A Lower Bound on the Probability of Error of Polar Codes over BMS Channels

Boaz Shuval, Ido Tal
Department of Electrical Engineering,
Technion, Haifa 32000, Israel.
Email: {bshuval@campus, idotal@ee}.technion.ac.il

***Abstract*—Consider a polar code designed for some binary memoryless symmetric channel. We develop a lower bound on the probability of error of this polar code under successive-cancellation decoding. The bound exploits the correlation between the various codeword bits and improves upon existing lower bounds.**

***Index Terms*—Channel polarization, channel upgrading, lower bounds, polar codes, probability of error.**

I. Introduction

Polar codes [1] achieve capacity on binary memoryless symmetric (BMS) channels with low-complexity construction, encoding, and decoding algorithms. Their error probability is given by a union of correlated events. The union bound, which ignores this correlation, is used to upper-bound the error probability. In this work, we exploit this correlation to develop a corresponding lower bound for any BMS channel.

The polar construction iteratively transforms $N = 2^n$ identical and independent channel uses into a set of correlated synthetic channels. Synthetic channel W_i has input u_i and output y_1^N, u_1^{i-1}, and assumes that the input bits of future channels are uniform. As $N \to \infty$ the synthetic channels *polarize* into "good" (almost noiseless) and "bad" (almost pure noise) channels. By determining which synthetic channels are "good" and which are "bad", one designs a polar code. Information is transmitted over the "good" synthetic channels and predetermined values over the "bad" synthetic channels. Since their values are predetermined, we call the "bad" synthetic channels *frozen*.

Decoding is accomplished via the successive-cancellation (SC) decoder, which decodes the synthetic channels in succession using previous bit decisions. Decision for a synthetic channel is either based on its likelihood or, if it is frozen, on its predetermined value. Previous bit decisions are assumed to be correct for non-frozen synthetic channels.

A *genie-aided SC decoder* has access to true values of previous input bits. The performance of polar codes under either SC or genie-aided SC decoding is identical [2, Lemma 1]. Henceforth, we assume a genie-aided SC decoder. Its probability of error is given by $\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\}$, where $\mathcal{E}_i$ denotes the event that channel W_i errs given that a genie had revealed to it the true previous bits.

The events $\mathcal{E}_i$ are correlated. Using the union bound, $\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\} \leq \sum_{i\in\mathcal{A}} \mathbb{P}\left\{\mathcal{E}_i\right\}$, Arıkan showed that polar codes achieve capacity [1]. To assess the tightness of this upper bound, we develop a lower bound on $\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\}$. A trivial lower bound is $\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\} \geq \max_{i\in\mathcal{A}} \mathbb{P}\left\{\mathcal{E}_i\right\}$. Tighter lower bounds may be obtained by considering pairs of error events: $\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\} \geq \max_{i,j\in\mathcal{A}} \mathbb{P}\left\{\mathcal{E}_i \cup \mathcal{E}_j\right\}$. A further improvement combines multiple pairs of error events [3]:

$$\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}} \mathcal{E}_i\right\} \geq \sum_{i\in\mathcal{A}} \mathbb{P}\left\{\mathcal{E}_i\right\} - \sum_{i,j\in\mathcal{A},\, i<j} \mathbb{P}\left\{\mathcal{E}_i \cap \mathcal{E}_j\right\}, \qquad (1)$$

which can also be cast in terms of unions of error events using $\mathbb{P}\left\{\mathcal{E}_i \cap \mathcal{E}_j\right\} = \mathbb{P}\left\{\mathcal{E}_i\right\} + \mathbb{P}\left\{\mathcal{E}_j\right\} - \mathbb{P}\left\{\mathcal{E}_i \cup \mathcal{E}_j\right\}$.

Computing probabilities of unions of error events requires the joint distribution of two synthetic channels. The size of the joint distribution's output alphabet is the product of each synthetic channel's alphabet size. A side effect of polarization is an exponential increase in output alphabet size, rendering the joint distributions infeasible to store. One remedy is to approximate the joint distribution.

Previous attempts at a lower bound [2], [4] were also based on (1). In [2], a density evolution approach was proposed. Due to increasing alphabet size, practical implementation of density evolution must involve quantization [5, Appendix B]. The probability of error derived from quantized joint distributions approximates, but does not generally bound, the real probability of error (except for the BEC, for which, as noted and analyzed in [2], no quantization is needed). In [4], the focus was the BEC. By tracking the joint probability of erasure the authors were able to show that the union bound is asymptotically tight for a BEC.

In this work, we develop an algorithm to compute lower bounds on the joint probability of error of two synthetic channels. Our technique applies to synthetic channels that are polar descendants of any BMS channel. Using (1), we lower-bound the probability of error of polar codes. For the BEC, our bounds recover the results of [2] and [4].

Our method is based on approximating the joint distribution by a stochastically upgraded joint distribution with a smaller output alphabet. However, key ideas that hold for a single channel no longer apply to joint distributions. For example, a degrading operation on a joint distribution may *improve* the performance of an SC decoder. Therefore, we develop methods that in one sense decouple the two synthetic channels yet in another sense couple them even further.

Due to space limitations, this paper contains an outline of the algorithm and statement of the results. A detailed presentation, complete with proofs, appears in the full version of the paper, available online [6].

A. Notation

We denote by $y_j^k = y_j, y_{j+1}, \ldots, y_k$ for $j < k$. For a logical expression `expr`, $[\![\texttt{expr}]\!]$ is 0 whenever `expr` is not true and is 1 otherwise.

II. Preliminaries

A. Degradation and Upgradation

Channel $W(y|u)$ is (stochastically) degraded with respect to $Q(z|u)$, denoted $W \preccurlyeq Q$, when there exists channel $P(y|z)$ such that $W(y|u) = \sum_z P(y|z)Q(z|u)$. If W is degraded with respect to Q, then Q is upgraded with respect to W. Degradation implies error probability ordering under optimal decoding [5, Chapter 4]: if $W \preccurlyeq Q$ then $P_e(W) \geq P_e(Q)$.

The output alphabets of Q and W may differ. In [7], methods of upgrading a BMS channel and reducing its output alphabet were introduced. These methods *do not* apply to joint distributions.

B. Joint Distribution of Two Synthetic Channels

Let W be a BMS channel that undergoes n polarization steps. Let a and b, $b > a$, be indices of two synthetic channels, $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$, respectively, where $y_a = (y_1^N, u_1^{a-1})$, $y_b = (y_1^N, u_1^{b-1})$, and $N = 2^n$. We respectively call the channels the *a-channel* and the *b-channel*. Their joint distribution is $W_{a,b}(y_a, y_b|u_a, u_b)$. With probability 1, the prefix of y_b is (y_a, u_a). Namely, y_b has the form $y_b = ((y_1^N, u_1^{a-1}), u_a, u_{a+1}^{b-1}) \equiv (y_a, u_a, y_r)$, where y_r denotes the remainder of y_b after removing y_a and u_a. Thus,

$$W_{a,b}(y_a, y_b|u_a, u_b) = 2W_b(y_b|u_b)\,[\![y_b = (y_a, u_a, y_r)]\!]\,. \quad (2)$$

The factor 2 stems from the uniform distribution of u_a. With some abuse of notation, we shall write

$$W_{a,b}(y_a, y_b|u_a, u_b) = W_{a,b}(y_a, u_a, y_r|u_a, u_b).$$

Observe from (2) that $W_b(y_a, u_a, y_r|u_b)$ is the joint distribution $W_{a,b}$ up to a constant factor. Indeed, we will use $W_b(y_a, u_a, y_r|u_b)$ to denote the joint channel where convenient.

III. Decoding Two Dependent Synthetic Channels

Let $W_a(y_a|u_a), W_b(y_b|u_b)$ be two jointly polar synthetic channels with joint distribution $W_{a,b}(y_a, y_b|u_a, u_b)$. The SC decoder for these channels makes a maximum-likelihood (ML) decision separately for each marginal. We call this decoder an *Individual Maximum Likelihood* (IML) decoder. A different decoder is the *Individual Minimum Joint* P_e (IMJP) decoder, which seeks decoders $\hat{u}_a = \phi_a(y_a), \hat{u}_b = \phi_b(y_b)$ that minimize the joint probability that at least one of the decoded bits is in error. We denote the probabilities of error of these decoders by $P_e^{\text{IML}}(W_{a,b})$ and $P_e^{\text{IMJP}}(W_{a,b})$, respectively.

The performance of the IMJP decoder by definition lower-bounds that of the SC decoder, and the decoders do not

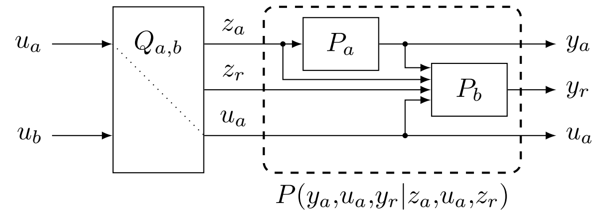

Fig. 1. The structure of proper degrading channels.

coincide in general. Since the IML decoder is suboptimal, its probability of error may, in fact, *decrease* after degradation. We demonstrate this in the following example.

Example 1. Let W be a BSC with crossover probability 0.4. We perform 2 polarization steps and consider joint channel $W_{1,4}$, i.e., $W_a = W^{--}$ and $W_b = W^{++}$. We have $0.6544 = P_e^{\text{IMJP}}(W_{1,4}) < P_e^{\text{IML}}(W_{1,4}) = 0.6976$. We degrade the a-channel to a useless channel by merging together all a-channel symbols. This results in degraded joint channel $W'_{1,4}$, for which the IML and IMJP decoder coincide, yielding $P_e^{\text{IMJP}}(W'_{1,4}) = P_e^{\text{IML}}(W'_{1,4}) = 0.6760$. Hence, for the degraded channel, the IML decoding error *decreases*.

Finding the IMJP decoder generally requires an exhaustive search. For polar codes, thanks to the successive structure of joint synthetic channels (2), we can explicitly find it.

Theorem 1. *Let $W_a(y_a|u_a)$ and $W_b(y_b|u_b)$ be two jointly polar synthetic channels. The IMJP decoder is given by setting ϕ_b as an ML decoder for W_b and ϕ_a according to*

$$\phi_a(y_a) = \arg\max_{u_a} \sum_{y_r} \max_{u_b} W_b(y_a, u_a, y_r|u_b).$$

A general degrading channel does not necessarily preserve the successive structure (2). We now define a subset of degrading channels that, by construction, preserves this structure.

Definition 1. A *proper* degrading channel has the form

$$P(y_a, u_a, y_r|z_a, u_a, z_r) = P_a(y_a|z_a)P_b(y_r|z_a, u_a, z_r, y_a),$$

illustrated in Figure 1. We write $Q \overset{p}{\succcurlyeq} W$ to denote that channel Q is upgraded from W with a proper degrading channel. An upgrading (degrading) procedure is proper if its degrading channel is proper.

Unlike the IML decoder, the probability of error of the IMJP decoder is guaranteed not to decrease after proper degradation. Intuitively, this is because the decoder for the original channel can simulate the degraded channel.

Lemma 2. *If $Q_{a,b} \overset{p}{\succcurlyeq} W_{a,b}$, then $P_e^{\text{IMJP}}(Q_{a,b}) \leq P_e^{\text{IMJP}}(W_{a,b})$.*

The SC probability of error of a polar code with non-frozen set $\mathcal{A}$ used on BMS channel W is $P_e^{\text{SC}}(W) = \mathbb{P}\left\{\bigcup_{a\in\mathcal{A}} \mathcal{E}_a^{\text{ML}}\right\}$, where $\mathcal{E}_a^{\text{ML}}$ is the error probability of synthetic channel W_a under ML decoding. The IMJP error probability lower-bounds the SC error probability.

Lemma 3.

$$P_e^{\mathrm{SC}}(W) \geq \max_{a,b\in\mathcal{A}} P_e^{\mathrm{IMJP}}(W_{a,b}) \geq \max_{a\in\mathcal{A}} \mathbb{P}\left\{\mathcal{E}_a^{\mathrm{ML}}\right\}.$$

If two channels are ordered by degradation, so are their polar transforms [8, Lemma 4.7]. This readily extends to joint channels. For joint channel $W_{a,b}$ we denote its jointly polarized versions by W_{a^α,b^β}, where $\alpha,\beta \in \{-,+\}$ denote the type of transform the a-channel and b-channel undergo, respectively.

Lemma 4. *If $Q_{a,b} \overset{p}{\succcurlyeq} W_{a,b}$ then $Q_{a^\alpha,b^\beta} \overset{p}{\succcurlyeq} W_{a^\alpha,b^\beta}$.*

Lemmas 2, 3, and 4 are the key to our lower bound: if we perform a sequence of polarization and upgrading operations on a joint distribution, and compute the IMJP probability of error for the resultant joint distribution, we achieve a lower bound on SC decoding that is tighter than the trivial one.

IV. Representations of Joint Synthetic Channels

A. D-value Representation of Joint Synthetic Channels

Two channels W and W' are called *equivalent* if $W \succcurlyeq W'$ and $W' \succcurlyeq W$. We denote this by $W \equiv W'$.

Definition 2. Joint channel $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ is in D-value representation if

$$d_b = \frac{W_b(y_a,u_a,d_b|0) - W_b(y_a,u_a,d_b|1)}{W_b(y_a,u_a,d_b|0) + W_b(y_a,u_a,d_b|1)}.$$

The following lemma affords a more convenient description of the joint channel, in which, in line with the IMJP decoder, the b-channel's ML decision is immediately apparent. It also greatly simplifies the expressions that follow.

Lemma 5. *Channels $W_{a,b}(y_a, u_a, y_r|u_a, u_b)$ and $W_{a,b}(y_a, u_a, d_b|u_a, u_b)$ are equivalent and the degrading channels from one to the other are proper.*

We use the same notation $W_{a,b}$ for both the regular and the D-value representations of the joint channel due to their equivalence. As in Section II-B, we will use $W_b(y_a, u_a, d_b|u_b)$ to denote the joint synthetic channel distribution. Under D-value representation, proper degrading channels admit the form

$$P(y_a, u_a, d_b|z_a, u_a, z_b) = P_a(y_a|z_a)P_b(d_b|z_a, y_a, u_a, z_b).$$

B. Symmetrization

Let $W_b(y_a, u_a, d_b|u_b)$ be a joint synthetic channel descendant from BMS W. From the symmetry properties of polar synthetic channels [1, Proposition 13], we conclude that for every y_a, d_b there exists $y_a^{(\mathsf{b})}$ such that

$$W_b(y_a, u_a, d_b|u_b) = W_b(y_a^{(\mathsf{b})}, u_a, -d_b|\bar{u}_b). \tag{3}$$

In general $y_a^{(\mathsf{b})} \neq y_a$, so (3) marks a departure from symmetry for BMS channels, for which $W(d|u) = W(-d|\bar{u})$. Turning to marginals, we observe that y_a and $y_a^{(\mathsf{b})}$ have the same a-channel D-value. Thus, an SC decoder cannot distinguish between y_a and $y_a^{(\mathsf{b})}$ when making its decision for the a-channel. A similar conclusion can be shown to hold for the IMJP decoder. Thus, if the a-channel were told only whether its output was one of $\{y_a, y_a^{(\mathsf{b})}\}$, either the SC or IMJP decoder would make the same decision had it been told its output was, say, y_a. Consequently, either decoder's probability of error is unaffected by obscuring the a-channel output in this manner. Let $\mathring{y}_a \triangleq \{y_a, y_a^{(\mathsf{b})}\}$ and define a *symmetrized* version of the joint synthetic channel distribution, $\mathring{W}_{a,b}$, as follows.

$$\begin{aligned}\mathring{W}_{a,b}(\mathring{y}_a, u_a, d_b|u_a, u_b) &= W_{a,b}(y_a, u_a, d_b|u_a, u_b)\\ &\quad + W_{a,b}(y_a^{(\mathsf{b})}, u_a, d_b|u_a, u_b).\end{aligned}$$

The marginal synthetic channels $\mathring{W}_a$ and $\mathring{W}_b$ satisfy

$$\begin{aligned}\mathring{W}_a(\mathring{y}_a|u_a) &= \mathring{W}_a(\bar{\mathring{y}}_a|\bar{u}_a),\\ \mathring{W}_b(\mathring{y}_a, u_a, d_b|u_b) &= \mathring{W}_b(\bar{\mathring{y}}_a, \bar{u}_a, d_b|u_b)\\ &= \mathring{W}_b(\mathring{y}_a, u_a, -d_b|\bar{u}_b)\\ &= \mathring{W}_b(\bar{\mathring{y}}_a, \bar{u}_a, -d_b|\bar{u}_b).\end{aligned} \tag{4}$$

Definition 3. A joint distribution whose marginals satisfy (4) is called *symmetrized*.

A symmetrized joint distribution remains symmetrized upon polarization.

Clearly, $\mathring{W}_{a,b}$ is *degraded* with respect to $W_{a,b}$, exactly the opposite of our main thrust. Nevertheless, by the above, both channels have the same error probability under SC and IMJP decoding. This is preserved under polarization and upgrading.

Proposition 6. *Let $W_{a,b}$ be the joint distribution of two synthetic channels. If $\mathring{Q}_{a,b} \overset{p}{\succcurlyeq} \mathring{W}_{a,b}$ then $P_e^{\mathrm{SC}}(W) \geq P_e^{\mathrm{IMJP}}(\mathring{Q}_{a,b})$.*

Due to Proposition 6, we henceforth assume that joint channel $W_{a,b}$ is symmetrized, and dispense with the $\mathring{(\cdot)}$ symbol. Replacing the joint channel with its symmetrized version need only be performed once, at the first instance the two channels go through different polarization transforms.

The great utility of symmetrization is that given u_a, y_a becomes *independent* of u_b, yielding the following lemma.

Lemma 7. *Let $W_b(y_a, u_a, d_b|u_b)$ be a symmetrized joint distribution. It admits the decomposition*

$$W_b(y_a, u_a, d_b|u_b) = \frac{1}{2}W_a(y_a|u_a)W_2(d_b|u_b; y_a, u_a). \tag{5}$$

For any y_a, u_a, W_2 is a BMS channel with input u_b and output d_b, i.e., $W_2(d_b|u_b; y_a, u_a) = W_2(-d_b|\bar{u}_b; y_a, u_a)$.

We call (5) a *decoupling decomposition* for W_b. We obtain W_a by marginalizing W_b. Once we know W_b and W_a, we can obtain W_2. We assume that W_2 is a valid channel[1], so if $W_a(y_a|u_a) = 0$ we set W_2 to an arbitrary BMS channel.

V. Upgrading Procedures for Joint Synthetic Channels

We now introduce two proper upgrading procedures for joint synthetic channels. Each reduces the alphabet size of one marginal without changing the other. Joint channel $W_{a,b}$

[1]Whence our notation $W_2(d_b|u_b; y_a, u_a)$ (with a semicolon, as opposed to $W_2(d_b|y_a, u_a, d_b)$).

is assumed to be symmetrized and in D-value representation. We do not distinguish symmetrized channels with any special symbol.

A. Upgrading Channel W_a

Let symmetrized joint channel $W_b(y_a, u_a, d_b|u_b)$ admit decoupling decomposition (5). Let $Q_b(z_a, u_a, z_b|u_b)$ be another symmetrized joint channel with decoupling decomposition

$$Q_b(z_a, u_a, z_b|u_b) = \frac{1}{2} Q_a(z_a|u_a) Q_2(z_b|u_b; z_a, u_a). \quad (6)$$

Theorem 8. *Let W_b and Q_b be symmetrized joint distributions satisfying (5) and (6). Then, $Q_b \overset{p}{\succcurlyeq} W_b$ if*

1) *$Q_a \succcurlyeq W_a$ with degrading channel $P_a(y_a|z_a)$.*
2) *$Q_2 \succcurlyeq W_2$ for all u_a, y_a, z_a such that $P_a(y_a|z_a) > 0$.*

The meaning of the second item is that, for fixed z_a, u_a, BMS channel $Q_2(z_b|u_b; z_a, u_a)$ is upgraded from a set of BMS channels $\{W_2(d_b|u_b; y_a, u_a)\}_{y_a}$.

A naive way to upgrade the a-channel using Theorem 8 is to upgrade the marginal W_a to Q_a and then find channel Q_2 that satisfies the second item of Theorem 8. However, this approach results in a trivial bound [6, Section VI.A].

The *upgrade-couple* transform enables upgrading the a-channel without changing the b-channel. It splits each a-channel symbol to several classes, according to the possible b-channel outputs. Symbols in a class share the same W_2 channel, so confining upgrade operations to a class inherently satisfies the second condition of Theorem 8, circumventing changes to the b-channel.

Let channel W_b have $2B$ possible D-values, $\pm d_{b1}, \pm d_{b2}, \ldots, \pm d_{bB}$.[2] For each a-channel symbol y_a we define B^2 upgrade-couple symbols $y_a^{i,j}$, $i, j \in \{1, 2, \ldots B\}$. The new symbols *couple* the outputs of the a- and b-channels: for a-channel output $y_a^{i,j}$, if $u_a = 0$, the b-channel output can only be $\pm d_{bi}$; if $u_a = 1$, the b-channel output can only be $\pm d_{bj}$.

The upgrade-couple channel $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ is defined by $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) \triangleq W_b(y_a, u_a, d_b|u_b) \cdot S_{i,j}(y_a, u_a, d_b)$, where

$$S_{i,j}(y_a, u_a, d_b) = \begin{cases} \displaystyle\sum_{d=\pm d_{bj}} W_2(d|u_b; y_a, \bar{u}_a) & \begin{matrix} u_a = 0, \\ d_b = \pm d_{bi} \end{matrix} \\ \displaystyle\sum_{d=\pm d_{bi}} W_2(d|u_b; y_a, \bar{u}_a) & \begin{matrix} u_a = 1, \\ d_b = \pm d_{bj} \end{matrix} \\ 0 & \text{otherwise,} \end{cases}$$

and W_2 is from the decoupling decomposition of W_b in (5). Note that channel $\check{W}_b$ is symmetrized and admits decoupling decomposition $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b) = \frac{1}{2}\check{W}_a(y_a^{i,j}|u_a)\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$. It can be shown [6, Lemma 19] that for every y_a,

$$\check{W}_2(d_b|u_b; y_a^{i,j}, u_a) = \begin{cases} \text{BSC}\left(\frac{1-d_{bi}}{2}\right) & u_a = 0 \\ \text{BSC}\left(\frac{1-d_{bj}}{2}\right) & u_a = 1. \end{cases} \quad (7)$$

[2]We assume that erasure symbols are duplicated. I.e., there is a "positive" erasure and a "negative" erasure, see [7, Lemma 4].

Definition 4. The canonical channel $W^*(d|u)$ of channel $W(y|u)$ has a single entry for each D-value. I.e., denoting by D_d the set of symbols y whose D-value is d, we have $W^*(d|u) = \sum_{D_d} W(y|u)$.

It can be shown that the canonical a-channel and b-channel marginals of $\check{W}_b$ coincide with those of W_b.

Definition 5. The *class* $C_{i,j}$ is the set of symbols $y_a^{i,j}$ with fixed i, j.

There are B^2 classes. The size of each class is the number of symbols y_a. By (7), $\check{W}_2(d_b|u_b; y_a^{i,j}, u_a)$ is the *same* BSC for all symbols of class $C_{i,j}$ and fixed u_a. Thus, the second item of Theorem 8 is immediately satisfied if we confine upgrading procedures to the same class $C_{i,j}$. In [7], two upgrading procedures were introduced. *Ugprade-merge-2* merges two conjugate symbol pairs into a single conjugate symbol pair; *upgrade-merge-3* merges three conjugate symbol pairs into two conjugate symbol pairs. It turns out that a symbol and its conjugate belong to different classes. Since upgrade-merge-2 combines symbols and their conjugates, it cannot be confined to a single class. Upgrade-merge-3 does not suffer from this, so this is the upgrade-merge procedure we use.

Theorem 9. *Let $W_b(y_a, u_a, d_b|u_b)$ be some joint distribution with marginals $W_a(y_a|u_a), W_b^*(d_b|u_b)$ and upgrade-couple counterpart $\check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$. Let $Q_a(z_a|u_a) \succcurlyeq W_a(y_a|u_a)$ obtained by an upgrade-merge-3 procedure. Then there exists joint distribution $\check{Q}_b(z_a^{i,j}, u_a, d_b|u_b) \overset{p}{\succcurlyeq} \check{W}_b(y_a^{i,j}, u_a, d_b|u_b)$ with canonical marginals $\check{Q}_a^*(z_a|u_a), \check{Q}_b^*(d_b|u_b)$ such that $\check{Q}_a^* = Q_a^*$ and $\check{Q}_b^* = W_b^*$.*

To use Theorem 9, one begins with a design parameter A that controls the output alphabet size. Working one class at a time, one applies upgrade operations in succession to reduce the class size to $2A$. This results in an a-channel with $2AB^2$ symbols, whose canonical version has at most $2A$ symbols.

B. Upgrading Channel W_b

The following theorem shows how to upgrade $W_{a,b}$ to $Q_{a,b}$ such that marginal $Q_b \succcurlyeq W_b$ and marginal $Q_a = W_a$.

Theorem 10. *Let $W_b(y_a, u_a, d_b|u_b)$ with canonical b-channel marginal $W_b^*(d_b|u_b)$. Let $Q_b^*(z_b|u_b) \succcurlyeq W_b^*(d_b|u_b)$ with degrading channel $P_b^*(d_b|z_b)$. Then there exists joint channel $Q_b(y_a, u_a, z_b|u_b)$ such that $Q_b(y_a, u_a, z_b|u_b) \overset{p}{\succcurlyeq} W_b(y_a, u_a, z_b|u_b)$ and $\sum_{y_a, u_a} Q_b(y_a, u_a, z_b|u_b) = Q_b^*(z_b|u_b)$.*

Omitting details to conserve space, we only state that whenever $W_b^*(d_b) > 0$ the upgraded joint channel is given by

$$Q_b(y_a, u_a, z_b|u_b) = Q_b^*(z_b|u_b) \sum_{d_b} \frac{P_b^*(d_b|z_b) W_b(y_a, u_a, d_b)}{W_b^*(d_b)},$$

where $W_b^*(d_b) = \sum_{u_b} W_b^*(d_b|u_b)$ and $W_b(y_a, u_a, d_b) = \sum_{u_b} W_b(y_a, u_a, d_b|u_b)$.

To use Theorem 10, one begins with design parameter B that controls the output alphabet size. The channel Q_b^*, with output alphabet of size $2B$, is obtained from W_b^* using a sequence of upgrade operations. To obtain upgraded joint channel Q_b, one uses the theorem to turn them into a sequence of upgrade operations to be performed on channel W_b.

VI. Lower Bound Procedure

The input to our procedure is BMS channel W, number of polarization steps n, indices a and b of the a-channel and b-channel, respectively, and parameters A and B that control the alphabet sizes of the a- and b-channels, respectively. The binary expansions of $a-1$ and $b-1$ are $\mathbf{a} = \langle a_1, a_2, \ldots, a_m \rangle$ and $\mathbf{b} = \langle b_1, b_2, \ldots, b_m \rangle$, respectively; they specify the order of polarization transforms to be performed.

Algorithm A provides a high-level description of the procedure. We first determine the first index m for which a_m and b_m differ (i.e., $a_\ell = b_\ell$ for $\ell < m$ and $a_m \neq b_m$). The first $m-1$ polarization steps are of a single channel. Since these are single channels, we utilize the upgrading procedures of [7] to reduce the output alphabet size. At the mth polarization step, the a- and b-channels differ. We perform joint polarization and symmetrize the channel. This symmetrization need only be performed once as subsequent polarizations maintain symmetrization. We then perform the b-channel upgrading procedure, which reduces the b-channel alphabet size to $2B$. Following that, we upgrade the a-channel by first upgrade-coupling the channel to generate B^2 classes and then upgrading each class separately to reduce its size to $2A$ elements. We continue in this manner until $\ell = n$. Finally, we compute the probability of error of the IMJP decoder for the resulting channel.

Algorithm A: Lower bound on SC error probability

Input: BMS channel W, number of polarization steps n, channel indices a,b, and control parameters A, B.
Output: A lower bound on the probability of error $W_{a,b}$.
$m \leftarrow$ `first_difference`$(\mathbf{a}, \mathbf{b})$
$Q \leftarrow$ `single_upgrade`$(W, \max\{A, B\})$
for $\ell = 1, 2, \ldots, n$ **do**
  if $\ell < m$ **then**
    $Q \leftarrow$ `single_polarize`(Q, a_ℓ)
    $Q \leftarrow$ `D-Value_representation`(Q)
    $Q \leftarrow$ `single_upgrade`$(Q, \max\{A, B\})$
  else
    $Q \leftarrow$ `jointly_polarize`(Q, a_ℓ, b_ℓ)
    $Q \leftarrow$ `D-Value_representation`(Q)
    if $\ell = m$ **then**
      $Q \leftarrow$ `symmetrize`(Q)
    $Q \leftarrow$ `b-channel_upgrade`(Q, B)
    $Q \leftarrow$ `upgrade_couple`(Q)
    foreach class $\in Q$ **do**
      $Q \leftarrow$ `a-channel_upgrade`$(Q, A,$ class$)$
return $P_e^{\text{IMJP}}(Q)$

The lower bound of this procedure compares favorably with the trivial lower bound, $\max\{\mathbb{P}\{\mathcal{E}_a\}, \mathbb{P}\{\mathcal{E}_b\}\}$, since the upgrading procedure only ever changes one marginal. By leveraging single channel upgrading transforms, the marginal channels obtained are the same as would be obtained on single channels using the same upgrading steps. Thus, by Lemma 3 this lower bound is at least as good as $\max\{\mathbb{P}\{\mathcal{E}_a\}, \mathbb{P}\{\mathcal{E}_b\}\}$.

Remark 1. An initial step of Algorithm A is to upgrade the channel W. This step enables us to apply our algorithm on continuous-output channels, see [7, Section VI].

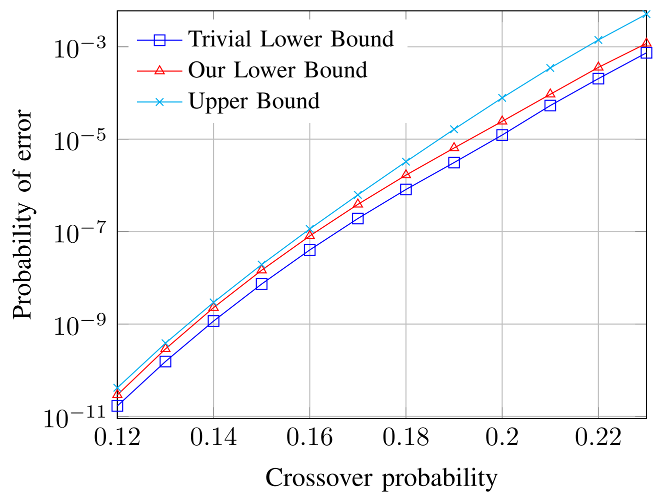

Fig. 2. Bounds on the probability of error of a rate 0.1 polar code of length 2^{10} designed for a BSC with crossover probability 0.2.

VII. Numerical Results

Figure 2 presents bounds on the error probability of a polar code of length $N = 2^{10}$ used over various BSCs. The code was designed for a BSC with crossover probability 0.2 using the techniques of [7]. The non-frozen set $\mathcal{A}$ consists of the 102 synthetic channels with smallest probability of error. The upper bound is an upper bound on $\sum_{a \in \mathcal{A}} P_e(W_a)$, and the trivial lower bound is a lower bound on $\max_{a \in \mathcal{A}} P_e(W_a)$; upper and lower bounds on $P_e(W_a)$ were obtained using the techniques of [7]. For our lower bound we used $2A = 32$ and $2B = 8$ for all possible pairs of the 10 worst channels in $\mathcal{A}$ and used (1) computed for the subset of these channels that yielded the highest bound.

References

[1] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
[2] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE Int. Symp. on Information Theory*, June 2009, pp. 1496–1500.
[3] F. M. Hoppe, "Iterating Bonferroni bounds," *Statistics & Probabiliy Letters*, vol. 3, no. 3, pp. 121–125, 1985.
[4] M. B. Parizi and E. Telatar, "On the correlation between polarized BECs," in *Proc. IEEE Int. Symp. on Information Theory*, July 2013, pp. 784–788.
[5] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.
[6] B. Shuval and I. Tal, "A lower bound on the probability of error of polar codes over bms channels," 2017. [Online]. Available: arXiv:1701.01628
[7] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. on Information Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.
[8] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédèrale de Lausanne, 2009.