# Achieving Marton's Region for Broadcast Channels Using Polar Codes

Marco Mondelli, S. Hamed Hassani, and Rüdiger Urbanke
School of Computer and Communication Sciences
EPFL, CH-1015 Lausanne, Switzerland
Email: {marco.mondelli, seyehamed.hassani, ruediger.urbanke}@epfl.ch

Igal Sason
Department of Electrical Engineering
Technion, Haifa 32000, Israel
Email: sason@ee.technion.ac.il

***Abstract*—We present polar coding schemes for the 2-user discrete memoryless broadcast channel (DM-BC) which achieve Marton's region with both common and private messages. This is the best achievable rate region up to date, and it is tight for all classes of 2-user DM-BCs whose capacity regions are known. Due to space limitations, this paper describes polar codes for the superposition strategy. The scheme for the achievability of Marton's region is presented in the longer version [1], and it is based on a combination of *superposition coding* and *binning*. We follow the lead of the recent work by Goela, Abbe, and Gastpar, who introduce polar codes emulating these two information-theoretic techniques. In order to align the polar indices, for both schemes, their solution involves some degradedness constraints that are assumed to hold between the auxiliary random variables and the channel outputs. To remove these constraints, we consider the transmission of $k$ blocks, and employ chaining constructions that guarantee the proper alignment of polarized indices. The techniques described in this work are quite general, and they can be adopted in many other multi-terminal scenarios whenever there is the need for the aligning of polar indices.**

## I. Introduction

Polar coding, introduced by Arıkan in [2], allows one to achieve the capacity of binary-input memoryless output-symmetric channels (BMSCs) with encoding and decoding complexity $\Theta(n \log n)$ and a block error probability decaying like $O(2^{-n^\beta})$, where $n$ is the block length of the code and $\beta \in (0, 1/2)$, see [3]. The original point-to-point scheme has been extended, amongst others, to lossless and lossy source coding and to several multi-user scenarios (see [1] and references therein).

Goela, Abbe, and Gastpar recently introduced polar coding schemes for the $m$-user deterministic broadcast channel and for the noisy discrete memoryless broadcast channel (DM-BC) [4]. For the second scenario, they considered two fundamental transmission strategies: *superposition coding*, in the version proposed by Bergmans [5], and *binning* [6]. In order to guarantee a proper alignment of the polar indices, in both the superposition and binning schemes, their solution involves some degradedness constraints that are assumed to hold between the auxiliary random variables and the channel outputs.

Due to space limitations, this paper is only focused on the construction of polar codes to achieve the whole superposition region. A more detailed presentation and a polar scheme to achieve Marton's region with both common and private messages is provided in the longer version [1]. The crucial point consists in removing the degradedness conditions on auxiliary random variables and channel outputs[1]. The ideas which make it possible to lift the constraints come from recent progress in constructing *universal* polar codes, which are capable of achieving the compound capacity of the whole class of BMSCs [8], [9]. The proposed schemes possess the standard properties of polar codes with respect to encoding and decoding, which can be performed with complexity $\Theta(n \log n)$, as well as with respect to the scaling of the block error probability as a function of the block length, which decays like $O(2^{-n^\beta})$ for any $\beta \in (0, 1/2)$.

The remainder of the paper is organized as follows: Section II reviews the rate region achievable by superposition coding from the information-theoretic perspective as well as via the polar construction proposed in [4], call it the AGG construction. In Section III, we review two "polar primitives" that are the basis of the AGG construction and of our extension. In Section IV, we describe our polar coding scheme that achieves the whole superposition region. We conclude with some final thoughts in Section V.

## II. Bergmans's Superposition Region

Let us start by recalling the rate region achievable by Bergmans's superposition scheme [10].

*Theorem 1 (Superposition Region):* Consider the transmission over a 2-user DM-BC $p_{Y_1,Y_2 \mid X}$, where $X$ denotes the input to the channel, and $Y_1$, $Y_2$ denote the outputs at the first and second receiver, respectively. Let $V$ be an auxiliary random variable. Then, for any joint distribution $p_{V,X}$ s.t. $V - X - (Y_1, Y_2)$ forms a Markov chain, a rate pair $(R_1, R_2)$ is achievable if

$$\begin{aligned} R_1 &< I(X; Y_1 \mid V), \\ R_2 &< I(V; Y_2), \\ R_1 + R_2 &< I(X; Y_1). \end{aligned} \tag{1}$$

Note that the above only describes "half" of the region actually achievable by superposition coding. We get the second "half" by swapping the roles of the two users, i.e., by swapping the indices 1 and 2. The actual achievable region is the convex hull of the closure of the union of these two subsets.

Let us compare (1) with the region achievable by the AGG construction [4]. We write $p \succ q$ to denote that the channel $q$ is degraded with respect to the channel $p$.

[1]Note that, in general, such kind of extra conditions make the achievable rate region strictly smaller, see [7].

*Theorem 2 (AGG Superposition Region):* Consider the transmission setting defined in Theorem 1, where the channel input alphabet is supposed to be binary. Assume further that $p_{Y_1 \mid V}(y_1 \mid v) \succ p_{Y_2 \mid V}(y_2 \mid v)$. Then, for any rate pair $(R_1, R_2)$ s.t.

$$R_1 < I(X; Y_1 \mid V), \quad R_2 < I(V; Y_2), \tag{2}$$

there exists a sequence of polar codes with increasing block length $n$ that achieves this rate pair with encoding and decoding complexity $\Theta(n \log n)$ and with a block error probability that decays like $O(2^{-n^\beta})$ for $\beta \in (0, 1/2)$.

As before, the region (2) describes "half" of the region actually achievable with polar codes by superposition coding. However, to achieve both "halves" of the region, we need that $p_{Y_1 \mid V}(y_1 \mid v) \succ p_{Y_2 \mid V}(y_2 \mid v)$ and $p_{Y_2 \mid V}(y_2 \mid v) \succ p_{Y_1 \mid V}(y_1 \mid v)$, which is equivalent to $p_{Y_1 \mid V}(y_1 \mid v) = p_{Y_2 \mid V}(y_2 \mid v)$. Note that the AGG superposition scheme is optimal for the class of stochastically-degraded broadcast channels. In addition, the alignment of indices which motivates the degradedness assumption also occurs for less noisy and, under suitable conditions, for more capable broadcast channels [11].

In order to simplify the description of our novel polar coding scheme, we reduce the achievability of the rate region (1) to the achievability of certain specific rate pairs. The following proposition is proved in [1].

*Proposition 1 (Equivalent Superposition Region):* To obtain the whole region (1), it suffices to describe a coding scheme that achieves the following rate pairs. First,

$$(R_1, R_2) = (I(X; Y_1 \mid V), \min_i I(V; Y_i)). \tag{3}$$

Second, if $I(V; Y_1) < I(V; Y_2) < I(X; Y_1)$, we need to achieve the rate pair

$$(R_1, R_2) = (I(X; Y_1) - I(V; Y_2), I(V; Y_2)). \tag{4}$$

## III. Polar Coding Primitives

First, we discuss the problem of lossless compression, with and without side information, considering the point of view of Arıkan in [12]. Then, we deal with the transmission over a general binary-input discrete memoryless channel, where the capacity achieving input distribution might not be the uniform one which is imposed by linear codes. Polar codes offer a direct and elegant solution to this problem [13].

*Notation:* In what follows, we assume that $n$ is a power of 2, say $n = 2^m$ for $m \in \mathbb{N}$, and we define $G_n$ to be the polar matrix given by $G_n = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m}$, where $\otimes$ denotes the Kronecker product of matrices. The set $\{1, \cdots, n\}$ is abbreviated as $[n]$ and, given a set $\mathcal{A} \subseteq [n]$, we denote by $\mathcal{A}^c$ its complement. We use $X^{i:j}$ as a shorthand for $(X^i, \cdots, X^j)$.

### A. Lossless Compression

Consider a binary random variable $X \sim p_X$. Then, given the random vector $X^{1:n} = (X^1, \cdots, X^n)$ consisting of $n$ i.i.d. copies of $X$, the aim is to compress $X^{1:n}$ in a lossless fashion into a binary codeword of size roughly $nH(X)$, which is the entropy of $X^{1:n}$.

Let $U^{1:n} = (U^1, \cdots, U^n)$ be defined as $U^{1:n} = X^{1:n} G_n$. Then, $U^{1:n}$ is a random vector whose components are polarized in the sense that either $U^i$ is approximately uniform and independent of $U^{1:i-1}$, or $U^i$ is approximately a deterministic function of $U^{1:i-1}$. Formally, for $\beta \in (0, 1/2)$, let $\delta_n = 2^{-n^\beta}$ and set

$$\begin{aligned} \mathcal{H}_X &= \{i \in [n] : Z(U^i \mid U^{1:i-1}) \geq 1 - \delta_n\}, \\ \mathcal{L}_X &= \{i \in [n] : Z(U^i \mid U^{1:i-1}) \leq \delta_n\}, \end{aligned} \tag{5}$$

where $Z$ denotes the Bhattacharyya parameter. Recall that, given $(T, V) \sim p_{T,V}$, where $T$ is binary and $V$ takes values in an arbitrary discrete alphabet $\mathcal{V}$, we define

$$Z(T \mid V) = 2 \sum_{v \in \mathcal{V}} \mathbb{P}_V(v) \sqrt{\mathbb{P}_{T \mid V}(0 \mid v) \mathbb{P}_{T \mid V}(1 \mid v)}.$$

Hence, for $i \in \mathcal{H}_X$, the bit $U^i$ is approximately uniformly distributed and independent of the past $U^{1:i-1}$; also, for $i \in \mathcal{L}_X$, the bit $U^i$ is approximately a deterministic function of $U^{1:i-1}$. Furthermore,

$$\lim_{n \to \infty} \frac{1}{n} |\mathcal{H}_X| = H(X), \qquad \lim_{n \to \infty} \frac{1}{n} |\mathcal{L}_X| = 1 - H(X). \tag{6}$$

Given the vector $x^{1:n}$ that we want to compress, the encoder computes $u^{1:n} = x^{1:n} G_n$ and outputs the values of $u^{1:n}$ in the positions $\mathcal{L}_X^c$, i.e., it outputs $\{u^i\}_{i \in \mathcal{L}_X^c}$. Then, the decoder can reconstruct $u^i$ for $i \in \mathcal{L}_X$, according to the rule

$$\hat{u}^i = \arg \max_{u \in \{0,1\}} \mathbb{P}_{U^i \mid U^{1:i-1}}(u \mid u^{1:i-1}). \tag{7}$$

Indeed, for $i \in \mathcal{L}_X$ the distribution $\mathbb{P}_{U^i \mid U^{1:i-1}}(u \mid u^{1:i-1})$ is highly biased towards the correct value $u_i$. The probabilities in (7) can be computed recursively with complexity $\Theta(n \log n)$ and the block error probability is $O(2^{-n^\beta})$.

Consider now the case with side information and let $(X, Y) \sim p_{X,Y}$ be a pair of random variables, where we think of $X$ as the source to be compressed and of $Y$ as a *side information* about $X$. Given the vector $(X^{1:n}, Y^{1:n})$ of $n$ independent samples from the distribution $p_{X,Y}$, the problem is to compress $X^{1:n}$ into a codeword of size roughly $nH(X \mid Y)$, where the side information $Y^{1:n}$ is available at the decoder.

Let $U^{1:n} = X^{1:n} G_n$ and consider the sets $\mathcal{H}_{X \mid Y}$ and $\mathcal{L}_{X \mid Y}$, defined by

$$\begin{aligned} \mathcal{H}_{X \mid Y} &= \{i \in [n] : Z(U^i \mid U^{1:i-1}, Y^{1:n}) \geq 1 - \delta_n\}, \\ \mathcal{L}_{X \mid Y} &= \{i \in [n] : Z(U^i \mid U^{1:i-1}, Y^{1:n}) \leq \delta_n\}, \end{aligned} \tag{8}$$

which, respectively, represent the positions s.t. $U^i$ is approximately uniform and independent of $(U^{1:i-1}, Y^{1:n})$ and the positions s.t. $U^i$ is approximately a deterministic function of $(U^{1:i-1}, Y^{1:n})$. Then,

$$\lim_{n \to \infty} \frac{1}{n} |\mathcal{H}_{X|Y}| = H(X|Y), \quad \lim_{n \to \infty} \frac{1}{n} |\mathcal{L}_{X|Y}| = 1 - H(X|Y). \tag{9}$$

Given a realization of $X^{1:n}$, namely $x^{1:n}$, the encoder constructs $u^{1:n} = x^{1:n} G_n$ and outputs $\{u^i\}_{i \in \mathcal{L}_{X \mid Y}^c}$ as the compressed version of $x^{1:n}$. The decoder, using the side information $y^{1:n}$ and a decoding rule similar to (7) is able to reconstruct $x^{1:n}$ with vanishing block error probability.

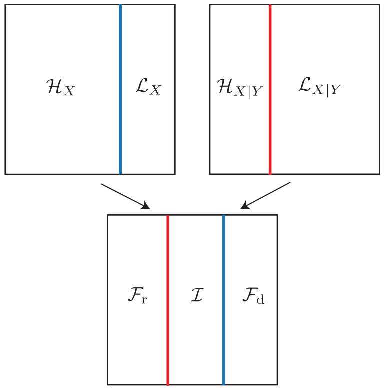



Figure 1. Graphical representation for the sets associated to the (asymmetric) channel coding problem. The two images on top represent how the set $[n]$ (the whole square) is partitioned by the source $X$ (top left), and by the source $X$ together with the output $Y$ assumed as a side information (top right). Since $\mathcal{H}_{X\,|\,Y} \subseteq \mathcal{H}_X$, $[n]$ can be partitioned into three subsets (bottom image): the information indices $\mathcal{I}$; the frozen indices $\mathcal{F}_\mathrm{r}$ filled with binary bits chosen uniformly at random; the frozen indices $\mathcal{F}_\mathrm{d}$ chosen according to a deterministic rule.

### B. Transmission over Binary-Input DMCs

Let $W$ be a discrete memoryless channel with input $X$ and output $Y$. Fix a distribution $p_X$ over $X$. The aim is to transmit over $W$ with a rate close to $I(X;Y)$.

Let $U^{1:n} = X^{1:n}G_n$, where $X^{1:n}$ is a vector of $n$ i.i.d. components drawn according to $p_X$. Consider the sets $\mathcal{H}_X$ and $\mathcal{L}_X$ defined in (5). Now, assume that the channel output $Y^{1:n}$ is given, and interpret this as side information on $X^{1:n}$. Consider the sets $\mathcal{H}_{X\,|\,Y}$ and $\mathcal{L}_{X\,|\,Y}$, as defined in (8). It is clear that

$$\mathcal{H}_{X\,|\,Y} \subseteq \mathcal{H}_X, \qquad \mathcal{L}_X \subseteq \mathcal{L}_{X\,|\,Y}. \tag{10}$$

To construct a polar code for the channel $W$ we proceed now as follows. We place the information in the positions indexed by $\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X\,|\,Y}$. Indeed, if $i \in \mathcal{I}$, then $U^i$ is approximately random given $U^{1:i-1}$, since $i \in \mathcal{H}_X$. This implies that $U^i$ is suitable to contain information. Further, $U^i$ is approximately a deterministic function if we are given $U^{1:i-1}$ *and* $Y^{1:n}$, since $i \in \mathcal{L}_{X\,|\,Y}$. This implies that it is also decodable in a successive manner given the channel output. By using (6), (9), and (10), it follows that

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}| = H(X) - H(X\,|\,Y) = I(X;Y). \tag{11}$$

Hence, our requirement on the transmission rate is met. The remaining positions are frozen. More precisely, they are divided into two subsets, namely $\mathcal{F}_\mathrm{r} = \mathcal{H}_X \cap \mathcal{L}^c_{X\,|\,Y}$ and $\mathcal{F}_\mathrm{d} = \mathcal{H}^c_X$. For $i \in \mathcal{F}_\mathrm{r}$, $U^i$ is independent of $U^{1:i-1}$, but cannot be reliably decoded using $Y^{1:n}$. We fill these positions with bits chosen uniformly at random, and this randomness is assumed to be shared between the transmitter and the receiver (i.e., the encoder and the decoder know the values associated to these positions). For $i \in \mathcal{F}_\mathrm{d}$, $U^i$ is approximately a deterministic function of $U^{1:i-1}$, and its value can be chosen according to $\arg\max_{u\in\{0,1\}} \mathbb{P}_{U^i\,|\,U^{1:i-1}}(u\,|\,u^{1:i-1})$. The situation is schematically represented in Figure 1.

## IV. Polar Codes for Superposition Region

The theorem below provides our main result about the achievability of the superposition region with polar codes.

*Theorem 3 (Polar Superposition Region):* Consider the transmission setting defined in Theorem 1, where the channel input alphabet is supposed to be binary. Then, for any rate pair $(R_1, R_2)$ satisfying the constraints in (1), there exists a sequence of polar codes with increasing block length $n$ which achieves this rate pair with encoding and decoding complexity $\Theta(n \log n)$ and a block error probability decaying like $O(2^{-n^\beta})$ for any $\beta \in (0, 1/2)$.

Let $(V, X) \sim p_{V,X} = p_V p_{X\,|\,V}$. We will show how to transmit over the 2-user DM-BC $p_{Y_1,Y_2\,|\,X}(y_1, y_2\,|\,x)$ achieving the rate pair (3) when $I(V;Y_1) < I(V;Y_2) < I(X;Y_1)$. Then, a slight modification of this scheme allows us to achieve, in addition, the rate pair (4). Therefore, by Proposition 1, we can achieve the whole region (1) and Theorem 3 is proved.

Set $U_2^{1:n} = V^{1:n}G_n$. As in the case of the transmission over an asymmetric channel with $V$ in place of $X$ and $Y_i$ ($i \in \{1, 2\}$) in place of $Y$, define the sets $\mathcal{H}_V$, $\mathcal{L}_V$, $\mathcal{H}_{V\,|\,Y_i}$, and $\mathcal{L}_{V\,|\,Y_i}$, which satisfy

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_V| = H(V), \qquad \lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_V| = 1 - H(V),$$
$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_{V|Y_i}| = H(V|Y_i), \quad \lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_{V|Y_i}| = 1 - H(V|Y_i).$$

Set $U_1^{1:n} = X^{1:n}G_n$. By thinking of $V$ as side information and by considering the transmission of $X$ over the (asymmetric) channel with output $Y_1$, define also the sets $\mathcal{H}_{X\,|\,V}$, $\mathcal{L}_{X\,|\,V}$, $\mathcal{H}_{X\,|\,V,Y_1}$, and $\mathcal{L}_{X\,|\,V,Y_1}$, which satisfy

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_{X|V}| = H(X|V), \quad \lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_{X|V}| = 1 - H(X|V),$$
$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{H}_{X|V,Y_1}| = H(X|V, Y_1),$$
$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{L}_{X|V,Y_1}| = 1 - H(X|V, Y_1).$$

First, consider only the point-to-point communication problem between the transmitter and the second receiver. As discussed in Section III-B, for this scenario the correct choice is to place the information in those positions of $U_2^{1:n}$ that are indexed by the set $\mathcal{I}^{(2)} = \mathcal{H}_V \cap \mathcal{L}_{V\,|\,Y_2}$. If, in addition, we restrict ourselves to positions in $\mathcal{I}^{(2)}$ which are also contained in $\mathcal{I}_v^{(1)} = \mathcal{H}_V \cap \mathcal{L}_{V\,|\,Y_1}$, then also the first receiver will be able to decode this message. Indeed, recall that in the superposition coding scheme the first user needs to decode the message intended for the second user, before decoding his own message. Consequently, the first user knows the vector $U_2^{1:n}$, and, hence, also the vector $V^{1:n}$. Now, consider the point-to-point communication problem between the transmitter and the first receiver, given the side information $V^{1:n}$ (as we just discussed, this vector is known to the first receiver). Then, the information has to be placed in those positions of $U_1^{1:n}$ that are indexed by $\mathcal{I}^{(1)} = \mathcal{H}_{X\,|\,V} \cap \mathcal{L}_{X\,|\,V,Y_1}$. The cardinalities of these information sets are given by

$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}^{(2)}| = I(V;Y_2), \quad \lim_{n\to\infty} \frac{1}{n}|\mathcal{I}_v^{(1)}| = I(V;Y_1),$$
$$\lim_{n\to\infty} \frac{1}{n}|\mathcal{I}^{(1)}| = I(X;Y_1\,|\,V).$$

Let us now get back to the broadcasting scenario and see how we can use the previous observations to construct a polar coding scheme. Remember that $X^{1:n}$ is transmitted, the second user only decodes "his" message, but the first user decodes both messages.

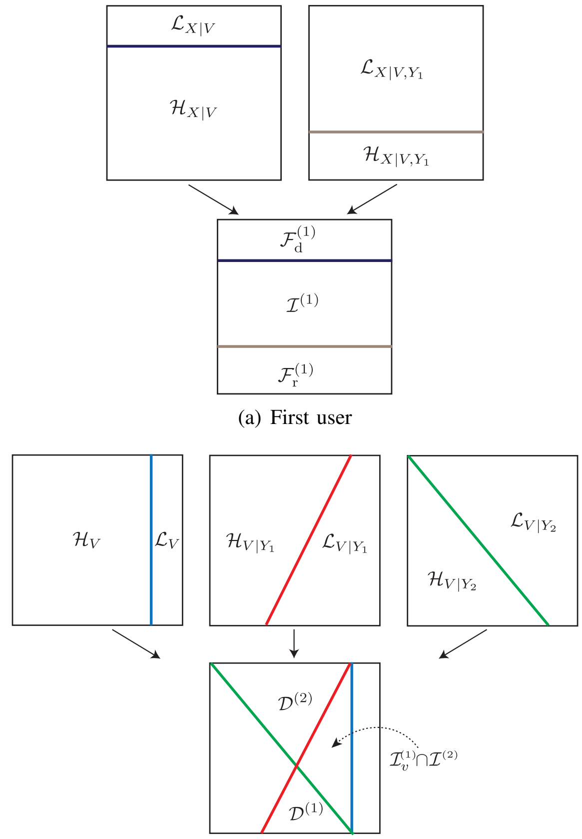
We start by reviewing the AGG scheme [4]. This scheme achieves the rate pair $(R_1, R_2) = (I(X; Y_1 \mid V), I(V; Y_2))$, assuming that $p_{Y_1 \mid V} \succ p_{Y_2 \mid V}$. Note that if $p_{Y_2 \mid V}$ is degraded with respect to $p_{Y_1 \mid V}$, then $\mathcal{L}_{V \mid Y_2} \subseteq \mathcal{L}_{V \mid Y_1}$, which implies that $\mathcal{I}^{(2)} \subseteq \mathcal{I}_v^{(1)}$. As a consequence, we can in fact use the point-to-point solutions outlined above, i.e., the second user can place his information in $\mathcal{I}^{(2)}$ and decode, and the first user will also be able to decode this message. Furthermore, once the message intended for the second user is known by the first user, the latter can decode his own information which is placed in the positions in $\mathcal{I}^{(1)}$.

Let us now see how to eliminate the restriction imposed by the degradedness condition $p_{Y_1 \mid V} \succ p_{Y_2 \mid V}$. Recall that we want to achieve the rate pair (3) when $I(V; Y_1) < I(V; Y_2) < I(X; Y_1)$. The information set for the first user is exactly the same as before, namely the positions of $U_1^{1:n}$ indexed by $\mathcal{I}^{(1)}$. The only difficulty lies in designing a scheme in which *both* receivers can decode the message intended for the second user.

First of all, one can use all the positions in $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, since the information in these positions is decodable by both users. Let us define $\mathcal{D}^{(2)} = \mathcal{I}^{(2)} \setminus \mathcal{I}_v^{(1)}$. If $p_{Y_1|V} \succ p_{Y_2|V}$, as before, then $\mathcal{D}^{(2)} = \emptyset$, i.e., all the positions decodable by the second user are also decodable by the first user. However, in the general case, i.e., if we no longer assume that $p_{Y_1|V} \succ p_{Y_2|V}$, then $\mathcal{D}^{(2)}$ is non-empty and those positions cannot be decoded by the first user. Note that there is a similar set, but with the roles of the two users exchanged, call it $\mathcal{D}^{(1)}$, namely, $\mathcal{D}^{(1)} = \mathcal{I}_v^{(1)} \setminus \mathcal{I}^{(2)}$. The set $\mathcal{D}^{(1)}$ contains the positions of $U_2^{1:n}$ which are decodable by the first user, but not by the second user. Observe further that $|\mathcal{D}^{(1)}| \le |\mathcal{D}^{(2)}|$ for sufficiently large $n$, since

$$\begin{aligned}\frac{1}{n}(|\mathcal{D}^{(2)}| - |\mathcal{D}^{(1)}|) &= \frac{1}{n}(|\mathcal{I}^{(2)}| - |\mathcal{I}_v^{(1)}|) \\ &= I(V; Y_2) - I(V; Y_1) + o(1) \ge 0.\end{aligned} \tag{12}$$

Assume at first that the sizes of the two sets are equal. The general case will require only a small modification. The idea is to consider the "chaining" construction introduced in [8] to define universal polar codes. Recall that we are only interested in the message intended for the second user, but that both receivers must be able to decode such a message. Our scheme consists in transmitting $k$ polar blocks and in repeating ("chaining") some information. More precisely, in block 1 fill $\mathcal{D}^{(1)}$ with information, but set the bits indexed by $\mathcal{D}^{(2)}$ to a fixed known sequence. In block $j$ ($j \in \{2, \cdots, k-1\}$), fill $\mathcal{D}^{(1)}$ again with information, and repeat the bits which were contained in the set $\mathcal{D}^{(1)}$ of block $j-1$ into the positions indexed by $\mathcal{D}^{(2)}$ of block $j$. In the final block $k$, put a known sequence in the positions indexed by $\mathcal{D}^{(1)}$, and repeat in the positions $\mathcal{D}^{(2)}$ the bits which were contained in the set $\mathcal{D}^{(1)}$ of block $k-1$. The remaining bits are frozen and, as in Section III-B, they are divided into the two subsets $\mathcal{F}_{\mathrm{d}}^{(2)} = \mathcal{H}_V^{\mathrm{c}}$ and $\mathcal{F}_{\mathrm{r}}^{(2)} = \mathcal{H}_V \cap \mathcal{L}_{V \mid Y_2}^{\mathrm{c}} \subset \mathcal{H}_V$. In the first case, $U_2^i$ is




(a) First user

(b) Second user

Figure 2. As concerns the first user, the set $[n]$ is partitioned into three subsets: the information indices $\mathcal{I}^{(1)}$, and the frozen indices $\mathcal{F}_{\mathrm{r}}^{(1)}$ and $\mathcal{F}_{\mathrm{d}}^{(1)}$. As concerns the second user, $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$ contains the indices which are decodable by both users; $\mathcal{D}^{(1)} = \mathcal{I}_v^{(1)} \setminus \mathcal{I}^{(2)}$ contains the indices decodable by the first user, but not by the second; $\mathcal{D}^{(2)} = \mathcal{I}^{(2)} \setminus \mathcal{I}_v^{(1)}$ contains the indices decodable by the second user, but not by the first.

approximately a deterministic function of $U_2^{1:i-1}$, while in the second case $U_2^i$ is approximately independent of $U_2^{1:i-1}$. Note that we lose some rate, since at the boundary we put a known sequence into some bits which were supposed to contain information. However, this rate loss decays like $1/k$ so, by choosing $k$ large, we achieve a rate as close as desired to the intended one.

In the above construction both users can decode all blocks, but the first receiver has to decode "forward", starting with block 1 and ending with block $k$, whereas the second receiver decodes "backwards", starting with block $k$ and ending with block 1. Let us discuss this procedure in some more detail. Look at the first user and start with block 1. By construction, information is only contained in the positions indexed by $\mathcal{D}^{(1)}$ as well as $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, while the positions indexed by $\mathcal{D}^{(2)}$ are set to known values. Hence, the first user can decode this block. For block $j$ ($j \in \{2, \cdots, k-1\}$) the situation is similar: the first user decodes the positions indexed by $\mathcal{D}^{(1)}$ and $\mathcal{I}_v^{(1)} \cap \mathcal{I}^{(2)}$, while the positions in $\mathcal{D}^{(2)}$ contain repeated information, which has been already decoded in the previous block. An analogous analysis applies to block $k$, in which the positions indexed by $\mathcal{D}^{(1)}$ are also fixed to a known sequence. The second user proceeds exactly in the same fashion, but it goes backwards.

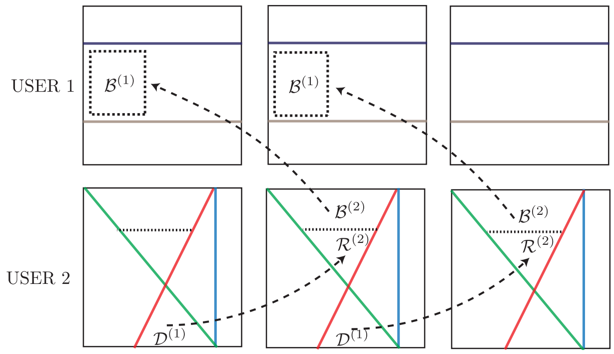



Figure 3. Graphical representation of the repetition construction for the superposition coding scheme with $k = 3$: the set $\mathcal{D}^{(1)}$ is repeated into the set $\mathcal{R}^{(2)}$ of the following block; the set $\mathcal{B}^{(2)}$ is repeated into the set $\mathcal{B}^{(1)}$ of the previous block (belonging to a different user).

To get to the general scenario, we need to discuss what happens when $|\mathcal{D}^{(1)}| < |\mathcal{D}^{(2)}|$. In this case, we do not have sufficiently many positions in $\mathcal{D}^{(1)}$ to repeat all the information contained in $\mathcal{D}^{(2)}$. To get around this problem, pick sufficiently many extra positions out of the vector $U_1^{1:n}$ indexed by $\mathcal{I}^{(1)}$, and repeat the extra information there. In order to specify this scheme, let us introduce some notation for the various sets. Recall that we "chain" the positions in $\mathcal{D}^{(1)}$ with an equal amount of positions in $\mathcal{D}^{(2)}$. It does not matter what subset of $\mathcal{D}^{(2)}$ we pick, but call the chosen subset $\mathcal{R}^{(2)}$. Now, we still have some positions left in $\mathcal{D}^{(2)}$, call them $\mathcal{B}^{(2)}$. More precisely, $\mathcal{B}^{(2)} = \mathcal{D}^{(2)} \setminus \mathcal{R}^{(2)}$. By (12), it follows that

$$\frac{1}{n}|\mathcal{B}^{(2)}| = I(V;Y_2) - I(V;Y_1) + o(1) \geq 0.$$

Let $\mathcal{B}^{(1)}$ be a subset of $\mathcal{I}^{(1)}$ s.t. $|\mathcal{B}^{(1)}| = |\mathcal{B}^{(2)}|$. Again, it does not matter what subset we pick. The existence of such a set $\mathcal{B}^{(1)}$ is ensured by the fact that $I(X;Y_1) > I(V;Y_2)$. As explained above, we place in $\mathcal{B}^{(1)}$ the value of those extra bits from $\mathcal{D}^{(2)}$ which will help the first user to decode the message of the second user in the next block. Operationally, we repeat the information contained in the positions indexed by $\mathcal{B}^{(2)}$ into the positions indexed by $\mathcal{B}^{(1)}$ of the previous block. Of course, by doing this, the first user pays a rate penalty of $I(V;Y_2) - I(V;Y_1)$ compared to his original rate given by $\frac{1}{n}|\mathcal{I}^{(1)}| = I(X;Y_1|V) + o(1)$. Hence, $R_1$ approaches the required rate $I(X;Y_1) - I(V;Y_2)$, as $k$ goes large.

To summarize, the first user puts information bits at positions $\mathcal{I}^{(1)} \setminus \mathcal{B}^{(1)}$, repeats in $\mathcal{B}^{(1)}$ the information bits in $\mathcal{B}^{(2)}$ for the next block, and freezes the rest. In the last block, the information set is the whole $\mathcal{I}^{(1)}$. The frozen positions are divided into the usual two subsets $\mathcal{F}_{\rm r}^{(1)} = \mathcal{H}_{X\,|\,V} \cap \mathcal{L}^c_{X\,|\,V,Y_1}$ and $\mathcal{F}_{\rm d}^{(1)} = \mathcal{H}^c_{X\,|\,V}$, which contain positions s.t. $U_1^i$ is or is not, respectively, approximately independent of $(U_1^{1:i-1}, V^{1:n})$. The situation is schematically represented in Figures 2 and 3.

If we let $\frac{1}{n}|\mathcal{B}^{(2)}|$ go from $I(V;Y_2) - I(V;Y_1) + o(1)$ to $o(1)$, by applying the same scheme, one obtains the line going from the rate pair $(I(X;Y_1) - I(V;Y_2), I(V;Y_2))$ to $(I(X;Y_1\,|\,V), I(V;Y_1))$ without time-sharing. Finally, to obtain the pair $(I(X;Y_1\,|\,V), I(V;Y_2))$ when $I(V;Y_2) \leq I(V;Y_1)$, it suffices to set $\mathcal{B}^{(2)} = \emptyset$ and switch the roles of $\mathcal{I}^{(2)}$ and $\mathcal{I}_v^{(1)}$ in the discussion concerning the second user.

## V. Conclusions

Extending the work in [4], we have shown how to construct polar codes for the 2-user DM-BC that achieve the whole rate region defined by Bergmans's superposition strategy. The description of the polar scheme for Marton's region with both common and private messages is provided in the extended version [1]. The current exposition is limited to the case of binary auxiliary random variables and binary inputs. However, there is no fundamental difficulty in extending the work to the $q$-ary case [14]. We conclude by remarking that the chaining constructions used to align the polarized indices do not rely on the specific structure of the broadcast channel. Indeed, similar techniques have been applied, independently of this work, to interference networks [15] and, in general, we believe that they can be adapted to the design of polar coding schemes for a variety of multi-user scenarios.


## Acknowledgment

The work of M. Mondelli was supported by grant No. 200020_146832/1 of the Swiss National Science Foundation. The work of I. Sason was supported by the Israeli Science Foundation (ISF), grant no. 12/12.


## References


[1] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," Feb. 2014, [Online]. Available: http://arxiv.org/pdf/1401.6060v2.pdf.

[2] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[3] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seoul, South Korea, Jul. 2009, pp. 1493–1495.

[4] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," Jan. 2013, [Online]. Available: http://arxiv.org/pdf/1301.6150v1.pdf.

[5] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[6] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[7] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," *47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 16 – 21, 2009.

[8] S. H. Hassani and R. Urbanke, "Universal polar codes," Dec. 2013, [Online]. Available: http://arxiv.org/pdf/1307.7223v2.pdf.

[9] E. Şaşoğlu and L. Wang, "Universal polarization," Dec. 2013, [Online]. Available: http://arxiv.org/pdf/1307.7495v2.pdf.

[10] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[11] D. Sutter and J. M. Renes, "Universal polar codes for more capable and less noisy channels and sources," Apr. 2014, [Online]. Available: http://arxiv.org/pdf/1312.5990v3.pdf.

[12] E. Arıkan, "Source polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Austin, Texas, Jun. 2010, pp. 899–903.

[13] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[14] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. of the IEEE Inform. Theory Workshop*, Taormina, Sicily, Oct. 2009, pp. 144–148.

[15] L. Wang and E. Şaşoğlu, "Polar coding for interference networks," Jan. 2014, [Online]. Available: http://arxiv.org/pdf/1401.7293v1.pdf.