A novel partial order for the information sets of polar codes over B-DMCs

Mine Alsan
Information Theory Laboratory
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: mine.alsan@epfl.ch

Abstract—We study partial orders on the information sets of polar codes designed for binary discrete memoryless channels. We show that the basic polarization transformations defined by Arıkan preserve ‘symmetric convex/concave orders’. While for symmetric channels this ordering turns out to be equivalent to the stochastic degradation ordering already known to order the information sets of polar codes, we show that a strictly weaker partial order is obtained when at least one of the channels is asymmetric. We also discuss two tools which can be useful for verifying ‘symmetric convex/concave ordering’: a criterion known as the cut criterion and channel symmetrization.

I. INTRODUCTION

The construction of the information set of a polar code for a given binary discrete memoryless channel (B-DMC) \( W \) is based on a very neat principle. First, independent copies of the channel \( W \) are combined and split by applying Arıkan’s basic polarization transformations \([1]\) in a recursive fashion: At step one the channels \( W^- \) and \( W^+ \) are synthesized, after, applying the polar transforms to these new channels, \( W^{--} := (W^-)^- \), \( W^{+-} := (W^-)^+ \) and \( W^{++} := (W^+)^- \), \( W^{+-} := (W^+)^+ \) are obtained, and so on. At stage \( n \), this procedure yields a set of \( N = 2^n \) channels \( \{ W^s : s \in \{+, -\}^n \} \). Then, the information set of a polar code of blocklength \( N \) for the channel \( W \), denoted as \( A_N(W) \), is specified by picking the channel indices from the set \( \{+, -\}^n \) which are good for uncoded transmission, i.e., \( A_N(W) = \{ s \in \{+, -\}^n : W^s \text{ is } \text{‘good’} \} \). (We used the qualifier “good” purposely at the expense of vagueness as later we will see that many quantifiers can be used as a replacement.)

Once the information set is constructed, the inputs of the bad channels are frozen to known values and uncoded data is only transmitted through the good channels. Arıkan \([1]\) shows that, in this way, by using an appropriate successive cancellation decoder knowing the channel, the information set, and the frozen values, codes whose rate \( |A_N(W)|/N \) approaches the channel’s symmetric capacity \( I(W) \) can be communicated reliably over the channel. We refer to \([1]\) for a more detailed account on polar codes.

A question of both theoretical and practical interest since the invention of polar codes has been: How large is \( A_N(W) \cap A_N(V) \), for two given channels \( W \) and \( V \), \( W \) and \( V \) have been pointed out in \([1]\) for the information sets of polar codes: Any binary erasure channel (BEC) provides good indices for all other B-DMCs having smaller Bhattacharyya parameters, and any channel which is degraded with respect to another B-DMC provides good indices for the upgraded channel. Another partial order was noticed in \([2, \text{Theorem 1}]\) showing that any symmetric B-DMC provides good indices for the BEC of the same variational distance.

In this work, we will show that all these three orderings can be studied in the context of stochastic orders known as convex ordering. We will introduce in Theorem II a new partial order for the information sets of polar codes over B-DMCs we refer as ‘symmetric convex/concave ordering’. Our main goal is to show that in general this is a strictly weaker condition than stochastic degradation between two B-DMCs. We will show that while the orders are equivalent when both channels are symmetric, the new relation is strictly weaker when at least one of the two channels is not symmetric. In particular, we will illustrate this by an example which studies both orderings between a Z-channel and a binary symmetric channel (BSC) whose inputs are used with equal frequency. Moreover, we will apply in the example two tools which can be useful for verifying ‘symmetric convex/concave ordering’: the cut criterion and channel symmetrization.

The rest of this paper is organized as follows. In the next section, we will give the necessary definitions and state the two main results in Theorem II and Corollary II. Subsequently in Section III some results related to convex ordering will be presented. In the final section, the theorems will be proved and the results will be explored.

II. DEFINITIONS AND STATEMENT OF THE RESULTS

Let \( W : \{0, 1\} \rightarrow Y \) be a B-DMC with transition probabilities denoted as \( W(y|0) \) and \( W(y|1) \), for all \( y \in Y \). We denote the normalized difference between these two transition probabilities by

\[
\Delta_W(y) = \frac{W(y|0) - W(y|1)}{W(y|0) + W(y|1)}.
\]

Assuming a uniform input distribution on the channel, \( \Delta_W(Y) \) defines a random variable taking values in \([-1, 1]\) with \( Y \) distributed according to \( q_W(y) = (W(y|0) + W(y|1))/2 \). The
variational distance between the two distributions $W(y|0)$ and $W(y|1)$ is then given by the expectation $\mathbb{E}[|\Delta W(y)|]$.

The basic polarization transformations synthesize two new binary input channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ from two independent copies of $W$. Their transition probabilities are given by \[ W^-(y_1 y_2 | u_1) = \sum_{u_2 \in F_2} \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2), \] \[ W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2). \]

Let $X$ be a random variable with distribution $F$ and $Y$ be a random variable with distribution $G$. \[\begin{align*}
\text{Definition 1:} & \quad X \text{ is smaller with respect to the increasing convex ordering than } Y, \text{ written } X \preceq_{icx} Y, \text{ if } \mathbb{E} [\phi(X)] \leq \mathbb{E} [\phi(Y)], \\
\text{Definition 2:} & \quad X \text{ is smaller with respect to the decreasing convex ordering than } Y, \text{ written } X \succeq_{dcx} Y, \text{ if } \mathbb{E} [\phi(X)] \geq \mathbb{E} [\phi(Y)].
\end{align*}\]

Theorem 1: Let $W$ and $V$ be two B-DMCs such that $|\Delta V| \preceq_{icx} |\Delta W|$ holds. Then, the basic polarization transformations preserve this ordering, i.e., $|\Delta V^+| \preceq_{icx} |\Delta W^+|$.

Remark 1: The ‘symmetric convex/concave ordering’ term we have been using so far actually refers to the ordering $|\Delta V| \preceq_{icx} |\Delta W|/|\Delta V| \gg_{dcx} |\Delta W|$. Proposition 2 in Section III will support our naming choice.

Proposition 1: \[\begin{align*}
\text{Definition 3:} & \quad \phi(\cdot) \text{ denotes a convex increasing function in the interval } [0,1] \text{ such that } \phi(0) = 0 \text{ and } \phi(1) = 1, \text{ and let } \epsilon \in (0,1). W \text{ is called } \text{’good’ for the channel parameter } \\
& \quad \mathbb{E} [\phi(\Delta W)] = \mathbb{E} \geq 1 - \epsilon \text{ holds. Accordingly, the information set definition is adapted as } \\
& \quad A_{\phi}^W(W) = \{ x \in \{+,-\}^N : \mathbb{E} \phi(\Delta W^+ \leq 1 - \epsilon \}. \]
\end{align*}\]

For instance, the particular choice of $\phi(x) = 1 - h(x)$, where $h(\cdot)$ denotes the binary entropy function, or $\phi(x) = 1 - \sqrt{1 - x^2}$, lead to information set definitions based on the values of the symmetric capacities and the Bhattacharyya parameters of the synthetic channels, respectively.

Using this refined definition, we get the following corollary to Theorem 1:
\[\begin{align*}
\text{Corollary 1:} & \quad \text{Let } W \text{ and } V \text{ be two B-DMCs such that } |\Delta V| \preceq_{icx} |\Delta W| \text{ holds. Then, for any } \epsilon \in (0,1) \text{ and any } \phi(\cdot) \text{ which is convex in the interval } [0,1], A_{\phi}^W(V) \subseteq A_{\phi}^W(W) \text{ holds for all } N = 2^n.
\end{align*}\]

Note that Korada [3] gave a proof for the fact that stochastic degradation is preserved under the basic polarization transformations in [3, Lemma 4.7], and also proved that degradation orders the Bhattacharyya parameters of the channels in [3, Lemma 1.8]. These show that, as stated by [1], the information sets of degraded channels are ordered for the particular choice of the function $\phi(x) = \sqrt{1 - x^2}$, whose expectation gives the Bhattacharyya parameter of the channel, i.e., $B(W) = \mathbb{E} [\sqrt{1 - |\Delta W|^2}]$.

Definition 4: $W$ is stochastically degraded with respect to $V$ if there exists a channel $P : \mathcal{Y} \to \mathcal{Y}$ such that \[ V(y|x) = \sum_{z \in \mathcal{Y}} W(z|x) P(y|z), \quad \text{for all } y \in \mathcal{Y}. \]

The final definition is of channel symmetrization.

Definition 5: [3, Definition 1.3] For any B-DMC $W : \mathcal{X} \to \mathcal{Y}$, the symmetrized B-DMC $W_s : \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$ is defined as \[ W_s(y, z|x) = 0.5W(y|x \oplus z) \]

### III. Convex Ordering

The material up to and including Theorem 1 is drawn from the book [4, Chapter 1, Section 1.3].

The following theorem discusses a special case of the increasing convex ordering.

Theorem 2: \[\begin{align*}
&\text{Theorem B} \quad \text{Suppose } X \text{ and } Y \text{ have equal mean values. } X \text{ is smaller with respect to the convex ordering than } Y, \text{ written } X \preceq_{cv} Y, \text{ if and only if } \\
&\mathbb{E} \phi(X) \leq \mathbb{E} \phi(Y), \quad \text{for all convex } \phi, \text{ for which the expectations exist.}
\end{align*}\]

An alternative description of the convex ordering due to Blackwell [5] is given in [4, Theorem C]. We will see later that this result establishes the equivalence between stochastic degradation and convex ordering.

Theorem 3: \[\begin{align*}
&\text{Suppose } T \text{ is a mean value preserving Markov kernel. Then, } X \preceq_{cv} Y \text{ if and only if } G = TF.
\end{align*}\]

In the next theorem a ‘simple’ criterion, known as the Karlin-Novikoff [6] cut criterion, for two random variables to satisfy increasing convex ordering is given.

Theorem 4: \[\begin{align*}
&\text{Theorem C} \quad \text{Suppose that for } X, Y \text{ with finite first moments } m_X, m_Y, \text{ respectively, we have } m_X \leq m_Y \text{ and } \\
&F(x) \leq G(x), \quad \text{for } x \leq \delta, \quad F(x) \geq G(x), \quad \text{for } x > \delta,
\end{align*}\]

for some $\delta \in \mathbb{R}$, then $X \preceq_{icx} Y$.

A random variable $X$ is called symmetric if the distribution of $X$ satisfies $F(x) = 1 - F(-x)$, for all $x \in \mathbb{R}$. A function $f(x)$ is called symmetric if $f(x) = f(-x)$, for all $x \in \mathbb{R}$. In the next two propositions, we exploit these symmetry properties.

Proposition 1: \[\begin{align*}
&\text{For symmetric } X \text{ and } Y, X \preceq_{icx} Y \text{ if and only if } |X| \preceq_{icx} |Y|.
\end{align*}\]

\textbf{Proof:} The ‘only if part’ follows by definition. So, we only need to prove the ‘if part’. Let $\phi(\cdot)$ be a function which is convex in $x \in \mathbb{R}$. As $X$ is symmetric, we can write \[ \mathbb{E} \phi(X) = \mathbb{E} \left[ \frac{\phi(x) + \phi(-x)}{2} \right] = \mathbb{E} \phi_s(X), \]

\[\text{where } \phi_s(\cdot) = \mathbb{E} \left[ \frac{\phi(x) + \phi(-x)}{2} \right]. \]
where 
\[ \phi_s(t) = \frac{\phi(t) + \phi(-t)}{2} \]  
(12)
is a convex symmetric function. In particular, \( \phi_s(.) \) is increasing on \( \mathbb{R}_+ \). As a result,
\[ \mathbb{E} [\phi(X)] = \mathbb{E} [\phi_s([X]]) \leq \mathbb{E} [\phi_s([Y])] = \mathbb{E} [\phi(Y)], \]  
(13)
where the inequality follows from \( |X| \prec_{icx} |Y| \).

By Definition 1, \( |X| \prec_{icx} |Y| \) holds if \( \mathbb{E}[\phi(|X|)] \leq \mathbb{E}[\phi(|Y|)] \), for all \( \phi \) convex in the interval \([0, \infty)\). Alternatively, this order can be described by only using the class of symmetric functions.

**Proposition 2:** \( X \) is smaller with respect to the ‘symmetric convex ordering’ than \( Y \), defined as \( |X| \prec_{icx} |Y| \) if
\[ \mathbb{E}[\phi_s(|X|)] \leq \mathbb{E}[\phi_s(|Y|)], \]  
(14)
for all convex symmetric functions \( \phi_s \), for which the expectations exist.

**Proof:** The proof follows by the fact that \( \phi_s(|x|) = \phi_s(x) \) holds for any symmetric function \( \phi_s(x) \).

We will use this characterization in the proof of Theorem 1.

IV. EXPLORATION

We start by proving Theorem 1.

**Proof of Theorem 1:** After applying the polar transformations to the channels, one can derive the following recursion
\[ \Delta_{W-}(Y_1Y_2) = \Delta_W(Y_1)\Delta_W(Y_2), \]  
(15)
\[ \Delta_{W+}(Y_1Y_2U_1) = \frac{\Delta_W(Y_1) + (-1)^y_1\Delta_W(Y_2)}{1 + (-1)^y_1\Delta_W(Y_1)\Delta_W(Y_2)}, \]  
(16)
where \( Y_1Y_2 \sim q_W(y_1)q_W(y_2) \), and
\[ Y_1Y_2U_1 \sim q_W(y_1)q_W(y_2)\frac{1 + (-1)^y_1\Delta_W(y_1)\Delta_W(y_2)}{2}. \]

See the proofs of [7, Lemma 2, Lemma 3] for a proof.

Let \( f(x) \) be a function which is convex and symmetric in \( x \in [-1,1] \). Note that in this case the function must be increasing in \( x \in [0,1] \). For the minus polar transformation, we can write
\[ \sum_{y_1y_2} q_W(y_1y_2)f(\Delta_{W-}(y_1y_2)) \]  
(17)
\[ = \sum_{y_1} q_W(y_1) \sum_{y_2} q_W(y_2)f^{-}(\Delta_W(y_1), \Delta_W(y_2)) \]  
(18)
where \( f^- (d_1, d_2) = f(d_1d_2) \), for \( d_1, d_2 \in [-1,1] \). As we assumed \( f(x) \) to be convex and symmetric in its argument, so is \( f^- \) in both of its arguments.

Similarly, for the plus polar transformation, we can write
\[ \sum_{y_1y_2u_1} q_W(y_1y_2u_1)f(\Delta_{W+}(y_1y_2u_1)) \]  
= \[ \sum_{y_1} q_W(y_1) \sum_{y_2} q_W(y_2)f^+(\Delta_W(y_1), \Delta_W(y_2)), \]  
(19)
where
\[ f^+(d_1, d_2) = \frac{1 + d_1d_2}{2} f\left(\frac{d_1 + d_2}{1 + d_1d_2}\right) + \frac{1 - d_1d_2}{2} f\left(\frac{d_1 - d_2}{1 - d_1d_2}\right), \]  
(20)
for \( d_1, d_2 \in [-1,1] \). Lemma 1 in the Appendix shows that \( f^+ \) is also a convex and symmetric function in both of its variables.

So, using the assumption \( |\Delta_V| \prec_{icx} |\Delta_W| \), we deduce that
\[ \sum_{y_1} q_V(y_1) \sum_{y_2} q_V(y_2)f^\pm(\Delta_V(y_1), \Delta_V(y_2)) \]  
(21)
\[ \leq \sum_{y_1} q_V(y_1) \sum_{y_2} q_V(y_2)f^\pm(\Delta_W(y_1), \Delta_W(y_2)) \]  
(22)
\[ = \sum_{y_1} q_W(y_1) \sum_{y_2} q_W(y_2)f^\pm(\Delta_W(y_1), \Delta_W(y_2)) \]  
(23)
\[ \leq \sum_{y_1} q_W(y_1) \sum_{y_2} q_W(y_2)f^\pm(\Delta_V(y_1), \Delta_V(y_2)). \]  
(24)
This proves our claim that both \( |\Delta_{V\pm}| \prec_{icx} |\Delta_{W\pm}| \) hold.

Now, we show that for symmetric channels Theorem 1 is equivalent to the stochastic degradation ordering. For that purpose, we first show that stochastically degraded channels satisfy \( \Delta_V \prec_{icx} \Delta_W \) ordering.

Let \( V \) be stochastically degraded with respect to \( W \). Then, by definition 1 we know there exists a channel \( P \) such that \( \Delta_V \prec_{icx} \Delta_W \) holds. In this case, one can derive that
\[ \Delta_V(y) = \frac{V(y|0) - V(y|1)}{V(y|0) + V(y|1)} = \sum_z P(z|y)\Delta_W(z), \]  
(25)
where
\[ P(z|y) = \frac{q_W(z)P(y|z)}{\sum_z q_W(z)P(y|z)}. \]  
(26)
corresponds to the inputs posterior probabilities given the output of the channel \( P \). So, for any convex function \( f(.) \), we obtain
\[ E_V[f(\Delta_V)] = \sum_y q_V(y)f(\Delta_V(y)) \]  
(27)
\[ = \sum_y \left( \sum_z q_W(z)P(y|z) \right) f\left( \sum_z P(z)\Delta_W(z) \right) \]  
(28)
\[ \leq \sum_z \sum_y q_W(z)P(y|z)f(\Delta_W(z)) \]  
(29)
\[ = \sum_z q_W(z)f(\Delta_W(z)) = E_W[f(\Delta_W)], \]  
(30)
where the inequality follows by Jensen’s inequality for convex functions. In particular, the ordering holds for the function \( f(x) = x \) with equality. Hence, degradation preserves the mean value, i.e., \( E_W[|\Delta_W|] = E_V[|\Delta_V|] \). By Theorem 1 we conclude the order relation \( \Delta_V \prec_{icx} \Delta_W \) holds for
stochastically degraded channels. Moreover, Theorem 3 shows that indeed the two orders are equivalent.

We saw that all the channels $W$ and $V$ which satisfy $\Delta_V \prec_{sc} \Delta_W$ ordering are in fact stochastically degraded channels with respect to each other. This is also clear by definition that the ordering implies the ‘symmetric convex ordering’ introduced in Theorem 1. So, we need to study if the reverse implication is also true or not in order to decide whether Theorem 1 is a strictly weaker condition than convex ordering. At this point, by recalling the equivalence stated in Proposition 2 we notice that this is not the case for symmetric channels as the two orders $\Delta_V \prec_{sc} \Delta_W$ and $|\Delta_V| \prec_{sc} |\Delta_W|$ are equivalent for symmetric channels. Observe also that for any symmetric channel $W$, we have $E_W[|\Delta_W|] = 0$.

Nevertheless, the exciting part is that no such equivalence exists if one of the two channels is asymmetric. So, if we can find a pair of B-DMCs that are not stochastically degraded, but satisfy the ‘symmetric convex ordering’ relation, we will be done. This is illustrated in the next example.

Example 1: Let $W$ be a $Z$-channel with crossover probability $p \in [0, 1]$ and $V$ be a BSC with crossover probability $\epsilon \in [0, 0.5]$. In this example, we will answer the following three questions:

1. Suppose $V$ is a stochastically degraded version of $W$. What is the best possible BSC which satisfies this condition?

2. Suppose instead that the channels satisfy the ‘symmetric convex ordering’ $|\Delta_V| \prec_{sc} |\Delta_W|$. What is the best possible BSC which satisfies this condition?

3. Suppose we first symmetrize $W$ according to Definition 5 to construct $W_\ast$. Suppose now $V$ is a stochastically degraded version of $W_\ast$. What is the best possible BSC which satisfies this condition?

Then, we will compare the three BSCs to decide which ordering results in a better channel. Here are the answers.

(a1) Stochastic degradation: Let us derive the range of possible values of $\epsilon$ in terms of $p$ under this assumption. For this purpose, we define the asymmetric binary channel $P$ degrading $W$ to $V$:

$$\begin{align*}
V(y|x) &= \sum_{z \in \{0, 1\}} W(z|x)P(y|z).
\end{align*}$$

First we note that $P(0|0) = 1 - \epsilon$ and $P(0|1) = \epsilon$ are the only possibilities. Let $P(0|1) = \alpha$. Then, using (31), we get

$$\begin{align*}
V(0|1) &= \epsilon = \frac{(1 - p)\alpha + p(1 - \epsilon)}{1 + p},
\end{align*}$$

which implies

$$\begin{align*}
\epsilon &= \frac{p + (1 - p)\alpha}{1 + p}.
\end{align*}$$

Noting that the right-hand side of (33) is increasing in $\alpha \in [0, 1]$, we conclude that

$$\begin{align*}
\frac{p}{1 + p} \leq \epsilon \leq \frac{1}{1 + p}
\end{align*}$$

whenever we impose stochastic degradation on the channels. Picking the BSC having the smallest crossover probability $\epsilon = p/(1 + p)$ answers the first question.

(a2) $|\Delta_V| \prec_{sc} |\Delta_W|$: Now, we will derive the range of possible values of $\epsilon$ in terms of $p$ under this assumption by using the cut-criterion given in Theorem 1. We start by computing the values of $E[|\Delta_V|]$ and $E[|\Delta_W|]$ in terms of the channel parameters. For the BSC, we have $E[|\Delta_V|] = 1 - \epsilon$. For the Z-channel, we have

$$\begin{align*}
|\Delta_W(y)| &= \begin{cases}
\frac{1 - p}{1 + p}, & \text{if } y = 0, \\
1, & \text{if } y = 1,
\end{cases}
\end{align*}$$

$q_W(0) = (1 + p)/2$, and $q_W(0) = (1 - p)/2$. So, we compute $E[|\Delta_W]| = 1 - p$. As any B-DMC $W'$ together with a BSC will always satisfy the conditions (9) and (10) of Theorem 2 with $\delta = 1 - 2\epsilon$ and $F$ being the probability mass function of the BSC, we can see by the theorem’s statement that the condition $E[|\Delta_V|] \leq E[|\Delta_W|]$ is necessary for $|\Delta_V| \prec_{sc} |\Delta_W|$ to hold. This in turn implies that $\epsilon \geq p/2$ in our example. Hence, the best possible BSC in this case has crossover probability $\epsilon = p/2$. This answers the second question.

(a3) Channel symmetrization: We first note a more general result: a given B-DMC $W'$ and its symmetrized version $W'_\ast$ always satisfy $|\Delta_W(y, z)| = |\Delta_W'(y, z)|$ with $|\Delta_W'(y, z)| \sim 0.5q_W(y)$, for $z \in \{0, 1\}$. Therefore, for any function $f(x)$ defined for $x \in [0, 1]$, we have

$$\begin{align*}
E[f(|\Delta_W'|)] &= E[f(|\Delta_W'|)]^2.
\end{align*}$$

We conclude that for any two B-DMCs $W'$ and $V'$: $|\Delta_V'| \prec_{sc} |\Delta_W'|$ if and only if $|\Delta_V'| \prec_{sc} |\Delta_W'|$. Moreover, as the channels in this last condition are symmetric, we know the condition holds if and only if $|\Delta_V'| \prec_{sc} |\Delta_W'|$. So, we have the same answer as in the previous case: the best possible BSC in this case has crossover probability $\epsilon = p/2$.

Let us compare the results. Noting that $p/2 \leq p/(1 + p)$ holds for any $p \in [0, 1]$, and with equality if and only if $p = \{0, 1\}$, we conclude that, for $p \in (0, 1)$, the better BSC is found by the ‘symmetric convex ordering’, and this BSC is not stochastically degraded with respect to the Z-channel. For instance when $p = 0.5$, the crossover probabilities of the best found BSC is 0.25 in the second case compared to $1/3$ in the first one. Finally, we also showed that one can verify the ‘symmetric convex ordering’ by first symmetrizing the asymmetric channels and then checking for stochastic degradation.

The previous example proves that for general B-DMCs the ‘symmetric convex ordering’ is strictly weaker than stochastic degradation. Moreover, we can see that the information set of the polar code designed for the best possible BSC which is smaller with respect to the symmetric convex ordering than the Z channel may be significantly larger than the set designed for the best possible BSC which is stochastically degraded with respect to the Z-channel.
In addition, the example uncovered an advantage of the channel symmetrization operation for symmetric channels before polarization; we showed that it matters during the design whether the channel is directly approximated by a degraded channel or the channel is first symmetrized and then approximated by a degraded one. Although channel symmetrization or stochastic degradation are not novel ideas for the theory of channel polarization, the implications we discussed in this paper on the partial orderings of the information sets are novel.

Finally, we discuss the two other known orderings related to BECs. For simplicity, we define $Z = |\Delta|$. Suppose a BEC $BEC$ with erasure probability $\epsilon \in [0, 1]$ and a B-DMC $W$ satisfy $E[Z_W] \leq E[Z_{BEC}]$. As $Z_{BEC}$ is $\{0, 1\}$ valued with distribution $P(Z_{BEC} = 0) = \epsilon$, $Z_{BEC}$ and any arbitrary random variable $Z$ satisfy the conditions (9) and (10) of Theorem 4 with $G$ being the probability mass function of $Z_{BEC}$. As a result, the assumption $E[Z_W] \leq E[Z_{BEC}]$ implies $Z_W \preceq_{acx} Z_{BEC}$. Using Theorem 1 we recover the ordering proved in [2, Theorem 1] for channels with equal variational distance $E[Z_W] = E[Z_{BEC}]$.

Similarly, another special case of increasing convex ordering happens when $BEC$ and $W$ are such that the Bhattacharyya parameters of the channels satisfy $B(W) \leq B(BEC)$. Let us define the random variable $B_W = \sqrt{1 - Z_W}$. Then, $B(W) = E[B_W]$. Hence, the channels satisfy $E[B_W] \leq E[B_{BEC}]$. One more time, letting $G$ denote the probability mass function of $B_{BEC}$ in Theorem 2 we see that $E[B_W] \leq E[B_{BEC}]$ implies $B_W \preceq_{acx} B_{BEC}$. Finally, it is well known from [1, Proposition 6] that this ordering is also preserved under the polarization transformations.

**APPENDIX**

**Lemma 1:** Let $f(x)$ be a convex and symmetric function in $x \in [-1, 1]$. Then, the function defined in (20) is also a convex and symmetric function.

**Proof:** For simplicity, we first define

\[
    f_1(d_1, d_2) = (1 + d_1 d_2) f\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right),
\]

\[
    f_2(d_1, d_2) = (1 - d_1 d_2) f\left( \frac{d_1 - d_2}{1 - d_1 d_2} \right),
\]

for $d_1, d_2 \in [-1, 1]$. Hence, (20) equals to

\[
    f_+(d_1, d_2) = \frac{1}{2} f_1(d_1, d_2) + \frac{1}{2} f_2(d_1, d_2).
\]

By the permutation symmetry of the function in the variables $d_1$ and $d_2$, i.e., $f_+(d_1, d_2) = f_+(d_2, d_1)$, it is sufficient to prove the lemma for one of the variables.

One can easily prove that the function is symmetric in $d_1 \in [-1, 1]$, i.e., $f_+(d_1, d_2) = f_+(d_1, -d_2)$ by using the symmetry of the function $f(x)$ in $x \in [-1, 1]$.

Let $f_1''$ and $f_2''$ denote the first and second derivatives of $f(x)$ with respect to the variable $x$. Then, taking the first and second derivatives of $f_1(d_1, d_2)$ with respect to $d_1$, we get

\[
    \frac{\partial}{\partial d_1} f_1(d_1, d_2) = d_2 f\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) + \frac{1 - d_2^2}{1 + d_1 d_2} f_1\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right),
\]

and

\[
    \frac{\partial^2}{\partial d_1^2} f_1(d_1, d_2) = \frac{\partial}{\partial d_1} f_1\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) - d_2 \frac{1 - d_2^2}{1 + d_1 d_2} f_1\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) + \frac{1}{1 + d_1^2} f_1\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) \frac{1 - d_2^2}{1 + d_1 d_2}.
\]

Similarly, we obtain the derivatives of $f_2(d_1, d_2)$ with respect to $d_1$ as

\[
    \frac{\partial}{\partial d_1} f_2(d_1, d_2) = \frac{1}{2} f_1\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) + \frac{1}{2} f_1\left( \frac{d_1 - d_2}{1 + d_1 d_2} \right),
\]

\[
    \frac{\partial^2}{\partial d_1^2} f_2(d_1, d_2) = \frac{1}{2} f_2\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) + \frac{1}{2} f_2\left( \frac{d_1 - d_2}{1 + d_1 d_2} \right).
\]

So, we get

\[
    \frac{\partial^2}{\partial d_1^2} f_+(d_1, d_2) = \frac{1}{2} \left( \frac{1 - d_2^2}{1 + d_1 d_2} \right) f_1''\left( \frac{d_1 + d_2}{1 + d_1 d_2} \right) + \frac{1}{2} \left( \frac{1 - d_2^2}{1 + d_1 d_2} \right) f_2''\left( \frac{d_1 - d_2}{1 + d_1 d_2} \right) \geq 0,
\]

where the sign of $f_+(d_1, d_2)$ can be deduced from the convexity of the function $f(x)$ in $x \in [-1, 1]$. This proves that $f_+(d_1, d_2)$ is convex in $d_1 \in [-1, 1]$ and completes the proof.

**Acknowledgment**

This work was supported by Swiss National Science Foundation under grant number 200021-125347/1.

**References**

[1] E. Arkin, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[2] M. Alas, “Properties of the polarization transformations for the likelihood ratios of symmetric B-DMCs,” in *Information Theory (CIIT), 2013 13th Canadian Workshop on*, 2013, pp. 22–27.

[3] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, Lausanne, 2009.
[4] R. Szekli, *Stochastic ordering and dependence in applied probability*, ser. Lecture notes in statistics. Springer-Verlag, 1995.

[5] D. Blackwell, “Equivalent comparisons of experiments,” *The Annals of Mathematical Statistics*, vol. 24, no. 2, pp. 265–272, 1953.

[6] S. Karlin and A. Novikoff, *Generalized Convex Inequalities*. Pacific J. Math, 1963.

[7] M. Alsan, “Extremality properties for the basic polarization transformations,” *eprint arXiv:1301.5258*, 2013.