Alignment of Polarized Sets

Joseph M. Renes*, David Sutter*, and S. Hamed Hassani†
*Institute for Theoretical Physics, ETH Zurich, Switzerland
†Department of Computer Science, ETH Zurich, Switzerland
Email: {renes, suttedav}@phys.ethz.ch, hamed@inf.ethz.ch

***Abstract*—Arıkan's polar coding technique is based on the idea of synthesizing n channels from the n instances of the physical channel by a simple linear encoding transformation. Each synthesized channel corresponds to a particular input to the encoder. For large n, the synthesized channels become either essentially noiseless or almost perfectly noisy, but in total carry as much information as the original n channels. Capacity can therefore be achieved by transmitting messages over the essentially noiseless synthesized channels. Unfortunately, the set of inputs corresponding to reliable synthesized channels is poorly understood, in particular how the set depends on the underlying physical channel. In this work, we present two analytic conditions sufficient to determine if the reliable inputs corresponding to different discrete memoryless channels are *aligned* or not, i.e. if one set is contained in the other. Understanding the alignment of the polarized sets is important as it is directly related to universality properties of the induced polar codes, which are essential in particular for network coding problems. Finally we show that these conditions imply that the simple quantum polar coding scheme of Renes *et al.* [Phys. Rev. Lett. 109, 050504 (2012)] requires entanglement assistance for general channels, but also show such assistance to be unnecessary in many cases of interest.**

I. Introduction

In Arıkan's celebrated *polarization phenomenon* [1], applying a specific linear transformation called the *polar transform* to n instances of a binary-input output-symmetric discrete memoryless channel (DMC) W induces n synthesized channels which become either ideal or useless channels as n grows large. More precisely, when assigned with an index, the n induced synthesized channels can be classified into two categories, defining two index sets: the set $\mathcal{D}(\mathsf{W})$ of indices corresponding to good channels and the set $\mathcal{R}(\mathsf{W})$ of indices that belong to bad channels. Polarization is the property that the sizes of these sets satisfy $\lim_{n\to\infty}\frac{1}{n}|\mathcal{D}(\mathsf{W})| = I(\mathsf{W})$ and $\lim_{n\to\infty}\frac{1}{n}|\mathcal{R}(\mathsf{W})| = 1 - I(\mathsf{W})$, and this ensures that polar codes are capacity achieving [1].

However, the structure of $\mathcal{D}(\mathsf{W})$ and $\mathcal{R}(\mathsf{W})$ is poorly understood. In particular, the dependency on W is difficult to analyze in general. For V a binary-input output-symmetric DMC different from W, it is unclear if $\mathcal{D}(\mathsf{W})$ and $\mathcal{D}(\mathsf{V})$ are *aligned* or not, i.e. whether $\mathcal{D}(\mathsf{W}) \subseteq \mathcal{D}(\mathsf{V})$ or $\mathcal{D}(\mathsf{W}) \supseteq \mathcal{D}(\mathsf{V})$. An exception is the case when V is assumed to be a *degraded* version of W which implies that $\mathcal{D}(\mathsf{V}) \subseteq \mathcal{D}(\mathsf{W})$ [2]. The methods introduced in [3] can be used to detect nonalignment of $\mathcal{D}(\mathsf{W})$ and $\mathcal{D}(\mathsf{V})$, but not their alignment.

Understanding the structure (and the relation) of the polarized sets $\mathcal{D}(\mathsf{W})$ and $\mathcal{D}(\mathsf{V})$ is important in several respects. First, this is directly linked to the universality of polar codes, if one fixed code can be used for reliable communication over each member of a given class of channels $\mathcal{W}$. Universal codes are important in different coding scenarios, for instance when the statistics of the actual channel are not known precisely. Second, several different channels are simultaneously involved in network coding tasks such as wiretap or broadcast channels, and alignment is helpful in designing efficient polar coding schemes. Third, knowledge of the structure and relation of polarized sets can be helpful in other aspects of polar coding, e.g. in the construction of polar codes (see [4, Chapter 5]).

Polar coding with successive cancellation (SC) decoding is not universal in general [3]. However, universality holds for certain classes of channels with a specific ordering, such as less noisy comparable channels as explained in [5]. There has been recent progress in slightly modifying standard polar codes such that they become universal, however at the cost of larger blocklengths [6], [7]. Therefore it is of interest to have a computationally efficient way to determine if for a given class of channels $\mathcal{W}$ standard polar codes using SC decoding are universal on $\mathcal{W}$ or not.

Contributions.— In this article, we introduce a condition for alignment (Theorem 5) and a condition for nonalignment (Theorem 3) of two arbitrary binary input symmetric channels. Applied to several examples of interest, we show that these conditions are sometimes close in the sense that it can be conclusively determined if there is an alignment of the polarized sets or not. The proof of the alignment bounds is based on the uncertainty principle of quantum mechanics.

Since aligned polarized sets imply that the corresponding polar codes are universal with SC decoding, our conditions can be used to determine if for a given set of DMCs polar codes are universal or not. We also discuss how the alignment bounds derived in this paper can be used to determine if quantum polar codes [8] require entanglement assistance or not.

Notation.— Let $[k] := \{1, \dots, k\}$ for $k \in \mathbb{Z}^+$. For $x \in \mathbb{Z}_2^k$ and $\mathcal{I} \subseteq [k]$ we have $x[\mathcal{I}] = [x_i : i \in \mathcal{I}]$, $x^i = [x_1, \dots, x_i]$ and $x_j^i = [x_j, \dots, x_i]$ for $j \leq i$. For two sets $\mathcal{A}, \mathcal{B} \subseteq [n]$ we write $\mathcal{A} \dot{\subseteq} \mathcal{B}$ meaning that $\mathcal{A}$ is essentially contained in $\mathcal{B}$ or more precisely $|\mathcal{A}\backslash\mathcal{B}| = o(n)$. The complement of a set $\mathcal{A} \subseteq [n]$ is denoted by $\bar{\mathcal{A}} := [n]\backslash\mathcal{A}$. All logarithms in this article are with respect to the basis 2. For $\alpha \in [0, 1]$, $H_{\mathrm{b}}(\alpha) := -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ denotes the binary entropy function. We denote the Bhattacharyya parameter of a binary-input discrete memoryless channel $\mathsf{W} : \{0, 1\} \to \mathcal{Y}$ by $Z(\mathsf{W}) := \sum_{y\in\mathcal{Y}} \sqrt{\mathsf{W}(y|0)\mathsf{W}(y|1)} \in [0, 1]$. For some binary

string $b \in \{0,1\}^k$ we denote its binary complement by $\bar{b}$. The logical *and* is denoted by $\wedge$ and the logical *or* by $\vee$. The binary symmetric channel with transition probability $\alpha \in [0, \frac{1}{2}]$ is abbreviated by BSC(α) and the binary erasure channel with erasure probability $\beta \in [0,1]$ is denoted by BEC(β). The space of all Hermitian operators on a finite dimensional Hilbert space $\mathcal{H}$ is denoted by H. We denote the set of density operators on a Hilbert space $\mathcal{H}$ by $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathrm{H} : \rho \geq 0, \mathrm{tr}[\rho] = 1\}$. For a density operator $\rho \in \mathcal{S}(\mathcal{H})$ we define its von Neumann entropy by $H(\rho) := -\mathrm{tr}[\rho \log \rho]$. The Pauli matrices are denoted by σ_X, σ_Y and σ_Z. For a matrix $A \in \mathbb{C}^{m \times n}$ the trace norm is defined as $\|A\|_{\mathrm{tr}} := \mathrm{tr}[\sqrt{A^\dagger A}]$. For two maps $\Phi : A \to B$ and $\Theta : B \to C$ the map $\Theta \circ \Phi : A \to C$ denotes the concatenation of Φ with Θ.

II. Preliminaries

Given a binary-input output-symmetric DMC $\mathsf{W} : \{0,1\} \to \mathcal{Y}$, following [1] we define a *channel splitting* map $(\mathsf{W}, \mathsf{W}) \to (\mathsf{W}_0, \mathsf{W}_1)$ where the synthesized channels $\mathsf{W}_0 : \{0,1\} \to \mathcal{Y}^2$ and $\mathsf{W}_1 : \{0,1\} \to \{0,1\} \times \mathcal{Y}^2$ are given by

$$\mathsf{W}_0(y_1, y_2|u_1) = \sum_{u_2 \in \{0,1\}} \frac{1}{2} \mathsf{W}(y_1|u_1 \oplus u_2) \mathsf{W}(y_2|u_2) \quad (1)$$

and

$$\mathsf{W}_1(y_1, y_2, u_1|u_2) = \frac{1}{2} \mathsf{W}(y_1|u_1 \oplus u_2) \mathsf{W}(y_2|u_2), \quad (2)$$

where u_1, u_2 are (for symmetric channels) assumed to be i.i.d. Bernoulli($\frac{1}{2}$) distributed. The channel splitting map outputs two synthesized channels where W_0 is more noisy and W_1 more reliable than the original channel W. By applying the transform $k = \log n$ times we get n synthesized channels such that in the limit $n \to \infty$ essentially all synthesized channels are either almost noiseless or very noisy [1]. A recursive application of the channel splitting can be visualized in a *polarization tree* that defines the notation of the synthesized channels (cf. Figure 1).

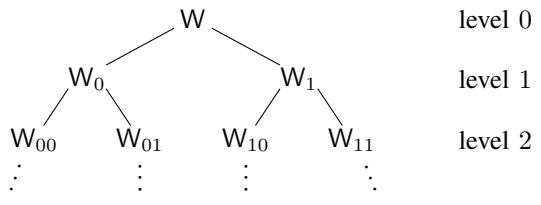

Fig. 1. Polarization tree up to level 2.

Let X^n be a vector with i.i.d. Bernoulli(p) distributed entries for $p \in [0,1]$ and $n = 2^k$ with $k \in \mathbb{N}$. Then, define $U^n = G_n X^n$, where $G_n := \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)^{\otimes \log n}$ denotes the polarization (or polar) transform. Furthermore, let $Y^n = \mathsf{W}^n X^n$, where W^n denotes n independent uses of a DMC $\mathsf{W} : \mathcal{X} \to \mathcal{Y}$ and let $Z^n = \mathsf{V}^n X^n$, where $\mathsf{V} : \mathcal{X} \to \mathcal{Z}$ denotes another DMC. For any $\varepsilon \in (0,1)$ we consider the four sets

$$\mathcal{D}_\varepsilon^n(\mathsf{W}) := \left\{i \in [n] : Z(\mathsf{W}_{b(i-1)}) \leq \varepsilon\right\} \quad (3a)$$

$$\mathcal{R}_\varepsilon^n(\mathsf{W}) := \left\{i \in [n] : Z(\mathsf{W}_{b(i-1)}) \geq 1 - \varepsilon\right\} \quad (3b)$$

$$\mathcal{D}_\varepsilon^n(\mathsf{V}) := \left\{i \in [n] : Z(\mathsf{V}_{b(i-1)}) \leq \varepsilon\right\} \quad (3c)$$

$$\mathcal{R}_\varepsilon^n(\mathsf{V}) := \left\{i \in [n] : Z(\mathsf{V}_{b(i-1)}) \geq 1 - \varepsilon\right\}, \quad (3d)$$

where $b(i)$ for $i \in [n]$ denotes the binary representation of the integer i with $\log n$ bits. The sets $\mathcal{D}_\varepsilon^n(\mathsf{W})$ and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ define a polar code for W respectively V that is reliable using SC decoding. Within this article the parameter $\varepsilon \in (0,1)$ can be arbitrary. As discussed in [1] the error probability of the polar codes for W and V will decay faster for small ε. Therefore this parameter should be chosen as small as possible. As a result, for most applications it is convenient to assume that $\varepsilon = O(2^{-n^\nu})$ for some $\nu < \frac{1}{2}$. We note that in general for an arbitrary DMC W and $\varepsilon \in (0, \frac{1}{2})$ we have $\overline{\mathcal{D}_\varepsilon^n(\mathsf{W})} = \mathcal{R}_{1-\varepsilon}^n(\mathsf{W}) \supsetneq \mathcal{R}_\varepsilon^n(\mathsf{W})$.

Recall that we call two sets, e.g., $\mathcal{D}_\varepsilon^n(\mathsf{W})$ and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ being aligned if $\mathcal{D}_\varepsilon^n(\mathsf{W}) \subseteq \mathcal{D}_\varepsilon^n(\mathsf{V})$ or $\mathcal{D}_\varepsilon^n(\mathsf{W}) \supseteq \mathcal{D}_\varepsilon^n(\mathsf{V})$. We say that these two sets are *essentially aligned* if $\mathcal{D}_\varepsilon^n(\mathsf{W}) \dot{\subseteq} \mathcal{D}_\varepsilon^n(\mathsf{V})$ or $\mathcal{D}_\varepsilon^n(\mathsf{W}) \dot{\supseteq} \mathcal{D}_\varepsilon^n(\mathsf{V})$.

III. Alignment of Polarized Sets

In this section we will state and prove our main results (Theorems 3 and 5), which are two sufficient conditions for the sets $\mathcal{D}_\varepsilon^n(\mathsf{W})$ and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ being aligned or being not aligned (not even essentially). The conditions can be applied to arbitrary DMCs W and V. The first criterion, that is derived in Section III-A and can be used to conclude that $\mathcal{D}_\varepsilon^n(\mathsf{W})$ and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ are not aligned, is based on a simple counting argument using the polarization phenomenon. The second criterion derived in Section III-C that implies that two polarized sets $\mathcal{D}_\varepsilon^n(\mathsf{W})$ and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ are aligned, is more elaborate and uses a particular property of the polarization transformation together with an uncertainty relation from quantum mechanics for which the (classical) channel has to be embedded into a quantum-mechanical channel as explained in Section III-B.

For this reason we have to introduce some basic quantum information theoretic concepts and notations. For a general overview, see [9]. A binary-input classical-quantum (cq) channel $\mathsf{W} : \{0,1\} \ni x \mapsto \rho_x \in \mathcal{S}(\mathcal{H})$ prepares a quantum state ρ_x at the output, depending on a classical input bit x. The analog of the Bhattacharyya parameter for classical channels is the *fidelity* of a cq channel that is defined as $F(\mathsf{W}) := \left\|\sqrt{\rho_0}\sqrt{\rho_1}\right\|_{\mathrm{tr}}$. The symmetric Holevo information is defined as $I(\mathsf{W}) := H(\frac{1}{2}(\rho_0 + \rho_1)) - \frac{1}{2}(H(\rho_0) + H(\rho_1))$. It is straightforward to verify that in case W is a classical binary-input discrete memoryless channel $F(\mathsf{W}) = Z(\mathsf{W})$ and that the symmetric Holevo information coincides with the symmetric mutual information. The polarization process for cq channels works similarly as for classical DMCs [10]. We can define a channel splitting map $(\mathsf{W}, \mathsf{W}) \to (\mathsf{W}_0, \mathsf{W}_1)$, where the synthesized channels $\mathsf{W}_0 : \{0,1\} \to \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ and $\mathsf{W}_1 : \{0,1\} \to \{0,1\} \otimes \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ are properly defined in [10].

Proposition 1. *For two binary-input cq channels* W *and* V *such that* $F(\mathsf{W})+F(\mathsf{V})\leq 1$ *we have* $F(\mathsf{W}_0)+F(\mathsf{V}_1)\leq 1$ *and* $F(\mathsf{W}_1)+F(\mathsf{V}_0)\leq 1$.

Proof: See [5]. ∎

Applying Proposition 1 recursively to the polarization tree given in Figure 1 proves the following corollary.

Corollary 2. *Consider two binary-input cq channels* W *and* V *such that* $F(\mathsf{W})+F(\mathsf{V})\leq 1$. *Then* $F(\mathsf{W}_b)+F(\mathsf{V}_{\bar{b}})\leq 1$ *for all* $b\in\{0,1\}^{\log n}$.[1]

Remark 1. For two binary-input discrete memoryless channels W and V such that $1-I(\mathsf{W})+I(\mathsf{V})\geq 1$, we have $|\mathcal{R}_\varepsilon^n(\mathsf{W})\cap\mathcal{D}_\varepsilon^n(\mathsf{V})|=\Omega(n)$.

Remark 1 follows by the polarization phenomenon [1], [11] which ensures that $n(1-I(\mathsf{W}))=|\mathcal{R}_\varepsilon^n(\mathsf{W})|+o(n)$ and $nI(\mathsf{V})=|\mathcal{D}_\varepsilon^n(\mathsf{V})|+o(n)$. By replacing W and V the same argument shows that $I(\mathsf{W})+1-I(\mathsf{V})\geq 1$ implies $|\mathcal{D}_\varepsilon^n(\mathsf{W})\cap\mathcal{R}_\varepsilon^n(\mathsf{V})|=\Omega(n)$.

A. Sufficient conditions for nonalignment

Let W and V be two binary-input discrete memoryless channels. Remark 1 can be used to derive sufficient conditions for $|\mathcal{R}_\varepsilon^n(\mathsf{W})\cap\mathcal{D}_\varepsilon^n(\mathsf{V})|=\Omega(n)$ and $|\mathcal{D}_\varepsilon^n(\mathsf{W})\cap\mathcal{R}_\varepsilon^n(\mathsf{V})|=\Omega(n)$. In the following we will state the conditions for $|\mathcal{R}_\varepsilon^n(\mathsf{W})\cap\mathcal{D}_\varepsilon^n(\mathsf{V})|=\Omega(n)$ as the conditions for $|\mathcal{D}_\varepsilon^n(\mathsf{W})\cap\mathcal{R}_\varepsilon^n(\mathsf{V})|=\Omega(n)$ follow by the same argument by swapping W and V. We can derive conditions on every level of the polarization tree.

Theorem 3 (Level k condition for no alignment)**.** *Let* $k\in\mathbb{N}_0$ *and* $\varepsilon\in(0,1)$. *If* $1-I(\mathsf{W}_b)+I(\mathsf{V}_b)\geq 1$ *for some* $b\in\{0,1\}^k$, *then* $|\mathcal{R}_\varepsilon^n(\mathsf{W})\cap\mathcal{D}_\varepsilon^n(\mathsf{V})|=\Omega(n)$.

Proof: The level 0 statement follows directly from Remark 1. Remark 1 can be applied at every step of the polarization tree which proves the assertion. ∎

Remark 2 (Criterion for nonalignment cannot get worse for higher levels)**.** Using the identiy $I(\mathsf{W}_0)+I(\mathsf{W}_1)=2I(\mathsf{W})$ we obtain $2\left(1-I(\mathsf{W})+I(\mathsf{V})\right)=1-I(\mathsf{W}_0)+I(\mathsf{V}_0)+1-I(\mathsf{W}_1)-I(\mathsf{V}_1)$, which shows that if the conditions that imply no alignment (cf. Theorem 3) at level k are satisfied they are also satisfied for all levels $\ell\leq k$.[2]

B. Counterpart of a channel

In order to prove the sufficient conditions for alignment of the polarized sets given in Theorem 5, we need the concept of a *quantum counterpart* of a DMC. The quantum counterpart is useful because its information tranmission capabilities are directly related to those of the original channel by uncertainty relations. Such counterpart channels were defined generally in [12, Sec. IIA] and we give a slightly different presentation here.

[1]Recall that for some binary string $b\in\{0,1\}^k$, we denote its complement by $\bar{b}$.

[2]The opposite is not true. Oftentimes the criterion for no alignment becomes strictly better by considering higher levels.

Suppose we are given a binary-input DMC $\mathsf{W}:\{0,1\}\to\mathcal{Y}$ characterized by the transition probabilities $P_{Y|X}(y|x)$ for $x\in\{0,1\}$ and $y\in\mathcal{Y}$. To the input and output alphabets we may associate orthonormal bases of finite-dimensional vector spaces, which we regard as the state spaces of quantum systems. Let the input alphabet correspond to the basis $|x\rangle^A$ of system A and the output alphabet correspond to the basis $|y\rangle^B$ of system B. By defining the quantum states $\varphi_x=\sum_{y\in\mathcal{Y}}P_{Y|X}(y|x)\,|y\rangle\langle y|^B$, it is always possible to embed W into a quantum channel as

$$\mathsf{W}:|x\rangle\langle x|^A\mapsto\varphi_x^B.$$

Indeed, there are many quantum channels with this action, as we have not specified the mapping for quantum states not diagonal in the basis $\{|x\rangle\}$. Since we are modelling a classical channel, the output at B should always be a convex combination of the states φ_x^B, a condition we will take care to enforce in the construction below.

Once in the quantum setting, we may consider the description of W in terms of the *Stinespring dilation* (see [9, Chap. 8]). Let C and D be additional quantum systems isomorphic to B and define the states $|\varphi_x\rangle^{BC}=\sum_{y\in\mathcal{Y}}\sqrt{P_{Y|X}(y|x)}\,|y\rangle^B|y\rangle^C$, which satisfy $\varphi_x^B=\mathrm{tr}_C\Big(|\varphi_x\rangle\langle\varphi_x|^{BC}\Big)$. Then, a Stinespring dilation of W is the partial isometry $U_{\mathsf{W}}^{A\to BCD}$ from A to $B\otimes C\otimes D$ such that

$$U_{\mathsf{W}}^{A\to BCD}|x\rangle^A=|\varphi_x\rangle^{BC}|x\rangle^D. \tag{4}$$

The action of the channel can be expressed in terms of the dilation as mapping any quantum state ρ to $\mathrm{tr}_{CD}[U_{\mathsf{W}}^{A\to BCD}\rho^A(U_{\mathsf{W}}^{A\to BCD})^\dagger]$. The presence of the additional $|x\rangle^D$ ensures that the output states at B are convex combinations of the φ_x, as required.

Using $U_{\mathsf{W}}^{A\to BCD}$ we can define the quantum counterpart to W as

$$\begin{aligned}\mathsf{W}^{\mathsf{c}}&:\{0,1\}\to\mathcal{S}(\mathcal{H})\\ &x\mapsto\sigma_x^{CD}:=\mathrm{tr}_B\Big(U_{\mathsf{W}}^{A\to BCD}|\tilde{x}\rangle\langle\tilde{x}|^A(U_{\mathsf{W}}^{A\to BCD})^\dagger\Big)\end{aligned} \tag{5}$$

for $|\tilde{x}\rangle=\frac{1}{\sqrt{2}}\sum_{z\in\{0,1\}}(-1)^{xz}|z\rangle$. These are the same output states defined in [12, Eq. 6]. The isometry is not unique, but all possible isometries are related by isometries involving the additional systems C and D only, and therefore these isometries do not change the distinguishability of the outputs of the counterpart channel. Up to this freedom, the counterpart channel is essentially unique. An equivalent means of defining the counterpart is via the *channel state*. Define the quantum state

$$|\psi_{\mathsf{W}}\rangle=\tfrac{1}{\sqrt{2}}\sum_{z\in\{0,1\}}|z\rangle^A|\varphi_z\rangle^{BC}|z\rangle^D \tag{6a}$$

$$=\tfrac{1}{\sqrt{2}}\sum_{x\in\{0,1\}}|\tilde{x}\rangle^A|\sigma_x\rangle^{BCD}, \tag{6b}$$

and denote the associated density operator by simply ψ_{W}^{ABCD}. In the second expression we have used $|\sigma_x\rangle^{BCD}=$

$\frac{1}{\sqrt{2}}\sum_z (-1)^{xz}|\varphi_z\rangle^{BC}|z\rangle^D$ for the purification $|\sigma_x\rangle^{BCD}$ of σ_x^{CD}. Then the outputs of W are just $\varphi_z^B = 2\mathrm{tr}_{ACD}[|z\rangle\langle z|^A \psi_W{}^{ABCD}]$, while the outputs of the counterpart W^{c} are $\sigma_x^{CD} = 2\mathrm{tr}_{AB}[|\tilde{x}\rangle\langle\tilde{x}|^A \psi_W{}^{ABCD}]$.

Although defined completely independently, the counterpart and channel synthesis operations in fact have a particular relation to each other. This relation is the basis of the quantum polar coding technique of [8], [12]. For n systems, consider the channel state

$$|\xi_{\mathsf{W}}\rangle = \tfrac{1}{\sqrt{2^n}} \sum_{z^n\in\{0,1\}^n} |z^n\rangle^A |\varphi_{G_n z^n}\rangle^{BC} |G_n z^n\rangle^D \tag{7a}$$

$$= \tfrac{1}{\sqrt{2^n}} \sum_{x^n\in\{0,1\}^n} |\tilde{x}^n\rangle^A \left|\sigma_{G_n^T x^n}\right\rangle^{BCD}. \tag{7b}$$

The action of W_b is $z_j \to \frac{1}{2^{n-1}}\sum_{\bar{z}_i} \left|z_1^{j-1}\right\rangle\left\langle z_1^{j-1}\right|^{A_1^{j-1}} \otimes \varphi^B_{G_n z^n}$ for the $j \in [n]$ such that the binary expansion of $j+1$ is b, where the summation runs over all $z_k \in \{0,1\}$ for $k \neq j$ [12]. Observe that the output is obtained from ξ_{W} by projecting the jth system of A onto $|z_j\rangle$, tracing out $A_{j+1}^n CD$ but keeping the first $j-1$ systems of A. In [8], [12] it is shown that the polar transform is transposed for the counterpart, which has the effect of reversing the ordering of inputs. That is, the same position j corresponds to $(\mathsf{W}^{\mathsf{c}})_{\bar{b}}$, and the discussion subsequent to Equation 25 of [12] shows that its action is $x_j \to \frac{1}{2^n}\sum_{\bar{x}_j} \left|\tilde{x}_{j+1}^n\right\rangle\left\langle\tilde{x}_{j+1}^n\right|^{A_{j+1}^n} \otimes U_{\mathrm{enc}}^D \sigma^{CD}_{G_n^T x^n} (U_{\mathrm{enc}}^D)^\dagger$, where U_{enc} is the polar transform as a unitary operation: $U_{\mathrm{enc}}|z^n\rangle = |G_n z^n\rangle$. Up to this unitary, which is irrelevant for the counterpart channel, this output is obtained from ξ_{W} by projecting system A_j onto $|\tilde{x}_j\rangle$, measuring the subsequent $n-j$ systems of A in the $|\tilde{x}\rangle$ basis and tracing out $A_1^{j-1}B$.

On the other hand, the counterpart of W_b involves the mapping

$$|\tilde{x}_j\rangle \to \tfrac{1}{\sqrt{2^n}} \sum_{z^n\in\{0,1\}^n} (-1)^{xz_j} \left|z_1^{j-1}\right\rangle^{A_1^{j-1}} \left|z_1^{j-1}\right\rangle^{D_1^{j-1}} \otimes |z_j\rangle^{D_j} \left|z_{j+1}^n\right\rangle^{D_{j+1}^n} |\varphi_{G_n z^n}\rangle^{BC} \tag{8}$$

where systems $A_1^{j-1}B$ are the outputs of the original channel and CD are the outputs of the counterpart. The output of the counterpart can be obtained from ξ_{W} by again projecting A_j onto $|\tilde{x}_j\rangle$, tracing out $A_1^{j-1}B$, but now leaving the remaining A systems untouched rather than measuring them. This shows that $(\mathsf{W}^{\mathsf{c}})_{\bar{b}}$ is a degraded version of $(\mathsf{W}_b)^{\mathsf{c}}$, since we can measure the systems A_{j+1}^n of the latter to obtain the former.

A useful uncertainty relation constrains the fidelities of the two channels:

Proposition 4. *Let W be a binary-input discrete memoryless channel and W^{c} be its counterpart as defined above. Then for every $b \in \{0,1\}^{\log n}$ we have $F(\mathsf{W}_b) + F((\mathsf{W}^{\mathsf{c}})_{\bar{b}}) \geq 1$.*

Proof: See [5]. ∎

Remark 3. In [5] we explain in detail how to derive the counterpart for three classical DMCs. In particular, we show that

(i) the counterpart of a $\mathsf{W} = \mathrm{BEC}(\beta)$ with $\beta \in [0,1]$ is $\mathsf{W}^{\mathsf{c}} = \mathrm{BEC}(1-\beta)$ and therefore $F(\mathsf{W}^{\mathsf{c}}) = 1-\beta$.
(ii) for $\mathsf{W} = \mathrm{BSC}(\alpha)$ with $\alpha \in [0, \frac{1}{2}]$, $F(\mathsf{W}^{\mathsf{c}}) = 1-2\alpha$.
(iii) for $\mathsf{W} = \mathrm{BEC}(\beta) \circ \mathrm{BSC}(\alpha)$ with $(\alpha,\beta) \in [0,\frac{1}{2}] \times [0,1]$, $F(\mathsf{W}^{\mathsf{c}}) = (1-\beta)(1-2\alpha)$.

C. Sufficient conditions for alignment

Given two binary-input discrete memoryless channels W and V we can use Corollary 2 and Proposition 4 to derive sufficient conditions for $\mathcal{R}_\varepsilon^n(\mathsf{W}) \subseteq \mathcal{R}_\varepsilon^n(\mathsf{V})$ or similarly $\mathcal{R}_\varepsilon^n(\mathsf{W}) \supseteq \mathcal{R}_\varepsilon^n(\mathsf{V})$ by swapping the role of W and V. We can derive such conditions on every level of the polarization tree. With V^{c} we denote the counterpart of channel V as defined in Section III-B.

Theorem 5 (Level k condition for alignment)**.** *Let $k \in \mathbb{N}_0$ and $0 < \varepsilon < 1$. If $F(\mathsf{W}_b) + F((\mathsf{V}^{\mathsf{c}})_{\bar{b}}) \leq 1$ for all $b \in \{0,1\}^k$, then $\mathcal{R}_\varepsilon^n(\mathsf{W}) \subseteq \mathcal{R}_\varepsilon^n(\mathsf{V})$.*

Proof: Consider $n \geq k$ and suppose $d \in \{0,1\}^n$ is such that the synthesized channel W_d is noisy, i.e. $F(\mathsf{W}_d) \geq 1-\varepsilon$. According to Corrolary 2 together with the assumption of the theorem this implies that $F((\mathsf{V}^{\mathsf{c}})_{\bar{d}}) \leq \varepsilon$. Proposition 4 then ensures that $F(\mathsf{V}_d) \geq 1-\varepsilon$. This implies that $\mathcal{R}_\varepsilon^n(\mathsf{W}) \subseteq \mathcal{R}_\varepsilon^n(\mathsf{V})$. ∎

Remark 4 (Criterion for alignment cannot get worse for higher levels)**.** Suppose the sufficient conditions at level 1 in Theorem 5 are satisfied. Then using the inequality $F(\mathsf{W}_0) \leq 2F(\mathsf{W}) - F(\mathsf{W})^2$ and the identity $F(\mathsf{W}_1) = F(\mathsf{W})^2$ [13, Prop. 17] and $0 \leq F(\mathsf{W}) \leq 1$, we obtain

$$F(\mathsf{W}) + F(\mathsf{V}^{\mathsf{c}}) \leq \sqrt{F(\mathsf{W}_1)} + 1 - \sqrt{1 - F(\mathsf{V}_0^{\mathsf{c}})} \leq 1, \tag{9}$$

where the last inequality uses $F(\mathsf{W}_1) + F((\mathsf{V}^{\mathsf{c}})_0) \leq 1$ which is given by assumption. This argument can be applied to each level and thus shows that if the assumptions in Theorem 5 at level k are satisfied they are also satisfied for all levels $\ell \leq k$.

Remark 5 (No improvement after level 0 for BECs)**.** In case W or V is a BEC, the sufficient conditions in Theorem 5 cannot be improved by going to higher levels than level 0. Let W be a $\mathrm{BEC}(\alpha)$. The level 0 condition requires that $\alpha \geq F(\mathsf{V}^{\mathsf{c}})$. One condition of the first level is $Z(\mathsf{W}_0) + F((\mathsf{V}^{\mathsf{c}})_1) \leq 1$. Since W is a BEC we know that $Z(\mathsf{W}_0) = 2Z(\mathsf{W}) - Z(\mathsf{W})^2 = 1-\beta^2$. Moreover $F((\mathsf{V}^{\mathsf{c}})_1) = F(\mathsf{V}^{\mathsf{c}})^2$ and thus as $\beta \in [0,1]$ the condition from level 1 coincides with the one from level 0. This argument carries over to higher levels. Note that in case V is a BEC the same justification can be applied as the counterpart channel of a BEC is a BEC again (see Remark 3).

In [5] it is explained how to apply the alignment bounds for DMCs with a non-uniform input distribution.

IV. Applications

In this section we demonstrate the performance of Theorem 3 and Theorem 5 on the example of a BSC/BEC pair with a uniform input distribution. Let $\mathsf{W} : \mathcal{X} \to \mathcal{Y}$ be a $\mathrm{BSC}(\alpha)$ for $\alpha \in [0, \frac{1}{2}]$ and $\mathsf{V} : \mathcal{X} \to \mathcal{Z}$ be a $\mathrm{BEC}(\beta)$ for $\beta \in [0,1]$.

Consider a uniform input distribution, i.e., $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. According to [14, Ex. 5.4, p. 121] and [5, Prop. 2.2] we know that for $\beta \leq 4\alpha(1-\alpha)$ the channel V is less noisy than W and hence $\mathcal{D}_\varepsilon^n(\mathsf{W}) \dot{\subseteq} \mathcal{D}_\varepsilon^n(\mathsf{V})$ and $\mathcal{R}_\varepsilon^n(\mathsf{W}) \dot{\supseteq} \mathcal{R}_\varepsilon^n(\mathsf{V})$. To determine a region where $\mathcal{R}_\varepsilon^n(\mathsf{W}) \subseteq \mathcal{R}_\varepsilon^n(\mathsf{V})$ we can use the technique derived in Section III-C which ensures that this is the case if $Z(\mathsf{W}) - Z(\mathsf{V}^{\mathsf{c}}) \leq 1$. (This is the condition at level 0.) Recalling that $\mathsf{V}^{\mathsf{c}} = \text{BEC}(1-\beta)$ (see Remark 3) then gives $\beta \geq 2\sqrt{\alpha(1-\alpha)}$. As discussed in Remark 5 this criterion cannot be improved by considering higher levels as the channel V is a BEC. Using the technique explained in Section III-A (cf. Theorem 3) we can determine regions where $|\mathcal{R}_\varepsilon^n(\mathsf{W}) \cap \mathcal{D}_\varepsilon^n(\mathsf{V})| = \Omega(n)$ or $|\mathcal{D}_\varepsilon^n(\mathsf{W}) \cap \mathcal{R}_\varepsilon^n(\mathsf{V})| = \Omega(n)$. Figure 2 summarizes the results about the alignment properties of the polarized sets $\mathcal{R}_\varepsilon^n(\mathsf{W})$, $\mathcal{R}_\varepsilon^n(\mathsf{V})$, $\mathcal{D}_\varepsilon^n(\mathsf{W})$, and $\mathcal{D}_\varepsilon^n(\mathsf{V})$ for all pairs $(\alpha, \beta) \in [0, \frac{1}{2}] \times [0, 1]$.

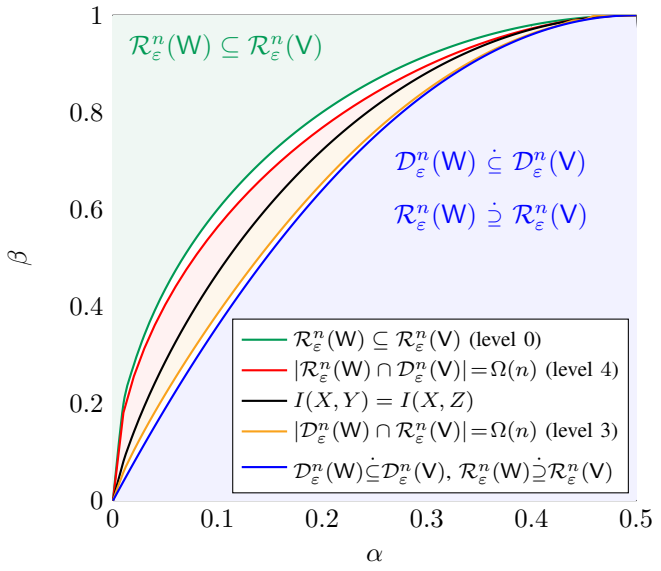

Fig. 2. Alignment of the polarized sets defined in (3) for $\mathsf{W} = \text{BSC}(\alpha)$, $\mathsf{V} = \text{BEC}(\beta)$ with $\alpha \in [0, \frac{1}{2}]$ and $\beta \in [0, 1]$ and a uniform input distribution. The black line shows the region where the two channels have the same capacity, $\beta = H_{\text{b}}(\alpha)$. In the blue region, V is less noisy than W and hence [5, Prop. 2.2] ensures $\mathcal{D}_\varepsilon^n(\mathsf{W}) \dot{\subseteq} \mathcal{D}_\varepsilon^n(\mathsf{V})$ and $\mathcal{R}_\varepsilon^n(\mathsf{W}) \dot{\supseteq} \mathcal{R}_\varepsilon^n(\mathsf{V})$. The remaining colored regions are determined using the conditions given in Theorems 3 and 5 evaluated for different levels.

In [5] we explain how the alignment bounds (i.e., Theorems 3 and 5) can be applied to the setup of a BSC-BEC and a BEC-BSC wiretap channel. We further discuss a BSC/BEC broadcast channel.

V. Entanglement Assistance for Quantum Polar Codes

Recently, the polarization phenomenon has been used to construct efficient codes, quantum polar codes, for transmitting quantum information. These codes inherit several desirable features of (classical) polar codes. In particular, quantum polar codes achieve high rates while allowing for an efficient encoding and decoding [8], [10]. An important open question regards the necessity of *preshared entanglement*: Specifically, whether the coding scheme requires the sender and receiver to share a nonzero amount of maximally entangled states before the protocol begins. In [5], we show that the alignment bounds derived in this paper can be used to determine whether quantum polar codes require entanglement assistance or not. We provide examples of quantum channels where no preshared entanglement is needed (e.g., a low-noise BB84 channel) and examples where entanglement assistance provably is needed (e.g., a high-noise depolarizing channel).

VI. Conclusion

We derived two analytical conditions that can be used to determine the alignment of polarized sets between different DMCs. The condition of Theorem 3 that recognizes situations where there is no alignment (not even essentially) uses a simple counting argument. The condition of Theorem 5, which identifies scenarios where there is an alignment of the polarized sets, is based on the uncertainty principle of quantum mechanics.

Acknowledgments

The authors thank Dominik Waldburger for helpful discussions. JMR and DS were supported by the Swiss National Science Foundation (through the National Centre of Competence in Research 'Quantum Science and Technology') and by the European Research Council (grant No. 258932). SHH thanks ERC Starting Grant under grant number 307036.

References

[1] 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, Jul 2009.

[2] S. B. Korada, "Polar codes for channel and source coding," *PhD thesis, EPFL*, 2009.

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

[4] S. H. Hassani, "Polarization and spatial coupling: Two techniques to boost performance," *PhD thesis, EPFL*, 2013.

[5] J. M. Renes, D. Sutter, and H. Hassani, "Alignment of polarized sets," Nov 2014. available at `arXiv:1411.7925`.

[6] S. H. Hassani and R. Urbanke, "Universal polar codes," July 2013. available at `arXiv:1307.7223`.

[7] E. Sasoglu and L. Wang, "Universal polarization," July 2013. available at `arXiv:1307.7495`.

[8] J. M. Renes, F. Dupuis, and R. Renner, "Efficient polar coding of quantum information," *Physical Review Letters*, vol. 109, p. 050504, Aug 2012.

[9] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, September 2000.

[10] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, pp. 1175 –1187, Feb 2013.

[11] E. Arıkan, "Source polarization," *Proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 899 –903, Jun 2010.

[12] J. M. Renes and M. M. Wilde, "Polar codes for private and quantum communication over arbitrary channels," *IEEE Transactions on Information Theory*, vol. 60, pp. 3090–3103, June 2014.

[13] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner, "Efficient quantum channel coding scheme requiring no preshared entanglement," *Proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 354–358, July 2013. extended version available at `arXiv:1307.1136`.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, Jan 2012.