Channel Polarization and Blackwell Measures

Maxim Raginsky

***Abstract*—The Blackwell measure of a binary-input channel (BIC) is the distribution of the posterior probability of 0 under the uniform input distribution. This paper gives an explicit characterization of the evolution of the Blackwell measure of an arbitrary symmetric BIC under Arıkan's polarization transform, and uses this characterization to provide a unifying set of techniques for studying the polarization phenomenon.**

***Index Terms*—Channel polarization, Blackwell measure, convex functionals of channels.**

I. Introduction

Polar codes, introduced by Arıkan [1], are a structured family of codes with low encoding and decoding complexity that provably achieve capacity for any symmetric binary-input channel (BIC). The salient feature of polar codes, from which they derive their name, is the polarization phenomenon: n independent copies of a given symmetric BIC are transformed in such a way that most of the resulting n channels are either nearly completely noisy (with capacity near zero) or nearly completely noiseless (with capacity near 1 bit).

In the original paper of Arıkan [1], the polarization phenomenon was demonstrated for two channel parameters: the capacity and the Bhattacharyya parameter. Subsequent results by Alsan and Telatar [2]–[4] have revealed other channel parameters that polarize – for example, Gallager's E_0 function, which is related to various error exponents and cutoff rates. However, the proofs of polarization are usually tailored to the particular channel parameter of interest, and there does not appear to be a common technique underlying them.

The objective of this paper is to offer such a technique that relies on viewing each BIC through the lens of its *Blackwell measure* [5], [6], i.e., the distribution of the posterior probability of input symbol being 0 under the uniform input distribution. We exploit two key facts pertaining to this representation: (a) the Blackwell measure of $\mathrm{BSC}(p)$ assigns equal masses to the points p and $1-p$, and (b) each symmetric BIC can be decomposed into a mixture of BSCs with input-independent weights. Using these facts, we can easily track the evolution of the Blackwell measures under the polarization transform, which in turn gives us a handle on the polarization phenomenon for a very broad class of channel characteristics.

Notation. For $p,q \in [0,1]$, we let $\bar{p} = 1-p$ (for $p \in \{0,1\}$, this is the Boolean NOT) and $p \star q = p\bar{q} + \bar{p}q$. For $a,b \in \mathbb{R}$, we let $a \wedge b \triangleq \min(a,b)$ and $a \vee b \triangleq \max(a,b)$. Given a random object U, we will denote by $\mathcal{L}(U)$ its probability law.

The author is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, USA. E-mail: maxim@illinois.edu.

This work was supported in part by Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF–0939370.

II. BICs and their Blackwell measures

A discrete binary-input channel (BIC) is a pair (Y, W), where Y is the finite output alphabet and $W = \big(W(\cdot|0), W(\cdot|1)\big)$ is a pair of probability distributions on Y. For $x \in \{0,1\}$, $W(\cdot|x)$ is the probability distribution of the channel output when the channel input is equal to x. Given two BICs (Y_1, W_1) and (Y_2, W_2), we define the product BIC $(\mathsf{Y}_1 \times \mathsf{Y}_2, W_1 \times W_2)$ by

$$(W_1 \times W_2)(y_1,y_2|x) \triangleq W_1(y_1|x)W_2(y_2|x)$$

for all $x \in \{0,1\}$ and all $(y_1,y_2) \in \mathsf{Y}_1 \times \mathsf{Y}_2$. In other words, $W_1 \times W_2$ is the parallel broadcast channel formed by W_1 and W_2 [7]. A BIC (Y, W) is symmetric [1] if there exists a bijection $\pi : \mathsf{Y} \to \mathsf{Y}$, such that $\pi^{-1} = \pi$ and $W(\pi y|0) = W(y|1)$ for all $y \in \mathsf{Y}$.

Our analysis hinges on a certain representation of channels that originates in the work of Blackwell on the comparison of statistical experiments [5], [6] (see [8] for a modern synthesis). We particularize it to BICs as follows. Given a BIC (Y, W), let (X,Y) be a random couple with $P_X = \mathrm{Bern}(1/2)$ and $P_{Y|X} = W$. The random variable

$$S = S_W(Y) \triangleq \frac{W(Y|0)}{W(Y|0) + W(Y|1)},$$

which is equal to the posterior probability of $X=0$ given Y, takes values in the unit interval $[0,1]$ and has mean $1/2$. The *Blackwell measure* of W, which we will denote by m_W, is the probability law of S. For example:

$$\mathsf{m}_{\mathrm{BSC}(p)} = \frac{1}{2}\delta_p + \frac{1}{2}\delta_{\bar{p}}, \quad \mathsf{m}_{\mathrm{BEC}(\varepsilon)} = \frac{\bar{\varepsilon}}{2}\delta_0 + \frac{\bar{\varepsilon}}{2}\delta_1 + \varepsilon\delta_{1/2},$$

where δ_s denotes the Dirac measure at s. Given two BICs (Y, W) and (Y, W'), we say that W dominates W' (or is more informative than W') if there exists a random transformation K from Y to Y', such that $W' = K \circ W$ (i.e., W' is stochastically degraded with respect to W). In that case, we write $W \succeq W'$. We say that W and W' are equivalent if $W \succeq W'$ and $W' \succeq W$. In that case, we write $W \equiv W'$.

Theorem 1 (Blackwell–Sherman–Stein)**.** *Consider two BICs W and W'. Then:*

1) *$W \equiv W'$ if and only if $\mathsf{m}_W = \mathsf{m}_{W'}$ (that is, the Blackwell measure specifies the channel uniquely up to equivalence). Moreover, let $\mathcal{M}$ denote the collection of all Borel probability measures on $[0,1]$ with mean $1/2$. Then for any $\mathsf{m} \in \mathcal{M}$ there exists a BIC W, unique up to equivalence, such that $\mathsf{m} = \mathsf{m}_W$.*[1]

[1]The BIC W has a finite output alphabet if and only if m_W has finite support. This is precisely the setting of this paper.

2) *$W \succeq W'$ if and only if*

$$\int_{[0,1]} f \mathrm{d}\mathsf{m}_W \geq \int_{[0,1]} f \mathrm{d}\mathsf{m}_{W'}$$

for every convex $f : [0,1] \to \mathbb{R}$.

Remark 1. There are one-to-one correspondences between the Blackwell measure m_W and other probabilistic objects associated to the BIC W, such as its L- and D-distributions [9, Ch. 4] (see, e.g., [8, pp. 603–606]). The Blackwell measure is also a special case of the so-called α-representation [10].

Let (Y_i, W_i), $i \in [m] \triangleq \{1, \ldots, m\}$, be a collection of BICs, and let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be a probability distribution on $[m]$. A *compound channel* with subchannels $\{W_i\}$ and mixing distribution λ is a BIC W with output (I, Y_I), where

$$W(i, y_i|x) = \lambda_i W_i(y|x), \qquad \forall x \in \{0,1\}, i \in [m], y \in \mathsf{Y}_i.$$

In this case, we write $W = \bigoplus_{i=1}^m \lambda_i W_i$. We have the following structural result (see, e.g., [11, Thm. 2.1]):

Theorem 2. *For any symmetric BIC W, there exist a positive integer n, a probability vector $\lambda = (\lambda_1, \ldots, \lambda_n)$, and error parameters $p_1, \ldots, p_n \in [0,1]$, such that*

$$W \equiv \bigoplus_{i=1}^n \lambda_i \mathrm{BSC}(p_i). \tag{1}$$

It is not difficult to show that the Blackwell measure of a compound channel is given by the mixture of the Blackwell measures of the constituent subchannels:

$$\mathsf{m}_W = \sum_{i=1}^m \lambda_i \mathsf{m}_{W_i}.$$

Thus, Theorem 2 shows that the Blackwell measure of any symmetric BIC is a mixture of Blackwell measures of BSCs. For example, $\mathrm{BEC}(\varepsilon) \equiv \bar{\varepsilon}\,\mathrm{BSC}(0) \oplus \varepsilon\,\mathrm{BSC}(1/2)$, which is also evident from the decomposition $\mathsf{m}_{\mathrm{BEC}(\varepsilon)} = \bar{\varepsilon}\,\mathsf{m}_{\mathrm{BSC}(0)} + \varepsilon\,\mathsf{m}_{\mathrm{BSC}(1/2)}$.

A. Functionals of BICs

Any measurable function $f : [0,1] \to \mathbb{R}$ induces a functional I_f on the collection of all BICs via

$$\mathrm{I}_f(W) \triangleq \int_{[0,1]} f \mathrm{d}\mathsf{m}_W = \mathbb{E}[f(S)],$$

where $S \sim \mathsf{m}_W$. A variety of channel characteristics can be expressed in this way; for example:

- With $f(s) = \log 2 - h_2(s)$, where $h_2(\cdot)$ is the binary entropy function, $\mathrm{I}_f(W)$ is equal to the *symmetric capacity* $I(W)$ of W [1], i.e., the mutual information $I(P, W)$ with $P = \mathrm{Bern}(1/2)$.
- With $f(s) = 2\sqrt{s(1-s)}$, we get

$$\mathrm{I}_f(W) = Z(W) \triangleq \sum_{y \in \mathsf{Y}} \sqrt{W(y|0)W(y|1)},$$

the *Bhattacharyya parameter* of W [1]. More generally, with $f(s) = 2s^\alpha(1-s)^{1-\alpha}$ for $\alpha \in [0,1]$, $\mathrm{I}_f(W)$ is equal to the *Hellinger affinity* of order α:

$$\mathcal{H}_\alpha(W) \triangleq \sum_{y \in \mathsf{Y}} W(y|0)^\alpha W(y|1)^{1-\alpha}.$$

- Taking

$$f(s) = 2^{-\rho}\left(s^{\frac{1}{1+\rho}} + (1-s)^{\frac{1}{1+\rho}}\right)^{1+\rho}$$

for some $\rho \geq 0$, we can relate $\mathrm{I}_f(W)$ to Gallager's function $E_0(\rho, W) = E_0(\rho, \mathrm{Bern}(\frac{1}{2}), W)$ [12] evaluated at the uniform input distribution:

$$\mathrm{I}_f(W) = \exp\left(-E_0(\rho, W)\right).$$

- With $f(s) = |2s - 1|$, we get $\mathrm{I}_f(W) = 1 - 2P_{\mathrm{e,ML}}(W)$, where $P_{\mathrm{e,ML}}(W)$ denotes the probability of error of maximum-likelihood decoding of a single equiprobable bit sent through the channel W [4, Ch. 5].
- With $f_\lambda(s) = \bar{\lambda} \wedge \lambda - (\bar{\lambda}s) \wedge (\lambda\bar{s})$ for $0 < \lambda < 1$, $\mathrm{I}_f(W)$ is equal to

$$B_\lambda(W) \triangleq \frac{1}{2}|1 - 2\lambda| - \frac{1}{2}\|\lambda W(\cdot|1) - \bar{\lambda}W(\cdot|0)\|_{\mathrm{TV}},$$

the Bayesian information gain about $X \sim \mathrm{Bern}(\lambda)$ based on a single observation through W. Since any convex $f : [0,1] \to \mathbb{R}$ can be approximated by a positive linear combination of such f_λ's [8], it follows that $W \succeq W'$ if and only if $B_\lambda(W) \geq B_\lambda(W')$ for all $\lambda \in (0,1)$.

It follows directly from definitions that

$$\mathrm{I}_f(\mathrm{BSC}(p)) = \frac{1}{2}f(p) + \frac{1}{2}f(\bar{p}).$$

Moreover, if W is a symmetric BIC that admits the decomposition (1), then

$$\mathrm{I}_f(W) = \sum_{i=1}^n \lambda_i \mathrm{I}_f(\mathrm{BSC}(p_i)) = \sum_{i=1}^n \frac{\lambda_i\left[f(p_i) + f(\bar{p}_i)\right]}{2}.$$

B. Two operations on Blackwell measures

Given two Blackwell measures $\mathsf{m}_1, \mathsf{m}_2 \in \mathcal{M}$, we define two probability measures $\mathsf{m}_1 \boxast \mathsf{m}_2$ and $\mathsf{m}_1 \circledast \mathsf{m}_2$ on $[0,1]$ as follows. Let $S_1 \sim \mathsf{m}_1$ and $S_2 \sim \mathsf{m}_2$ be two independent random variables. For any continuous $f : [0,1] \to \mathbb{R}$, let[2]

$$\int_{[0,1]} f \mathrm{d}(\mathsf{m}_1 \boxast \mathsf{m}_2) = \mathbb{E}[f(1 - S_1 \star S_2)] \tag{2}$$

and

$$\int_{[0,1]} f \mathrm{d}(\mathsf{m}_1 \circledast \mathsf{m}_2) = \mathbb{E}\Bigg[(1 - S_1 \star S_2) f\left(\frac{S_1 S_2}{1 - S_1 \star S_2}\right) + (S_1 \star S_2) f\left(\frac{\bar{S}_1 S_2}{S_1 \star S_2}\right)\Bigg]. \tag{3}$$

[2]Since the interval $[0,1]$ is compact, every continuous $f : [0,1] \to \mathbb{R}$ is bounded by the Weierstrass theorem.

Lemma 1. *The probability measures* $\mathsf{m}_1 \boxast \mathsf{m}_2$ *and* $\mathsf{m}_1 \circledast \mathsf{m}_2$ *are also Blackwell measures.*

Proof. Using $f(s) = s$ in Eqs. (2) and (3), and recalling that S_1 and S_2 are independent and both have mean $\frac{1}{2}$, we get

$$\int_{[0,1]} s(\mathsf{m}_1 \boxast \mathsf{m}_2)(\mathrm{d}s) = \mathbb{E}[1 - S_1 \star S_2] = \frac{1}{2}$$

and

$$\int_{[0,1]} s(\mathsf{m}_1 \circledast \mathsf{m}_2)(\mathrm{d}s) = \mathbb{E}[S_2] = \frac{1}{2}.$$

Thus, both $\mathsf{m}_1 \boxast \mathsf{m}_2$ and $\mathsf{m}_1 \circledast \mathsf{m}_2$ are in $\mathcal{M}$. □

III. Evolution of Blackwell Measures under the Polarization Transform

The polarization transform maps a pair of BICs (Y_1, W_1) and (Y_2, W_2) into another pair of BICs $(\mathsf{Y}_1 \times \mathsf{Y}_2, W_1 \boxast W_2)$ and $(\mathsf{Y}_1 \times \mathsf{Y}_2 \times \{0,1\}, W_1 \circledast W_2)$ as follows:

$$(W_1 \boxast W_2)(y_1, y_2|x) \triangleq \frac{1}{2} \sum_{u \in \{0,1\}} W_1(y_1|u \oplus x) W_2(y_2|u) \tag{4a}$$

$$(W_1 \circledast W_2)(y_1, y_2, u|x) \triangleq \frac{1}{2} W_1(y_1|u \oplus x) W_2(y_2|x) \tag{4b}$$

for all $x, u \in \{0,1\}$ and all $(y_1, y_2) \in \mathsf{Y}_1 \times \mathsf{Y}_2$, where $\oplus$ is the Boolean XOR. The Blackwell measures of $W_1 \boxast W_2$ and $W_1 \circledast W_2$ can be computed from those of W_1 and W_2:

Theorem 3. *The Blackwell measures of the BICs* $W_1 \boxast W_2$ *and* $W_1 \circledast W_2$ *are given by*

$$\mathsf{m}_{W_1 \boxast W_2} = \mathsf{m}_{W_1} \boxast \mathsf{m}_{W_2} \quad \textit{and} \quad \mathsf{m}_{W_1 \circledast W_2} = \mathsf{m}_{W_1} \circledast \mathsf{m}_{W_2}.$$

Remark 2. This result can be derived as a special case of Prop. 3 in [10], but we give a self-contained proof below.

Proof. The formula for $W_1 \boxast W_2$ follows from a straightforward computation. Now, from the definition (4) it follows that

$$W_1 \circledast W_2 \equiv \frac{1}{2} W^{(0)} \oplus \frac{1}{2} W^{(1)} \tag{5}$$

with $W^{(0)} \triangleq W_1 \times W_2$ and $W^{(1)} \triangleq \bar{W}_1 \times W_2$, where $\bar{W}_1$ is the BIC related to W_1 via $\bar{W}_1(\cdot|x) = W_1(\cdot|\bar{x})$. Then the random variables $S^{(s)} \sim \mathsf{m}_{W^{(s)}}$, $s \in \{0,1\}$, evidently satisfy

$$\mathbb{E}[f(S^{(0)})] = 2\,\mathbb{E}\left[(1 - S_1 \star S_2) f\left(\frac{S_1 S_2}{1 - S_1 \star S_2}\right)\right]$$

and

$$\mathbb{E}[f(S^{(1)})] = 2\,\mathbb{E}\left[(S_1 \star S_2) f\left(\frac{\bar{S}_1 S_2}{S_1 \star S_2}\right)\right]$$

for every continuous $f : [0,1] \to \mathbb{R}$. From this and (5), we get

$$\begin{aligned}\int_{[0,1]} f \mathrm{d}\mathsf{m}_{W_1 \circledast W_2} &= \frac{1}{2}\mathbb{E}[f(S^{(0)})] + \frac{1}{2}\mathbb{E}[f(S^{(1)})] \\ &= \mathbb{E}\Bigg[(1 - S_1 \star S_2) f\left(\frac{S_1 S_2}{1 - S_1 \star S_2}\right) \\ &\qquad + (S_1 \star S_2) f\left(\frac{\bar{S}_1 S_2}{S_1 \star S_2}\right)\Bigg] \\ &= \int_{[0,1]} f \mathrm{d}(\mathsf{m}_{W_1} \circledast \mathsf{m}_{W_2}).\end{aligned}$$

Since f is arbitrary, we obtain the formula for $\mathsf{m}_{W_1 \circledast W_2}$. □

Using this theorem, we can explicitly write down the image of a pair of BSCs under the polarization transformation:

Corollary 1 (Polarization of BSCs)**.** *For any pair of crossover probabilities* $(p, q) \in [0, \frac{1}{2}] \times [0, \frac{1}{2}]$,

$$\mathrm{BSC}(p) \boxast \mathrm{BSC}(q) \equiv \mathrm{BSC}(p \star q) \tag{6}$$

and

$$\mathrm{BSC}(p) \circledast \mathrm{BSC}(q) \equiv \mathrm{BSC}(p) \times \mathrm{BSC}(q). \tag{7}$$

Proof. A random variable $S \sim \mathsf{m}_{\mathrm{BSC}(p)}$ takes two equiprobable values, p and $\bar{p}$. Therefore, if $S_1 \sim_{\mathrm{BSC}(p)}$ and $S_2 \sim_{\mathrm{BSC}(q)}$ are independent, the random variable $1 - S_1 \star S_2$ takes two equiprobable values, $p \star q$ and $1 - p \star q = \bar{p} \star q$ — that is,

$$\begin{aligned}\mathsf{m}_{\mathrm{BSC}(p) \boxast \mathrm{BSC}(q)} &= \mathsf{m}_{\mathrm{BSC}(p)} \boxast \mathsf{m}_{\mathrm{BSC}(q)} \\ &= \frac{1}{2}\delta_{p \star q} + \frac{1}{2}\delta_{1 - p \star q}\end{aligned}$$

which proves (6). Next, from the definition (3), for any continuous $f : [0,1] \to \mathbb{R}$ we have

$$\begin{aligned}&\int_{[0,1]} f \mathrm{d}\mathsf{m}_{\mathrm{BSC}(p) \circledast \mathrm{BSC}(q)} \\ &= \frac{1}{2}(1 - p \star q) f\left(\frac{pq}{1 - p \star q}\right) + \frac{1}{2}(1 - p \star q) f\left(\frac{\bar{p}\bar{q}}{1 - p \star q}\right) \\ &\quad + \frac{1}{2}(p \star q) f\left(\frac{p\bar{q}}{p \star q}\right) + \frac{1}{2}(p \star q) f\left(\frac{\bar{p}q}{p \star q}\right).\end{aligned}$$

On the other hand, a straightforward computation shows that

$$\begin{aligned}\mathsf{m}_{\mathrm{BSC}(p) \times \mathrm{BSC}(q)} &= (p \star q)\left(\frac{1}{2}\delta_{\frac{p\bar{q}}{p \star q}} + \frac{1}{2}\delta_{\frac{\bar{p}q}{p \star q}}\right) \\ &\quad + (1 - p \star q)\left(\frac{1}{2}\delta_{\frac{pq}{1 - p \star q}} + \frac{1}{2}\delta_{\frac{\bar{p}\bar{q}}{1 - p \star q}}\right),\end{aligned}$$

which proves (7). □

We can also characterize the image of a pair of arbitrary symmetric BICs under the polarization transformation:

Corollary 2. *For any pair* W_1 *and* W_2 *of symmetric BICs, there exist positive integers* m, n*, probability vectors*

$\lambda = (\lambda_1, \ldots, \lambda_m), \mu = (\mu_1, \ldots, \mu_n)$, *and tuples* $p = (p_1, \ldots, p_m) \in [0,1]^m, q = (q_1, \ldots, q_n) \in [0,1]^n$, *such that*

$$\mathsf{m}_{W_1 \boxast W_2} = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \mu_j \mathsf{m}_{\mathrm{BSC}(p_i \star q_j)} \tag{8}$$

and

$$\mathsf{m}_{W_1 \circledast W_2} = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \mu_j \mathsf{m}_{\mathrm{BSC}(p_i) \times \mathrm{BSC}(q_j)}. \tag{9}$$

Proof. From Theorem 2, we can write

$$\mathsf{m}_{W_1} = \sum_{i=1}^{m} \lambda_i \mathsf{m}_{\mathrm{BSC}(p_i)} = \sum_{i=1}^{m} \frac{\lambda_i (\delta_{p_i} + \delta_{\bar{p}_i})}{2}$$

and

$$\mathsf{m}_{W_2} = \sum_{j=1}^{n} \mu_j \mathsf{m}_{\mathrm{BSC}(q_j)} = \sum_{j=1}^{n} \frac{\mu_j (\delta_{q_j} + \delta_{\bar{q}_j})}{2}$$

for some choices of (m, λ, p) and (n, μ, q). Therefore, for two independent r.v.'s $S_1 \sim \mathsf{m}_{W_1}$ and $S_2 \sim \mathsf{m}_{W_2}$ and for any continuous $f : [0,1] \to \mathbb{R}$ we have

$$\begin{aligned}
&\int_{[0,1]} f \mathrm{d}\mathsf{m}_{W_1 \boxast W_2} = \mathbb{E}[f(1 - S_1 \star S_2)] \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \mu_j \left(\frac{1}{2} f(p_i \star q_j) + \frac{1}{2} f(1 - p_i \star q_j) \right)
\end{aligned}$$

and

$$\begin{aligned}
\int_{[0,1]} f \mathrm{d}\mathsf{m}_{W_1 \circledast W_2} &= \mathbb{E}\Bigg[(1 - S_1 \star S_2) f\left(\frac{S_1 S_2}{1 - S_1 \star S_2} \right) \\
&\qquad + (S_1 \star S_2) f\left(\frac{\bar{S}_1 S_2}{S_1 \star S_2} \right) \Bigg] \\
= \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \mu_j \Bigg\{ &\frac{1}{2}(1 - p_i \star q_j) f\left(\frac{p_i q_j}{1 - p_i \star q_j} \right) \\
&+ \frac{1}{2}(1 - p_i \star q_j) f\left(\frac{\bar{p}_i \bar{q}_j}{1 - p_i \star q_j} \right) \\
&+ \frac{1}{2}(p_i \star q_j) f\left(\frac{p_i \bar{q}_j}{p_i \star q_j} \right) + \frac{1}{2}(p_i \star q_j) f\left(\frac{\bar{p}_i q_j}{p_i \star q_j} \right) \Bigg\}.
\end{aligned}$$

Applying Corollary 1, we obtain Eqs. (8) and (9). □

IV. Polarization is generic

Informally speaking, the polar transform (4) replaces the original pair of BICs W_1 and W_2 with another pair, where one BIC $W_1 \circledast W_2$ is "better" than W_1 or W_2 and another BIC $W_1 \boxast W_2$ is "worse" than both W_1 and W_2. This is precisely the polarization effect that is responsible for the capacity-achieving performance of polar codes. To be more precise, let us fix a class $\mathcal{W}$ of BICs and a functional Φ that associates a real number $\Phi(W)$ to each BIC W. Then we will say that Φ *polarizes* on $\mathcal{W}$ if, for any two BICs $W_1, W_2 \in \mathcal{W}$,

$$\begin{aligned}
\Phi(W_1 \boxast W_2) &\le \Phi(W_1) \wedge \Phi(W_2) \\
&\le \Phi(W_1) \vee \Phi(W_2) \le \Phi(W_1 \circledast W_2).
\end{aligned} \tag{10}$$

Of course, this property has an operational meaning only if $\Phi(W) \ge \Phi(W')$ corresponds to W being somehow "better" than W'. For example, Arıkan has shown that the functionals $\Phi(W) = I(W)$ and $\Phi(W) = -Z(W)$ polarize [1]. Since then, the polarization property has been demonstrated for other channel parameters, such as Gallager's E_0 [4]. We now show that the polarization phenomenon is generic, i.e., Eq. (10) holds for a very broad class of channel functionals:

Theorem 4. *All channel functionals* I_f *with a convex* $f : [0,1] \to \mathbb{R}$ *polarize on the class of all symmetric BICs. That is, if* W_1, W_2 *are two symmetric BICs, then*

$$\begin{aligned}
\mathrm{I}_f(W_1 \boxast W_2) &\le \mathrm{I}_f(W_1) \wedge \mathrm{I}_f(W_2) \\
&\le \mathrm{I}_f(W_1) \vee \mathrm{I}_f(W_2) \le \mathrm{I}_f(W_1 \circledast W_2).
\end{aligned}$$

Proof. Let $S_1 \sim \mathsf{m}_{W_1}$ and $S_2 \sim \mathsf{m}_{W_2}$ be independent. Then, using Theorem 3, we can write

$$\begin{aligned}
\mathrm{I}_f(W_1 \boxast W_2) &= \int_{[0,1]} f \mathrm{d}\mathsf{m}_{W_1 \boxast W_2} \\
&= \int_{[0,1]} f \mathrm{d}(\mathsf{m}_{W_1} \boxast \mathsf{m}_{W_2}) \\
&= \mathbb{E}[f(S_1 S_2 + (1 - S_1)(1 - S_2))] \\
&= \mathbb{E}\Big[\mathbb{E}[f(S_1 S_2 + (1 - S_1)(1 - S_2)) | S_2] \Big] \\
&\overset{(a)}{\le} \mathbb{E}\Big[\mathbb{E}[S_2 f(S_1) + (1 - S_2) f(1 - S_1) | S_2] \Big] \\
&\overset{(b)}{=} \frac{1}{2} \mathbb{E}[f(S_1)] + \frac{1}{2} \mathbb{E}[f(1 - S_1)] \\
&\overset{(c)}{=} \mathrm{I}_f(W_1),
\end{aligned}$$

where (a) is by Jensen's inequality, (b) follows from the fact that S_1 and S_2 are independent with $\mathbb{E}[S_1] = \mathbb{E}[S_2] = \frac{1}{2}$, and (c) follows from the symmetry of W_1, which is equivalent to $\mathcal{L}(S_1) = \mathcal{L}(1 - S_1)$. This shows that $\mathrm{I}_f(W_1 \boxast W_2) \le \mathrm{I}_f(W_1)$. Conditioning on S_1 instead of S_2, we prove that $\mathrm{I}_f(W_1 \boxast W_2) \le \mathrm{I}_f(W_2)$.

Using Theorem 3 and Jensen's inequality, we obtain

$$\begin{aligned}
\mathrm{I}_f(W_1 \circledast W_2) &= \int_{[0,1]} f \mathrm{d}\mathsf{m}_{W_1 \circledast W_2} \\
&= \int_{[0,1]} f \mathrm{d}(\mathsf{m}_{W_1} \circledast \mathsf{m}_{W_2}) \\
&= \mathbb{E}\Bigg[(1 - S_1 \star S_2) f\left(\frac{S_1 S_2}{1 - S_1 \star S_2} \right) \\
&\qquad + (S_1 \star S_2) f\left(\frac{\bar{S}_1 S_2}{S_1 \star S_2} \right) \Bigg] \\
&\ge \mathbb{E}[f(S_2)] = \mathrm{I}_f(W_2).
\end{aligned}$$

By symmetry, the channels $W_1 \circledast W_2$ and $W_2 \circledast W_1$ are equivalent, so we also have $\mathrm{I}_f(W_1 \circledast W_2) \ge \mathrm{I}_f(W_1)$. □

Corollary 3.

$$\begin{aligned}
W_1 \boxast W_2 \preceq_{\mathrm{B}} W_1 \preceq_{\mathrm{B}} W_1 \circledast W_2 \\
W_1 \boxast W_2 \preceq_{\mathrm{B}} W_2 \preceq_{\mathrm{B}} W_1 \circledast W_2.
\end{aligned}$$

Proof. Use the Blackwell–Sherman–Stein theorem. □

All existing proofs of polarization (for capacity, Bhattacharyya parameter, Gallager's E_0, etc.) emerge as special cases of Theorem 4. Moreover, it reveals that the polarization phenomenon is naturally linked to channel domination in the sense of Blackwell. In particular, it implies the polarization phenomenon for the Bayes error functionals $B_\lambda(\cdot)$ (see Sec. II-A), which, to the best of our knowledge, has not been discussed previously.

V. The Submartingale Property

Another salient feature of the polarization process noted by Arıkan [1] is the following: Fix a BIC W and let $\{B_n\}_{n=1}^\infty$ be a sequence of i.i.d. $\mathrm{Bern}(\frac{1}{2})$ r.v.'s. Let $W_0 = W$, and

$$W_n = \begin{cases} W_{n-1} \boxast W_{n-1}, & \text{if } B_n = 0 \\ W_{n-1} \circledast W_{n-1}, & \text{if } B_n = 1 \end{cases}$$

for $n \geq 1$. Define the random processes $\{I_n\}_{n=0}^\infty$ and $\{Z_n\}_{n=0}^\infty$ via $I_n = I(W_n)$ and $Z_n = Z(W_n)$. Then $\{I_n\}$ is a nonnegative martingale, while $\{Z_n\}$ is a nonnegative supermartingale, both with respect to the natural filtration generated by $\{B_n\}$. These properties are a consequence of the following relations: for any pair W_1, W_2 of symmetric BICs,

$$I(W_1 \boxast W_2) + I(W_1 \circledast W_2) = I(W_1) + I(W_2)$$

and

$$Z(W_1 \boxast W_2) + Z(W_1 \circledast W_2) \leq Z(W_1) + Z(W_2).$$

Motivated by this, we introduce the following definition: Let $f : [0,1] \to \mathbb{R}$ be a convex function. We say that the polarization process is *f-improving* on a given collection $\mathcal{W}$ of BICs if, for any $W_1, W_2 \in \mathcal{W}$,

$$\mathrm{I}_f(W_1 \boxast W_2) + \mathrm{I}_f(W_1 \circledast W_2) \geq \mathrm{I}_f(W_1) + \mathrm{I}_f(W_2). \qquad (11)$$

If (11) holds, then the random process $\{I_n^{(f)}\}_{n=0}^\infty$ with $I_n^{(f)} = \mathrm{I}_f(W_n)$ is a submartingale. The following result shows that, in order to establish (11) for all symmetric BICs, it suffices to verify it on the subclass consisting only of BSCs:

Theorem 5. *The polarization process is f-improving on the class of all symmetric BICs if and only if* (11) *holds for all pairs* $(W_1, W_2) = (\mathrm{BSC}(p), \mathrm{BSC}(q))$, $(p,q) \in [0, \frac{1}{2}] \times [0, \frac{1}{2}]$.

Proof. If (11) holds for all symmetric BICs, then it holds for all BSCs. To prove the converse, fix two symmetric BICs W_1, W_2. By Theorem 2 we can write

$$W_1 = \bigoplus_{i=1}^m \lambda_i \, \mathrm{BSC}(p_i) \quad \text{and} \quad W_2 = \bigoplus_{j=1}^n \mu_j \, \mathrm{BSC}(q_j).$$

By Corollary 2,

$$\mathsf{m}_{W_1 \boxast W_2} = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \mathsf{m}_{\mathrm{BSC}(p_i \star q_j)}$$

and

$$\mathsf{m}_{W_1 \circledast W_2} = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \mathsf{m}_{\mathrm{BSC}(p_i) \times \mathrm{BSC}(q_j)}.$$

Consequently, using definitions and the assumption that (11) holds for all BSCs, we have

$$\begin{aligned}
&\mathrm{I}_f(W_1 \boxast W_2) + \mathrm{I}_f(W_1 \circledast W_2) \\
&= \int f \mathrm{d}\mathsf{m}_{W_1 \boxast W_2} + \int f \mathrm{d}\mathsf{m}_{W_1 \circledast W_2} \\
&= \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \int f \left(\mathrm{d}\mathsf{m}_{\mathrm{BSC}(p_i \star q_j)} + \mathrm{d}\mathsf{m}_{\mathrm{BSC}(p_i) \times \mathrm{BSC}(q_j)} \right) \\
&= \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \Big(\mathrm{I}_f(\mathrm{BSC}(p_i \star q_j)) + \mathrm{I}_f(\mathrm{BSC}(p_i) \times \mathrm{BSC}(q_j)) \Big) \\
&\geq \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \Big(\mathrm{I}_f(\mathrm{BSC}(p_i)) + \mathrm{I}_f(\mathrm{BSC}(q_j)) \Big) \\
&= \sum_{i=1}^m \lambda_i \mathrm{I}_f(\mathrm{BSC}(p_i)) + \sum_{j=1}^n \mu_j \mathrm{I}_f(\mathrm{BSC}(q_j)) \\
&= \mathrm{I}_f(W_1) + \mathrm{I}_f(W_2).
\end{aligned}$$

□

It is tempting to conjecture that this property holds for all convex f; if this were the case, then, by the Blackwell–Sherman–Stein theorem, the channel $\frac{1}{2}(W_1 \boxast W_2) \oplus \frac{1}{2}(W_1 \circledast W_2)$ would dominate the channel $\frac{1}{2} W_1 \oplus \frac{1}{2} W_2$. However, this conjecture turns out to be false. As a counterexample, consider the function $f_\lambda(s) = \bar{\lambda} \wedge \lambda - (\bar{\lambda} s) \wedge (\lambda \bar{s})$ for $\lambda \in (0,1)$. Then a simple calculation shows that, for instance,

$$\begin{aligned}
\mathrm{I}_{f_{1/3}}(\mathrm{BSC}(1/4) \boxast \mathrm{BSC}(1/4)) + \mathrm{I}_{f_{1/3}}(\mathrm{BSC}(1/4) \circledast \mathrm{BSC}(1/4)) \\
< 2\mathrm{I}_{f_{1/3}}(\mathrm{BSC}(1/4)).
\end{aligned}$$

Acknowledgment

The author would like to thank E. Arıkan for his comments.

References

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

[2] M. Alsan, "Extremal properties of Gallager's E_0 under the basic polarization transformations," *IEEE Trans. Inform. Theory*, vol. 60, no. 3, pp. 1582–1591, March 2014.

[3] M. Alsan and E. Telatar, "Polarization improves E_0," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2714–2719, May 2014.

[4] M. Alsan, "Re-proving channel polarization theorems: An extremality and robustness analysis," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2014.

[5] D. Blackwell, "Comparison of experiments," in *Proc. 2nd Berkeley Symp. Math. Statist. Prob.*, 1951, pp. 93–102.

[6] ——, "Equivalent comparisons of experiments," *Ann. Math. Statist.*, vol. 24, pp. 265–272, 1953.

[7] I. Sutskover, S. Shamai (Shitz), and J. Ziv, "Extremes of information combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1313–1325, April 2005.

[8] E. Torgersen, *Comparison of Statistical Experiments*. Cambridge Univ. Press, 1991.

[9] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge Univ. Press, 2008.

[10] E. Arıkan, "Varentropy decreases under the polar transform," *IEEE Trans. Inform. Theory*, 2016, to appear.

[11] I. Land and J. Huber, "Information combining," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 3, pp. 227–330, November 2006.

[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.