Polar Codes for Broadcast Channels

Naveen Goela[†], Emmanuel Abbe[♯], and Michael Gastpar[†]

Abstract

Polar codes are introduced for discrete memoryless broadcast channels. For m-user *deterministic* broadcast channels, polarization is applied to map uniformly random message bits from m independent messages to one codeword while satisfying broadcast constraints. The polarization-based codes achieve rates on the boundary of the private-message capacity region. For two-user *noisy* broadcast channels, polar implementations are presented for two information-theoretic schemes: i) Cover's superposition codes; ii) Marton's codes. Due to the structure of polarization, constraints on the auxiliary and channel-input distributions are identified to ensure proper alignment of polarization indices in the multi-user setting. The codes achieve rates on the capacity boundary of a few classes of broadcast channels (e.g., binary-input stochastically degraded). The complexity of encoding and decoding is $O(n \log n)$ where n is the block length. In addition, polar code sequences obtain a stretched-exponential decay of $O(2^{-n^{\beta}})$ of the average block error probability where $0 < \beta < \frac{1}{2}$. Reproducible experiments for finite block lengths $n = 512, 1024, 2048$ corroborate the theory.

Index Terms

Polar Codes, Deterministic Broadcast Channel, Cover's Superposition Codes, Marton's Codes.

I. INTRODUCTION

[†]N. Goela and M. C. Gastpar are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720-1770 USA (e-mail: {ngoela, gastpar}@eecs.berkeley.edu) and also with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland (e-mail: {naveen.goela, michael.gastpar}@epfl.ch).

[♯]E. Abbe is currently with the School of Engineering and Applied Sciences, Princeton University, Princeton, NJ, 08544 USA (e-mail: {eabbe@princeton.edu}).

INTRODUCED by T. M. Cover in 1972, the broadcast problem consists of a single source transmitting m independent private messages to m receivers through a single discrete, memoryless, broadcast channel (DM-BC) [1]. The private-message capacity region is known if the channel structure is *deterministic*, *degraded*, *less-noisy*, or *more-capable* [2]. For general classes of DM-BCs, there exist inner bounds such as Marton's inner bound [3] and outer bounds such as the Nair-El-Gamal outer bound [4]. One difficult aspect of the broadcast problem is to design an encoder which maps m independent messages to a single codeword of symbols which are transmitted simultaneously to all receivers. Several codes relying on *random binning*, *superposition*, and *Marton's strategy* have been analyzed in the literature (see e.g., the overview in [5]).

A. Overview of Contributions

The present paper focuses on low-complexity codes for broadcast channels based on polarization methods. Polar codes were invented originally by Arıkan and were shown to achieve the capacity of binary-input, symmetric, point-to-point channels with $O(n \log n)$ encoding and decoding complexity where n is the code length [6]. In this paper, we obtain the following results.

- Polar codes for deterministic, linear and non-linear, binary-output, m-user DM-BCs (cf. [7]). The capacity-achieving broadcast codes implement low-complexity *random binning*, and are related to polar codes for other multi-user scenarios such as Slepian-Wolf distributed source coding [8], [9], and multiple-access channel (MAC) coding [10]. For deterministic DM-BCs, the polar transform is applied to channel *output* variables. Polarization is useful for shaping uniformly random message bits from m independent messages into non-equiprobable codeword symbols in the presence of hard broadcast constraints. As discussed in Section I-B1 and referenced in [11]–[13], it is difficult to design low-complexity parity-check (LDPC) codes or belief propagation algorithms for the deterministic DM-BC due to multi-user broadcast constraints.

- Polar codes for general two-user DM-BCs based on *superposition coding* [1], [14]. In the multi-user setting, constraints on the auxiliary and channel-input distributions are placed to ensure alignment of polarization indices. The achievable rates lie on the boundary of the capacity region for certain classes of DM-BCs such as binary-input stochastically degraded channels.

- Polar codes for general two-user DM-BCs based on *Marton's coding* strategy. In the multi-user setting, due to the structure of polarization, constraints on the auxiliary and channel-input distributions are identified to ensure alignment of polarization indices. The achievable rates lie on the boundary of the capacity region for certain classes of DM-BCs such as binary-input semi-deterministic channels.

- For the above broadcast polar codes, the asymptotic decay of the average error probability under successive cancelation decoding at the broadcast receivers is established to be $O(2^{-n^\beta})$ where $0 < \beta < \frac{1}{2}$. The error probability is analyzed by averaging over polar code ensembles. In addition, properties such as the chain rule of the Kullback-Leibler divergence between discrete probability measures are exploited.

- Reproducible experiments are provided for finite block lengths $n = 512, 1024, 2048$. The results of the experiments corroborate the theory.

Throughout the paper, for different broadcast coding strategies, a systems-level block diagram of the communication channel and polar transforms is provided.

B. Relation to Prior Work

1) Deterministic Broadcast Channels: The deterministic broadcast channel has received considerable attention in the literature (e.g. due to related extensions such as secure broadcast, broadcasting with side information, and index coding [15], [16]). Several *practical* codes have been designed. For example, the authors of [11] propose sparse linear coset codes to emulate random binning and survey propagation to enforce broadcast channel constraints. In [12], the authors propose enumerative source coding and Luby-Transform codes for deterministic DM-BCs specialized to interference-management scenarios. Additional research includes reinforced belief propagation with non-linear coding [13]. To our knowledge, polarization-based codes provide provable guarantees for achieving rates on the capacity-boundary with practical codes in the general case.

2) Polar Codes for Multi-User Settings: Subsequent to the derivation of channel polarization in [6] and the refined rate of polarization in [17], polarization methods have been extended to analyze multi-user information theory problems. In [10], a joint polarization method is proposed for m-user MACs with connections to matroid theory. Polar codes were extended for several other multi-user settings: arbitrarily-permuted parallel channels [18], degraded relay channels [19], cooperative relaying [20], and wiretap channels [21]–[23]. In addition, several binary multi-user communication scenarios including the Gelfand-Pinsker problem, and Wyner-Ziv problem were analyzed in [24, Chapter 4]. Polar codes for lossless and lossy source compression were investigated respectively in [8] and [25]. In [8], source polarization was extended to the Slepian-Wolf problem involving distributed sources. The approach is based on an "onion-peeling" encoding of sources, whereas a joint encoding is proposed in [26]. In [9], a unified approach is provided for the Slepian-Wolf problem based on generalized monotone chain rules of entropy. To our knowledge, the design of polarization-based broadcast codes is relatively new.
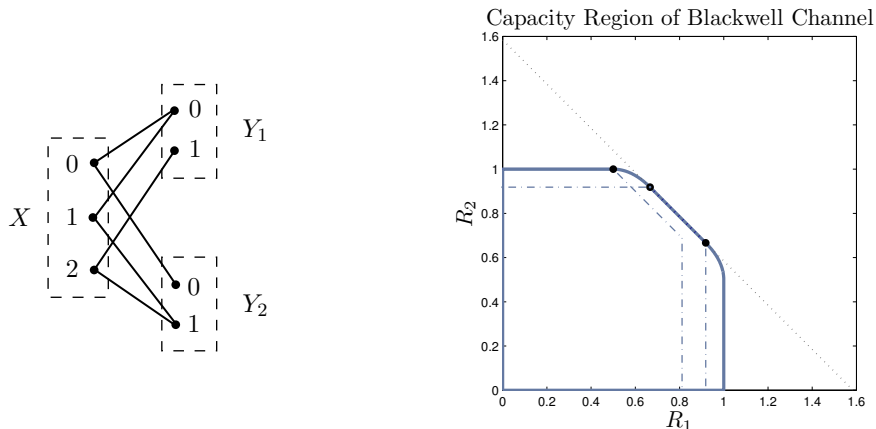
Fig. 1. *Blackwell Channel:* An example of a *deterministic* broadcast channel with $m = 2$ broadcast users. The channel is defined as $Y_1 = f_1(X)$ and $Y_2 = f_2(X)$ where the *non-linear* functions $f_1(x) = \max(x - 1, 0)$ and $f_2(x) = \min(x, 1)$. The private-message capacity region of the Blackwell channel is drawn. For different input distributions $P_X(x)$, the achievable rate points are contained within corresponding polyhedrons in $\mathbb{R}_+^m$.

3) Binary vs. q-ary Polarization: The broadcast codes constructed in the present paper for DM-BCs are based on polarization for binary random variables. However, in extending to arbitrary alphabet sizes, a large body of prior work exists and has focused on generalized constructions and kernels [27], and generalized polarization for q-ary random variables and q-ary channels [28]–[31]. The reader is also referred to the monograph in [32] containing a clear overview of polarization methods.

C. Notation

An index set $\{1, 2, \ldots, m\}$ is abbreviated as $[m]$. An $m \times n$ matrix array of random variables is comprised of variables $Y_i(j)$ where $i \in [m]$ represents the row and $j \in [n]$ the column. The notation $Y_i^{k:\ell} \triangleq \{Y_i(k), Y_i(k + 1), \ldots, Y_i(\ell)\}$ for $k \leq \ell$. When clear by context, the term Y_i^n represents $Y_i^{1:n}$. In addition, the notation for the random variable $Y_i(j)$ is used interchangeably with Y_i^j. The notation $f(n) = O(g(n))$ means that there exists a constant κ such that $f(n) \leq \kappa g(n)$ for sufficiently large n. For a set $\mathcal{S}$, $\mathrm{cl}(\mathcal{S})$ represents set closure, $\mathrm{conv}(\mathcal{S})$ indicates the convex hull operation over set $\mathcal{S}$, and $\mathcal{S}^{\complement}$ denotes the complement of $\mathcal{S}$ with respect to a universal set. Let $h_b(x) = -x \log_2(x) - (1 - x) \log_2(1 - x)$ denote the binary entropy function. Let $a * b \triangleq (1 - a)b + a(1 - b)$.

II. Model

Definition 1 (Discrete, Memoryless Broadcast Channel): The discrete memoryless broadcast channel (DM-BC) with m broadcast receivers consists of a discrete input alphabet $\mathcal{X}$, discrete output alphabets

$\mathcal{Y}_i$ for $i \in [m]$, and a conditional distribution $P_{Y_1,Y_2,\ldots,Y_m|X}(y_1, y_2, \ldots, y_m|x)$ where $x \in \mathcal{X}$ and $y_i \in \mathcal{Y}_i$.

Definition 2 (Private Messages): For a DM-BC with m broadcast receivers, there exist m private messages $\{W_i\}_{i \in [m]}$ such that each message W_i is composed of nR_i bits and $(W_1, W_2, \ldots, W_m)$ is uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}] \times \cdots \times [2^{nR_m}]$.

Definition 3 (Channel Encoding and Decoding): For the DM-BC with independent messages, let the vector of rates $\vec{R} \triangleq \begin{bmatrix} R_1 & R_2 & \ldots & R_m \end{bmatrix}^T$. An $(\vec{R}, n)$ code for the DM-BC consists of one encoder

$$x^n : [2^{nR_1}] \times [2^{nR_2}] \times \cdots \times [2^{nR_m}] \to \mathcal{X}^n,$$

and m decoders specified by $\hat{W}_i : \mathcal{Y}_i^n \to [2^{nR_i}]$ for $i \in [m]$. Based on received observations $\{Y_i(j)\}_{j \in [n]}$, each decoder outputs a decoded message $\hat{W}_i$.

Definition 4 (Average Probability of Error): The average probability of error $P_e^{(n)}$ for a DM-BC code is defined to be the probability that the decoded message at all receivers is not equal to the transmitted message,

$$P_e^{(n)} = \mathbb{P}\left(\bigcup_{i=1}^m \left\{ \hat{W}_i\left(\{Y_i(j)\}_{j \in [n]} \right) \neq W_i \right\} \right).$$

Definition 5 (Private-Message Capacity Region): If there exists a sequence of $(\vec{R}, n)$ codes with $P_e^{(n)} \to 0$, then the rates $\vec{R} \in \mathbb{R}_+^m$ are achievable. The private-message capacity region is the closure of the set of achievable rates.

III. DETERMINISTIC BROADCAST CHANNELS

Definition 6 (Deterministic DM-BC): Define m deterministic functions $f_i(x) : \mathcal{X} \to \mathcal{Y}_i$ for $i \in [m]$. The deterministic DM-BC with m receivers is defined by the following conditional distribution

$$P_{Y_1,Y_2,\ldots,Y_m|X}(y_1, y_2, \ldots, y_m|x) = \prod_{i=1}^m \mathbb{1}\{y_i = f_i(x)\}. \tag{1}$$

A. Capacity Region

Proposition 1 (Marton [33], Pinsker [34]): The capacity region of the deterministic DM-BC includes those rate-tuples $\vec{R} \in \mathbb{R}_+^m$ in the region

$$\mathfrak{C}_{DET-BC} \triangleq \text{conv}\left(\text{cl}\left(\bigcup_{X, \{Y_i\}_{i \in [m]}} \mathfrak{R}\left(X, \{Y_i\}_{i \in [m]} \right) \right) \right), \tag{2}$$

where the polyhedral region $\mathfrak{R}(X, \{Y_i\}_{i \in [m]})$ is given by

$$\mathfrak{R} \triangleq \left\{ \vec{R} \in \mathbb{R}_+^m \,\middle|\, \sum_{i \in \mathcal{S}} R_i < H(\{Y_i\}_{i \in \mathcal{S}}), \ \forall \mathcal{S} \subseteq [m] \right\}. \tag{3}$$

The union in Eqn. (2) is over all random variables $X, Y_1, Y_2, \ldots, Y_m$ with joint distribution induced by $P_X(x)$ and $Y_i = f_i(X)$. The notations $\mathrm{conv}(\cdot)$ and $\mathrm{cl}(\cdot)$ were defined in Section I-C.

Example 1 (Blackwell Channel): In Figure 1, the Blackwell channel is depicted with $\mathcal{X} = \{0, 1, 2\}$ and $\mathcal{Y}_i = \{0, 1\}$. For any fixed distribution $P_X(x)$, it is seen that $P_{Y_1 Y_2}(y_1, y_2)$ has *zero mass* for the pair $(1, 0)$. Let $\alpha \in [\frac{1}{2}, \frac{2}{3}]$. Due to the symmetry of this channel, the capacity region is the union of two regions,

$$\{(R_1, R_2) \in \mathbb{R}_+^2 : \ R_1 \leq h_b(\alpha), R_2 \leq h_b(\frac{\alpha}{2}), R_1 + R_2 \leq h_b(\alpha) + \alpha\},$$

$$\{(R_1, R_2) \in \mathbb{R}_+^2 : \ R_1 \leq h_b(\frac{\alpha}{2}), R_2 \leq h_b(\alpha), R_1 + R_2 \leq h_b(\alpha) + \alpha\},$$

where the first region is achieved with input distribution $P_X(0) = P_X(1) = \frac{\alpha}{2}$, and the second region is achieved with $P_X(1) = P_X(2) = \frac{\alpha}{2}$ [2, Lec. 9]. The sum rate is maximized for a uniform input distribution which yields a pentagonal achievable rate region: $R_1 \leq h_b(\frac{1}{3})$, $R_2 \leq h_b(\frac{1}{3})$, $R_1 + R_2 \leq \log_2 3$. Figure 1 illustrates the capacity region.

B. Main Result

Theorem 1 (Polar Code for Deterministic DM-BC): Consider an m-user deterministic DM-BC with arbitrary discrete input alphabet $\mathcal{X}$, and binary output alphabets $\mathcal{Y}_i \in \{0, 1\}$. Fix input distribution $P_X(x)$ where $x \in \mathcal{X}$ and constant $0 < \beta < \frac{1}{2}$. Let $\pi : [m] \to [m]$ be a permutation on the index set of receivers. Let R_i for $i \in [m]$ be the rate for each receiver. Define rate-vector

$$\vec{R} \triangleq \begin{bmatrix} R_1 & R_2 & \ldots & R_m \end{bmatrix}^T.$$

There exists a sequence of polar broadcast codes over n channel uses which achieves rate-vectors $\vec{R}$ where the rate $R_{\pi(i)}$ is bounded as

$$0 \leq R_{\pi(i)} < H\left(Y_{\pi(i)} | \{Y_{\pi(k)}\}_{k=1:i-1}\right).$$

The average error probability of this code sequence decays as $P_e^{(n)} = O(2^{-n^\beta})$. The complexity of encoding and decoding is $O(n \log n)$.
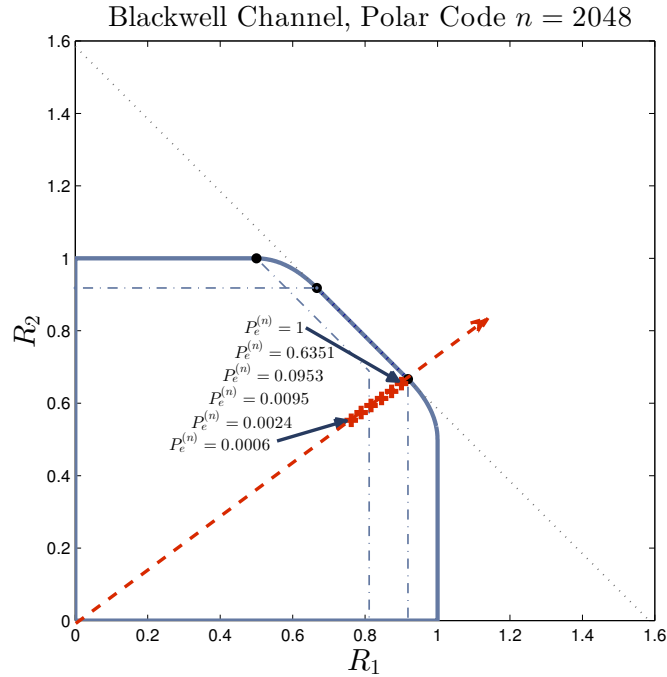
Fig. 2. *Polar Code for Blackwell Channel:* Broadcast code approaching the capacity boundary point of $(R_1, R_2) = (h_b(\frac{2}{3}), \frac{2}{3})$.

Remark 1: To prove the existence of *low-complexity* broadcast codes, a successive randomized protocol is introduced in Section V-A which utilizes $o(n)$ bits of randomness at the encoder. A deterministic encoding protocol is also presented.

Remark 2: The achievable rates for a fixed input distribution $P_X(x)$ are the vertex points of the polyhedral rate region defined in (3). To achieve non-vertex points, the following coding strategies could be applied: time-sharing; rate-splitting for the deterministic DM-BC [35]; polarization by Arıkan utilizing generalized chain rules of entropy [9]. For certain input distributions $P_X(x)$, as illustrated in Figure 1 for the Blackwell channel, a subset of the achievable vertex points lie on the capacity boundary.

Remark 3: Polarization of channels and sources extends to q-ary alphabets (see e.g. [28]). Similarly, it is entirely possible to extend Theorem 1 to include DM-BCs with q-ary output alphabets.

TABLE I

$P_e^{(n)}$ FOR DIFFERENT RATE PAIRS ACHIEVED FOR THE BLACKWELL CHANNEL

(R_1, R_2)	n 512	n 1024	n 2048	n 4096
$(0.73, 0.53)$	0.106	0.0518	0.0195	0.0051
$(0.76, 0.55)$	0.201	0.1356	0.0631	0.0194
$(0.79, 0.57)$	0.3799	0.3177	0.2246	0.1188
$(0.82, 0.59)$	0.5657	0.5606	0.5079	0.4070
$(0.85, 0.61)$	0.7849	0.8181	0.8286	0.8133
$(0.87, 0.63)$	0.9454	0.9757	0.9866	0.9936
$(0.90, 0.65)$	0.9986	1.0000	1.0000	1.0000

C. Experimental Results For The Blackwell Channel

As an experiment for the Blackwell channel described in Example 1, the target rate pair on the capacity boundary is selected to be $(R_1, R_2) = (h_b(\frac{2}{3}), \frac{2}{3})$. Note that $R_1 + R_2 = \log_2 3$ which is the maximum sum rate possible for any input distribution. To achieve the target rate pair, the input distribution $P_X(x)$ is uniform. The output distribution is then $P_{Y_1 Y_2}(0,0) = P_{Y_1 Y_2}(0,1) = P_{Y_1 Y_2}(1,1) = \frac{1}{3}$. For the output distribution, $H(Y_1) = h_b(\frac{2}{3})$ and $H(Y_2|Y_1) = \frac{2}{3}$. According to Theorem 1, these distributions permit polar codes approaching the target boundary rate pair. Figure 2 shows the average probability of error $P_e^{(n)}$ for block length $n = 2048$ with selected rate pairs approaching the capacity boundary. The broadcast code employs a deterministic rule as opposed to a randomized rule at the encoder as described in Section V-A. Table I provides results of experiments for different block lengths for a randomized rule at the encoder. While the randomized rule is important for the proof, the deterministic rule provides better error results in practice. All data points for error probabilities were generated using 10^4 codeword transmissions.

Remark 4 (Zero Error vs. Average Error): A zero-error coding scheme is trivial for rate pairs (R_1, R_2) within the triangle: $(0,0), (0,1), (1,0)$. Beyond the triangular region, it is possible to achieve zero-error throughout the whole capacity region by purging the polar code-book of any codewords causing error at the encoder. However, unless there exists an efficient method to enumerate the code-book, the purging process is not feasible with low-complexity since there exist an exponential number of codewords.

IV. OVERVIEW OF POLARIZATION METHOD

FOR DETERMINISTIC DM-BCS

For the proof of Theorem 1, we utilize binary polarization theorems. In contrast to polar codes for point-to-point channels, our polar codes for deterministic DM-BCs apply the polar transform to the *output* random variables of the channel.

A. Polar Transform

Consider an input distribution $P_X(x)$ to the deterministic DM-BC. Over n channel uses, the input random variables to the channel are given by

$$X^{1:n} = \{X^1, X^2, \ldots, X^n\},$$

where $X^j \sim P_X$ are independent and identically distributed (*i.i.d.*) random variables. The channel output variables are given by $Y_i(j) = f_i(X(j))$ where $f_i(\cdot)$ are the deterministic functions to each broadcast receiver. Denote the random matrix of channel output variables by

$$\mathbf{Y} \triangleq \begin{bmatrix} Y_1^1 & Y_1^2 & Y_1^3 & \ldots & Y_1^n \\ Y_2^1 & Y_2^2 & Y_2^3 & \ldots & Y_2^n \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ Y_m^1 & Y_m^2 & Y_m^3 & \ldots & Y_m^n \end{bmatrix}, \tag{4}$$

where $\mathbf{Y} \in \mathbb{F}_2^{m \times n}$. For $n = 2^\ell$ and $\ell \geq 1$, the polar transform is defined as the following invertible linear transformation,

$$\mathbf{U} = \mathbf{Y}\mathbf{G}_n \tag{5}$$

$$\text{where } \mathbf{G}_n \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes \log_2 n} \mathbf{B}_n.$$

The matrix $\mathbf{G}_n \in \mathbb{F}_2^{n \times n}$ is formed by multiplying a matrix of successive Kronecker matrix-products (denoted by $\otimes$) with a bit-reversal matrix $\mathbf{B}_n$ introduced by Arıkan [8]. The polarized random matrix $\mathbf{U} \in \mathbb{F}_2^{m \times n}$ is indexed as

$$\mathbf{U} \triangleq \begin{bmatrix} U_1^1 & U_1^2 & U_1^3 & \ldots & U_1^n \\ U_2^1 & U_2^2 & U_2^3 & \ldots & U_2^n \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ U_m^1 & U_m^2 & U_m^3 & \ldots & U_m^n \end{bmatrix}. \tag{6}$$

B. Joint Distribution of Polarized Variables

Consider the channel output distribution $P_{Y_1 Y_2 \cdots Y_m}$ of the deterministic DM-BC induced by input distribution $P_X(x)$. The j-th *column* of the random matrix $\mathbf{Y}$ is distributed as $(Y_1^j, Y_2^j, \cdots, Y_m^j) \sim P_{Y_1 Y_2 \cdots Y_m}$. Due to the memoryless property of the channel, the joint distribution of all output variables is

$$P_{Y_1^n Y_2^n \cdots Y_m^n}\left(y_1^n, y_2^n, \cdots, y_m^n\right) = \prod_{j=1}^{n} P_{Y_1 Y_2 \cdots Y_m}\left(y_1^j, y_2^j, \ldots, y_m^j\right). \tag{7}$$

The joint distribution of the matrix variables in $\mathbf{Y}$ is characterized easily due to the $i.i.d.$ structure. The polarized random matrix $\mathbf{U}$ does *not* have an $i.i.d.$ structure. However, one way to define the joint distribution of the variables in $\mathbf{U}$ is via the polar transform equation (5). An alternate representation is via a decomposition into conditional distributions as follows[1].

$$P_{U_1^n U_2^n \cdots U_m^n}\left(u_1^n, u_2^n, \cdots u_m^n\right) = \prod_{i=1}^{m} \prod_{j=1}^{n} P\left(u_i(j) \Big| u_i^{1:j-1}, \{u_k^{1:n}\}_{k \in [1:i-1]}\right). \tag{8}$$

As derived by Arıkan in [8] and summarized in Section IV-E, the conditional probabilities in (8) and associated likelihoods may be computed using a dynamic programming method which "divides-and-conquers" the computations efficiently.

C. Polarization of Conditional Entropies

Proposition 2 (Source Polarization [8]): Following the statement of [8, Theorem 1], consider the pair of random matrices $(\mathbf{Y}, \mathbf{U})$ related through the polar transformation in (5). For $i \in [m]$ and any $\epsilon \in (0,1)$, define the set of indices

$$\mathcal{A}_i^{(n)} \triangleq \left\{ j \in [n] : H\left(U_i(j) \Big| U_i^{1:j-1}, \{Y_k^{1:n}\}_{k \in [1:i-1]}\right) \geq 1 - \epsilon \right\}. \tag{9}$$

Then for $n = 2^\ell$, in the limit as $\ell \to \infty$,

$$\frac{1}{n}\left|\mathcal{A}_i^{(n)}\right| \to H(Y_i | Y_1 Y_2 \cdots Y_{i-1}). \tag{10}$$

For sufficiently large n, Proposition 2 establishes that there exist approximately $nH(Y_i | Y_1 Y_2 \cdots Y_{i-1})$ indices per row $i \in [m]$ of random matrix $\mathbf{U}$ for which the conditional entropy is close to 1. The total number of indices in $\mathbf{U}$ for which the conditional entropy terms polarize to 1 is approximately $nH(Y_1 Y_2 \cdots Y_m)$. The polarization phenomenon is illustrated in Figure 3.

---

[1] The abbreviated notation of the form $P(a|b)$ which appears in (8) indicates $P_{A|B}(a|b)$, i.e. the conditional probability $\mathbb{P}\{A = a | B = b\}$ where A and B are random variables.
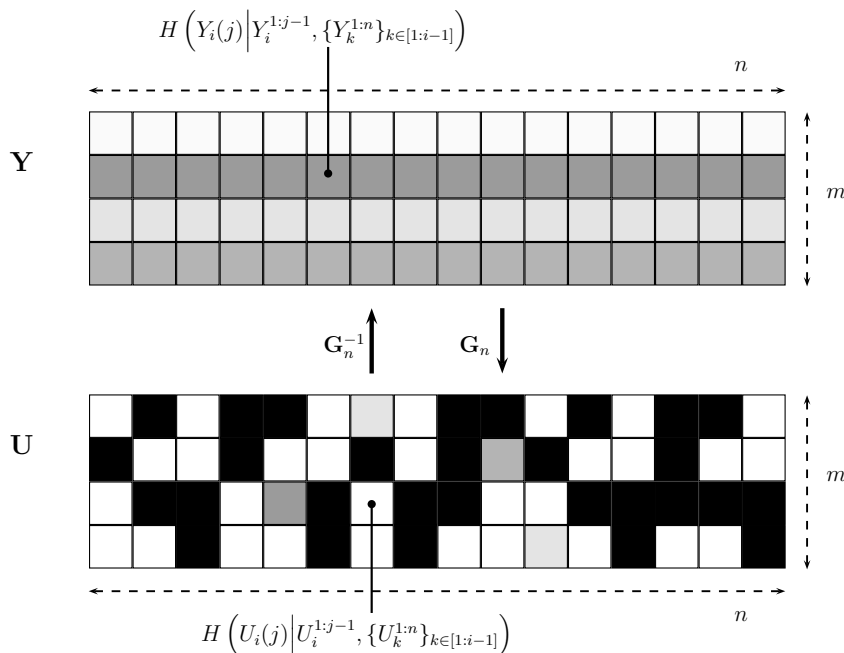
$$H\left(Y_i(j)\Big|Y_i^{1:j-1}, \{Y_k^{1:n}\}_{k\in[1:i-1]}\right)$$

Fig. 3. The polar transform applied to random matrix $\mathbf{Y}$ with $i.i.d.$ structure results in a polarized random matrix $\mathbf{U}$.

Remark 5: Since the polar transform $\mathbf{G}_n$ is invertible, $\{U_k^{1:n}\}_{k\in[1:i-1]}$ are in one-to-one correspondence with $\{Y_k^{1:n}\}_{k\in[1:i-1]}$. Therefore the conditional entropies $H\left(U_i(j)\big|U_i^{1:j-1}, \{U_k^{1:n}\}_{k\in[1:i-1]}\right)$ also polarize to 0 or 1.

D. Rate of Polarization

The Bhattacharyya parameter of random variables is closely related to the conditional entropy. The parameter is useful for characterizing the rate of polarization.

Definition 7 (Bhattacharyya Parameter): Let $(T,V) \sim P_{T,V}$ where $T \in \{0,1\}$ and $V \in \mathcal{V}$ where $\mathcal{V}$ is an arbitrary discrete alphabet. The Bhattacharyya parameter $Z(T|V) \in [0,1]$ is defined

$$Z(T|V) = 2\sum_{v\in\mathcal{V}} P_V(v)\sqrt{P_{T|V}(0|v)P_{T|V}(1|v)}. \tag{11}$$

As shown in Lemma 16 of Appendix A, $Z(T|V) \to 1$ implies $H(T|V) \to 1$, and similarly $Z(T|V) \to 0$ implies $H(T|V) \to 0$ for T a binary random variable. Based on the Bhattacharyya parameter, the following theorem specifies sets $\mathcal{M}_i^{(n)} \subset [n]$ that will be called *message* sets.

Proposition 3 (Rate of Polarization [17]): Consider the pair of random matrices $(\mathbf{Y}, \mathbf{U})$ related through the polar transformation in (5). Fix constants $i \in [m]$, $\tau > 0$, and $0 < \beta < \frac{1}{2}$. Let $\delta_n = 2^{-n^\beta}$ be the rate

of polarization. Define the set

$$\mathcal{M}_i^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_i(j) \Big| U_i^{1:j-1}, \{Y_k^{1:n}\}_{k \in [1:i-1]}\right) \geq 1 - \delta_n \right\}. \tag{12}$$

Then there exists an $N_0 = N_0(\beta, \tau)$ such that

$$\frac{1}{n}\left|\mathcal{M}_i^{(n)}\right| \geq H(Y_i|Y_1 Y_2 \cdots Y_{i-1}) - \tau, \tag{13}$$

for all $n > N_0$ where $n = 2^\ell$ is a power of two.

Proposition 3 is established by defining a super-martingale with respect to the Bhattacharyya parameters and applying the Martingale Convergence Theorem (see e.g., [8, Theorem 1 and 2] for sources and [6] for channel polarization). The rate of polarization was characterized by Arıkan and Telatar for channel polarization in [17, Theorem 1]. A similar statement for asymmetric channels and sources may be found in [36, Theorem 1].

Remark 6: The message sets $\mathcal{M}_i^{(n)}$ are computed "offline" only once during a code construction phase. The sets do not depend on the realization of random variables. In the following Section IV-E, a Monte Carlo sampling approach for estimating Bhattacharyya parameters is reviewed. Other highly efficient algorithms are known in the literature for finding the message indices (see e.g. Tal and Vardy [37]).

E. Estimating Bhattacharyya Parameters

As shown in Lemma 11 in Appendix A, one way to estimate the Bhattacharyya parameter $Z(T|V)$ is to sample from the distribution $P_{T,V}(t,v)$ and evaluate $\mathbb{E}_{T,V} \sqrt{\varphi(T,V)}$. The function $\varphi(t,v)$ is defined based on likelihood ratios

$$L(v) \triangleq \frac{P_{T|V}(0|v)}{P_{T|V}(1|v)},$$

$$L^{-1}(v) \triangleq \frac{P_{T|V}(1|v)}{P_{T|V}(0|v)}.$$

Similarly, to determine the indices in the message sets $\mathcal{M}_i^{(n)}$ defined in Proposition 3, the Bhattacharyya parameters $Z\left(U_i(j) \Big| U_i^{1:j-1}, \{Y_k^{1:n}\}_{k \in [i-1]}\right)$ must be estimated efficiently. For $n \geq 2$, define the likelihood ratio

$$L_n^{(i,j)}\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k \in [1:i-1]}\right) \triangleq \frac{\mathbb{P}\left(U_i(j) = 0 \Big| U_i^{1:j-1} = u_i^{1:j-1}, \{Y_k^{1:n} = y_k^{1:n}\}_{k \in [1:i-1]}\right)}{\mathbb{P}\left(U_i(j) = 1 \Big| U_i^{1:j-1} = u_i^{1:j-1}, \{Y_k^{1:n} = y_k^{1:n}\}_{k \in [1:i-1]}\right)}. \tag{14}$$
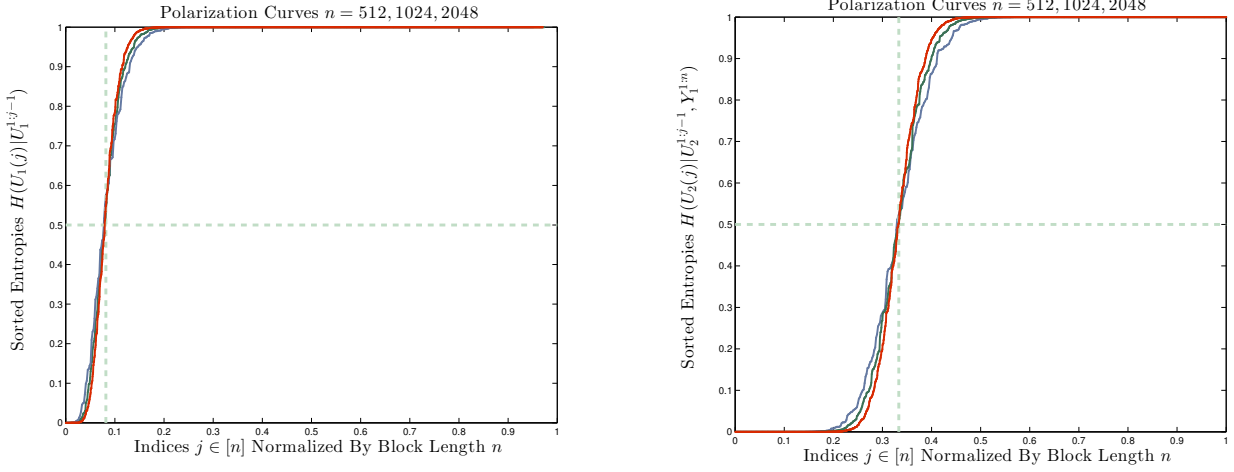
Fig. 4.   *Polarization Curves:* (a) Bernoulli source polarization; (b) Polarization of conditional entropies.

The dynamic programming method given in [8] allows for a recursive computation of the likelihood ratio. Define the following sub-problems

$$\Xi_1 = L_{\frac{n}{2}}^{(i,j)} \left(u_{i,o}^{1:2j-2} \oplus u_{i,e}^{1:2j-2}, \{y_k^{1:\frac{n}{2}}\}_{k \in [1:i-1]} \right),$$

$$\Xi_2 = L_{\frac{n}{2}}^{(i,j)} \left(u_{i,e}^{1:2j-2}, \{y_k^{\frac{n}{2}+1:n}\}_{k \in [1:i-1]} \right),$$

where the notation $u_{i,o}^{1:2j-2}$ and $u_{i,e}^{1:2j-2}$ represents the odd and even indices respectively of the sequence $u_i^{1:2j-2}$. The recursive computation of the likelihoods is characterized by

$$L_n^{(i,2j-1)} \left(u_i^{1:2j-2}, \{y_k^{1:n}\}_{k \in [1:i-1]} \right) = \frac{\Xi_1 \Xi_2 + 1}{\Xi_1 + \Xi_2}.$$

$$L_n^{(i,2j)} \left(u_i^{1:2j-1}, \{y_k^{1:n}\}_{k \in [1:i-1]} \right) = (\Xi_1)^\gamma \Xi_2,$$

where $\gamma = 1$ if $u_i(2j-1) = 0$ and $\gamma = -1$ if $u_i(2j-1) = 1$. In the above recursive computations, the base case is for sequences of length $n = 2$.

The dynamic programming technique may be modified to estimate target probabilities, Bhattacharyya parameters, and also conditional entropies directly. Figure 4 shows the results of polarizing a joint distribution $P_{Y_1 Y_2}(y_1, y_2)$ when $P_{Y_1 Y_2}(0,0) = P_{Y_1 Y_2}(0,1) = P_{Y_1 Y_2}(1,1) = \frac{1}{3}$. In the plot to the left, a single-source polarization result is shown for an *i.i.d.* Bernoulli source $P_{Y_1}(0) = \frac{2}{3}$. In the plot to the right, a conditional polarization result is given for $P_{Y_1 Y_2}(y_1, y_2)$. The block lengths are $n = 512, 1024, 2048$.

V. PROOF OF THEOREM 1

The proof of Theorem 1 is based on binary polarization theorems as discussed in Section IV. The random coding arguments of C. E. Shannon prove the existence of capacity-achieving codes for point-to-point channels. Furthermore, random binning and joint-typicality arguments suffice to prove the existence of capacity-achieving codes for the deterministic DM-BC. However, it is shown in this section that there exist capacity-achieving *polar codes* for the binary-output deterministic DM-BC.

A. Broadcast Code Based on Polarization

The ordering of the receivers' rates in $\vec{R}$ is arbitrary due to symmetry. Therefore, let $\pi(i) = i$ be the identity permutation which denotes the successive order in which the message bits are allocated for each receiver. The encoder must map m independent messages $(W_1, W_2, \ldots, W_m)$ uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}] \times \cdots \times [2^{nR_m}]$ to a codeword $x^n \in \mathcal{X}^n$. To construct a codeword for broadcasting m independent messages, the following binary sequences are formed at the encoder: $u_1^{1:n}, u_2^{1:n}, \ldots, u_m^{1:n}$. To determine a particular bit $u_i(j)$ in the binary sequence $u_i^{1:n}$, if $j \in \mathcal{M}_i^{(n)}$, the bit is selected as a uniformly distributed message bit intended for receiver $i \in [m]$. As defined in (12) of Proposition 3, the message set $\mathcal{M}_i^{(n)}$ represents those indices for bits transmitted to receiver i. The remaining *non-message* indices in the binary sequence $u_i^{1:n}$ for each user $i \in [m]$ are computed either according to a deterministic or random mapping.

1) Deterministic Mapping: Consider a class of deterministic boolean functions indexed by $i \in [m]$ and $j \in [n]$:

$$\psi^{(i,j)} : \{0,1\}^{n(\max\{0,i-1\})+j-1} \to \{0,1\}. \tag{15}$$

As an example, consider the deterministic boolean function based on the *maximum a posteriori* polar coding rule.

$$\psi_{MAP}^{(i,j)}\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]}\right) \triangleq \arg\max_{u\in\{0,1\}}\left\{ \mathbb{P}\left(U_i(j) = u \middle| U_i^{1:j-1} = u_i^{1:j-1}, \{Y_k^{1:n} = y_k^{1:n}\}_{k\in[1:i-1]}\right) \right\}. \tag{16}$$

2) Random Mapping: Consider a class of random boolean functions indexed by $i \in [m]$ and $j \in [n]$:

$$\Psi^{(i,j)} : \{0,1\}^{n(\max\{0,i-1\})+j-1} \to \{0,1\}. \tag{17}$$

As an example, consider the random boolean function

$$\Psi_{RAND}^{(i,j)}\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]}\right) \triangleq \begin{cases} 0, & \text{w. p. } \lambda_0\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]}\right), \\ 1, & \text{w. p. } 1 - \lambda_0\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]}\right), \end{cases} \tag{18}$$

where

$$\lambda_0\left(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]}\right) \triangleq \mathbb{P}\left(U_i(j) = 0 \middle| U_i^{1:j-1} = u_i^{1:j-1}, \{Y_k^{1:n} = y_k^{1:n}\}_{k\in[1:i-1]}\right).$$

For a fixed i and j, the random boolean function $\Psi_{RAND}^{(i,j)}$ may be thought of as a vector of independent Bernoulli random variables. Each Bernoulli random variable of the vector has a fixed probability of being one or zero that is well-defined.

3) Mapping From Messages To Codeword: The binary sequences $u_i^{1:n}$ for $i \in [m]$ are formed *successively* bit by bit. If $j \in \mathcal{M}_i^{(n)}$, then the bit $u_i(j)$ is one message bit from the uniformly distributed message W_i intended for user i. If $j \notin \mathcal{M}_i^{(n)}$, $u_i(j) = \psi_{MAP}^{(i,j)}(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]})$ in the case of a deterministic mapping, or $u_i(j) = \Psi_{RAND}^{(i,j)}(u_i^{1:j-1}, \{y_k^{1:n}\}_{k\in[1:i-1]})$ in the case of a random mapping. The encoder then applies the *inverse* polar transform for each sequence: $y_i^{1:n} = u_i^{1:n}\mathbf{G}_n^{-1}$. The codeword x^n is formed symbol-by-symbol as follows:

$$x(j) \in \bigcap_{i=1}^{m} f_i^{-1}\left(y_i(j)\right).$$

If the intersection set is empty, the encoder declares a block error. A block error only occurs at the encoder.

4) Decoding at Receivers: If the encoder succeeds in transmitting a codeword x^n, each receiver obtains the sequence $y_i^{1:n}$ noiselessly and applies the polar transform $\mathbf{G}_n$ to recover $u_i^{1:n}$ exactly. Since the message indices $\mathcal{M}_i^{(n)}$ are known to each receiver, the message bits in $u_i^{1:n}$ are decoded correctly by receiver i.

B. Total Variation Bound

While the deterministic mapping $\psi_{MAP}^{(i,j)}$ performs well in practice, the average probability of error $P_e^{(n)}$ of the coding scheme is more difficult to analyze in theory. The random mapping $\Psi_{RAND}^{(i,j)}$ at the encoder is more amenable to analysis via the probabilistic method. Towards that goal, consider the following

probability measure defined on the space of tuples of binary sequences[2].

$$Q\Big(u_1^n, u_2^n, \cdots, u_m^n\Big) \triangleq \prod_{i=1}^{m} \prod_{j=1}^{n} Q\Big(u_i(j)\Big| u_i^{1:j-1}, \{u_k^{1:n}\}_{k\in[1:i-1]}\Big). \tag{19}$$

where the conditional probability measure

$$Q\left(u_i(j)\Big| u_i^{1:j-1}, \{u_k^{1:n}\}_{k\in[1:i-1]}\right) \triangleq \begin{cases} \frac{1}{2}, & \text{if } j \in \mathcal{M}_i^{(n)}, \\ P\left(u_i(j)\Big| u_i^{1:j-1}, \{u_k^{1:n}\}_{k\in[1:i-1]}\right), & \text{otherwise.} \end{cases}$$

The probability measure Q defined in (19) is a perturbation of the joint probability measure P defined in (8) for the random variables $U_i(j)$. The only difference in definition between P and Q is due to those indices in message set $\mathcal{M}_i^{(n)}$. The following lemma provides a bound on the total variation distance between P and Q.

Lemma 1: (*Total Variation Bound*) Let probability measures P and Q be defined as in (8) and (19) respectively. Let $0 < \beta < 1$. The total variation distance between P and Q is bounded as

$$\sum_{\{u_k^{1:n}\}_{k\in[m]}} \Big| P(\{u_k^{1:n}\}_{k\in[m]}) - Q(\{u_k^{1:n}\}_{k\in[m]}) \Big| = O(2^{-n^\beta}).$$

Proof: See Section B of the Appendices. ∎

C. Analysis of the Average Probability of Error

For the m-user deterministic DM-BC, an error event occurs at the encoder if a codeword x^n is unable to be constructed symbol by symbol according to the broadcast protocol described in Section V-A. Define the following set consisting of m-tuples of binary sequences,

$$\mathcal{T} \triangleq \Big\{ (y_1^n, y_2^n, \ldots, y_m^n) : \exists j \in [n], \bigcap_{i=1}^{m} f_i^{-1}(y_i(j)) = \emptyset \Big\}. \tag{20}$$

The set $\mathcal{T}$ consists of those m-tuples of binary output sequences which are *inconsistent* due to the properties of the deterministic channel. In addition, due to the one-to-one correspondence between sequences $u_i^{1:n}$ and $y_i^{1:n}$, denote by $\tilde{\mathcal{T}}$ the set of m-tuples $(u_1^n, u_2^n, \ldots, u_m^n)$ that are inconsistent.

---

[2]A related proof technique was provided for lossy source coding based on polarization in a different context [25]. In the present paper, a different proof is supplied that utilizes the chain rule for KL-divergence. In addition, deterministic and random mappings were used by Honda *et al.* in [36] to extend Arıkan's results on polar code ensembles to asymmetric point-to-point channels.

For the broadcast protocol, the rate $R_i = \frac{1}{n}\big|\mathcal{M}_i^{(n)}\big|$ for each receiver. Let the total sum rate for all broadcast receivers be $R_\Sigma = \sum_{i \in [m]} R_i$. If the encoder uses a fixed deterministic map $\psi^{(i,j)}$ in the broadcast protocol, the average probability of error is

$$
P_e^{(n)}\Big[\{\psi^{(i,j)}\}\Big] = \frac{1}{2^{nR_\Sigma}} \sum_{\{u_k^{1:n}\}_{k \in [m]}} \bigg[\mathbb{1}\Big\{(u_1^n, u_2^n, \ldots, u_m^n) \in \tilde{\mathcal{T}}\Big\}
$$

$$
\cdot \prod_{\substack{i \in [m] \\ j \in [n]: j \notin \mathcal{M}_i^{(n)}}} \mathbb{1}\Big\{\psi^{(i,j)}\Big(u_i^{1:j-1}, \{y_k^{1:n}\}_{k \in [1:i-1]}\Big) = u_i(j)\Big\}\bigg]. \tag{21}
$$

In addition, if the random maps $\Psi^{(i,j)}$ are used at the encoder, the average probability of error is a random quantity given by

$$
P_e^{(n)}\Big[\{\Psi^{(i,j)}\}\Big] = \frac{1}{2^{nR_\Sigma}} \sum_{\{u_k^{1:n}\}_{k \in [m]}} \bigg[\mathbb{1}\Big\{(u_1^n, u_2^n, \ldots, u_m^n) \in \tilde{\mathcal{T}}\Big\}
$$

$$
\cdot \prod_{\substack{i \in [m] \\ j \in [n]: j \notin \mathcal{M}_i^{(n)}}} \mathbb{1}\Big\{\Psi^{(i,j)}\Big(u_i^{1:j-1}, \{y_k^{1:n}\}_{k \in [1:i-1]}\Big) = u_i(j)\Big\}\bigg]. \tag{22}
$$

Instead of characterizing $P_e^{(n)}$ directly for deterministic maps, the analysis of $P_e^{(n)}[\{\Psi^{(i,j)}\}]$ leads to the following lemma.

Lemma 2: Consider the broadcast protocol of Section V-A. Let $R_i = \frac{1}{n}\big|\mathcal{M}_i^{(n)}\big|$ for $i \in [m]$ be the broadcast rates selected according to the criterion given in (12) in Proposition 3. Then for $0 < \beta < 1$,

$$
\mathbb{E}_{\{\Psi^{(i,j)}\}}\Big[P_e^{(n)}[\{\Psi^{(i,j)}\}]\Big] = O(2^{-n^\beta}).
$$

Proof:

$$
\mathbb{E}_{\{\Psi^{(i,j)}\}}\Big[P_e^{(n)}[\{\Psi^{(i,j)}\}]\Big]
$$

$$
= \frac{1}{2^{nR_\Sigma}} \sum_{\{u_k^{1:n}\}_{k \in [m]}} \bigg[\mathbb{1}\Big\{(u_1^n, u_2^n, \ldots, u_m^n) \in \tilde{\mathcal{T}}\Big\} \cdot
$$

$$
\prod_{\substack{i \in [m] \\ j \in [n]: j \notin \mathcal{M}_i^{(n)}}} \mathbb{P}\Big\{\Psi^{(i,j)}\Big(u_i^{1:j-1}, \{y_k^{1:n}\}_{k \in [1:i-1]}\Big) = u_i(j)\Big\}\bigg]
$$

$$
= \sum_{\{u_k^{1:n}\}_{k \in [m]} \in \tilde{\mathcal{T}}} Q\big(\{u_k^{1:n}\}_{k \in [m]}\big) \tag{23}
$$

$$
= \sum_{\{u_k^{1:n}\}_{k \in [m]} \in \tilde{\mathcal{T}}} \Big|P\big(\{u_k^{1:n}\}_{k \in [m]}\big) - Q\big(\{u_k^{1:n}\}_{k \in [m]}\big)\Big| \tag{24}
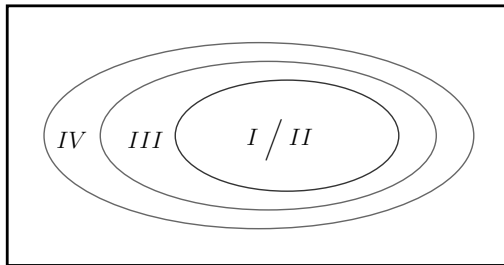$$

$$
= O(2^{-n^\beta}). \tag{25}
$$

Fig. 5. The special classes of noisy broadcast channels as described in Section VI-A. Class I represents stochastically degraded DM-BCs. Class II represents broadcast channels for which $V - X - (Y_1, Y_2)$ and $P_{Y_2|V}(y_2|v) \succ P_{Y_1|V}(y_1|v)$ for all $P_{X|V}(x|v)$. Class II is equivalent to Class I. Class III represents less-noisy DM-BCs. Class IV represents broadcast channels with the more capable property.

Step (23) follows since the probability measure Q matches the desired calculation exactly. Step (24) is due to the fact that the probability measure P has *zero mass* over m-tuples of binary sequences that are inconsistent. Step (25) follows directly from Lemma 1. Lastly, since the expectation over random maps $\{\Psi^{(i,j)}\}$ of the average probability of error decays stretched-exponentially, there must exist a set of deterministic maps which exhibit the same behavior. ■

VI. NOISY BROADCAST CHANNELS
SUPERPOSITION CODING

Coding for noisy broadcast channels is now considered using polarization methods. By contrast to the deterministic case, a decoding error event occurs at the receivers on account of the randomness due to noise. For the remaining sections, it is assumed that there exist $m = 2$ users in the DM-BC. The private-message capacity region for the DM-BC is unknown even for binary input, binary output two-user channels such as the skew-symmetric DM-BC. However, the private-message capacity region is known for specific classes.

A. Special Classes of Noisy DM-BCs

Definition 8: The two-user *physically degraded* DM-BC is a channel $P_{Y_1 Y_2|X}(y_1, y_2|x)$ for which $X - Y_1 - Y_2$ forms a Markov chain, i.e. one of the receivers is statistically stronger than the other:

$$P_{Y_1 Y_2|X}(y_1, y_2|x) = P_{Y_1|X}(y_1|x) P_{Y_2|Y_1}(y_2|y_1). \tag{26}$$

Definition 9: A two-user DM-BC $P_{Y_1Y_2|X}(y_1, y_2|x)$ is *stochastically degraded* if its conditional marginal distributions are the same as that of a physically degraded DM-BC, i.e., if there exists a distribution $\tilde{P}_{Y_2|Y_1}(y_2|y_1)$ such that

$$P_{Y_2|X}(y_2|x) = \sum_{y_1 \in \mathcal{Y}_1} P_{Y_1|X}(y_1|x)\tilde{P}_{Y_2|Y_1}(y_2|y_1). \tag{27}$$

If (27) holds for two conditional distributions $P_{Y_1|X}(y_1|x)$ and $P_{Y_2|X}(y_2|x)$ defined over the same input, then the property is denoted as follows: $P_{Y_1|X}(y_1|x) \succeq P_{Y_2|X}(y_2|x)$.

Definition 10: A two-user DM-BC $P_{Y_1Y_2|X}(y_1, y_2|x)$ for which $V - X - (Y_1, Y_2)$ forms a Markov chain is said to be *less noisy* if

$$\forall P_{VX}(v, x) : I(V; Y_1) \geq I(V; Y_2). \tag{28}$$

Definition 11: A two-user DM-BC $P_{Y_1Y_2|X}(y_1, y_2|x)$ is said to be *more capable* if

$$\forall P_X(x) : I(X; Y_1) \geq I(X; Y_2). \tag{29}$$

The following lemma relates the properties of the special classes of noisy broadcast channels. A more comprehensive treatment of special classes is given by C. Nair in [38].

Lemma 3: Consider a two-user DM-BC $P_{Y_1Y_2|X}(y_1, y_2|x)$. Let $V - X - (Y_1, Y_2)$ form a Markov chain, $|\mathcal{V}| > 1$, and $P_V(v) > 0$. The following implications hold:

$$X - Y_1 - Y_2$$

$$\Rightarrow P_{Y_1|X}(y_1|x) \succeq P_{Y_2|X}(y_2|x) \tag{30}$$

$$\Leftrightarrow \forall P_{X|V}(x|v) : P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v) \tag{31}$$

$$\Rightarrow \forall P_{VX}(v, x) : I(V; Y_1) \geq I(V; Y_2) \tag{32}$$

$$\Rightarrow \forall P_X(x) : I(X; Y_1) \geq I(X; Y_2). \tag{33}$$

The converse statements for (30), (32), and (33) do *not* hold in general. Figure 5 illustrates the different types of broadcast channels as a hierarchy.

Proof: See Section E of the Appendices. ∎

B. Superposition Region

Superposition coding involves one auxiliary random variable V which conveys a "cloud center" or a coarse message decoded by both receivers [1], [14]. One of the receivers then decodes an additional "satellite codeword" conveyed through X containing a fine-grain message that is superimposed upon
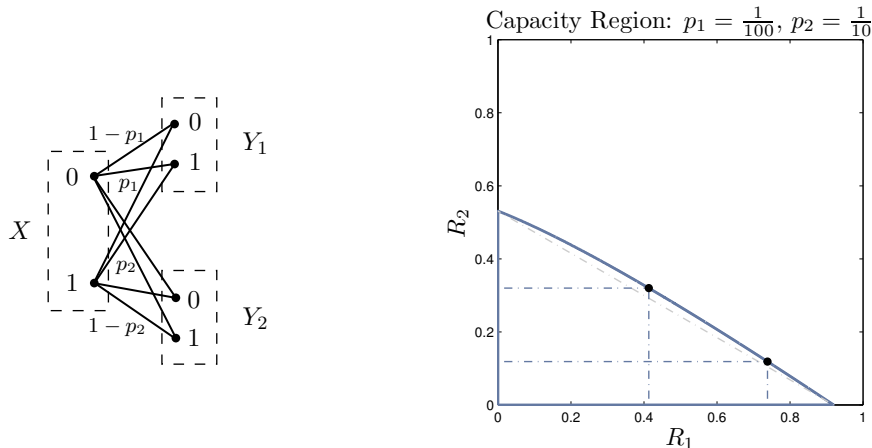
Fig. 6. *DM-BC with BSCs:* The classic two-user broadcast channel consisting of a $\mathrm{BSC}(p_1 = \frac{1}{100})$ and a $\mathrm{BSC}(p_2 = \frac{1}{10})$. The private-message capacity region is equivalent to the superposition coding inner bound. For a fixed auxiliary and input distribution $P_{VX}(v, x)$, the superposition inner bound is plotted as a rectangle in $\mathbb{R}^2_+$ for $\alpha = \frac{1}{10}$ and $\alpha = \frac{1}{4}$ as described in Example 2. For this example, polar codes achieve all points on the capacity boundary.

the coarse information. Recent research has also uncovered distinctions between superposition coding schemes [39]. The following superposition region is standard as given in modern textbooks [2].

Proposition 4 (Superposition Inner Bound [1], [14]): For any two-user DM-BC, the rates $(R_1, R_2) \in \mathbb{R}^2_+$ are achievable in the region $\mathfrak{R}(X, V, Y_1, Y_2)$ where

$$\mathfrak{R}(X, V, Y_1, Y_2) \triangleq \Big\{ (R_1, R_2) \in \mathbb{R}^2_+ \ \Big| \ R_1 \leq I(X; Y_1 | V),$$

$$R_2 \leq I(V; Y_2),$$

$$R_1 + R_2 \leq I(X; Y_1) \Big\}. \tag{34}$$

and where random variables X, V, Y_1, Y_2 obey the Markov chain $V - X - (Y_1, Y_2)$.

Remark 7: The superposition inner bound is applicable for *any* broadcast channel. By symmetry, the following rate region is also achievable: $\{R_1, R_2 \mid R_2 \leq I(X; Y_2 | V), R_1 \leq I(V; Y_1), R_1 + R_2 \leq I(X; Y_2)\}$ for random variables obeying the Markov chain $V - X - (Y_1, Y_2)$.

Remark 8: The inner bound is the capacity region for degraded, less-noisy, and more-capable DM-BCs (i.e. Class I through Class IV as shown in Figure 5). For the degraded and less-noisy special classes, the capacity region is simplified to $\{R_1, R_2 \mid R_1 \leq I(X; Y_1 | V), R_2 \leq I(V; Y_2)\}$. To see this, note that $I(V; Y_2) \leq I(V; Y_1)$ which implies $I(V; Y_2) + I(X; Y_1 | V) \leq I(V; Y_1) + I(X; Y_1 | V) = I(X; Y_1)$. Therefore the sum-rate constraint $R_1 + R_2 \leq I(X; Y_1)$ of the rate-region in (34) is automatically satisfied.

Example 2 (Binary Symmetric DM-BC): The two-user binary symmetric DM-BC consists of a binary

symmetric channel with flip probability p_1 denoted as $\text{BSC}(p_1)$ and a second channel $\text{BSC}(p_2)$. Assume that $p_1 < p_2 < \frac{1}{2}$ which implies stochastic degradation as defined in (27). For $\alpha \in [0, \frac{1}{2}]$, the superposition inner bound is the region,

$$\Big\{(R_1, R_2) \in \mathbb{R}_+^2 \ \Big| \ R_1 \le h_b(\alpha * p_1) - h_b(p_1),$$
$$R_2 \le 1 - h_b(\alpha * p_2)\Big\} \tag{35}$$

The above inner bound is determined by evaluating (34) where V is a Bernoulli random variable with $P_V(v) = \frac{1}{2}$, $X = V \oplus S$, and S is a Bernoulli random variable with $P_S(1) = \alpha$. Figure 6 plots this rectangular inner bound for two different values $\alpha = \frac{1}{10}$ and $\alpha = \frac{1}{4}$. The corner points of this rectangle given in (35) lie on the capacity boundary.

Example 3 (DM-BC with BEC(ϵ) and BSC(p) [38]): Consider a two-user DM-BC comprised of a $\text{BSC}(p)$ from X to Y_1 and a $\text{BEC}(\epsilon)$ from X to Y_2. Then it can be shown that the following cases hold:

- $0 < \epsilon \le 2p$: Y_1 is degraded with respect to Y_2.
- $2p < \epsilon \le 4p(1 - p)$: Y_2 is less noisy than Y_1 but Y_1 is not degraded with respect to Y_2.
- $4p(1 - p) < \epsilon \le h_b(p)$: Y_2 is more capable than Y_1 but not less noisy.
- $h_b(p) < \epsilon < 1$: The channel does not belong to the special classes.

The capacity region for all channel parameters for this example is achieved using superposition coding.

C. Main Result

Theorem 2 (Polarization-Based Superposition Code): Consider any two-user DM-BC with binary input alphabet $\mathcal{X} = \{0, 1\}$ and arbitrary output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$. There exists a sequence of polar broadcast codes over n channel uses which achieves the following rate region

$$\mathfrak{R}(V, X, Y_1, Y_2) \triangleq \Big\{(R_1, R_2) \in \mathbb{R}_+^2 \ \Big| \ R_1 \le I(X; Y_1|V),$$
$$R_2 \le I(V; Y_2)\Big\}, \tag{36}$$

where random variables V, X, Y_1, Y_2 have the following listed properties:

- V is a binary random variable.
- $P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v)$.
- $V - X - (Y_1, Y_2)$ forms a Markov chain.

For $0 < \beta < \frac{1}{2}$, the average error probability of this code sequence decays as $P_e^{(n)} = O(2^{-n^\beta})$. The complexity of encoding and decoding is $O(n \log n)$.
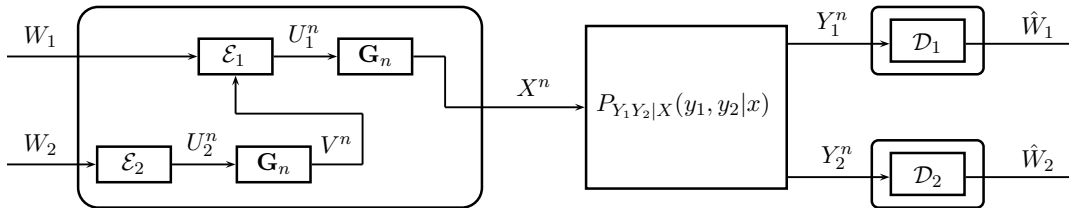
Fig. 7. Block diagram of a polarization-based superposition code for a two-user noisy broadcast channel.

Remark 9: The requirement that auxiliary V is a binary random variable is due to the use of binary polarization theorems in the proof. Indeed, the auxiliary V may need to have a larger alphabet in the case of broadcast channels. An extension to q-ary random variables is entirely possible if q-ary polarization theorems are utilized.

Remark 10: The requirement that $V - X - (Y_1, Y_2)$ holds is standard for superposition coding over noisy channels. However, the listed property $P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v)$ is due to the structure of polarization and is used in the proof to guarantee that polarization indices are *aligned*. If both receivers are able to decode the coarse message carried by the auxiliary random variable V, the polarization indices for the coarse message must be nested for the two receivers' channels.

VII. PROOF OF THEOREM 2

The block diagram for polarization-based superposition coding is given in Figure 7. Similar to random codes in Shannon theory, polarization-based codes rely on n-length *i.i.d.* statistics of random variables; however, a specific polarization structure based on the chain rule of entropy allows for efficient encoding and decoding. The key idea of Cover and Bergmans is to superimpose two messages of information onto one codeword [1], [14].

A. Polar Transform

Consider the *i.i.d.* sequence of random variables $(V^j, X^j, Y_1^j, Y_2^j) \sim P_V(v) P_{X|V}(x|v) P_{Y_1 Y_2|X}(y_1, y_2|x)$ where the index $j \in [n]$. Let the n-length sequence of auxiliary and input variables (V^j, X^j) be organized into the random matrix

$$\mathbf{\Omega} \triangleq \begin{bmatrix} X^1 & X^2 & X^3 & \dots & X^n \\ V^1 & V^2 & V^3 & \dots & V^n \end{bmatrix}. \tag{37}$$

Applying the polar transform to $\mathbf{\Omega}$ results in the random matrix $\mathbf{U} \triangleq \mathbf{\Omega G}_n$. Let the random variables in the random matrix $\mathbf{U}$ be indexed as follows:

$$\mathbf{U} = \left[\begin{array}{ccccc} U_1^1 & U_1^2 & U_1^3 & \cdots & U_1^n \\ U_2^1 & U_2^2 & U_2^3 & \cdots & U_2^n \end{array} \right]. \tag{38}$$

The above definitions are consistent with the block diagram given in Figure 7 (and noting that $\mathbf{G}_n = \mathbf{G}_n^{-1}$). The polar transform extracts the randomness of $\mathbf{\Omega}$. In the transformed domain, the joint distribution of the random variables in $\mathbf{U}$ is given by

$$P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) \triangleq P_{X^n V^n}\left(u_1^n \mathbf{G}_n, u_2^n \mathbf{G}_n\right). \tag{39}$$

For polar coding purposes, the joint distribution is decomposed as follows,

$$P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) = P_{U_2^n}(u_2^n) P_{U_1^n | U_2^n}\left(u_1^n | u_2^n\right) = \prod_{j=1}^n P\left(u_2(j) \big| u_2^{1:j-1}\right) P\left(u_1(j) \big| u_1^{1:j-1}, u_2^n\right). \tag{40}$$

The conditional distributions may be computed efficiently using recursive protocols as already mentioned. The polarized variables in $\mathbf{U}$ are *not necessarily i.i.d.* random variables.

B. Polarization Theorems Revisited

Definition 12 (Polarization Sets for Superposition Coding): Let V^n, X^n, Y_1^n, Y_2^n be the sequence of random variables as introduced in Section VII-A. In addition, let $U_1^n = X^n \mathbf{G}_n$ and $U_2^n = V^n \mathbf{G}_n$. Let $\delta_n = 2^{-n^\beta}$ for $0 < \beta < \frac{1}{2}$. The following *polarization sets* are defined:

$$\mathcal{H}_{X|V}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_1(j) \big| U_1^{1:j-1}, V^n\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{H}_{X|VY_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_1(j) \big| U_1^{1:j-1}, V^n, Y_1^n\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{L}_{X|VY_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_1(j) \big| U_1^{1:j-1}, V^n, Y_1^n\right) \leq \delta_n \right\},$$

$$\mathcal{L}_{V|Y_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \big| U_2^{1:j-1}, Y_1^n\right) \leq \delta_n \right\},$$

$$\mathcal{H}_V^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \big| U_2^{1:j-1}\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{L}_{V|Y_2}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \big| U_2^{1:j-1}, Y_2^n\right) \leq \delta_n \right\}.$$

Remark 11: The polarization sets denoted by calligraphic letters $\mathcal{H}$ indicate "high-entropy" sets, whereas sets denoted by $\mathcal{L}$ indicate "low-entropy" sets. For example, $\mathcal{H}_{X|V}^{(n)}$ denotes the "high-entropy" set of indices corresponding to a conditional polarization of the variables $X^{1:n}$ (polarized to $U_1^{1:n}$) given all variables $V^{1:n}$. Similarly, $\mathcal{L}_{V|Y_2}^{(n)}$ denotes the "low-entropy" set of indices corresponding to a conditional polarization of the variables $V^{1:n}$ (polarized to $U_2^{1:n}$) given all variables $Y_2^{1:n}$.

Definition 13 (Message Sets for Superposition Coding): In terms of the polarization sets given in Definition 12, the following *message sets* are defined:

$$\mathcal{M}_{1v}^{(n)} \triangleq \mathcal{H}_V^{(n)} \cap \mathcal{L}_{V|Y_1}^{(n)}, \tag{41}$$

$$\mathcal{M}_1^{(n)} \triangleq \mathcal{H}_{X|V}^{(n)} \cap \mathcal{L}_{X|VY_1}^{(n)}, \tag{42}$$

$$\mathcal{M}_2^{(n)} \triangleq \mathcal{H}_V^{(n)} \cap \mathcal{L}_{V|Y_2}^{(n)}. \tag{43}$$

Proposition 5 (Polarization): Consider the polarization sets given in Definition 12 and the message sets given in Definition 13 with parameter $\delta_n = 2^{-n^\beta}$ for $0 < \beta < \frac{1}{2}$. Then the asymptotic cardinality of the message sets is given by

$$\lim_{n \to \infty} \frac{1}{n} \left| \mathcal{M}_1^{(n)} \right| = H(X|V) - H(X|VY_1), \tag{44}$$

$$\lim_{n \to \infty} \frac{1}{n} \left| \mathcal{M}_2^{(n)} \right| = H(V) - H(V|Y_2). \tag{45}$$

Proof: We prove Eqn. (44). Eqn. (45) is derived in an identical manner. Consider the polarization sets given in Definition 12. Using set notation defined in Section I-C, define $\mathcal{B}_{X|VY_1}^{(n)} \triangleq \left(\mathcal{H}_{X|VY_1}^{(n)} \cup \mathcal{L}_{X|VY_1}^{(n)} \right)^{\mathsf{C}}$ where the set complement is with respect to the universal set $[n]$. The cardinality of set $\mathcal{M}_1^{(n)}$ is bounded as follows.

$$\frac{1}{n} \left| \mathcal{M}_1^{(n)} \right| = \frac{1}{n} \left| \mathcal{H}_{X|V}^{(n)} \cap \mathcal{L}_{X|VY_1}^{(n)} \right| \tag{46}$$

$$= \frac{1}{n} \left| \mathcal{H}_{X|V}^{(n)} \cap \left(\mathcal{H}_{X|VY_1}^{(n)} \cup \mathcal{B}_{X|VY_1}^{(n)} \right)^{\mathsf{C}} \right| \tag{47}$$

$$= \frac{1}{n} \left| \mathcal{H}_{X|V}^{(n)} \cap \left(\mathcal{H}_{X|VY_1}^{(n)} \right)^{\mathsf{C}} \cap \left(\mathcal{B}_{X|VY_1}^{(n)} \right)^{\mathsf{C}} \right| \tag{48}$$

$$\geq \frac{1}{n} \left| \mathcal{H}_{X|V}^{(n)} \cap \left(\mathcal{H}_{X|VY_1}^{(n)} \right)^{\mathsf{C}} \right| - \frac{1}{n} \left| \mathcal{B}_{X|VY_1}^{(n)} \right| \tag{49}$$

$$= \frac{1}{n} \left| \mathcal{H}_{X|V}^{(n)} \right| - \frac{1}{n} \left| \mathcal{H}_{X|VY_1}^{(n)} \right| - \frac{1}{n} \left| \mathcal{B}_{X|VY_1}^{(n)} \right| \tag{50}$$

Step (46) follows from the definition of the message set; Step (47) holds by definition of the set complement and because the union $\mathcal{B}_{X|VY_1}^{(n)} \cup \mathcal{L}_{X|VY_1}^{(n)} \cup \mathcal{H}_{X|VY_1}^{(n)} = [n]$; Step (48) and the inequality in (49) are due to standard set operations. The final equality in Step (50) holds due to $\mathcal{H}_{X|VY_1}^{(n)} \subseteq \mathcal{H}_{X|V}^{(n)}$. The inclusion $\mathcal{H}_{X|VY_1}^{(n)} \subseteq \mathcal{H}_{X|V}^{(n)}$ holds due to Lemma 16 and the fact that conditioning reduces entropy, i.e., $H(U_1(j)|U_1^{1:j-1}, V^{1:n}, Y_1^{1:n}) \leq H(U_1(j)|U_1^{1:j-1}, V^{1:n})$ for $j \in [n]$. Taking the limit $n \to \infty$ on both sides, one can apply Arıkan's conditional source polarization theorems [8, Theorem 1] with the rate of polarization given in [17, Theorem 1]. Since the set $\mathcal{B}_{X|VY_1}^{(n)}$ contains $o(n)$ indices, $\lim_{n\to\infty} \frac{1}{n} \left| \mathcal{M}_1^{(n)} \right| \geq$

$H(X|V) - H(X|VY_1)$. Disregarding the set $\mathcal{B}_{X|VY_1}^{(n)}$ provides a matching upper bound for the limit. This concludes the proof. ∎

Lemma 4: Consider the message sets defined in Definition 13. If the property $P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v)$ holds for conditional distributions $P_{Y_1|V}(y_1|v)$ and $P_{Y_2|V}(y_2|v)$, then the Bhattacharyya parameters

$$Z\left(U_2(j) \middle| U_2^{1:j-1}, Y_1^n\right) \leq Z\left(U_2(j) \middle| U_2^{1:j-1}, Y_2^n\right)$$

for all $j \in [n]$. As a result,

$$\mathcal{M}_2^{(n)} \subseteq \mathcal{M}_{1v}^{(n)}.$$

Proof: The proof follows from Lemma 12 and repeated application of Lemma 13 in Appendix A.

∎

C. Broadcast Encoding Blocks: $(\mathcal{E}_1, \mathcal{E}_2)$

The polarization theorems of the previous section are useful for defining a multi-user communication system as diagrammed in Figure 7. The broadcast encoder must map two independent messages (W_1, W_2) uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}]$ to a codeword $x^n \in \mathcal{X}^n$ in such a way that the decoding at each separate receiver is successful. The achievable rates for a particular block length n are

$$R_1 = \frac{1}{n}\left|\mathcal{M}_1^{(n)}\right|,$$
$$R_2 = \frac{1}{n}\left|\mathcal{M}_2^{(n)}\right|.$$

To construct a codeword, the encoder first produces two binary sequences $u_1^n \in \{0,1\}^n$ and $u_2^n \in \{0,1\}^n$. To determine $u_1(j)$ for $j \in \mathcal{M}_1^{(n)}$, the bit is selected as a uniformly distributed message bit intended for the first receiver. To determine $u_2(j)$ for $j \in \mathcal{M}_2^{(n)}$, the bit is selected as a uniformly distributed message bit intended for the second receiver. The remaining *non-message* indices of u_1^n and u_2^n are computed according to deterministic or random functions which are *shared* between the encoder and decoder.

1) Deterministic Mapping: Consider the following deterministic boolean functions indexed by $j \in [n]$:

$$\psi_1^{(j)} : \{0,1\}^{n+j-1} \to \{0,1\}, \tag{51}$$

$$\psi_2^{(j)} : \{0,1\}^{j-1} \to \{0,1\}. \tag{52}$$

As an example, consider deterministic boolean functions based on the *maximum a posteriori* polar coding rule (with ties broken in favor of decoding to a 0 versus a 1).

$$\psi_1^{(j)}\left(u_1^{1:j-1}, v^n\right) \triangleq \underset{u \in \{0,1\}}{\arg \max}\left\{\mathbb{P}\left(U_1(j) = u \Big| U_1^{1:j-1} = u_1^{1:j-1}, V^n = v^n\right)\right\}. \tag{53}$$

$$\psi_2^{(j)}\left(u_2^{1:j-1}\right) \triangleq \underset{u \in \{0,1\}}{\arg \max}\left\{\mathbb{P}\left(U_2(j) = u \Big| U_2^{1:j-1} = u_2^{1:j-1}\right)\right\}. \tag{54}$$

2) Random Mapping: Consider the following class of random boolean functions indexed by $j \in [n]$:

$$\Psi_1^{(j)} : \{0,1\}^{n+j-1} \rightarrow \{0,1\}, \tag{55}$$

$$\Psi_2^{(j)} : \{0,1\}^{j-1} \rightarrow \{0,1\}. \tag{56}$$

As an example, consider the random boolean functions

$$\Psi_1^{(j)}\left(u_1^{1:j-1}, v^n\right) \triangleq \begin{cases} 0, & \text{w. p. } \lambda_0\left(u_1^{1:j-1}, v^n\right), \\ 1, & \text{w. p. } 1 - \lambda_0\left(u_1^{1:j-1}, v^n\right), \end{cases} \tag{57}$$

$$\Psi_2^{(j)}\left(u_2^{1:j-1}\right) \triangleq \begin{cases} 0, & \text{w. p. } \lambda_0\left(u_2^{1:j-1}\right), \\ 1, & \text{w. p. } 1 - \lambda_0\left(u_2^{1:j-1}\right), \end{cases} \tag{58}$$

where

$$\lambda_0\left(u_2^{1:j-1}\right) \triangleq \mathbb{P}\left(U_2(j) = 0 \big| U_2^{1:j-1} = u_2^{1:j-1}\right),$$

$$\lambda_0\left(u_1^{1:j-1}, v^n\right) \triangleq \mathbb{P}\left(U_1(j) = 0 \big| U_1^{1:j-1} = u_1^{1:j-1}, V^n = v^n\right).$$

For fixed j, the random boolean function $\Psi_1^{(j)}$ (or $\Psi_2^{(j)}$) may be thought of as a vector of 2^{n+j-1} (or respectively 2^{j-1}) independent Bernoulli random variables. Each Bernoulli random variable of the vector is zero or one with a fixed probability.

3) Protocol: The encoder constructs the sequence u_2^n first using the message bits W_2 and either (54) or (58). Next, the sequence $v^n = u_2^n \mathbf{G}_n$ is created. Finally, the sequence u_1^n is constructed using the message bits W_1, the sequence v^n, and either the deterministic maps defined in (53) or the randomized maps in (57). The transmitted codeword is $x^n = u_1^n \mathbf{G}_n$.

D. Broadcast Decoding Based on Polarization

1) Decoding At First Receiver: Decoder $\mathcal{D}_1$ decodes the binary sequence $\hat{u}_2^n$ first using its observations y_1^n. It then reconstructs $\hat{v}^n = \hat{u}_2^n \mathbf{G}_n$. Using the sequence $\hat{v}^n$ and observations y_1^n, the decoder reconstructs

$\hat{u}_1^n$. The message W_1 is located at the indices $j \in \mathcal{M}_1^{(n)}$ in the sequence $\hat{u}_1^n$. More precisely, define the following deterministic polar decoding functions:

$$\xi_v^{(j)} \left(u_2^{1:j-1}, y_1^n \right) \triangleq \arg\max_{u \in \{0,1\}} \left\{ \mathbb{P} \left(U_2(j) = u \Big| U_2^{1:j-1} = u_2^{1:j-1}, Y_1^n = y_1^n \right) \right\}. \tag{59}$$

$$\xi_{u_1}^{(j)} \left(u_1^{1:j-1}, v^n, y_1^n \right) \triangleq \arg\max_{u \in \{0,1\}} \left\{ \mathbb{P} \left(U_1(j) = u \Big| U_1^{1:j-1} = u_1^{1:j-1}, V^n = v^n, Y_1^n = y_1^n \right) \right\}. \tag{60}$$

The decoder $\mathcal{D}_1$ reconstructs $\hat{u}_2^n$ bit-by-bit successively as follows using the *identical* shared random mapping $\Psi_2^{(j)}$ (or possibly the identical shared mapping $\psi_2^{(j)}$) used at the encoder:

$$\hat{u}_2(j) = \begin{cases} \xi_v^{(j)} \left(\hat{u}_2^{1:j-1}, y_1^n \right), & \text{if } j \in \mathcal{M}_2^{(n)}, \\ \Psi_2^{(j)} \left(\hat{u}_2^{1:j-1} \right), & \text{otherwise.} \end{cases} \tag{61}$$

If Lemma 4 holds, note that $\mathcal{M}_2^{(n)} \subseteq \mathcal{M}_{1v}^{(n)}$. With $\hat{u}_2^n$, decoder $\mathcal{D}_1$ reconstructs $\hat{v}^n = \hat{u}_2^n \mathbf{G}_n$. Then the sequence $\hat{u}_1^n$ is constructed bit-by-bit successively as follows using the *identical* shared random mapping $\Psi_1^{(j)}$ (or possibly the identical shared mapping $\psi_1^{(j)}$) used at the encoder:

$$\hat{u}_1(j) = \begin{cases} \xi_{u_1}^{(j)} \left(\hat{u}_1^{1:j-1}, \hat{v}^n, y_1^n \right), & \text{if } j \in \mathcal{M}_1^{(n)}, \\ \Psi_1^{(j)} \left(\hat{u}_1^{1:j-1}, \hat{v}^n \right), & \text{otherwise.} \end{cases} \tag{62}$$

2) Decoding At Second Receiver: The decoder $\mathcal{D}_2$ decodes the binary sequence $\hat{u}_2^n$ using observations y_2^n. The message W_2 is located at the indices $j \in \mathcal{M}_2^{(n)}$ of the sequence $\hat{u}_2^n$. More precisely, define the following polar decoding functions

$$\xi_v^{(j)} \left(u_2^{1:j-1}, y_2^n \right) \triangleq \arg\max_{u \in \{0,1\}} \left\{ \mathbb{P} \left(U_2(j) = u \Big| U_2^{1:j-1} = u_2^{1:j-1}, Y_2^n = y_2^n \right) \right\}. \tag{63}$$

The decoder $\mathcal{D}_2$ reconstructs $\hat{u}_2^n$ bit-by-bit successively as follows using the *identical* shared random mapping $\Psi_2^{(j)}$ (or possibly the identical shared mapping $\psi_2^{(j)}$) used at the encoder:

$$\hat{u}_2(j) = \begin{cases} \xi_v^{(j)} \left(\hat{u}_2^{1:j-1}, y_2^n \right), & \text{if } j \in \mathcal{M}_2^{(n)}, \\ \Psi_2^{(j)} \left(\hat{u}_2^{1:j-1} \right), & \text{otherwise.} \end{cases} \tag{64}$$

Remark 12: The encoder and decoders execute *the same* protocol for reconstructing bits at the *non-message* indices. This is achieved by applying the same deterministic maps $\psi_1^{(j)}$ and $\psi_2^{(j)}$ or randomized maps $\Psi_1^{(j)}$ and $\Psi_2^{(j)}$.

E. Total Variation Bound

To analyze the average probability of error $P_e^{(n)}$ via the probabilistic method, it is assumed that both the encoder and decoder share the *randomized* mappings $\Psi_1^{(j)}$ and $\Psi_2^{(j)}$. Define the following probability measure on the space of tuples of binary sequences.

$$Q\left(u_1^n, u_2^n\right) \triangleq Q\left(u_2^n\right) Q\left(u_1^n \big| u_2^n\right) = \prod_{j=1}^{n} Q\left(u_2(j) \Big| u_2^{1:j-1}\right) Q\left(u_1(j) \Big| u_1^{1:j-1}, u_2^n\right). \tag{65}$$

In (65), the conditional probability measures are defined as

$$Q\left(u_2(j) \Big| u_2^{1:j-1}\right) \triangleq \begin{cases} \frac{1}{2}, & \text{if } j \in \mathcal{M}_2^{(n)}, \\ P\left(u_2(j) \Big| u_2^{1:j-1}\right), & \text{otherwise.} \end{cases}$$

$$Q\left(u_1(j) \Big| u_1^{1:j-1}, u_2^n\right) \triangleq \begin{cases} \frac{1}{2}, & \text{if } j \in \mathcal{M}_1^{(n)}, \\ P\left(u_1(j) \Big| u_1^{1:j-1}, u_2^n\right), & \text{otherwise.} \end{cases}$$

The probability measure Q defined in (65) is a perturbation of the joint probability measure $P_{U_1^n U_2^n}(u_1^n, u_2^n)$ in (40). The only difference in definition between P and Q is due to those indices in message sets $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$. The following lemma provides a bound on the total variation distance between P and Q. The lemma establishes the fact that inserting uniformly distributed message bits in the proper indices $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$ at the encoder *does not* perturb the statistics of the n-length random variables too much. This occurs because $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$ denote "high-entropy" indices for which the associated conditional distributions are close to uniform.

Lemma 5: (*Total Variation Bound*) Let probability measures P and Q be defined as in (40) and (65) respectively. Let $0 < \beta < 1$. The total variation distance between P and Q is bounded as

$$\sum_{\substack{u_1^n \in \{0,1\}^n \\ u_2^n \in \{0,1\}^n}} \left| P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) - Q\left(u_1^n, u_2^n\right) \right| = O(2^{-n^\beta}).$$

Proof: See Section C of the Appendices. ∎

F. Error Sequences

The decoding protocols for $\mathcal{D}_1$ and $\mathcal{D}_2$ were established in Section VII-D. To analyze the probability of error of successive cancelation (SC) decoding, consider the sequences u_1^n and u_2^n formed at the encoder, and the resulting observations y_1^n and y_2^n received by the decoders. It is convenient to group the sequences together and consider all tuples $(u_1^n, u_2^n, y_1^n, y_2^n)$.

Decoder $\mathcal{D}_1$ makes an SC decoding error on the j-th bit for the following tuples:

$$\mathcal{T}_{1v}^j \triangleq \Big\{ \big(u_1^n, u_2^n, y_1^n, y_2^n\big):$$

$$P_{U_2^j \big| U_2^{1:j-1} Y_1^n} \big(u_2(j) \big| u_2^{1:j-1}, y_1^n\big) \leq$$

$$P_{U_2^j \big| U_2^{1:j-1} Y_1^n} \big(u_2(j) \oplus 1 \big| u_2^{1:j-1}, y_1^n\big) \Big\},$$

$$\mathcal{T}_1^j \triangleq \Big\{ \big(u_1^n, u_2^n, y_1^n, y_2^n\big):$$

$$P_{U_1^j \big| U_1^{1:j-1} V^n Y_1^n} \big(u_1(j) \big| u_1^{1:j-1}, u_2^n \mathbf{G}_n, y_1^n\big) \leq$$

$$P_{U_1^j \big| U_1^{1:j-1} V^n Y_1^n} \big(u_1(j) \oplus 1 \big| u_1^{1:j-1}, u_2^n \mathbf{G}_n, y_1^n\big) \Big\}. \tag{66}$$

The set $\mathcal{T}_{1v}^j$ represents those tuples causing an error at $\mathcal{D}_1$ in the case $u_2(j)$ is inconsistent with respect to observations y_1^n and the decoding rule. The set $\mathcal{T}_1^j$ represents those tuples causing an error at $\mathcal{D}_1$ in the case $u_1(j)$ is inconsistent with respect to $v^n = u_2^n \mathbf{G}_n$, observations y_1^n, and the decoding rule. Similarly, decoder $\mathcal{D}_2$ makes an SC decoding error on the j-th bit for the following tuples:

$$\mathcal{T}_2^j \triangleq \Big\{ \big(u_1^n, u_2^n, y_1^n, y_2^n\big): P_{U_2 \big| U_2^{1:j-1} Y_2^n} \big(u_2 \big| u_2^{1:j-1}, y_2^n\big) \leq$$

$$P_{U_2 \big| U_2^{1:j-1} Y_2^n} \big(u_2 \oplus 1 \big| u_2^{1:j-1}, y_2^n\big) \Big\}.$$

The set $\mathcal{T}_2^j$ represents those tuples causing an error at $\mathcal{D}_2$ in the case $u_2(j)$ is inconsistent with respect to observations y_2^n and the decoding rule. Since both decoders $\mathcal{D}_1$ and $\mathcal{D}_2$ only declare errors for those indices in the message sets, the set of tuples causing an error is

$$\mathcal{T}_{1v} \triangleq \bigcup_{j \in \mathcal{M}_2^{(n)} \subseteq \mathcal{M}_{1v}^{(n)}} \mathcal{T}_{1v}^j, \tag{67}$$

$$\mathcal{T}_1 \triangleq \bigcup_{j \in \mathcal{M}_1^{(n)}} \mathcal{T}_1^j, \tag{68}$$

$$\mathcal{T}_2 \triangleq \bigcup_{j \in \mathcal{M}_2^{(n)}} \mathcal{T}_2^j. \tag{69}$$

The complete set of tuples causing a broadcast error is

$$\mathcal{T} \triangleq \mathcal{T}_{1v} \cup \mathcal{T}_1 \cup \mathcal{T}_2. \tag{70}$$

The goal is to show that the probability of choosing tuples of error sequences in the set $\mathcal{T}$ is small under the distribution induced by the broadcast code.

G. Average Error Probability

Denote the total sum rate of the broadcast protocol as $R_\Sigma = R_1 + R_2$. Consider first the use of fixed deterministic maps $\psi_1^{(j)}$ and $\psi_2^{(j)}$ shared between the encoder and decoders. Then the probability of error of broadcasting the two messages at rates R_1 and R_2 is given by

$$
P_e^{(n)}\left[\{\psi_1^{(j)}, \psi_2^{(j)}\}\right] = \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} \left[P_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n \big| u_1^n, u_2^n\right) \right.
$$
$$
\cdot \frac{1}{2^{nR_2}} \prod_{j \in [n]: j \notin \mathcal{M}_2^{(n)}} \mathbb{1}\left\{\psi_2^{(j)}\left(u_2^{1:j-1}\right) = u_2(j)\right\}
$$
$$
\left. \cdot \frac{1}{2^{nR_1}} \prod_{j \in [n]: j \notin \mathcal{M}_1^{(n)}} \mathbb{1}\left\{\psi_1^{(j)}\left(u_1^{1:j-1}, u_2^n \mathbf{G}_n\right) = u_1(j)\right\} \right].
$$

If the encoder and decoders share randomized maps $\Psi_1^{(j)}$ and $\Psi_2^{(j)}$, then the average probability of error is a random quantity determined as follows

$$
P_e^{(n)}\left[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}\right] = \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} \left[P_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n \big| u_1^n, u_2^n\right) \right.
$$
$$
\cdot \frac{1}{2^{nR_2}} \prod_{j \in [n]: j \notin \mathcal{M}_2^{(n)}} \mathbb{1}\left\{\Psi_2^{(j)}\left(u_2^{1:j-1}\right) = u_2(j)\right\}
$$
$$
\left. \cdot \frac{1}{2^{nR_1}} \prod_{j \in [n]: j \notin \mathcal{M}_1^{(n)}} \mathbb{1}\left\{\Psi_1^{(j)}\left(u_1^{1:j-1}, u_2^n \mathbf{G}_n\right) = u_1(j)\right\} \right].
$$

By averaging over the randomness in the encoders and decoders, the expected block error probability is upper bounded in the following lemma.

Lemma 6: Consider the polarization-based superposition code described in Section VII-C and Section VII-D. Let R_1 and R_2 be the broadcast rates selected according to the Bhattacharyya criterion given in Proposition 5. Then for $0 < \beta < 1$,

$$
\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}]\right] = O(2^{-n^\beta}).
$$

Proof: See Section C of the Appendices. ∎

Remark 13: If the average probability of error decays to zero in expectation over the random maps $\{\Psi_1^{(j)}\}$ and $\{\Psi_2^{(j)}\}$, then there must exist at least one fixed set of deterministic maps for which $P_e^{(n)} \to 0$. While the result guarantees the existence of a low-complexity polar code, it does not guarantee that any specific deterministic map chosen from the ensemble will have a low probability of decoding error. The deterministic maps $\{\psi_1^{(j)}\}$ and $\{\psi_2^{(j)}\}$ must be tried experimentally.

VIII. Noisy Broadcast Channels

Marton's Coding Scheme

A. Marton's Inner Bound

For general noisy broadcast channels, Marton's inner bound involves two correlated auxiliary random variables V_1 and V_2 [3]. The intuition behind the coding strategy is to identify two "virtual" channels, one from V_1 to Y_1, and the other from V_2 to Y_2. Somewhat surprisingly, although the broadcast messages are independent, the auxiliary random variables V_1 and V_2 may be correlated to increase rates to both receivers. While there exist generalizations of Marton's strategy, the basic version of the inner bound is presented in this section[3].

Proposition 6 (Marton's Inner Bound): For any two-user DM-BC, the rates $(R_1, R_2) \in \mathbb{R}_+^2$ in the pentagonal region $\mathfrak{R}(X, V_1, V_2, Y_1, Y_2)$ are achievable where

$$\mathfrak{R}(X, V_1, V_2, Y_1, Y_2) \triangleq \Big\{ (R_1, R_2) \in \mathbb{R}_+^2 \;\Big|\; R_1 \leq I(V_1; Y_1),$$

$$R_2 \leq I(V_2; Y_2),$$

$$R_1 + R_2 \leq I(V_1; Y_1) + I(V_2; Y_2) - I(V_1; V_2) \Big\}. \qquad (71)$$

and where X, V_1, V_2, Y_1, Y_2 have a joint distribution given by $P_{V_1 V_2}(v_1, v_2) P_{X|V_1 V_2}(x|v_1, v_2) P_{Y_1 Y_2|X}(y_1, y_2|x)$.

Remark 14: It can be shown that for Marton's inner bound there is no loss of generality if $P_{X|V_1 V_2}(x|v_1, v_2) = \mathbb{1}\{x = \phi(v_1, v_2)\}$ where $\phi(v_1, v_2)$ is a deterministic function [2, Section 8.3]. Thus, by allowing a larger alphabet size for the auxiliaries, X may be a deterministic function of auxiliaries (V_1, V_2). Marton's inner bound is tight for the class of *semi-deterministic* DM-BCs for which one of the outputs is a deterministic function of the input.

B. Main Result

Theorem 3 (Polarization-Based Marton Code): Consider any two-user DM-BC with arbitrary input and output alphabets. There exist sequences of polar broadcast codes over n channel uses which achieve

---

[3]In addition, it is difficult even to evaluate Marton's inner bound for general channels due to the need for proper cardinality bounds on the auxiliaries [40]. These issues lie outside the scope of the present paper.
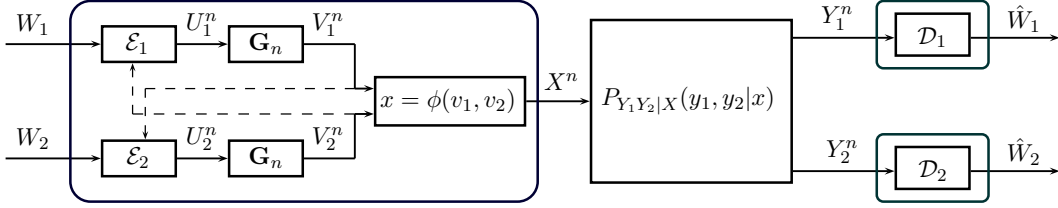
Fig. 8. Block diagram of a polarization-based Marton code for a two-user noisy broadcast channel.

the following rate region

$$\mathfrak{R}(V_1, V_2, X, Y_1, Y_2) \triangleq \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \ \middle| \ R_1 \leq I(V_1; Y_1), \right.$$
$$\left. R_2 \leq I(V_2; Y_2) - I(V_1; V_2) \right\}, \tag{72}$$

where random variables V_1, V_2, X, Y_1, Y_2 have the following listed properties:

- V_1 and V_2 are binary random variables.
- $P_{Y_2|V_2}(y_2|v_2) \succeq P_{V_1|V_2}(v_1|v_2)$.
- For a deterministic function $\phi : \{0,1\}^2 \to \mathcal{X}$, the joint distribution of all random variables is given by

$$P_{V_1 V_2 X Y_1 Y_2}(v_1, v_2, x, y_1, y_2) = P_{V_1 V_2}(v_1, v_2) \mathbb{1}\{x = \phi(v_1, v_2)\} P_{Y_1 Y_2|X}(y_1, y_2|x).$$

For $0 < \beta < \frac{1}{2}$, the average error probability of this code sequence decays as $P_e^{(n)} = O(2^{-n^\beta})$. The complexity of encoding and decoding is $O(n \log n)$.

Remark 15: The listed property $P_{Y_2|V_2}(y_2|v_2) \succeq P_{V_1|V_2}(v_1|v_2)$ is necessary in the proof due to polarization-based codes requiring an *alignment* of polarization indices. The property is a *natural* restriction since it also implies that $I(Y_2; V_2) > I(V_1; V_2)$ so that $R_2 > 0$. However, certain joint distributions on random variables are not permitted using the analysis of polarization presented here. It is not clear whether there is another approach that obviates the need for alignment of indices. The recent work of Mondelli *et al.* focuses on removing the alignment restrictions for broadcast channels [41].

Remark 16: By symmetry, the rate tuple $(R_1, R_2) = (I(V_1; Y_1) - I(V_1; V_2), I(V_2, Y_2))$ is achievable with low-complexity codes under similar constraints on the joint distribution of V_1, V_2, X, Y_1, Y_2. The rate tuple is a corner point of the pentagonal rate region of Marton's inner bound given in (71).

IX. PROOF OF THEOREM 3

The block diagram for polarization-based Marton coding is given in Figure 8. Marton's strategy differs form Cover's superposition coding with the presence of two auxiliaries and the function $\phi(v_1, v_2)$ which forms the codeword symbol-by-symbol. The polar transform is applied to each n-length $i.i.d.$ sequence of auxiliary random variables.

A. Polar Transform

Consider the $i.i.d.$ sequence of random variables

$$(V_1^j, V_2^j, X^j, Y_1^j, Y_2^j) \sim P_{V_1 V_2}(v_1, v_2) P_{X|V_1 V_2}(x|v_1, v_2) P_{Y_1 Y_2|X}(y_1, y_2|x),$$

where the index $j \in [n]$. For the particular coding strategy analyzed in this section, $P_{X|V_1 V_2}(x|v_1, v_2) = \mathbb{1}\{x = \phi(v_1, v_2)\}$. Let the n-length sequence of auxiliary variables (V_1^j, V_2^j) be organized into the random matrix

$$\mathbf{\Omega} \triangleq \left[\begin{array}{ccccc} V_1^1 & V_1^2 & V_1^3 & \dots & V_1^n \\ V_2^1 & V_2^2 & V_2^3 & \dots & V_2^n \end{array} \right]. \tag{73}$$

Applying the polar transform to $\mathbf{\Omega}$ results in the random matrix $\mathbf{U} \triangleq \mathbf{\Omega} \mathbf{G}_n$. Index the random variables of $\mathbf{U}$ as follows:

$$\mathbf{U} = \left[\begin{array}{ccccc} U_1^1 & U_1^2 & U_1^3 & \dots & U_1^n \\ U_2^1 & U_2^2 & U_2^3 & \dots & U_2^n \end{array} \right]. \tag{74}$$

The above definitions are consistent with the block diagram given in Figure 8 (and noting that $\mathbf{G}_n = \mathbf{G}_n^{-1}$). The polar transform extracts the randomness of $\mathbf{\Omega}$. In the transformed domain, the joint distribution of the variables in $\mathbf{U}$ is given by

$$P_{U_1^n U_2^n}(u_1^n, u_2^n) \triangleq P_{V_1^n V_2^n}(u_1^n \mathbf{G}_n, u_2^n \mathbf{G}_n). \tag{75}$$

However, for polar coding purposes, the joint distribution is decomposed as follows,

$$P_{U_1^n U_2^n}(u_1^n, u_2^n) = P_{U_1^n}(u_1^n) P_{U_2^n|U_1^n}(u_2^n|u_1^n) = \prod_{j=1}^{n} P(u_1(j)|u_1^{1:j-1}) P(u_2(j)|u_2^{1:j-1}, u_1^n). \tag{76}$$

The above conditional distributions may be computed efficiently using recursive protocols. The polarized random variables of $\mathbf{U}$ do *not necessarily* have an $i.i.d.$ distribution.
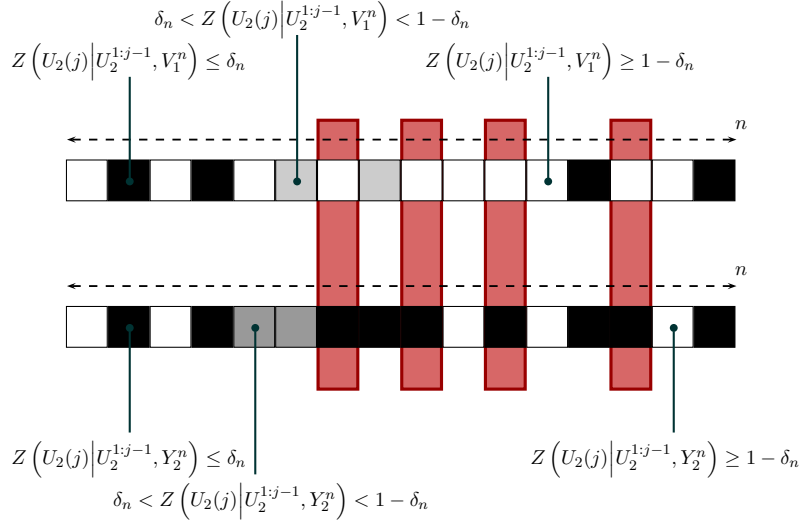
Fig. 9. The alignment of polarization indices for Marton coding over noisy broadcast channels with respect to the second receiver. The message set $\mathcal{M}_2^{(n)}$ is highlighted by the vertical red rectangles. At finite code length n, exact alignment is not possible due to partially-polarized indices pictured in gray.

B. Effective Channel

Marton's achievable strategy establishes virtual channels for the two receivers via the function $\phi(v_1, v_2)$. The virtual channel is given by

$$P_{Y_1 Y_2 | V_1 V_2}^{\phi}\left(y_1, y_2 \Big| v_1, v_2\right) \triangleq P_{Y_1 Y_2 | X}\left(y_1, y_2 \Big| \phi(v_1, v_2)\right).$$

Due to the memoryless property of the DM-BC, the effective channel between auxiliaries and channel outputs is given by

$$P_{Y_1^n Y_2^n | V_1^n V_2^n}^{\phi}\left(y_1^n, y_2^n \Big| v_1^n, v_2^n\right) \triangleq \prod_{i=1}^{n} P_{Y_1 Y_2 | X}\left(y_1(i), y_2(i) \Big| \phi(v_1(i), v_2(i))\right).$$

The polarization-based Marton code establishes a different effective channel between polar-transformed auxiliaries and the channel outputs. The effective polarized channel is

$$P_{Y_1^n Y_2^n | U_1^n U_2^n}^{\phi}\left(y_1^n, y_2^n \Big| u_1^n, u_2^n\right) \triangleq P_{Y_1^n Y_2^n | V_1^n V_2^n}^{\phi}\left(y_1^n, y_2^n \Big| u_1^n \mathbf{G}_n, u_2^n \mathbf{G}_n\right). \tag{77}$$

C. Polarization Theorems Revisited

Definition 14 (Polarization Sets for Marton Coding): Let $V_1^n, V_2^n, X^n, Y_1^n, Y_2^n$ be the sequence of random variables as introduced in Section IX-A. In addition, let $U_1^n = V_1^n \mathbf{G}_n$ and $U_2^n = V_2^n \mathbf{G}_n$. Let

$\delta_n = 2^{-n^\beta}$ for $0 < \beta < \frac{1}{2}$. The following *polarization sets* are defined:

$$\mathcal{H}_{V_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_1(j) \middle| U_1^{1:j-1}\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{L}_{V_1|Y_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_1(j) \middle| U_1^{1:j-1}, Y_1^n\right) \leq \delta_n \right\},$$

$$\mathcal{H}_{V_2|V_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \middle| U_2^{1:j-1}, V_1^n\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{L}_{V_2|V_1}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \middle| U_2^{1:j-1}, V_1^n\right) \leq \delta_n \right\},$$

$$\mathcal{H}_{V_2|Y_2}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \middle| U_2^{1:j-1}, Y_2^n\right) \geq 1 - \delta_n \right\},$$

$$\mathcal{L}_{V_2|Y_2}^{(n)} \triangleq \left\{ j \in [n] : Z\left(U_2(j) \middle| U_2^{1:j-1}, Y_2^n\right) \leq \delta_n \right\}.$$

Definition 15 (Message Sets for Marton Coding): In terms of the polarization sets given in Definition 14, the following *message sets* are defined:

$$\mathcal{M}_1^{(n)} \triangleq \mathcal{H}_{V_1}^{(n)} \cap \mathcal{L}_{V_1|Y_1}^{(n)}, \tag{78}$$

$$\mathcal{M}_2^{(n)} \triangleq \mathcal{H}_{V_2|V_1}^{(n)} \cap \mathcal{L}_{V_2|Y_2}^{(n)}. \tag{79}$$

Proposition 7 (Polarization): Consider the polarization sets given in Definition 14 and the message sets given in Definition 15 with parameter $\delta_n = 2^{-n^\beta}$ for $0 < \beta < \frac{1}{2}$. Then the asymptotic cardinality of the message sets is given by

$$\lim_{n \to \infty} \frac{1}{n} \left| \mathcal{M}_1^{(n)} \right| = H(V_1) - H(V_1|Y_1), \tag{80}$$

$$\lim_{n \to \infty} \frac{1}{n} \left| \mathcal{M}_2^{(n)} \right| = H(V_2|V_1) - H(V_2|Y_2). \tag{81}$$

Proof: The proof of Eqn. (80) is identical to the proof of Proposition 5. To prove Eqn. (81), an identical proof to the proof of Proposition 5 applies; however, the exception is that the set inclusion $\mathcal{H}_{V_2|Y_2}^{(n)} \subseteq \mathcal{H}_{V_2|V_1}^{(n)}$ is required. This set inclusion for Marton coding is proven in Lemma 7. ∎

Lemma 7: Consider the polarization sets defined in Proposition 7. If the property $P_{Y_2|V_2}(y_2|v_2) \succeq P_{V_1|V_2}(v_1|v_2)$ holds for conditional distributions $P_{Y_2|V_2}(y_2|v_2)$ and $P_{V_1|V_2}(v_1|v_2)$, then $I(V_2; Y_2) > I(V_1; V_2)$ and the Bhattacharyya parameters

$$Z\left(U_2(j) \middle| U_2^{1:j-1}, Y_2^n\right) \leq Z\left(U_2(j) \middle| U_2^{1:j-1}, V_1^n\right)$$

for all $j \in [n]$. As a result,

$$\mathcal{L}_{V_2|V_1}^{(n)} \subseteq \mathcal{L}_{V_2|Y_2}^{(n)},$$

$$\mathcal{H}_{V_2|Y_2}^{(n)} \subseteq \mathcal{H}_{V_2|V_1}^{(n)}.$$

Proof: The proof follows from Lemma 12 and repeated application of Lemma 13 in Appendix A. ∎

Remark 17: The alignment of polarization indices characterized by Lemma 7 is diagrammed in Figure 9. The alignment ensures the existence of polarization indices in the set $\mathcal{M}_2^{(n)}$ for the message W_2 to have a positive rate $R_2 > 0$. The indices in $\mathcal{M}_2^{(n)}$ represent those bits freely set at the broadcast encoder and simultaneously those bits that may be decoded by $\mathcal{D}_2$ given its observations.

D. Partially-Polarized Indices

As shown in Figure 9, for the Marton coding scheme, exact alignment of polarization indices is not possible. However, the alignment holds for all but $o(n)$ indices. The sets of partially-polarized indices shown in Figure 9 are defined as follows.

Definition 16 (Sets of Partially-Polarized Indices):

$$\Delta_1 \triangleq [n] \setminus \left(\mathcal{H}_{V_2|V_1}^{(n)} \cup \mathcal{L}_{V_2|V_1}^{(n)} \right), \tag{82}$$

$$\Delta_2 \triangleq [n] \setminus \left(\mathcal{H}_{V_2|Y_2}^{(n)} \cup \mathcal{L}_{V_2|Y_2}^{(n)} \right). \tag{83}$$

As implied by Arıkan's polarization theorems, the number of partially-polarized indices is negligible asymptotically as $n \to \infty$. For an arbitrarily small $\eta > 0$,

$$\frac{\left| \Delta_1 \cup \Delta_2 \right|}{n} \leq \eta, \tag{84}$$

for all n sufficiently large enough. As will be discussed, providing these $o(n)$ bits as side-information ("genie-given") bits to the decoders results in a rate penalty; however, the rate penalty is negligible for sufficiently large code lengths.

E. Broadcast Encoding Blocks: $(\mathcal{E}_1, \mathcal{E}_2)$

As diagrammed in Figure 8, the broadcast encoder must map two independent messages (W_1, W_2) uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}]$ to a codeword $x^n \in \mathcal{X}^n$ in such a way that the decoding at each separate receiver is successful. The achievable rates for a particular block length n are

$$R_1 = \frac{1}{n} \left| \mathcal{M}_1^{(n)} \right|,$$

$$R_2 = \frac{1}{n} \left| \mathcal{M}_2^{(n)} \right|.$$

To construct a codeword, the encoder first produces two binary sequences $u_1^n \in \{0,1\}^n$ and $u_2^n \in \{0,1\}^n$. To determine $u_1(j)$ for $j \in \mathcal{M}_1^{(n)}$, the bit is selected as a uniformly distributed message bit

intended for the first receiver. To determine $u_2(j)$ for $j \in \mathcal{M}_2^{(n)}$, the bit is selected as a uniformly distributed message bit intended for the second receiver. The remaining *non-message* indices of u_1^n and u_2^n are decided *randomly* according to the proper statistics as will be described in this section. The transmitted codeword is formed symbol-by-symbol via the ϕ function,

$$\forall j \in [n] : x(j) = \phi\big(v_1(j), v_2(j)\big)$$

where $v_1^n = u_1^n \mathbf{G}_n$ and $v_2^n = u_2^n \mathbf{G}_n$. A valid codeword sequence is always guaranteed to be formed unlike in the case of coding for deterministic broadcast channels.

1) Random Mapping: To fill in the non-message indices, we define the following random mappings. Consider the following class of random boolean functions where $j \in [n]$:

$$\Psi_1^{(j)} : \{0,1\}^{j-1} \to \{0,1\}, \tag{85}$$

$$\Psi_2^{(j)} : \{0,1\}^{n+j-1} \to \{0,1\}, \tag{86}$$

$$\Gamma : [n] \to \{0,1\}. \tag{87}$$

More concretely, we consider the following specific random boolean functions based on the statistics derived from polarization methods:

$$\Psi_1^{(j)}\left(u_1^{1:j-1}\right) \triangleq \begin{cases} 0, & \text{w.\,p.} \ \ \lambda_0\left(u_1^{1:j-1}\right), \\ 1, & \text{w.\,p.} \ \ 1 - \lambda_0\left(u_1^{1:j-1}\right), \end{cases} \tag{88}$$

$$\Psi_2^{(j)}\left(u_2^{1:j-1}, v_1^n\right) \triangleq \begin{cases} 0, & \text{w.\,p.} \ \ \lambda_0\left(u_2^{1:j-1}, v_1^n\right), \\ 1, & \text{w.\,p.} \ \ 1 - \lambda_0\left(u_2^{1:j-1}, v_1^n\right) \end{cases} \tag{89}$$

$$\Gamma(j) \triangleq \begin{cases} 0, & \text{w.\,p.} \ \ \frac{1}{2}, \\ 1, & \text{w.\,p.} \ \ \frac{1}{2}, \end{cases} \tag{90}$$

where

$$\lambda_0\left(u_1^{1:j-1}\right) \triangleq \mathbb{P}\left(U_1(j) = 0 \middle| U_1^{1:j-1} = u_1^{1:j-1}\right).$$

$$\lambda_0\left(u_2^{1:j-1}, v_1^n\right) \triangleq \mathbb{P}\left(U_2(j) = 0 \middle| U_2^{1:j-1} = u_2^{1:j-1}, V_1^n = v_1^n\right).$$

For a fixed $j \in [n]$, the random boolean function $\Psi_1^{(j)}$ (or $\Psi_2^{(j)}$) may be thought of as a vector of 2^{j-1} (or respectively 2^{n+j-1}) independent Bernoulli random variables. Each Bernoulli random variable of the vector is zero or one with a fixed well-defined probability that is efficiently computable. The random boolean function Γ may be thought of as an n-length vector of Bernoulli($\frac{1}{2}$) random variables.

2) Encoding Protocol: The broadcast encoder constructs the sequence u_1^n bit-by-bit successively,

$$u_1(j) = \begin{cases} W_1 \text{ message bit,} & \text{if } j \in \mathcal{M}_1^{(n)}, \\ \Psi_1^{(j)}(u_1^{1:j-1}), & \text{otherwise.} \end{cases} \tag{91}$$

The encoder then computes the sequence $v_1^n = u_1^n \mathbf{G}_n$. To generate v_2^n, the encoder constructs the sequence u_2^n (given v_1^n) as follows,

$$u_2(j) = \begin{cases} W_2 \text{ message bit,} & \text{if } j \in \mathcal{M}_2^{(n)}, \\ \Gamma(j), & \text{if } j \in \mathcal{H}_{V_2|V_1}^{(n)} \setminus \mathcal{M}_2^{(n)}, \\ \Psi_2^{(j)}(u_2^{1:j-1}, v_1^n), & \text{otherwise.} \end{cases} \tag{92}$$

Then the sequence $v_2^n = u_2^n \mathbf{G}_n$. The randomness in the above encoding protocol over non-message indices ensures that the pair of sequences (u_1^n, u_2^n) approximately has the correct statistics as if drawn from the joint distribution of (U_1^n, U_2^n). In the last step, the encoder transmits a codeword x^n formed symbol-by-symbol: $x(j) = \phi(v_1(j), v_2(j))$ for all $j \in [n]$.

Remark 18: Figure 9 illustrates the partial polarization and alignment of indices for Marton's coding scheme. For $j \in \Delta_2$, where Δ_2 is the set of partially-polarized indices defined in (83), the encoder records the realization of $u_2(j)$. These $o(n)$ indices are provided to the second receiver's decoder $\mathcal{D}_2$ as side-information or "genie-given" bits. The $o(n)$ indices do not affect the achieved rates significantly as explained further in Sec. IX-J.

Remark 19: Marton's coding scheme also motivates the reason for the function $\Gamma : [n] \to \{0,1\}$ at the encoder. The second receiver only observes y_2^n. The second receiver does *not* have access to v_1^n nor does it reconstruct $\hat{v}_1^n$. Since the second receiver cannot utilize the map $\Psi_2^{(j)}(u_2^{1:j-1}, v_1^n)$, the Γ function is necessary for a subset of the indices defined in Eqn. (92) at the encoder. The overall alignment of indices is drawn in Figure 9 for both the encoder and the second receiver's decoder.

F. Broadcast Decoding Based on Polarization

1) Decoding At First Receiver: Decoder $\mathcal{D}_1$ decodes the binary sequence $\hat{u}_1^n$ using its observations y_1^n. The message W_1 is located at the indices $j \in \mathcal{M}_1^{(n)}$ in the sequence $\hat{u}_1^n$. More precisely, we define the following deterministic polar decoding function for the j-th bit:

$$\xi_{u_1}^{(j)}\left(u_1^{1:j-1}, y_1^n\right) \triangleq \underset{u \in \{0,1\}}{\arg\max}\left\{\mathbb{P}\left(U_1(j) = u \middle| U_1^{1:j-1} = u_1^{1:j-1}, Y_1^n = y_1^n\right)\right\}. \tag{93}$$

Decoder $\mathcal{D}_1$ reconstructs $\hat{u}_1^n$ bit-by-bit successively as follows using the identical random mapping $\Psi_1^{(j)}$ at the encoder:

$$
\hat{u}_1(j) = \begin{cases} \xi_{u_1}^{(j)}\big(\hat{u}_1^{1:j-1}, y_1^n\big), & \text{if } j \in \mathcal{M}_1^{(n)}, \\ \Psi_1^{(j)}\big(\hat{u}_1^{1:j-1}\big), & \text{otherwise.} \end{cases} \tag{94}
$$

Given that all previous bits $\hat{u}_1^{1:j-1}$ have been decoded correctly, decoder $\mathcal{D}_1$ makes a mistake on the j-th bit $\hat{u}_1(j)$ only if $j \in \mathcal{M}_1^{(n)}$. For the remaining indices, the decoder produces the same bit produced at the encoder due to the shared random maps.

2) Decoding At Second Receiver: The decoder $\mathcal{D}_2$ decodes the binary sequence $\hat{u}_2^n$ using observations y_2^n. The message W_2 is located at the indices $j \in \mathcal{M}_2^{(n)}$ of the sequence $\hat{u}_2^n$. Define the following deterministic polar decoding functions

$$
\xi_{u_2}^{(j)}\big(u_2^{1:j-1}, y_2^n\big) \triangleq \arg\max_{u \in \{0,1\}} \Big\{ \mathbb{P}\Big(U_2(j) = u \Big| U_2^{1:j-1} = u_2^{1:j-1}, Y_2^n = y_2^n\Big) \Big\}. \tag{95}
$$

Decoder $\mathcal{D}_2$ reconstructs $\hat{u}_2^n$ bit-by-bit successively as follows using the *identical* shared random mapping Γ used at the encoder. Including all but $o(n)$ of the indices,

$$
\hat{u}_2(j) = \begin{cases} \xi_{u_2}^{(j)}\big(\hat{u}_2^{1:j-1}, y_2^n\big), & \text{if } j \in \mathcal{L}_{V_2|Y_2}^{(n)}, \\ \Gamma(j), & \text{if } j \in \mathcal{H}_{V_2|Y_2}^{(n)}. \end{cases} \tag{96}
$$

For those indices $j \in \Delta_2$ where Δ_2 is the set of partially-polarized indices defined in (83), the decoder $\mathcal{D}_2$ is provided with side-information ("genie-given") bits from the encoder. Thus, all bits are decoded, and $\mathcal{D}_2$ only makes a successive cancelation error for those indices $j \in \mathcal{L}_{V_2|Y_2}^{(n)}$. Communicating the side-information bits from the encoder to decoder results in a rate penalty. However, since the number of side-information bits scales asymptotically as $o(n)$, the rate penalty can be made arbitrarily small.

Remark 20: It is notable that decoder $\mathcal{D}_2$ reconstructs $\hat{u}_2^n$ using only the observations y_2^n. At the encoder, the sequence u_2^n was generated with the realization of a sequence v_1^n as given in (92). However, decoder $\mathcal{D}_2$ does *not* reconstruct the sequence $\hat{v}_1^n$. From this operational perspective, Marton's scheme differs crucially from Cover's superposition scheme because there does not exist the notion of a "stronger" receiver which reconstructs all the sequences decoded at the "weaker" receiver.

G. Total Variation Bound

To analyze the average probability of error $P_e^{(n)}$, it is assumed that both the encoder and decoder share the *randomized* mappings $\Psi_1^{(j)}$, $\Psi_2^{(j)}$, and Γ (where $\Psi_2^{(j)}$ is not utilized at decoder $\mathcal{D}_2$). Define the

following probability measure on the space of tuples of binary sequences.

$$Q(u_1^n, u_2^n) \triangleq Q(u_1^n) Q(u_1^n | u_2^n) = \prod_{j=1}^{n} Q(u_1(j) | u_1^{1:j-1}) Q(u_2(j) | u_2^{1:j-1}, u_1^n), \tag{97}$$

where the conditional probability measures are defined as

$$Q(u_1(j) | u_1^{1:j-1}) \triangleq \begin{cases} \frac{1}{2}, & \text{if } j \in \mathcal{M}_1^{(n)}, \\ P(u_1(j) | u_1^{1:j-1}), & \text{otherwise.} \end{cases}$$

$$Q(u_2(j) | u_2^{1:j-1}, u_1^n) \triangleq \begin{cases} \frac{1}{2}, & \text{if } j \in \mathcal{H}_{V_2|V_1}^{(n)}, \\ P(u_2(j) | u_2^{1:j-1}, u_1^n), & \text{otherwise.} \end{cases}$$

The probability measure Q defined in (97) is a perturbation of the joint probability measure $P_{U_1^n U_2^n}(u_1^n, u_2^n)$ in (76). The only difference in definition between P and Q is due to those indices in message sets $\mathcal{M}_1^{(n)}$ and $\mathcal{H}_{V_2|V_1}^{(n)}$ (note: $\mathcal{M}_2^{(n)} \subseteq \mathcal{H}_{V_2|V_1}^{(n)}$). The following lemma provides a bound on the total variation distance between P and Q. The lemma establishes the fact that inserting uniformly distributed message bits in the proper indices $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$ (or the entire set $\mathcal{H}_{V_2|V_1}^{(n)}$) at the encoder *does not* perturb the statistics of the n-length random variables too much.

Lemma 8: (*Total Variation Bound*) Let probability measures P and Q be defined as in (76) and (97) respectively. Let $0 < \beta < 1$. The total variation distance between P and Q is bounded as

$$\sum_{\substack{u_1^n \in \{0,1\}^n \\ u_2^n \in \{0,1\}^n}} \left| P_{U_1^n U_2^n}(u_1^n, u_2^n) - Q(u_1^n, u_2^n) \right| = O(2^{-n^\beta}).$$

Proof: Omitted. The proof follows via the chain rule for KL-divergence and is identical to the previous proofs of Lemma 1 and Lemma 5. ∎

H. Error Sequences

The decoding protocols for $\mathcal{D}_1$ and $\mathcal{D}_2$ were established in Section IX-F. To analyze the probability of error of successive cancellation (SC) decoding, consider the sequences u_1^n and u_2^n formed at the encoder, and the resulting observations y_1^n and y_2^n received by the decoders. The effective polarized channel $P_{Y_1^n Y_2^n | U_1^n U_2^n}^\phi(y_1^n, y_2^n | u_1^n, u_2^n)$ was defined in (77) for a fixed ϕ function. It is convenient to group the sequences together and consider all tuples $(u_1^n, u_2^n, y_1^n, y_2^n)$.

Decoder $\mathcal{D}_1$ makes an SC decoding error on the j-th bit for the following tuples:

$$\mathcal{T}_1^j \triangleq \Big\{ \big(u_1^n, u_2^n, y_1^n, y_2^n\big):$$

$$P_{U_1^j | U_1^{1:j-1} Y_1^n}\big(u_1(j) \big| u_1^{1:j-1}, y_1^n\big) \leq$$

$$P_{U_1^j | U_1^{1:j-1} Y_1^n}\big(u_1(j) \oplus 1 \big| u_1^{1:j-1}, y_1^n\big) \Big\}. \tag{98}$$

The set $\mathcal{T}_1^j$ represents those tuples causing an error at $\mathcal{D}_1$ in the case $u_1(j)$ is inconsistent with respect to observations y_1^n and the decoding rule. Similarly, decoder $\mathcal{D}_2$ makes an SC decoding error on the j-th bit for the following tuples:

$$\mathcal{T}_2^j \triangleq \Big\{ \big(u_1^n, u_2^n, y_1^n, y_2^n\big):$$

$$P_{U_2 | U_2^{1:j-1} Y_2^n}\big(u_2 \big| u_2^{1:j-1}, y_2^n\big) \leq$$

$$P_{U_2 | U_2^{1:j-1} Y_2^n}\big(u_2 \oplus 1 \big| u_2^{1:j-1}, y_2^n\big) \Big\}.$$

The set $\mathcal{T}_2^j$ represents those tuples causing an error at $\mathcal{D}_2$ in the case $u_2(j)$ is inconsistent with respect to observations y_2^n and the decoding rule. The set of tuples causing an error is

$$\mathcal{T}_1 \triangleq \bigcup_{j \in \mathcal{M}_1^{(n)}} \mathcal{T}_1^j, \tag{99}$$

$$\mathcal{T}_2 \triangleq \bigcup_{j \in \mathcal{L}_{V_2 | Y_2}^{(n)}} \mathcal{T}_2^j, \tag{100}$$

$$\mathcal{T} \triangleq \mathcal{T}_1 \cup \mathcal{T}_2. \tag{101}$$

The goal is to show that the probability of choosing tuples of error sequences in the set $\mathcal{T}$ is small under the distribution induced by the broadcast code.

I. Average Error Probability

If the encoder and decoders share randomized maps $\Psi_1^{(j)}$, $\Psi_2^{(j)}$, and Γ, then the average probability of error is a random quantity determined as follows

$$
P_e^{(n)}\left[\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}\right] = \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} \left[P^{\phi}_{Y_1^n Y_2^n | U_1^n U_2^n}(y_1^n, y_2^n | u_1^n, u_2^n) \right.
$$
$$
\cdot \frac{1}{2^{nR_1}} \prod_{j \in [n]: j \notin \mathcal{M}_1^{(n)}} \mathbb{1}\left\{\Psi_1^{(j)}\left(u_1^{1:j-1}\right) = u_1(j)\right\}
$$
$$
\cdot \frac{1}{2^{nR_2}} \prod_{j \in \mathcal{H}_{V_2|V_1}^{(n)} \setminus \mathcal{M}_2^{(n)}} \mathbb{1}\{\Gamma(j) = u_2(j)\}
$$
$$
\left. \cdot \prod_{j \in [n]: j \notin \mathcal{H}_{V_2|V_1}^{(n)}} \mathbb{1}\left\{\Psi_2^{(j)}\left(u_2^{1:j-1}, u_1^n \mathbf{G}_n\right) = u_2(j)\right\} \right].
$$

By averaging over the randomness in the encoders and decoders, the expected block error probability is upper bounded in the following lemma.

Lemma 9: Consider the polarization-based Marton code described in Section IX-E and Section IX-F. Let R_1 and R_2 be the broadcast rates selected according to the Bhattacharyya criterion given in Proposition 7. Then for $0 < \beta < 1$,

$$
\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}]\right] = O(2^{-n^\beta}).
$$

Proof: See Section D of the Appendices. ∎

If the average probability of block error decays to zero in expectation over the random maps $\{\Psi_1^{(j)}\}$, $\{\Psi_2^{(j)}\}$, and Γ, then there must exist at least one fixed set of maps for which $P_e^{(n)} \to 0$. Hence, polar codes for Marton's inner bound exist under suitable restrictions on distributions and they achieve reliable transmission according to the advertised rates (except for a small set of $o(n)$ polarization indices as is discussed next).

J. Rate Penalty Due to Partial Polarization

Lemma 9 is true assuming that decoder $\mathcal{D}_2$ obtains side-information ("genie-given") bits for the set of indices Δ_2 defined in (83). The set Δ_2 represents those indices that are partially-polarized and which cause a slight misalignment of polarization indices in the Marton scheme. Fortunately, the set Δ_2 contains a vanishing fraction of indices: $\frac{1}{n}|\Delta_2| \leq \eta$ for $\eta > 0$ arbitrarily small and n sufficiently large. Therefore, a two-phase strategy suffices for sending the side-information bits. In the first phase of communication,

the encoder sends several n-length blocks while decoder $\mathcal{D}_2$ waits to decode. After accumulating several blocks of output sequences, the encoder transmits all the known bits in the set Δ_2 for all the first-phase transmissions. The encoder and decoder can use any reliable point-to-point polar code with non-vanishing rate for communication. Having received the "genie-aided" bits in the second-phase, the second receiver then decodes all the first-phase blocks. The number of blocks sent in the first-phase is $O(\frac{1}{\eta})$. The rate penalty is $O(\eta)$ where η can be made arbitrarily small. A similar argument was provided in [25] for designing polar codes for the Gelfand-Pinsker problem.

X. Conclusion

Coding for broadcast channels is fundamental to our understanding of communication systems. Broadcast codes based on polarization methods achieve rates on the capacity boundary for several classes of DM-BCs. In the case of m-user deterministic DM-BCs, polarization of random variables from the channel output provides the ability to extract uniformly random message bits while maintaining broadcast constraints at the encoder. As referenced in the literature, maintaining multi-user constraints for the DM-BC is a difficult task for traditional belief propagation algorithms and LDPC codes.

For two-user noisy DM-BCs, polar codes were designed based on Marton's coding strategy and Cover's superposition strategy. Constraints on auxiliary and input distributions were placed in both cases to ensure alignment of polarization indices in the multi-user setting. The asymptotic behavior of the average error probability was shown to be $P_e^{(n)} = O(2^{-n^\beta})$ with an encoding and decoding complexity of $O(n \log n)$. The next step is to supplement the theory with experimental evidence of the error-correcting capability of polar codes over simulated channels for finite code lengths. The results demonstrate that polar codes have a potential for use in several *network communication* scenarios.

Appendix A
Polar Coding Lemmas

The following lemmas provide a basis for proving polar coding theorems. A subset of the lemmas were proven in different contexts, e.g., channel vs. source coding, and contain citations to references.

Lemma 10: Consider two random variables X taking values in $\{0,1\}$ and Y taking values in $\mathcal{Y}$. Denote the joint distribution by $P_{XY}(x,y)$. Let $Q(x|y) = \frac{1}{2}$ denote a uniform conditional distribution for $x \in \{0,1\}$ and $y \in \mathcal{Y}$. Then the following identity holds.

$$D\left(P_{X|Y}(x|y)\middle\|Q(x|y)\right) = 1 - H(X|Y). \tag{102}$$

Proof: The identity follows from standard definitions of entropy and Kullback-Leibler distance.

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \{0,1\}} P_{X|Y}(x|y) \log_2 \frac{1}{P_{X|Y}(x|y)}$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \{0,1\}} P_{X|Y}(x|y) \log_2 \frac{1}{Q(x|y)}$$

$$- \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \{0,1\}} P_{X|Y}(x|y) \log_2 \frac{P_{X|Y}(x|y)}{Q(x|y)}$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \left[1 - \sum_{x \in \{0,1\}} P_{X|Y}(x|y) \log_2 \frac{P_{X|Y}(x|y)}{Q(x|y)} \right]$$

$$= 1 - D\left(P_{X|Y}(x|y) \middle\| Q(x|y) \right).$$

■

Lemma 11 (Estimating The Bhattacharyya Parameter): Let $(T, V) \sim P_{T,V}(t, v)$ where $T \in \{0, 1\}$ and $V \in \mathcal{V}$ where $\mathcal{V}$ is an arbitrary discrete alphabet. Define a likelihood function $L(v)$ and inverse likelihood function $L^{-1}(v)$ as follows.

$$L(v) \triangleq \frac{P_{T|V}(0|v)}{P_{T|V}(1|v)}, \quad L^{-1}(v) \triangleq \frac{P_{T|V}(1|v)}{P_{T|V}(0|v)}$$

To account for degenerate cases in which $P_{T|V}(t|v) = 0$, define the following function,

$$\varphi(t, v) \triangleq \begin{cases} 0 & \text{if } \mathbb{1}\{P_{T|V}(t|v) = 0\} \\ L(v) & \text{if } \mathbb{1}\{P_{T|V}(t|v) > 0\} \text{ and } \mathbb{1}\{t = 1\} \\ L^{-1}(v) & \text{if } \mathbb{1}\{P_{T|V}(t|v) > 0\} \text{ and } \mathbb{1}\{t = 0\} \end{cases}$$

In order to estimate $Z(T|V) \in [0, 1]$, it is convenient to sample from $P_{TV}(t, v)$ and express $Z(T|V)$ as an expectation over random variables T and V,

$$Z(T|V) = \mathbb{E}_{T,V} \sqrt{\varphi(T, V)}. \tag{104}$$

Proof: The following forms of the Bhattacharyya parameter are equivalent.

$$Z(T|V) \triangleq 2 \sum_{v \in \mathcal{V}} P_V(v) \sqrt{P_{T|V}(0|v) P_{T|V}(1|v)}$$

$$= 2 \sum_{v \in \mathcal{V}} \sqrt{P_{TV}(0,v) P_{TV}(1,v)}$$

$$= \sum_{v \in \mathcal{V}} P_V(v) \sum_{t \in \{0,1\}} \sqrt{P_{T|V}(t|v)(1 - P_{T|V}(t|v))}$$

$$= \sum_{t \in \{0,1\}} \sum_{\substack{v: P_{T|V}(t|v) > 0 \\ v \in \mathcal{V}}} P_{TV}(t,v) \sqrt{\frac{1 - P_{T|V}(t|v)}{P_{T|V}(t|v)}}$$

$$= \mathbb{E}_{T,V} \sqrt{\varphi(T,V)}.$$

∎

Lemma 12 (Stochastic Degradation (cf. [24])): Consider discrete random variables V, Y_1, and Y_2. Assume that $|\mathcal{V}| = 2$ and that discrete alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ have an *arbitrary* size. Then

$$P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v) \Rightarrow Z(V|Y_2) \geq Z(V|Y_1). \tag{105}$$

Proof: Beginning with the definition of the Bhattacharyya parameter leads to the following derivation:

$$Z(V|Y_2) \triangleq 2 \sum_{y_2} \sqrt{P_{VY_2}(0,y_2) P_{VY_2}(1,y_2)}$$

$$= 2 \sum_{y_2} \sqrt{P_V(0) P_V(1)} \sqrt{P_{Y_2|V}(y_2|0) P_{Y_2|V}(y_2|1)}$$

$$= 2\sqrt{P_V(0) P_V(1)} \sum_{y_2} \left[\sqrt{\sum_{y_1} P_{Y_1|V}(y_1|0) \tilde{P}_{Y_2|Y_1}(y_2|y_1)} \cdot \sqrt{\sum_{y_1} P_{Y_1|V}(y_1|1) \tilde{P}_{Y_2|Y_1}(y_2|y_1)} \right].$$

Then applying the Cauchy–Schwarz inequality yields

$$Z(V|Y_2) \geq 2\sqrt{P_V(0) P_V(1)} \sum_{y_2} \left[\sum_{y_1} \sqrt{P_{Y_1|V}(y_1|0) \tilde{P}_{Y_2|Y_1}(y_2|y_1)} \cdot \sqrt{P_{Y_1|V}(y_1|1) \tilde{P}_{Y_2|Y_1}(y_2|y_1)} \right]$$

$$= 2\sqrt{P_V(0) P_V(1)} \sum_{y_2} \left[\sum_{y_1} \tilde{P}_{Y_2|Y_1}(y_2|y_1) \cdot \sqrt{P_{Y_1|V}(y_1|0) P_{Y_1|V}(y_1|1)} \right].$$

Interchanging the order of summations yields

$$Z(V|Y_2) \geq 2\sqrt{P_V(0) P_V(1)} \left[\sum_{y_1} \sqrt{P_{Y_1|V}(y_1|0) P_{Y_1|V}(y_1|1)} \cdot \sum_{y_2} \tilde{P}_{Y_2|Y_1}(y_2|y_1) \right]$$

$$= Z(V|Y_1).$$

∎

Lemma 13 (Successive Stochastic Degradation (cf. [24])): Consider a binary random variable V, and discrete random variables Y_1 with alphabet $\mathcal{Y}_1$, and Y_2 with alphabet $\mathcal{Y}_2$. Assume that the joint distribution $P_{VY_1Y_2}$ obeys the constraint $P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v)$. Consider two *i.i.d.* random copies (V^1, Y_1^1, Y_2^1) and (V^2, Y_1^2, Y_2^2) distributed according to $P_{VY_1Y_2}$. Define two binary random variables $U^1 \triangleq V^1 \oplus V^2$ and $U^2 \triangleq V^2$. Then the following holds

$$Z\left(U^1\middle|Y_2^{1:2}\right) \geq Z\left(U^1\middle|Y_1^{1:2}\right), \tag{106}$$

$$Z\left(U^2\middle|U^1, Y_2^{1:2}\right) \geq Z\left(U^2\middle|U^1, Y_1^{1:2}\right). \tag{107}$$

Proof: Given the assumptions, the following stochastic degradation conditions hold:

$$P_{Y_1^1|V^1}(y_1^1|v^1) \succeq P_{Y_2^1|V^1}(y_2^1|v^1), \tag{108}$$

$$P_{Y_1^2|V^2}(y_1^2|v^2) \succeq P_{Y_2^2|V^2}(y_2^2|v^2). \tag{109}$$

The goal is to derive new stochastic degradation conditions for the polarized conditional distributions. The binary random variables U^1 and U^2 are not necessarily independent Bernoulli($\frac{1}{2}$) variables. Taking this into account,

$$P_{Y_2^1Y_2^2|U^1}\left(y_2^1, y_2^2\middle|u^1\right) = \frac{1}{P_{U^1}(u^1)} \sum_{u^2 \in \{0,1\}} P_{V^1Y_2^1}\left(u^1 \oplus u^2, y_2^1\right) P_{V^2Y_2^2}\left(u^2, y_2^2\right)$$

$$= \frac{1}{P_{U^1}(u^1)} \sum_{u^2 \in \{0,1\}} \left[P_{Y_2^1|V^1}\left(y_2^1\middle|u^1 \oplus u^2\right) P_{V^1}\left(u^1 \oplus u^2\right) \cdot P_{Y_2^2|V^2}\left(y_2^2\middle|u^2\right) P_{V^2}(u^2) \right].$$

Applying the property due to the assumption in (108),

$$P_{Y_2^1Y_2^2|U^1}\left(y_2^1, y_2^2\middle|u^1\right) = \frac{1}{P_{U^1}(u^1)} \sum_{u^2 \in \{0,1\}} \left[P_{V^1}\left(u^1 \oplus u^2\right) P_{V^2}(u^2) \right.$$

$$\cdot \sum_{a \in \mathcal{Y}_1} P_{Y_1^1|V^1}\left(a\middle|u^1 \oplus u^2\right) \tilde{P}_{Y_2^1|Y_1^1}\left(y_2^1\middle|a\right)$$

$$\left. \cdot \sum_{b \in \mathcal{Y}_1} P_{Y_1^2|V^2}\left(b\middle|u^2\right) \tilde{P}_{Y_2^2|Y_1^2}\left(y_2^2\middle|b\right) \right].$$

Interchanging the order of summations and grouping the terms representing $P_{Y_1^1Y_1^2|U^1}\left(y_1^1, y_1^2\middle|u^1\right)$ yields the following

$$P_{Y_2^1Y_2^2|U^1}\left(y_2^1, y_2^2\middle|u^1\right) = \sum_{a \in \mathcal{Y}_1, b \in \mathcal{Y}_1} P_{Y_1^1Y_1^2|U^1}\left(a, b\middle|u^1\right) \tilde{P}_{Y_2^1|Y_1^1}\left(y_2^1\middle|a\right) \tilde{P}_{Y_2^2|Y_1^2}\left(y_2^2\middle|b\right).$$

The above derivation proves that

$$P_{Y_1^1Y_1^2|U^1}\left(y_1^1, y_1^2\middle|u^1\right) \succeq P_{Y_2^1Y_2^2|U^1}\left(y_2^1, y_2^2\middle|u^1\right).$$

Combined with Lemma 12, this concludes the proof for the ordering of the Bhattacharyya parameters given in (106).

In a similar way, it is possible to show that

$$P_{Y_2^1 Y_2^2 U^1 | U^2}\left(y_2^1, y_2^2, u^1 \middle| u^2\right) = \frac{1}{P_{U^2}(u^2)} P_{V^1 Y_2^1}\left(u^1 \oplus u^2, y_2^1\right) P_{V^2 Y_2^2}\left(u^2, y_2^2\right)$$

$$= \frac{1}{P_{U^2}(u^2)}\left[P_{Y_2^1 | V^1}\left(y_2^1 \middle| u^1 \oplus u^2\right) P_{V^1}\left(u^1 \oplus u^2\right) \cdot P_{Y_2^2 | V^2}\left(y_2^2 \middle| u^2\right) P_{V^2}(u^2)\right].$$

Applying the property due to the assumption in (109),

$$P_{Y_2^1 Y_2^2 U^1 | U^2}\left(y_2^1, y_2^2, u^1 \middle| u^2\right) = \frac{1}{P_{U^2}(u^2)}\left[P_{V^1}\left(u^1 \oplus u^2\right) P_{V^2}(u^2)\right.$$

$$\cdot \sum_{a \in \mathcal{Y}_1} P_{Y_1^1 | V^1}\left(a \middle| u^1 \oplus u^2\right) \tilde{P}_{Y_2^1 | Y_1^1}\left(y_2^1 \middle| a\right)$$

$$\left. \cdot \sum_{b \in \mathcal{Y}_1} P_{Y_1^2 | V^2}\left(b \middle| u^2\right) \tilde{P}_{Y_2^2 | Y_1^2}\left(y_2^2 \middle| b\right)\right].$$

Interchanging the order of the terms and grouping the terms representing $P_{Y_1^1 Y_1^2 U^1 | U^2}\left(y_1^1, y_1^2, u^1 \middle| u^2\right)$ yields the following

$$P_{Y_2^1 Y_2^2 U^1 | U^2}\left(y_2^1, y_2^2, u^1 \middle| u^2\right)$$

$$= \sum_{a \in \mathcal{Y}_1, b \in \mathcal{Y}_1}\left[P_{Y_1^1 Y_1^2 U^1 | U^2}\left(a, b, u^1 \middle| u^2\right) \tilde{P}_{Y_2^1 | Y_1^1}\left(y_2^1 \middle| a\right) \tilde{P}_{Y_2^2 | Y_1^2}\left(y_2^2 \middle| b\right)\right],$$

$$= \sum_{a \in \mathcal{Y}_1, b \in \mathcal{Y}_1, c \in \{0,1\}}\left[P_{Y_1^1 Y_1^2 U^1 | U^2}\left(a, b, c \middle| u^2\right) \tilde{P}_{Y_2^1 | Y_1^1}\left(y_2^1 \middle| a\right) \tilde{P}_{Y_2^2 | Y_1^2}\left(y_2^2 \middle| b\right) \mathbb{1}\left\{u^1 = c\right\}\right].$$

The above derivation proves that

$$P_{Y_1^1 Y_1^2 U^1 | U^2}\left(y_1^1, y_1^2, u^1 \middle| u^2\right) \succeq P_{Y_2^1 Y_2^2 U^1 | U^2}\left(y_2^1, y_2^2, u^1 \middle| u^2\right).$$

Combined with Lemma 12, this concludes the proof for the ordering of the Bhattacharyya parameters given in (107). ∎

Lemma 14 (Pinsker's Inequality): Consider two discrete probability measures $P(y)$ and $Q(y)$ for $y \in \mathcal{Y}$. The following inequality holds for a constant $\kappa \triangleq 2 \ln 2$,

$$\sum_{y \in \mathcal{Y}}\left|P(y) - Q(y)\right| \leq \sqrt{\kappa D\left(P \| Q\right)},$$

where $D(P \| Q)$ is the Kullback-Leibler divergence defined with logarithm $\log_2(\cdot)$.

Lemma 15 (Arıkan [8]): Consider two discrete random variables $X \in \{0,1\}$ and $Y \in \mathcal{Y}$. The Bhattacharyya parameter and conditional entropy are related as follows.

$$Z(X|Y)^2 \leq H(X|Y)$$

$$H(X|Y) \leq \log_2(1 + Z(X|Y))$$

Lemma 16 (Bhattacharyya vs. Entropy Parameters): Consider two discrete random variables $X \in \{0,1\}$ and $Y \in \mathcal{Y}$. For any $0 < \delta < \frac{1}{2}$,

$$Z(X|Y) \geq 1 - \delta \Rightarrow H(X|Y) \geq 1 - 2\delta.$$

$$Z(X|Y) \leq \delta \Rightarrow H(X|Y) \leq \frac{\delta}{\ln 2}.$$

Proof: Due to Lemma 15, $H(X|Y) \geq Z(X|Y)^2 \geq (1-\delta)^2 \geq 1 - 2\delta + \delta^2 \geq 1 - 2\delta$. It follows that if $Z(X|Y) \geq 1 - \delta$ and $\delta \to 0$, then $H(X|Y) \to 1$ as well. Similarly, due to Lemma 15, if $Z(X|Y) \leq \delta$ then $H(X|Y) \leq \log_2(1+\delta) = \left(\frac{1}{\ln 2}\right) \ln(1+\delta) \leq \frac{\delta}{\ln 2}$. To establish the last inequality, note that the function $f(x) \triangleq \ln(1+x)$ is concave for $x \geq 0$ and thus upper bounded by its tangent at point $(x, f(x)) = (0,0)$. It follows that if $Z(X|Y) \leq \delta$ and $\delta \to 0$, then $H(X|Y) \to 0$ as well. ∎

APPENDIX B

PROOF OF LEMMA 1

The total variation bound of Lemma 1 is decomposed in a simple way due to the chain rule for Kullback-Leibler distance between discrete probability measures. The joint probability measures P and Q were defined in (8) and (19) respectively. According to definition, if $P\big(\{u_i^{1:n}\}_{i \in [m]}\big) > 0$ then $Q\big(\{u_i^{1:n}\}_{i \in [m]}\big) >$

0. Therefore the Kullback-Leibler divergence $D(P\|Q)$ is well-defined and upper bounded as follows.

$$D\left(P_{\{U_i^{1:n}\}_{i\in[m]}}\middle\|Q_{\{U_i^{1:n}\}_{i\in[m]}}\right)$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}\left[D\left(P_{U_i(j)\middle|U_i^{1:j-1},\{U_k^{1:n}\}_{k\in[1:i-1]}}\middle\|Q_{U_i(j)\middle|U_i^{1:j-1},\{U_k^{1:n}\}_{k\in[1:i-1]}}\right)\right] \tag{110}$$

$$=\sum_{i=1}^{m}\sum_{j\in\mathcal{M}_i^{(n)}}\left[D\left(P_{U_i(j)\middle|U_i^{1:j-1},\{U_k^{1:n}\}_{k\in[1:i-1]}}\middle\|Q_{U_i(j)\middle|U_i^{1:j-1},\{U_k^{1:n}\}_{k\in[1:i-1]}}\right)\right] \tag{111}$$

$$=\sum_{i=1}^{m}\sum_{j\in\mathcal{M}_i^{(n)}}1-H\left(U_i(j)\middle|U_i^{1:j-1},\{U_k^{1:n}\}_{k\in[1:i-1]}\right) \tag{112}$$

$$=\sum_{i=1}^{m}\sum_{j\in\mathcal{M}_i^{(n)}}1-H\left(U_i(j)\middle|U_i^{1:j-1},\{Y_k^{1:n}\}_{k\in[1:i-1]}\right) \tag{113}$$

$$\leq\sum_{i=1}^{m}2\delta_n\left|\mathcal{M}_i^{(n)}\right|. \tag{114}$$

The equality in (110) is due to the chain rule for Kullback-Leibler distance. The equality in (111) is valid because for indices $j\notin\mathcal{M}_i^{(n)}$, $P\left(u_i(j)\middle|u_i^{1:j-1},\{u_k^{1:n}\}_{k\in[1:i-1]}\right)=Q\left(u_i(j)\middle|u_i^{1:j-1},\{u_k^{1:n}\}_{k\in[1:i-1]}\right)$. The equality in (112) is valid due to Lemma 10 and the fact that $Q\left(u_i(j)\middle|u_i^{1:j-1},\{u_k^{1:n}\}_{k\in[1:i-1]}\right)=\frac{1}{2}$ for indices $j\in\mathcal{M}_i^{(n)}$. The equality in (113) follows due to the one-to-one correspondence between variables $\{U_k^{1:n}\}_{k\in[1:i-1]}$ and $\{Y_k^{1:n}\}_{k\in[1:i-1]}$. The last inequality (114) follows from Lemma 16 due to the fact that $Z\left(U_i(j)\middle|U_i^{1:j-1},\{Y_k^{1:n}\}_{k\in[1:i-1]}\right)\geq1-\delta_n$ for indices $j\in\mathcal{M}_i^{(n)}$.

To finish the proof of Lemma 1,

$$\sum_{\{u_k^{1:n}\}_{k\in[m]}}\left|P(\{u_k^{1:n}\}_{k\in[m]})-Q(\{u_k^{1:n}\}_{k\in[m]})\right|$$

$$\leq\sqrt{\kappa D\left(P_{\{U_k^{1:n}\}_{k\in[m]}}\middle\|Q_{\{U_k^{1:n}\}_{k\in[m]}}\right)} \tag{115}$$

$$\leq\sqrt{\kappa\sum_{i=1}^{m}2\delta_n\left|\mathcal{M}_i^{(n)}\right|} \tag{116}$$

$$\leq\sqrt{(2\kappa)(m\cdot n)(2^{-n^{\beta'}})}.$$

The inequality in (115) is due to Pinsker's inequality given in Lemma 14. The inequality in (116) was proven in (114). Finally for $\beta'\in(\beta,\frac{1}{2})$, $\sqrt{(2\kappa)(m\cdot n)(2^{-n^{\beta'}})}<2^{-n^{\beta}}$ for sufficiently large n. Hence the total variation distance is bounded by $O(2^{-n^{\beta}})$ for any $0<\beta<\frac{1}{2}$.

APPENDIX C

SUPERPOSITION CODING

The total variation bound of Lemma 5 is decomposed in a simple way due to the chain rule for Kullback-Leibler distance between discrete probability measures. The joint probability measures P and Q were defined in (40) and (65) respectively. According to definition, if $P_{U_1^n U_2^n}(u_1^n, u_2^n) > 0$ then $Q(u_1^n, u_2^n) > 0$. Therefore the Kullback-Leibler divergence $D(P\|Q)$ is well-defined. Applying the chain rule,

$$
\begin{aligned}
D\left(P_{U_1^n U_2^n}\middle\|Q_{U_1^n U_2^n}\right) &= \sum_{j=1}^n D\left(P_{U_1(j)\middle|U_1^{1:j-1}}\middle\|Q_{U_1(j)\middle|U_1^{1:j-1}}\right) \\
&+ \sum_{j=1}^n D\left(P_{U_2(j)\middle|U_2^{1:j-1},U_1^{1:n}}\middle\|Q_{U_2(j)\middle|U_2^{1:j-1},U_1^{1:n}}\right) \\
&= \sum_{j\in\mathcal{M}_1^{(n)}} D\left(P_{U_1(j)\middle|U_1^{1:j-1}}\middle\|Q_{U_1(j)\middle|U_1^{1:j-1}}\right) \\
&+ \sum_{j\in\mathcal{M}_2^{(n)}} D\left(P_{U_2(j)\middle|U_2^{1:j-1},U_1^{1:n}}\middle\|Q_{U_2(j)\middle|U_2^{1:j-1},U_1^{1:n}}\right).
\end{aligned}
$$

Applying Lemma 10, the one-to-one relation between U_1^n and V^n, and Lemma 16 leads to the following result.

$$
\begin{aligned}
&D\left(P_{U_1^n U_2^n}\middle\|Q_{U_1^n U_2^n}\right) \\
&= \sum_{j\in\mathcal{M}_1^{(n)}} \left[1 - H\left(U_1(j)\middle|U_1^{1:j-1}\right)\right] + \sum_{j\in\mathcal{M}_2^{(n)}} \left[1 - H\left(U_2(j)\middle|U_2^{1:j-1}U_1^n\right)\right] \\
&= \sum_{j\in\mathcal{M}_1^{(n)}} \left[1 - H\left(U_1(j)\middle|U_1^{1:j-1}\right)\right] + \sum_{j\in\mathcal{M}_2^{(n)}} \left[1 - H\left(U_2(j)\middle|U_2^{1:j-1}V^n\right)\right] \\
&\leq 2\delta_n \left[\left|\mathcal{M}_1^{(n)}\right| + \left|\mathcal{M}_2^{(n)}\right|\right].
\end{aligned}
$$

Using identical arguments as applied in the proof of Lemma 1, the total variation distance between P and Q is bounded as $O(2^{-n^\beta})$.

To prove Lemma 6, the expectation of the average probability of error of the polarization-based

superposition code is written as

$$\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}]\right] =$$

$$\sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}}\left[P_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n | u_1^n, u_2^n\right) \right.$$

$$\cdot \frac{1}{2^{nR_2}} \prod_{j \in [n]: j \notin \mathcal{M}_2^{(n)}} \mathbb{P}\left\{\Psi_2^{(j)}\left(u_2^{1:j-1}\right) = u_2(j)\right\}$$

$$\left. \cdot \frac{1}{2^{nR_1}} \prod_{j \in [n]: j \notin \mathcal{M}_1^{(n)}} \mathbb{P}\left\{\Psi_1^{(j)}\left(u_1^{1:j-1}, u_2^n \mathbf{G}_n\right) = u_1(j)\right\}\right].$$

From the definitions of the random boolean functions $\Psi_1^{(j)}$ in (57) and $\Psi_2^{(j)}$ in (58), it follows that

$$\mathbb{P}\left\{\Psi_1^{(j)}\left(u_1^{1:j-1}, u_2^n \mathbf{G}_n\right) = u_1(j)\right\}$$

$$= \mathbb{P}\left\{U_1(j) = u_1(j) \middle| U_1^{1:j-1} = u_1^{1:j-1}, V^n = u_2^n \mathbf{G}_n\right\}$$

$$= \mathbb{P}\left\{U_1(j) = u_1(j) \middle| U_1^{1:j-1} = u_1^{1:j-1}, U_2^n = u_2^n\right\},$$

$$\mathbb{P}\left\{\Psi_2^{(j)}\left(u_2^{1:j-1}\right) = u_2(j)\right\}$$

$$= \mathbb{P}\left\{U_2(j) = u_2(j) \middle| U_2^{1:j-1} = u_2^{1:j-1}\right\}.$$

The expression for the expected average probability of error is then simplified by substituting the definition for $Q(u_1^n, u_2^n)$ provided in (65) as follows,

$$\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}]\right] = \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}}\left[P_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n | u_1^n, u_2^n\right) Q(u_1^n, u_2^n)\right].$$

The next step in the proof is to split the error term $\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}]\right]$ into *two* main parts, one part due to the error caused by polar decoding functions, and the other part due to the total variation

distance between probability measures.

$$\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}\}} \left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}\}] \right]$$

$$= \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} \left[P_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n \big| u_1^n, u_2^n\right) \right.$$

$$\left. \cdot \left(Q\left(u_1^n, u_2^n\right) - P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) + P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) \right) \right]$$

$$\leq \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} P_{U_1^n U_2^n Y_1^n Y_2^n}\left(u_1^n, u_2^n, y_1^n, y_2^n\right)$$

$$+ \sum_{\substack{u_1^n \in \{0,1\}^n \\ u_2^n \in \{0,1\}^n}} \left| P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) - Q\left(u_1^n, u_2^n\right) \right|. \tag{117}$$

Lemma 5 established that the error term due to the total variation distance is upper bounded as $O(2^{-n^\beta})$. Therefore, it remains to upper bound the error term due to the polar decoding functions. Towards this end, note first that $\mathcal{T} = \mathcal{T}_{1v} \cup \mathcal{T}_1 \cup \mathcal{T}_2$, $\mathcal{T}_{1v} = \cup_j \mathcal{T}_{1v}^j$ for $j \in \mathcal{M}_2^{(n)} \subseteq \mathcal{M}_{1v}^{(n)}$, $\mathcal{T}_1 = \cup_j \mathcal{T}_1^j$ for $j \in \mathcal{M}_1^{(n)}$, and $\mathcal{T}_2 = \cup_j \mathcal{T}_2^j$ for $j \in \mathcal{M}_2^{(n)}$. It is convenient to bound each type of error bit by bit successively at both decoder $\mathcal{D}_1$ and $\mathcal{D}_2$ as follows.

$$\mathcal{E}_{1v}^j \triangleq \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}_{1v}^j} P_{U_1^n U_2^n Y_1^n Y_2^n}\left(u_1^n, u_2^n, y_1^n, y_2^n\right)$$

$$= \sum_{(u_2^{1:j}, y_1^n) \in \{0,1\}^j \times \mathcal{Y}_1^n} P_{U_2^{1:j} Y_1^n}\left(u_2^{1:j}, y_1^n\right)$$

$$\cdot \mathbb{1}\left\{ P_{U_2^j | U_2^{1:j-1} Y_1^n}\left(u_2(j) \big| u_2^{1:j-1}, y_1^n\right) \leq \right.$$

$$\left. P_{U_2^j | U_2^{1:j-1} Y_1^n}\left(u_2(j) \oplus 1 \big| u_2^{1:j-1}, y_1^n\right) \right\}.$$

In this form, it is possible to upper bound the error term $\mathcal{E}_{1v}^j$ with the corresponding Bhattacharyya

parameter as follows,

$$
\mathcal{E}_{1v}^j = \sum_{\substack{u_2^{1:j} \in \{0,1\}^j \\ y_1^n \in \mathcal{Y}_1^n}} P\big(u_2^{1:j-1}, y_1^n\big) P\big(u_2^j \big| u_2^{1:j-1}, y_1^n\big)
$$

$$
\cdot \mathbb{1} \Bigg\{ P_{U_2^j | U_2^{1:j-1} Y_1^n}\big(u_2(j) \big| u_2^{1:j-1}, y_1^n\big) \leq
$$

$$
P_{U_2^j | U_2^{1:j-1} Y_1^n}\big(u_2(j) \oplus 1 \big| u_2^{1:j-1}, y_1^n\big) \Bigg\},
$$

$$
\leq \sum_{\substack{u_2^{1:j} \in \{0,1\}^j \\ y_1^n \in \mathcal{Y}_1^n}} P\big(u_2^{1:j-1}, y_1^n\big) P\big(u_2^j \big| u_2^{1:j-1}, y_1^n\big)
$$

$$
\cdot \sqrt{\frac{P_{U_2^j | U_2^{1:j-1} Y_1^n}\big(u_2(j) \oplus 1 \big| u_2^{1:j-1}, y_1^n\big)}{P_{U_2^j | U_2^{1:j-1} Y_1^n}\big(u_2(j) \big| u_2^{1:j-1}, y_1^n\big)}}
$$

$$
= Z\big(U_2^j \big| U_2^{1:j-1}, Y_1^n\big).
$$

Using identical arguments, the following upper bounds apply for the individual bit-by-bit error terms caused by successive decoding at both $\mathcal{D}_1$ and $\mathcal{D}_2$.

$$
\mathcal{E}_{1v}^j \leq Z\big(U_2^j \big| U_2^{1:j-1}, Y_1^n\big), \tag{118}
$$

$$
\mathcal{E}_1^j \leq Z\big(U_1^j \big| U_1^{1:j-1}, V^n, Y_1^n\big), \tag{119}
$$

$$
\mathcal{E}_2^j \leq Z\big(U_2^j \big| Y_2^n\big). \tag{120}
$$

Therefore, the total error due to decoding at the receivers is upper bounded as

$$
\mathcal{E} \triangleq \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} P_{U_1^n U_2^n Y_1^n Y_2^n}\big(u_1^n, u_2^n, y_1^n, y_2^n\big)
$$

$$
\leq \sum_{j \in \mathcal{M}_2^{(n)} \subseteq \mathcal{M}_{1v}^{(n)}} Z\big(U_2^j \big| U_2^{1:j-1}, Y_1^n\big)
$$

$$
+ \sum_{j \in \mathcal{M}_1^{(n)}} Z\big(U_1^j \big| U_1^{1:j-1}, V^n, Y_1^n\big)
$$

$$
+ \sum_{j \in \mathcal{M}_2^{(n)}} Z\big(U_2^j \big| Y_2^n\big)
$$

$$
\leq \delta_n \Big[\big|\mathcal{M}_{1v}^{(n)}\big| + \big|\mathcal{M}_1^{(n)}\big| + \big|\mathcal{M}_2^{(n)}\big| \Big]
$$

$$
\leq 3n\delta_n.
$$

This concludes the proof demonstrating that the expected average probability of error is upper bounded as $O(2^{-n^\beta})$.

APPENDIX D

MARTON CODING

To prove Lemma 9, the expectation of the average probability of error of the polarization-based Marton code is written as

$$
\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}]\right] =
$$

$$
\sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} \Bigg[P^\phi_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n | u_1^n, u_2^n\right)
$$

$$
\cdot \frac{1}{2^{nR_1}} \prod_{j \in [n]: j \notin \mathcal{M}_1^{(n)}} \mathbb{P}\left\{\Psi_1^{(j)}\left(u_1^{1:j-1}\right) = u_1(j)\right\}
$$

$$
\cdot \frac{1}{2^{nR_2}} \prod_{j \in \mathcal{H}_{V_2|V_1}^{(n)} \setminus \mathcal{M}_2^{(n)}} \mathbb{P}\left\{\Gamma(j) = u_2(j)\right\}
$$

$$
\cdot \prod_{j \in [n]: j \notin \mathcal{H}_{V_2|V_1}^{(n)}} \mathbb{P}\left\{\Psi_2^{(j)}\left(u_2^{1:j-1}, u_1^n \mathbf{G}_n\right) = u_2(j)\right\} \Bigg].
$$

The expression is then simplified by substituting the definition of $Q(u_1^n, u_2^n)$ provided in (97), and then splitting the error term into two parts:

$$
\mathbb{E}_{\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}}\left[P_e^{(n)}[\{\Psi_1^{(j)}, \Psi_2^{(j)}, \Gamma\}]\right] =
$$

$$
\sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} P^\phi_{Y_1^n Y_2^n | U_1^n U_2^n}\left(y_1^n, y_2^n | u_1^n, u_2^n\right) Q(u_1^n, u_2^n)
$$

$$
\leq \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} P_{U_1^n U_2^n Y_1^n Y_2^n}\left(u_1^n, u_2^n, y_1^n, y_2^n\right)
$$

$$
+ \sum_{\substack{u_1^n \in \{0,1\}^n \\ u_2^n \in \{0,1\}^n}} \left| P_{U_1^n U_2^n}\left(u_1^n, u_2^n\right) - Q\left(u_1^n, u_2^n\right) \right|.
$$

The error term pertaining to the total variation distance was already upper bounded as in Lemma 8. The error due to successive cancelation decoding at the receivers is upper bounded as follows.

$$
\begin{aligned}
\mathcal{E} &\triangleq \sum_{\{u_1^n, u_2^n, y_1^n, y_2^n\} \in \mathcal{T}} P_{U_1^n U_2^n Y_1^n Y_2^n}\left(u_1^n, u_2^n, y_1^n, y_2^n\right) \\
&\leq \sum_{j \in \mathcal{M}_1^{(n)}} Z\left(U_1^j \big| U_1^{1:j-1}, Y_1^n\right) + \sum_{j \in \mathcal{L}_{V_2|Y_2}^{(n)}} Z\left(U_2^j \big| U_2^{1:j-1}, Y_2^n\right), \\
&\leq \delta_n \left[\left|\mathcal{M}_1^{(n)}\right| + \left|\mathcal{L}_{V_2|Y_2}^{(n)}\right|\right] \\
&\leq 2n\delta_n.
\end{aligned}
$$

This concludes the proof demonstrating that the expectation of the average probability of block error is upper bounded as $O(2^{-n^\beta})$.

APPENDIX E

PROOF OF LEMMA 3

The implication in (30) follows since $X - Y_1 - Y_2$ means that $P_{Y_2|X}(y_2|x) = \sum_{y_1} P_{Y_1|X}(y_1|x) P_{Y_2|Y_1}(y_2|y_1)$. The implication in (31) follows by observing that

$$
\begin{aligned}
P_{Y_2|V}(y_2|v) &= \sum_{y_1 \in \mathcal{Y}_1} P_{Y_1 Y_2|V}(y_1, y_2|v) \\
&= \sum_{x \in \mathcal{X}} \sum_{y_1 \in \mathcal{Y}_1} P_{X|V}(x|v) P_{Y_1 Y_2|X}(y_1, y_2|x) \\
&= \sum_{x \in \mathcal{X}} P_{X|V}(x|v) \sum_{y_1 \in \mathcal{Y}_1} P_{Y_1 Y_2|X}(y_1, y_2|x) \\
&= \sum_{x \in \mathcal{X}} P_{X|V}(x|v) P_{Y_2|X}(y_2|x) \\
&= \sum_{x \in \mathcal{X}} P_{X|V}(x|v) \sum_{y_1 \in \mathcal{Y}_1} P_{Y_1|X}(y_1|x) \tilde{P}_{Y_2|Y_1}(y_2|y_1) \qquad (121) \\
&= \sum_{y_1 \in \mathcal{Y}_1} \sum_{x \in \mathcal{X}} P_{X|V}(x|v) P_{Y_1|X}(y_1|x) \tilde{P}_{Y_2|Y_1}(y_2|y_1) \\
&= \sum_{y_1 \in \mathcal{Y}_1} P_{Y_1|V}(y_1|v) \tilde{P}_{Y_2|Y_1}(y_2|y_1).
\end{aligned}
$$

In step (121), the assumed stochastic degraded condition $P_{Y_1|X}(y_1|x) \succeq P_{Y_2|X}(y_2|x)$ ensures the existence of the distribution $\tilde{P}_{Y_2|Y_1}(y_2|y_1)$. The converse to (31) follows since it is possible to select $P_{X|V}(x|v) =$

$\mathbb{1}\{x = v\}$ where the alphabet $\mathcal{V} = \mathcal{X}$. In this case, for any $v \in \mathcal{X}$,

$$P_{Y_2|V}(y_2|v) = \sum_{x \in \mathcal{X}} P_{X|V}(x|v)P_{Y_2|X}(y_2|x)$$

$$= \sum_{x \in \mathcal{X}} \mathbb{1}\{x = v\}P_{Y_2|X}(y_2|x)$$

$$= P_{Y_2|X}(y_2|v).$$

Similarly, $P_{Y_1|V}(y_1|v) = P_{Y_1|X}(y_1|v)$ for any $v \in \mathcal{X}$. Due to the assumed stochastic degradedness condition $P_{Y_2|V}(y_2|v) = \sum_{y_1} P_{Y_1|V}(y_1|v)\tilde{P}_{Y_2|Y_1}(y_2|y_1)$, for any $v \in \mathcal{X}$,

$$P_{Y_2|X}(y_2|v) = P_{Y_2|V}(y_2|v)$$

$$= \sum_{y_1} P_{Y_1|V}(y_1|v)\tilde{P}_{Y_2|Y_1}(y_2|y_1)$$

$$= \sum_{y_1} P_{Y_1|X}(y_1|v)\tilde{P}_{Y_2|Y_1}(y_2|y_1).$$

Therefore the stochastic degradedness property $P_{Y_1|X}(y_1|x) \succeq P_{Y_2|X}(y_2|x)$ must hold as well. The statement of (31) means that Class I and Class II are equivalent as shown in Figure 5. The implication in (32) follows because assuming the stochastic degradedness property $P_{Y_1|V}(y_1|v) \succeq P_{Y_2|V}(y_2|v)$ holds for all $P_{X|V}(x|v)$, there exists a $\tilde{Y}_1$ such that $V - \tilde{Y}_1 - Y_2$ forms a Markov chain and $P_{\tilde{Y}_1|V}(\tilde{y}_1|v) = P_{Y_1|V}(\tilde{y}_1|v)$ for all $P_{X|V}(x|v)$. By the data processing inequality, $I(V; \tilde{Y}_1) \geq I(V; Y_2)$. If $P_{\tilde{Y}_1|V}(\tilde{y}_1|v) = P_{Y_1|V}(\tilde{y}_1|v)$, then $P_{V\tilde{Y}_1}(v, \tilde{y}_1) = P_{VY_1}(v, \tilde{y}_1)$ for all $P_V(v)$. It follows that for all $P_{VX}(v, x)$, the mutual information $I(V; \tilde{Y}_1) = I(V; Y_1)$. The implication in (33) follows by setting $P_{VX}(v, x) = \mathbb{1}\{v = x\}P_X(x)$ and letting $\mathcal{V} = \mathcal{X}$. Then for any $v \in \mathcal{X}$,

$$P_{VY_1}(v, y_1) = \sum_{x \in \mathcal{X}} P_{VX}(v, x)P_{Y_1|X}(y_1|x)$$

$$= \sum_{x \in \mathcal{X}} \mathbb{1}\{v = x\}P_X(x)P_{Y_1|X}(y_1|x)$$

$$= P_X(v)P_{Y_1|X}(y_1|v)$$

$$= P_{XY_1}(v, y_1).$$

Similarly for any $v \in \mathcal{X}$, $P_{VY_2}(v, y_2) = P_{XY_2}(v, y_2)$. Therefore for the particular choice of $P_{VX}(v, x) = \mathbb{1}\{v = x\}P_X(x)$, $I(V; Y_1) = I(X; Y_1)$ and $I(V; Y_2) = I(X; Y_2)$. The converse statements for (30), (32), and (33) do *not* hold due to a counterexample involving a DM-BC comprised of a binary erasure channel BEC(ϵ) and a binary symmetric channel BSC(p) as described in Example 3.

REFERENCES

[1] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, January 1972.

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

[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 306–311, May 1979.

[4] C. Nair and A. E. Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Transactions on Information Theory*, vol. 53, pp. 350–355, January 2007.

[5] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, pp. 2524–2530, October 1998.

[6] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, pp. 3051–3073, July 2009.

[7] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for the deterministic broadcast channel," in *Proc. International Zurich Seminar on Communications*, (Zurich, Switzerland), pp. 51–54, February 2012.

[8] E. Arıkan, "Source polarization," in *Proc. of the IEEE International Symposium on Information Theory*, June 2010.

[9] E. Arıkan, "Polar coding for the slepian-wolf problem based on monotone chain rules," in *Proc. of the International Symposium on Information Theory*, July 2012.

[10] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Transactions on Information Theory*, vol. 58, pp. 5437 –5448, August 2012.

[11] W. Yu and M. Aleksic, "Coding for the blackwell channel: A survey propagation approach," in *Proc. of the IEEE International Symposium on Information Theory*, (Adelaide, Australia), September 2005.

[12] T. P. Coleman, E. Martinian, M. Effros, and M. Medard, "Interference management via capacity-achieving codes for the deterministic broadcast channel," in *Proc. IEEE Information Theory Workshop*, pp. 23–27, September 2005.

[13] A. Braunstein, F. Kayhan, G. Montorsi, and R. Zecchina, "Encoding for the blackwell channel with reinforced belief propagation," in *Proc. of the IEEE International Symposium on Information Theory*, (Nice, France), 2007.

[14] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, pp. 197–207, Mar 1973.

[15] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.

[16] S. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3187–3195, 2010.

[17] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. of the IEEE International Symposium on Information Theory*, (Seoul, South Korea), July 2009.

[18] E. Hof, I. Sason, and S. Shamai, "Polar coding for reliable communications over parallel channels," *CoRR*, vol. abs/1005.2770, 2010.

[19] M. Karzand, "Polar codes for degraded relay channels," in *Proc. International Zurich Seminar on Communications*, (Zurich, Switzerland), pp. 59–62, February 2012.

[20] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, "Polar codes for cooperative relaying," *IEEE Transactions on Communications*, vol. 60, pp. 3263–3273, November 2012.

[21] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Communications Letters*, vol. 14, pp. 752–754, August 2010.

[22] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Transactions on Information Theory*, vol. 57, pp. 6428–6443, October 2011.

[23] O. Koyluoglu and H. El Gamal, "Polar coding for secure transmission and key agreement," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 1472–1483, October 2012.

[24] S. B. Korada, *Polar Codes for Channel and Source Coding*. PhD thesis, EPFL, 2009.

[25] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, 2010.

[26] E. Abbe, "Randomness and dependencies extraction via polarization," in *Proc. of the Information Theory and Applications (ITA) Workshop*, (San Diego, California), February 2011.

[27] S. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions on Information Theory*, vol. 56, pp. 6253–6264, December 2010.

[28] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," *CoRR*, vol. abs/0908.0302, 2009.

[29] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," in *Proc. IEEE International Symposium on Information Theory*, pp. 894–898, June 2010.

[30] A. Sahebi and S. Pradhan, "Multilevel polarization of polar codes over arbitrary discrete memoryless channels," in *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1718–1725, September 2011.

[31] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Transactions on Information Theory*, vol. 59, pp. 955–969, February 2013.

[32] E. Şaşoğlu, "Polarization and polar codes," *Foundations and Trends in Comm. and Information Theory*, vol. 8, no. 4, pp. 259–381, 2012.

[33] K. Marton, "The capacity region of deterministic broadcast channels," in *Proc. of the IEEE International Symposium on Information Theory*, (Paris-Cachan), 1977.

[34] M. S. Pinsker, "Capacity of noiseless broadcast channels," *Probl. Inform. Transm.*, pp. 97–102, June 1978.

[35] T. P. Coleman, M. Effros, E. Martinian, and M. Medard, "Rate-splitting for the deterministic broadcast channel," in *Proc. of the IEEE International Symposium on Information Theory*, (Adelaide, Australia), September 2005.

[36] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, 2013.

[37] I. Tal and A. Vardy, "How to construct polar codes," *CoRR*, vol. abs/1105.6164, 2011.

[38] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Transactions on Information Theory*, vol. 56, pp. 4207–4214, September 2010.

[39] L. Wang, E. Şaşoğlu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. of the IEEE International Symposium on Information Theory*, pp. 2970–2974, IEEE, 2013.

[40] A. Gohari and V. Anantharam, "Evaluation of marton's inner bound for the general broadcast channel," *IEEE Transactions on Information Theory*, vol. 58, pp. 608–619, February 2012.

[41] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving the superposition and binning regions for broadcast channels using polar codes," *CoRR*, vol. abs/1401.6060, 2014.