [12]. The MMI decoder has another interesting feature: it does not even require the knowledge of the compound set to be defined. In that sense, the MMI decoder is a “doubly universal” decoder. However, practical use of MMI decoders are voided by complexity considerations, and similarly for other existing universal decoders. Note that in this paper, we are primarily concerned with the achievable rate rather than error exponent.

### B. Linearity

A linear (or additive) decoder is defined to have the following structure. Upon receiving each  $n$ -symbol output  $y$ , the decoder computes a score (decoding metric)  $d^n(x_m, y)$  for each codeword  $x_m, m = 1, 2, \dots, 2^{nR}$  and declares the codeword with the highest score as estimate of the sent codeword (ties are resolved arbitrarily). Moreover, the  $n$ -symbol decoding metric has the following additive structure.

$$d^n(x, y) = \sum_{i=1}^n d(x(i), y(i)), \quad \forall x, y \in \mathcal{X}^n, \mathcal{Y}^n$$

where  $d : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  is a single letter decoding metric. We call such decoders as linear decoders since the decoding metric

it computes is indeed linear in the joint empirical distribution between the codeword and the received word:

$$d^n(x, y) = n \sum_{u \in \mathcal{X}, v \in \mathcal{Y}} \hat{P}_{(x,y)}(u, v) d(u, v)$$

where  $\hat{P}_{(x,y)}(u, v)$  denotes the joint empirical distribution of  $(x, y)$ . We then say that the linear decoder is induced by the single letter metric  $d$ . A significant body of work on the topic of linear decoders exist in the literature. We refer to [2],[9], and the references within, for a detailed review of such decoders. Examples of linear decoders are the maximum likelihood (ML) or the maximum a posteriori (MAP) decoders, whereas the MMI decoder is not linear.

The main advantage of a linear decoder is that, when used with appropriately structured codes, it allows significant reduction in the decoding complexity. In this regard, a convenient structure is that of linear encoders, which produce codewords out of a linear transformation

$$x = Gu,$$

where  $u \in \mathcal{X}^{nR}$  contains the information bits and  $G \in \mathcal{X}^{n \times nR}$  is the generating matrix. With such encoders, linear decoders allow the use of techniques such as the Viterbi algorithm, where significant complexity reduction is made possible in the optimum decoding (e.g. maximum likelihood sequence decoding (MLSE)) of convolutional codes, or the message-passing (belief propagation) algorithms adopted for the decoding of several modern coding schemes [21].

The expected reduction in decoding complexity discussed so far is possible only when the code is appropriately structured. However, for the proof of existence of linear universal decoders in this paper, we rely on the random coding argument, where one establish the existence of a deterministic code which yields good performance, without explicitly constructing the code. For the complexity claims, one still needs to investigate on whether appropriately structured encoders, can realize the performance guaranteed by the random coding argument. However, from an argument of Elias [11], we already know that this is possible for binary symmetric channels, where it is sufficient to consider random binary codes to achieve the performance of arbitrary random codes; this argument is based on the fact that pairwise independence between the codewords is a sufficient assumption which has been further generalized in [6].

A class of decoders slightly more general than linear decoders and called *generalized linear decoders* in [2] is defined as follows. Instead of requiring the condition that the decoding metric is additive, it is only required that the score function breaks into the *maximization* of a finitely many additive metrics, cf. Definition 2 in [2]. The purpose of studying generalized linear decoders is that all properties mentioned above for linear decoders still hold for generalized decoders.

### C. Linear universal decoders

In view of constructing universal codes of manageable complexity, a first legitimate question is to ask whether linear

universal decoders can be constructed. Not surprisingly, some compound channels do not admit a linear universal decoder. It is known that, when the set  $\mathcal{S}$  is convex and compact, the maximum likelihood, tuned to the channel offering the least mutual information (for the optimal input distribution in (1)) serves as a compound capacity achieving decoder [9]. In [1], [2], it is shown that this result still holds if the set  $\mathcal{S}$  is non convex but one-sided, cf. Definition 4 in [2]. Moreover, authors in [2] show that if  $\mathcal{S}$  is a finite union of one-sided sets, a generalized linear decoder that achieves the compound capacity can be constructed.

In this paper, we construct a generalized linear decoder which achieve 94% of the compound capacity on any BMC compound set, by using the same two metrics, chosen irrespective of the given compound set.

The remainder of this paper is organized as follows. We review known results for DMC's and then introduce the notations in the next section. The problem statement is discussed in section III. We then present the main results for BMC's in section IV.

## II. KNOWN RESULTS FOR DMC

We consider discrete memoryless channels with input alphabet  $\mathcal{X}$  and output alphabet  $\mathcal{Y}$ . A DMC is described by a probability transition matrix  $W$ , each row of which is the conditional probability distribution of the output  $\mathcal{Y}$  given input  $\mathcal{X}$ . We denote by  $\mathcal{S}$ , a compound set of DMC's. While the set of channels is known to both the transmitter and the receiver, the exact channel of communication, denoted by  $W_0$ , is unknown to them.

We assume that the transmitter and receiver operate synchronously over blocks of  $n$  symbols. In each block, a message  $m \in \{1, 2, \dots, 2^{nR}\}$  is mapped by an encoder

$$F_n : \{1, 2, \dots, 2^{nR}\} \rightarrow \mathcal{X}^n = \{0, 1\}^n$$

to  $F_n(m) = x_m$ , referred to as the  $m^{\text{th}}$  codeword. The receiver upon observing a word, drawn from the distribution

$$W^n(y|x_m) = \prod_{i=1}^n W(y(i)|x_m(i)),$$

applies a decoding map

$$G_n : \mathcal{Y}^n \rightarrow \{1, 2, \dots, 2^{nR}\}.$$

The average probability of error, averaged over a given code  $(F_n, G_n)$  for a specific channel  $W$ , is expressed as

$$P_e(F_n, G_n, W) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y: G_n(y) \neq m} W^n(y|x_m).$$

We say that a rate  $R$  is achievable for a given compound set  $\mathcal{S}$ , if for any  $\epsilon > 0$ , there exist a block length  $n$  and code with rate at least  $R$ , such that, for all  $W \in \mathcal{S}$ ,  $P_e(F_n, G_n, W) < \epsilon$ . The supremum of such available rates is called the *compound channel capacity*, denoted as  $C(\mathcal{S})$ . Blackwell et al. [5], formulated the expression for compound channel capacity as

$$C(\mathcal{S}) \triangleq \max_P \inf_{W \in \mathcal{S}} I(P, W).$$

Before proceeding with the problem statement, we will introduce a few notations and some useful definitions.

*Definition 1:* (Generalized Linear Decoder)

Let  $d_1, d_2, \dots, d_K$  be  $K$  single letter metrics, where  $K$  is a finite integer. A generalized linear decoder induced by these metrics is defined by the decoding map:

$$\begin{aligned} G_n(y) &= \arg \max_m \vee_{k=1}^K \sum_{i=1}^n d_k(x_m(i), y(i)) \\ &= \arg \max_m \vee_{k=1}^K \mathbb{E}_{\hat{P}_{(x_m, y)}}[d] \end{aligned}$$

where  $\vee$  denotes the *maximum* operator and  $\hat{P}_{(x_m, y)}$  denotes the joint empirical distribution of  $(x_m, y)$ .

An example of generalized linear decoder is the generalized likelihood ratio test (GLRT) which, for a given collection of channels  $W_1, \dots, W_K$ , is induced by the metrics

$$\log W_k(v|u), \quad \forall u \in \mathcal{X}, v \in \mathcal{Y}, k = 1, \dots, K.$$

We now denote by  $P$  an input distribution. For a channel denoted by  $W_k$ , we use  $\mu_k$  to denote the joint distribution of an input and output pair through  $W_k$ , i.e.,  $\mu_k = P \circ W_k$ . We denote by  $\mu_k^p$  the product measure of the marginal input distribution  $P$  and the marginal output distribution  $(\mu_k)_Y(y) = \sum_{x \in \mathcal{X}} \mu_k(x, y)$ . Hence,

$$I(P, W_k) = D(\mu_k \| \mu_k^p).$$

*Lemma 1:* When the true channel is  $W_0$  and a generalized linear decoder induced by the single-letter metrics  $\{d_k\}_{k=1}^K$  is used, we can achieve the following rate

$$I_{\text{MIS}}(P_X, W_0, \{d_k\}_{k=1}^K) = \bigvee_{k=1}^K \min_{\mu \in \mathcal{A}_k} D(\mu \| \mu_0^p)$$

where

$$\mathcal{A}_k = \{\mu : \mu^p = \mu_0^p, \mathbb{E}_\mu[d_k] \geq \vee_{j=1}^K \mathbb{E}_{\mu_0}[d_j]\}, \quad \forall 1 \leq k \leq K.$$

In particular, if  $k = 1$ , this is the mismatched result in [9] and [19], and if  $k \geq 2$ , it is a consequence of [9],[19], as discussed in [2]. Extensive coverage of the *mismatched problem* [16] appears in the literature [3],[4],[9],[14],[15],[19] as well as [2].

*Definition 2 (one-sided sets):* A set  $\mathcal{S}$  of DMC's is one-sided with respect to an input distribution  $P$ , if  $W_S = \arg \min_{W \in \text{cl}(\mathcal{S})} I(P, W)$  is unique and if

$$D(\mu \| \mu_S^p) \geq D(\mu \| \mu_S) + D(\mu_S \| \mu_S^p)$$

for any  $\mu = P \circ W$ , where  $W \in \mathcal{S}$  and  $\mu_S = P \circ W_S$ . A set  $\mathcal{S}$  of DMCs is a union of one-sided sets if  $\mathcal{S} = \bigcup_{k=1}^K \mathcal{S}_k$  for some  $K < \infty$  and if the  $\mathcal{S}_k$ 's are one-sided with respect to  $P_* = \arg \max_P \inf_{W \in \mathcal{S}} I(P, W)$ .

*Theorem 1:* For a compound set  $\mathcal{S} = \bigcup_{k=1}^K \mathcal{S}_k$  which is a union of one-sided sets, the generalized linear decoder induced by the metrics

$$d_k = \log \frac{W_{\mathcal{S}_k}}{(\mu_{\mathcal{S}_k})_Y}, \quad 1 \leq k \leq K$$

where  $W_{\mathcal{S}_k} = \arg \min_{W \in \text{cl}(\mathcal{S}_k)} I(P_*, W)$ , is compound capacity achieving. Moreover, if the true channel  $W_0$  is known to belong to  $\mathcal{S}_k$ , this decoder allows to achieve the rate,

$$I_{\text{MIS}}(P_*, W_0, \log W_{\mathcal{S}_k}). \quad (2)$$

Notice that, this decoder requires the full knowledge of the compound set. In other words, knowing the compound capacity as well as the compound capacity achieving input distribution (which for instance, suffices for MMI) will not suffice to construct such a decoder. One can interpret this decoder as follows: the MMI decoder allows to achieve compound capacity on any compound set, by taking a ‘‘degenerate generalized linear’’ decoder with infinitely (and uncountably) many metrics: the a posteriori metrics of all possible DMCs with the given alphabets. Hence, there is no linear property (and consequent benefit) for such a decoder. However, what Theorem 1 says, is that, since we have the knowledge of the compound set, we can use it to tune a generalized linear decoder which will still achieve compound capacity by picking only the important channels and corresponding a posteriori metrics. Our goal in this paper is to investigate whether further simplification in the above generalized linear decoder can be made when restricted to BMC's. More specifically, we address the possibility of building a universal decoder tuned to metrics, chosen independently to the given compound BMC set.

Using the symmetry properties occurring in the BMC's, we have the following result (cf. [1]).

*Lemma 2:* Let  $P_1, P_2$  be two stochastic matrices of size  $2 \times 2$ , such that  $\det(P_1 P_2) > 0$ . Let  $C$  a set of binary vectors of length  $n$  with fixed composition. For any  $x_1, x_2 \in C$  and  $y \in \{0, 1\}^n$ , we respectively have

$$P_1(y|x_1) > (=) P_1(y|x_2) \Rightarrow P_2(y|x_1) > (=) P_2(y|x_2).$$

Note that the condition  $\det(P_1 P_2) > 0$  simply means that  $P_1$  and  $P_2$  have their maximal value within a column at the same place.

Hence, we would like to investigate whereas for BMC, the generalized linear decoder of Theorem 1 can be simplified, so as to require only few metrics and independently of the compound set.

### III. PROBLEM STATEMENT

#### A. The $\alpha$ and $\beta$ game

From now on, we only consider BMC's.

Let the parameter  $\alpha$  be defined as

$$\alpha = \max_P \inf_{W \in \text{BMC}} \frac{I(P, W)}{C(W)}. \quad (3)$$

and the distribution  $P$  which achieve  $\alpha$  is denoted by  $P_{\text{opt}}$ , i.e.,

$$P_{\text{opt}} := \arg \max_P \inf_{W \in \text{BMC}} \frac{I(P, W)}{C(W)}. \quad (4)$$

The term  $\alpha$  is the ratio of the maximum achievable rate to the channel capacity, for the worst possible channel in the compound set, when a single input distribution is chosen.

Let  $K \in \mathbb{Z}_+$  and let  $I_{\text{mis}}(P_{\text{opt}}, W_0, \{d_k\}_1^K)$  be the achievable mismatched rate on a channel  $W_0$ , using a generalized

linear decoder induced by the  $K$  metrics  $\{d_k\}_1^K$ . The expression of  $I_{\text{mis}}(P_{\text{opt}}, W_0, \{d_k\}_1^K)$  is given in the next section, and is proved to be an achievable rate in [9],[19] as well as discussed in [2]. Indeed, since we are working with binary input (and output) channels, this mismatched achievable rate is equal to the mismatched capacity. We then define another parameter  $\beta_K$  by

$$\beta_K := \inf_{W_0 \in \text{BMC}} \frac{I_{\text{MIS}}(P_{\text{opt}}, W_0, \{d_k\}_1^K)}{I(P_{\text{opt}}, W_0)}. \quad (5)$$

Clearly  $0 \leq \alpha, \beta \leq 1$ . The problem of finding  $\alpha$  has already been solved in [22], and the answer is  $\alpha \approx 0.942$ , as reviewed below. From Theorem 1 (second part), one can show that by taking a large enough  $K$ , we can make  $\beta_K$  arbitrarily close to 1. Indeed, this relates to the fact that we can approximate the set of DMC's by a covering of one-sided components (for the uniform input distribution). Hence, one can study the speed convergence of  $\beta_K$ , in  $K$ , to deduce how many metrics in magnitude need to be used to achieve a given performance. We believe this is an interesting problem, as it captures the cost (in the number of additive metrics) needed to "replace" the MMI decoder with a generalized linear decoder; and this problem can be addressed for any alphabet dimensions. However, as motivated in previous section with Proposition 2, we hope to get exactly  $\beta_2 = 1$  for the binary alphabets setting, in which case we do not need to investigate the speed convergence problem (this will indeed be the case).

#### IV. RESULTS

##### A. Optimal input distribution

The optimization problem for  $\alpha$  has been investigated in [22] with the following answer.

*Theorem 2 (Shulman and Feder):*  $\alpha \approx 0.942$  and  $P_{\text{opt}}$  is the uniform distribution.

The authors also identified the worst channel to be a Z-channel. This result is also a ramification of the fact that with uniform source distribution, the maximum loss for any channel is less than 5.8% of capacity, as originally reported by Majani and Rumsey [18].

##### B. Optimal Generalized Linear Decoder

We represent a BMC by a point in  $(a, b) \in [0, 1]^2$ , with the following mapping to specify the BMC

$$\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}, \quad 0 \leq a, b \leq 1.$$

*Definition 3:* Let  $B^- = \{(a, b) \in [0, 1]^2 | a + b < 1\}$  and  $B^+ = \{(a, b) \in [0, 1]^2 | a + b \geq 1\}$ , and let  $U$  denote the uniform binary distribution.

Note that  $B^-$  parameterizes the set of BMC's which are *flipping-like*, in the sense that, assuming the input and output alphabets to be  $\{0, 1\}$ , for any given output  $y$  of a BMC in  $B^-$ , it is more likely that the sent input is  $1 + y \pmod{2}$ . Similarly,  $B^+$  parameterizes the set of BMC's which are *non-flipping-like*, containing in particular the set of channels for

which  $a + b = 1$ , which are all the pure noise channels (zero mutual information).

*Proposition 1:* For any  $i \in \{-, +\}$  and any  $W_0 \in B^i$ , we have

$$I_{\text{MIS}}(U, W_0, \log W_1) = \begin{cases} I(U, W_0) & \text{if } W_1 \in B^i \\ 0 & \text{otherwise.} \end{cases}$$

This proposition tells us that, as long as the channel used for the decoding ( $W_1$ ) is in the same class ( $B^-$  or  $B^+$ ) as the true channel ( $W_0$ ), the mismatched mutual information is equal to the mutual information itself; i.e., equal to all of the mutual information being evaluated with the uniform input distribution. If instead the channel and the metrics each other hail from different class, then the mismatched mutual information is zero.

*Proof:* As defined in Proposition (1), we have

$$I_{\text{MIS}}(U, W_0, \log W_1) = \inf_{\mu \in \mathcal{A}} D(\mu || \mu_0^p) \quad (6)$$

where  $\mathcal{A} = \{\mu : \mu^p = \mu_0^p, \mathbb{E}_\mu \log W_1 \geq \mathbb{E}_{\mu_0} \log W_1\}$ . Note that the channels  $W$  which induce a  $\mu$  such that  $\mu^p = \mu_0^p$  is parameterized in  $[0, 1]^2$  by the line passing through  $\mu_0$  with a slope of 1. Hence, since  $\mu_0 \in \partial \mathcal{A}$  (the boundary of  $\mathcal{A}$ ), it is easy to verify that the region  $\mathcal{A}$  is the segment starting at  $\mu_0$  and going either up or down (with slope 1). This leads to two possibilities, either  $\mu_0^p \in \mathcal{A}$  and (6) is 0, or  $\mu_0^p \notin \mathcal{A}$  and the minimizer of (6) is  $\mu_0$ , implying the claims of Proposition 1. ■

*Proposition 2:* For any binary input/output channel  $W_0$  and for any binary symmetric channel  $W_1$ , i.e.,  $W_1(0|0) = W_1(1|1)$ , we have

$$I_{\text{MIS}}(U, W_0, \{\log W_1, \log \widetilde{W}_1\}) = I(U, W_0),$$

where  $\widetilde{W}_1$  is the BSC defined by  $\widetilde{W}_1(0|0) = 1 - W_1(0|0)$ .

*Proof:* As defined in Lemma (1), we have

$$I_{\text{MIS}}(U, W_0, \{\log W_1, \log \widetilde{W}_1\}) = \bigvee_{k=1}^2 \inf_{\mu \in \mathcal{A}_k} D(\mu || \mu_0^p), \quad (7)$$

where  $\mathcal{A}_k = \{\mu : \mu^p = \mu_0^p, \mathbb{E}_\mu \log W_k \geq \bigvee_{j=1}^2 \mathbb{E}_{\mu_0} \log W_j\}$ ,  $k = 1, 2$ , where  $W_2 = \widetilde{W}_1$ . Note that, although in general taking the likelihood metrics as opposed to the a posteriori metrics makes an important difference when defining a generalized linear decoder (cf. [2]), here it does not, since we are working with BSC channels for the metrics. Assume w.l.o.g. that  $W_0, W_1 \in B^-$ . Then, a straightforward computation shows that

$$\bigvee_{j=1}^2 \mathbb{E}_{\mu_0} \log W_j = \mathbb{E}_{\mu_0} \log W_1.$$

Hence

$$\inf_{\mu \in \mathcal{A}_1} D(\mu || \mu_0^p) = I_{\text{MIS}}(U, W_0, \log W_1),$$

and from Proposition 1

$$I_{\text{MIS}}(U, W_0, \log W_1) = I(U, W_0).$$

Moreover, note that for any channel  $W_0$ , if we define  $\widetilde{W}_0$  to be the reverse BSC (cf. Figure 1), and  $\mu_0, \widetilde{\mu}_0$  to be the corresponding measures, we have  $\mu_0^p = \widetilde{\mu}_0^p$ ,

$$\mathbb{E}_{\mu_0} \log W_1 = \mathbb{E}_{\widetilde{\mu}_0} \log \widetilde{W}_1$$

and

$$\begin{aligned} \mathcal{A}_2 &= \{\mu : \mu^p = \mu_0^p, \mathbb{E}_{\mu} \log W_2 = \mathbb{E}_{\mu_0} \log W_1\} \\ &= \{\mu : \mu^p = \widetilde{\mu}_0^p, \mathbb{E}_{\mu} \log \widetilde{W}_1 = \mathbb{E}_{\widetilde{\mu}_0} \log \widetilde{W}_1\}. \end{aligned}$$

Therefore,

$$\inf_{\mu \in \mathcal{A}_2} D(\mu \| \mu_0^p) = I_{\text{MIS}}(U, \widetilde{W}_0, \widetilde{W}_1) = I_{\text{MIS}}(U, W_0, W_1),$$

and both terms in the RHS of (7) are equal to  $I(U, W_0)$ . ■

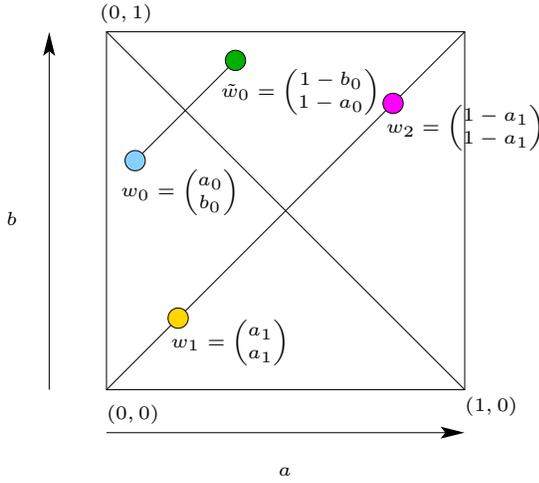


Fig. 1. Reverse channels in the BMC setting

An extended discussion and alternate proofs of previous results can be found in [20].

*Corollary 1:* We have  $\beta_2 = 1$ , which is achieved by picking any two metrics of the form  $d_1 = \log W_1$ ,  $d_2 = \log \widetilde{W}_1$ , as long as  $W_1$  is a BSC (and  $\widetilde{W}_1$  its reverse BSC).

*Corollary 2:* For any compound sets  $\mathcal{S}$ , 94.2% of the compound capacity can be achieved by using a generalized linear decoder induced by two metrics. Moreover, if the optimal input distribution (achieving compound capacity) is uniform, we can achieve 100% of the compound capacity with two metrics.

Note: if the optimal input distribution is non-uniform, it may still be possible to achieve 100% of the compound capacity with two metrics, but the above results will have to be adapted to the non-uniform input distribution case.

## V. DISCUSSION

In this paper, we have shown that, for binary input binary output memoryless channels, compound capacity achieving decoders can have a much simpler structure than the Maximum Mutual Information (MMI) decoder. These decoders, namely the generalized linear decoders, preserve many features of the MMI decoder. When the input distribution is chosen to be uniform, a generalized linear decoder, using channel independent metrics (i.e., the metrics are selected without

being aware of the channel rule) is shown to achieve the same rate as that of an optimum decoder (based on the exact channel rule). For the compound set consisting of all possible binary memoryless channels, this decoder is universal. On the other hand, for any arbitrary compound BMC, at least 94.2% of the compound capacity can be realized by such a decoder. Finally, a natural extension of this work would be to investigate the case of non-binary alphabets. Even for binary input and ternary outputs, it does not seem straightforward to establish whether  $\beta_K$  can be made exactly 1, for  $K$  large enough (although one can show that it must tend to 1 using results from [2]).

## ACKNOWLEDGMENT

Part of this work has been done in a master project at LCM-EPFL co-supervised by Bixio Rimoldi.

## REFERENCES

- [1] E. Abbe, "Local to geometric Methods in Information Theory, PhD thesis," Massachusetts Institute of Technology, 2008.
- [2] E. Abbe and L. Zheng, "Linear Universal Decoding for Compound Channels: a Local to Global Geometric Approach," arXiv:0809.5217v1 [cs.IT], Sep. 2008.
- [3] V. B. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rules," in Proc. 1st FrenchSoviet Workshop on Algebraic Coding (Lecture Notes in Computer Science 573), G. Cohen, S. Litsyn, A. Lobstein, and G. Zemor, Eds. Berlin, Germany: Springer-Verlag, pp. 142-150, Jul. 1991.
- [4] V. B. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," IEEE Trans. Inform. Theory, vol. 41, pp. 1889-1902, Nov. 1995.
- [5] D. Blackwell, L. Breiman and A.J. Thomasian, "The Capacity of a class of channels," *Ann. Mathe.Stat.*, vol. 30, pp. 1229-1241, Dec. 1959.
- [6] G. Como and F. Fagnani, "The capacity of abelian group codes over symmetric channels," IEEE Trans. Inform. Theory, submitted
- [7] T. M. Cover, "Broadcast channels," IEEE Trans. Inform. Theory, vol. IT-18, pp. 214, Jan. 1972.
- [8] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [9] I. Csiszar, P. Narayan, "Channel Capacity for a Given Decoding Metric", IEEE Trans. Inform. Theory, vol. 41, no.1, Jan. 1995.
- [10] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radio Eng. Electron.*, vol. 4, no. 12, pp. 18, 1959.
- [11] P. Elias, "Coding for two noisy channels", in Proc. 3rd London Symp. Information Theory, London, U.K., pp. 61-76, 1955.
- [12] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," IEEE Trans. Inform. Theory, vol. 44: pp. 1726-1745, Sep. 1998.
- [13] V. Goppa, "Nonprobabilistic mutual information without memory," *Problems of Control and Information Theory*, vol. 4, pp. 97-102, 1975.
- [14] J. Y. N. Hui, "Fundamental issues of multiple accessing," Ph.D. Thesis, Mass. Inst. Technol., Cambridge, MA, 1983.
- [15] A. Lapidoth, "Mismatched decoding and the multiple-access channel," IEEE Trans. Inform. Theory, vol. 42, pp. 1439-1452, Sep. 1996.
- [16] A. Lapidoth and P. Narayan, "Reliable Communication under channel uncertainty," IEEE Trans. Inform. Theory, vol.40, No.10, pp:2148-2177, Oct. 1998.
- [17] A. Lapidoth and I.E. Telatar, "The compound channel capacity of a class of finite state channels," IEEE Trans. Inform. Theory, Vol.44, No.3, May 1998.
- [18] E.E. Majani and H. Rumsey, "Two results on binary-input discrete memoryless channels," In Proc. IEEE Int. Symp. Information Theory, Budapest, Hungary, pp. 104, 1991.
- [19] N. Merhav, G. Kaplan, A. Lapidoth and S. Shamai, "On information rates for mismatched decoders," IEEE Trans. Inform. Theory, vol 40. No.6 pp. 953-1967, Nov. 1994.
- [20] R. Pulikkoonattu, "Linear Universal Decoder for Compound Discrete Memoryless Channels," MS Thesis, EPFL, Jul. 2009.
- [21] T. Richardson and R. Urbanke, *Modern Coding theory*, Cambridge University Press, 2007.

- [22] N. Shulman and M. Feder, "The uniform Distribution as a Universal Prior," IEEE Trans. Inform. Theory, vol.50.No.6, Jun. 2004.
- [23] J. Ziv, "Universal decoding for finite state channels," IEEE Trans. Inform. Theory, vol. 31, pp. 453-460, Jul. 1985.