Duality of channels and codes

Joseph M. Renes

Institute for Theoretical Physics, ETH Zürich, Switzerland

For any given channel W with classical inputs and possibly quantum outputs, a dual classical-input channel $W^\perp$ can be defined by embedding the original into a channel $\mathcal{N}$ with quantum inputs and outputs. Here we give new uncertainty relations for a general class of entropies that lead to very close relationships between the original channel and its dual. Moreover, we show that channel duality can be combined with duality of linear codes, whereupon the uncertainty relations imply that the performance of a given code over a given channel is entirely characterized by the performance of the dual code on the dual channel. This has several applications. In the context of polar codes, it implies that the rates of polarization to ideal and useless channels must be identical. Duality also relates the tasks of channel coding and privacy amplification, implying that the finite blocklength performance of extractors and codes is precisely linked, and that optimal rate extractors can be transformed into capacity-achieving codes, and vice versa. Finally, duality also extends to the EXIT function of any channel and code. Here it implies that for any channel family, if the EXIT function for a fixed code has a sharp transition, then it must be such that the rate of the code equals the capacity at the transition. This gives a different route to proving a code family achieves capacity by establishing sharp EXIT function transitions.

1 Introduction

Duality is an important concept in many branches of mathematics, often enabling given problems to be transformed into dual versions that are simpler to solve. Recently, the author and collaborators have introduced a dual channel in the context of quantum information processing and polar coding [1–4]. The dual construction applies to channels with classical inputs and classical or quantum outputs and is designed so that the original channel and its dual can both be embedded into the same quantum channel. Constraints on the form of quantum channels then lead to nontrivial constraints on the behavior of the channel and its dual.

Here we investigate the notion of duality more comprehensively. We find that it is entirely compatible with the duality of linear codes generally, as well as the notion of channel convolution appearing in belief propagation decoding and polar coding more specifically. Entropic uncertainty relations imply constraints between a wide variety of entropic functions of the channel and code, including EXIT functions. As the class of entropies is quite large, including Rényi entropies for instance, this essentially means that the behavior of a code over a channel is determined by that of the dual code over the dual channel.

Channel duality has several applications, which we briefly describe here by way of outlining the structure of the paper. In the next section we set the mathematical stage and define the class of entropies under consideration. Section 3 is then concerned with the definition and properties of dual channels themselves. In particular, duals of simple classical channels are given, and its relation with channel convolution is established in Theorem 1. The main result of §3, Theorem 2 is a tight entropic uncertainty relation between a channel and its dual. Some implications of this relation are given, such as the precise tradeoff of channel capacities and equality of channel dispersions. Perhaps more importantly, Theorem 2 also implies that the rates of polarization of arbitrary channels to either the ideal or useless channel are in fact identical; this is stated precisely in Corollary 6.

Section 4 considers duality for codes and channels. After examining the notion of code duality in the quantum-mechanical setting, the dual of a encoder and channel combination is shown to be related to randomized encoding of the dual channel in Proposition 4. The tight entropic relation for channels is then extended to the channel and code case in Theorem 3. This implies that channel coding and privacy amplification are closely related, so that randomness extractors can be used to create channel codes, and vice versa, where the error probability of the code is precisely related to the quality of the extracted key. Morever, the optimal finite-blocklength sizes of channel codes and randomness extractors sum precisely to the blocklength, as seen in Corollary 7. This can be used to sharpen bounds on finite blocklength bounds on randomness extraction as illustrated in an example. Finally, duality for EXIT functions is shown in Theorem 4. Combined with capacity duality, it implies that sharp transitions in the EXIT function must occur "at capacity", i.e. at a noise parameter such that the corresponding capacity equals the rate of the chosen code. This replaces the area theorem in locating the transition, as used for instance in the proof by Kudekar *et al.* that Reed-Muller codes achieve capacity over erasure channels [5].

2 Preliminaries
2.1 Mathematical setup

First let us fix the notation used to describe classical random variables, quantum states, and channels of all kinds. For a random variable X over alphabet $\mathcal{X}$, we denote its probability distribution by P_X and the size of its alphabet by $|\mathcal{X}|$. The Hilbert space associated with a quantum system A is denoted $\mathcal{H}_A$ and its dimension $|A|$. The set of density operators on $\mathcal{H}_A$, i.e. the positive semidefinite linear maps from $\mathcal{H}_A$ to itself having unit trace, is denoted $\mathcal{D}(\mathcal{H}_A)$. A channel W which takes $z \in \mathcal{Z}$ to $W(z) \in \mathcal{D}(\mathcal{H})$ is called a classical-quantum or CQ channel. A fully quantum channel, say from $\mathcal{D}(\mathcal{H}_A)$ to $\mathcal{D}(\mathcal{H}_B)$, will be denoted $\mathcal{E}_{B|A}$. This notation mimics the notation for conditional probability distributions, which are channels from classical systems (random variables) to classical systems. In the same spirit, just as P_X is the marginal probability of X when working in the context of the joint distribution P_{XY}, for quantum states ϱ_A is the marginal state when working in the context of a joint state $\varrho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, i.e. $\varrho_A = \mathrm{Tr}_B[\varrho_{AB}]$. We also occasionally abuse notation by referring to a pure state (density operator of rank one) by its nonzero eigenvector in situations calling for density operators, e.g. $|\sigma\rangle_A$ as the input to a channel.

The construction of the dual makes use of two conjugate bases of the input Hilbert space, which are defined using the discrete Fourier transform. Let $\{|z\rangle\}_{z=0}^{d-1}$ be an arbitrary basis of some $\mathcal{H}$ of dimension d and then define the conjugate basis with elements $|\tilde{x}\rangle = \frac{1}{\sqrt{d}} \sum_{z=0}^{d-1} \omega^{xz} |z\rangle$ for $x = 0, \ldots, d-1$ and ω a primitive dth root of unity. We will also make use of the operators $X = \sum_{z=0}^{d-1} |z+1\rangle\langle z|$ and $Z = \sum_{z=0}^{d-1} \omega^z |z\rangle\langle z|$, and we will refer to the $|z\rangle$ basis as the standard basis and $|\tilde{x}\rangle$ as the conjugate basis. Arithmetic inside kets is understood to be modulo d. Observe that $X = \sum_{x=0}^{d-1} \omega^{-x} |\tilde{x}\rangle\langle\tilde{x}|$. The canonical maximally entangled state on $\mathcal{H}_A \otimes \mathcal{H}_{A'}$ with $\mathcal{H}_{A'} \simeq \mathcal{H}_A$ associated with the standard basis is $|\Phi\rangle_{AA'} = \frac{1}{\sqrt{|A|}} \sum_z |z\rangle_A |z\rangle_{A'} = \frac{1}{\sqrt{|A|}} \sum_x |\tilde{x}\rangle_A |-\tilde{x}\rangle_{A'}$.

The use of the Fourier transform is related to treating the input alphabet $\mathcal{Z}$ of CQ channels as an Abelian group. In this setting, it is natural to consider a CQ channel W to be symmetric if there exists a set of unitary transformations U_z for $z \in \mathcal{Z}$ such that $U_{z'} W(z) U_{z'}^* = W(z + z')$ for all $z, z' \in \mathcal{Z}$. Here U^* denotes the adjoint of the map U. Two CQ channels $W : \mathcal{Z} \to \mathcal{D}(\mathcal{H})$ and $W' : \mathcal{Z} \to \mathcal{D}(\mathcal{H}')$ are said to be output equivalent if there exist quantum channels $\mathcal{E} : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H}')$ and $\mathcal{F} : \mathcal{D}(\mathcal{H}') \to \mathcal{D}(\mathcal{H})$ such that $\mathcal{E}(W(z)) = W'(z)$ and $\mathcal{F}(W'(z)) = W(z)$ for all $z \in \mathcal{Z}$. We will regard them as equivalent, denoted $W \simeq W'$, if there exists a bijection T on $\mathcal{Z}$ such that $W' \circ T$ is output equivalent to W. Equivalence in this sense is that of equivalence for coding purposes.

The properties of quantum channels will also be important in the construction of the dual channel. The most important of these is the Stinespring representation theorem, which states that any channel $\mathcal{E}_{B|A} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$ can be represented by an isometry from $\mathcal{H}_A$ to the joint system $\mathcal{H}_{BE} = \mathcal{H}_B \otimes \mathcal{H}_E$, followed by discarding E. Formally, $\mathcal{E}_{B|A}(\varrho_A) = \mathrm{Tr}_E[V_{BE|A} \varrho_A V_{BE|A}^*]$ for some isometry $V_{BE|A}$ and all $\varrho_A \in \mathcal{D}(\mathcal{H}_A)$. Using the representative isometry we can define the complementary channel $\mathcal{E}_{E|A}^\sharp$ by the action $\mathcal{E}_{E|A}^\sharp(\varrho_A) = \mathrm{Tr}_B[V_{BE|A} \varrho_A V_{BE|A}^*]$ for all $\varrho_A \in \mathcal{D}(\mathcal{H}_A)$. (Note that $\sharp$ does not operate on the system labels of the channel, but $*$ does.) The representative isometry is not unique, but for any two isometries $V_{BE|A}$ and $V'_{BE'|A}$ associated with the same channel there exists a partial isometry $U_{E'|E}$ such that $V'_{BE'|A} = U_{E'|E} V_{BE|A}$. (We only require $U_{E'|E}^* U_{E'|E}$ to be a projection onto the image of $V_{BE|A}$.) Thus, the complementary channels are also not unique, though essentially so, as they are all related by the action of partial isometries on the output system.

2.2 Dual entropies

Entropy duality will also play a crucial role in the results, which hold for a wide variety of entropy measures. Following [6], let $\mathsf{D}(\varrho, \sigma)$ for $\varrho \in \mathcal{D}(\mathcal{H})$ and σ a positive operator on $\mathcal{H}$ be a divergence measure which satisfies the following four properties:

1. Monotonicity, or the data-processing inequality: For any channel $\mathcal{E}$, $\mathsf{D}(\varrho, \sigma) \geqslant D(\mathcal{E}(\varrho), \mathcal{E}(\sigma))$,

2. Normalization: For $c > 0$, $\mathsf{D}(\varrho, c\sigma) = \mathsf{D}(\varrho, \sigma) - \log c$,

3. Dominance: For $\sigma' \geqslant \sigma$, $\mathsf{D}(\varrho, \sigma') \leqslant \mathsf{D}(\varrho, \sigma)$, and

4. Zero: $\mathsf{D}(\varrho, \varrho) = 0$.

Using any D we may define two conditional entropies

$$\mathsf{H}_\downarrow(A|B)_\varrho := -\mathsf{D}(\varrho_{AB}, \mathbb{I}_A \otimes \varrho_B), \qquad \text{and} \tag{1}$$

$$\mathsf{H}_\uparrow(A|B)_\varrho := \max_\sigma [-\mathsf{D}(\varrho_{AB}, \mathbb{I}_A \otimes \sigma_B)]. \tag{2}$$

Each has a dual, defined by $H_\downarrow^\perp(A|B)_\varrho := -H_\downarrow(A|C)_\varrho$ for pure ϱ_{ABC}, and similarly for $H_\uparrow^\perp(A|B)_\varrho$.

The standard von Neumann entropy $H(A|B)_\varrho$, defined using the relative entropy $D(\varrho, \sigma) = \text{Tr}[\varrho(\log \varrho - \log \sigma)]$, is self-dual. The optimal σ_B is the marginal ϱ_B, so $H_\downarrow(A|B)_\varrho = H_\uparrow(A|B)_\varrho$. We will also be interested in the dispersion $V(\varrho, \sigma) = \text{Tr}[\varrho(\log \varrho - \log \sigma)^2]$.

There are two especially useful versions of the Rényi entropy in the quantum setting, defined using either the Petz [7] or sandwiched [8, 9] Rényi divergences, respectively:

$$\bar{D}_\alpha(\varrho, \sigma) := \tfrac{1}{\alpha-1} \log \text{Tr}[\varrho^\alpha \sigma^{1-\alpha}] \quad \text{and} \tag{3}$$

$$\tilde{D}_\alpha(\varrho, \sigma) := \tfrac{1}{\alpha-1} \log \text{Tr}[(\sigma^{\frac{1-\alpha}{2\alpha}} \varrho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha]. \tag{4}$$

Both of these satisfy the four properties above (for an excellent overview, see [10]). Various duality relations are known for the various Rényi entropies [8, 11–14]. In particular,

$$(\bar{H}_\alpha^\downarrow)^\perp = \tilde{H}_{2-\alpha}^\downarrow \qquad \alpha \in [0, 2] \tag{5}$$

$$(\tilde{H}_\alpha^\uparrow)^\perp = \tilde{H}_{\alpha/(2\alpha-1)}^\uparrow \qquad \alpha \in [\tfrac{1}{2}, \infty] \tag{6}$$

$$(\bar{H}_\alpha^\uparrow)^\perp = \tilde{H}_{1/\alpha}^\downarrow \qquad \alpha \in [0, \infty]. \tag{7}$$

Especially useful are the min- and max-entropies, $H_{\min} = \tilde{H}_\infty^\uparrow$ and $H_{\max} = \tilde{H}_{1/2}^\uparrow$, which are dual to one another. These can also be directly defined by

$$H_{\min}(A|B)_\psi = \max_{\sigma \in \mathcal{D}(\mathcal{H}_B)} \sup\{\lambda \in \mathbb{R} : \psi_{AB} \leqslant 2^{-\lambda} \mathbb{I}_A \otimes \sigma_B\} \tag{8}$$

$$H_{\max}(A|B)_\psi = \max_{\sigma \in \mathcal{D}(\mathcal{H}_B)} \log |A| \, F(\psi_{AB}, \mu_A \otimes \sigma_B)^2, \tag{9}$$

where $F(\varrho, \sigma) = \|\sqrt{\varrho}\sqrt{\sigma}\|_1$ is the fidelity, the quantum analog of the Bhattacharyya parameter.

In all these examples, the dual entropy is itself known to be a divergence-based entropy for an appropriate choice of divergence. But we could also choose different variants of the Rényi divergence, for which the duals of the associated conditional entropies are not known to themselves come from a divergence, such as the maximal [15] or the reversed sandwiched relative entropy [16], respectively[1]

$$D_\alpha^{\text{maximal}} := \tfrac{1}{\alpha-1} \log \text{Tr}[\sigma^{1/2}(\sigma^{-1/2}\varrho\sigma^{-1/2})^\alpha \sigma^{1/2}] \qquad \text{and} \tag{10}$$

$$D_\alpha^{\text{reverse}} := \tfrac{1}{\alpha-1} \log \text{Tr}\Big[\Big(\varrho^{\frac{\alpha}{2(\alpha-1)}} \sigma \varrho^{\frac{\alpha}{2(\alpha-1)}}\Big)^{1-\alpha}\Big]. \tag{11}$$

Another example is the conditional entropy based on hypothesis testing [17]. Consider the the minimum type-II error in asymmetric hypothesis testing of ϱ versus σ with fixed type-I error,

$$\beta_\varepsilon(\varrho, \sigma) := \min\{\text{Tr}[\Lambda\sigma] : \text{Tr}[\Lambda\varrho] \geqslant 1 - \varepsilon, 0 \leqslant \Lambda \leqslant \mathbb{I}\}. \tag{12}$$

Then the divergence $D_h^\varepsilon(\varrho, \sigma) = -\log \frac{\beta_\varepsilon(\varrho,\sigma)}{1-\varepsilon}$ satisfies the four properties above [18].

Beyond the framework of entropy duality based on relative entropies, another dual pair is given by the *smooth* min- and max-entropies. To define them, first define the purification distance $P(\varrho, \sigma)$ between two states ϱ and σ to be $P(\varrho, \sigma) = \sqrt{1 - F(\varrho, \sigma)^2}$. Then denote by $\mathcal{B}_\varepsilon(\varrho)$ the set of states with distance no larger than ε from ϱ. Finally, we can define the smooth entropies:

$$H_{\min}^\varepsilon(A|B)_\varrho := \max_{\varrho' \in \mathcal{B}_\varepsilon(\varrho)} H_{\min}(A|B)_{\varrho'} \qquad \text{and} \tag{13}$$

$$H_{\max}^\varepsilon(A|B)_\varrho := \min_{\varrho' \in \mathcal{B}_\varepsilon(\varrho)} H_{\max}(A|B)_{\varrho'}. \tag{14}$$

Min- and max-entropies with identical smoothing parameters are dual to one another [19].

Mostly we will be interested in the entropy of a classical random variable conditional on a quantum system, say X given B. This is denoted $H(X|B)_\psi$ where the state ψ_{XB} is the CQ state $\psi_{XB} = \sum_x P_X(x)|x\rangle\langle x|_X \otimes (\varrho_x)_B$ corresponding to the ensemble $\{P_X(x), \varrho_x\}_x$. Often the classical random variable will be the result of measuring a quantum observable, say the observable Z on system A. Overloading notation somewhat, we denote this random variable Z_A and the conditional entropy by $H(Z_A|B)_\psi$, where now ψ denotes the state prior to the measurement. Formally, if $\varrho_{ZB} = \sum_z |z\rangle\langle z|_Z \otimes \text{Tr}_A[|z\rangle\langle z|_A \psi_{AB}]$, then $H(Z_A|B)_\psi = H(Z|B)_\varrho$.

---

[1]These can be shown to satisfy dominance using [6, Lemma 5].

The min-entropy of a CQ state is directly related to the optimal probability of guessing the value of the random variable by making a measurement of the quantum system [12]. Formally, for a CQ state ψ_{XB}, let

$$P(X|B)_\psi := \max_{\Lambda_x} \sum_x P_X(x) \text{Tr}[\Lambda_x \varrho_x], \tag{15}$$

where the optimization is over all POVMs $\{\Lambda_x\}$, i.e. sets of positive operators Λ_x on $\mathcal{H}_B$ such that $\sum_x \Lambda_x = \mathbb{I}$. Then

$$P(X|B)_\psi = 2^{-H_{\min}(X|B)_\psi} \tag{16}$$

Its dual, the max-entropy, is related to the quality of X as a secret key relative to B, as the fidelity measures how close the CQ state is to one in which X is uniform and completely independent of B. This is a form of "decoupling" of X from B. We will denote the decoupling quality as $Q(X|B)_\psi := \max_{\sigma \in \mathcal{D}(\mathcal{H}_B)} F(\psi_{XB}, \mu_X \otimes \sigma_B)^2$. Then, from (9) we have $2^{H_{\max}(X|B)_\psi} = |X|Q(X|B)_\psi$.

3 Dual channels

3.1 Definition and basic properties

The notion of a dual channel based on embedding both the original and dual channels into a single quantum channel is implicit in [1, 2, 20]. Here we follow and add detail to the more explicit presentation of [3, 4]. Consider an arbitrary CQ channel W with classical inputs in an alphabet $\mathcal{Z}$ and quantum outputs which are density operators on the Hilbert space $\mathcal{H}_B$. We can embed W in a quantum channel $\mathcal{N}_{B|A}$ from A to B which measures the quantum input A in the $|z\rangle$ basis and then produces the corresponding output φ_z in B. Formally, this is described by $\mathcal{N}_{B|A}(\varrho_A) = \sum_z \langle z|\varrho|z\rangle (\varphi_z)_B$. The dual channel comes from using the complement of $\mathcal{N}_{B|A}$, restricted to inputs diagonal in the conjugate basis. Formally,

$$W^\perp(x) := \mathcal{N}^\sharp_{E|A}(|\tilde{x}\rangle\langle\tilde{x}|_A). \tag{17}$$

As noted above, the complement is not unique, so the definition in (17) leads to a family of dual channels. Nevertheless, since complementary channels are all related by partial isometries, all possible dual channels are equivalent to one another. For a convenient concrete representation, let $|\varphi_z\rangle_{BD}$ be a purification of φ_z and define the isometry $V_{BCD|A}$ by

$$V_{BCD|A}|z\rangle_A = |z\rangle_C \otimes |\varphi_z\rangle_{BD}. \tag{18}$$

Here CD together form the dilation space E. Defining $|\theta_x\rangle_{BCD} := V_{BCD|A}|\tilde{x}\rangle_A$, the channel outputs are simply $W^\perp(x) = (\theta_x)_{CD}$.

It is also useful to note that we can generate the outputs of W and $W^\perp$ from the following maximally-entangled quantum state $|\psi\rangle_{ABCD}$ by measuring system A appropriately. Using the two expressions for $|\Phi\rangle_{AA'}$ in the standard and conjugate bases, we have

$$|\psi\rangle_{ABCD} = \mathbb{I}_A \otimes V_{BCD|A'}|\Phi\rangle_{AA'} \tag{19a}$$

$$= \frac{1}{\sqrt{d}} \sum_z |z\rangle_A |z\rangle_C |\varphi_z\rangle_{BD} \tag{19b}$$

$$= \frac{1}{\sqrt{d}} \sum_x |-\tilde{x}\rangle_A |\theta_x\rangle_{BCD}. \tag{19c}$$

where we now take V to act on $\mathcal{H}_{A'}$. By the definition of $|\Phi\rangle_{AA'}$, measurement of A in the standard basis $\{|z\rangle\}$ clearly yields $|z\rangle_C |\varphi_z\rangle_{BD}$ for outcome z, and subsequently tracing out CD gives $W(z)$. Meanwhile, measurement of A in the conjugate basis $\{|\tilde{x}\rangle\}$ yields $|\theta_x\rangle_{BCD}$ for outcome $-x$, and subsequently tracing out B gives $W^\perp(x)$. Indeed, we could just as well take this procedure of starting from (19) and measuring appropriately as the definition of the dual, and we will frequently make use of this formulation in the remainder of the paper.

Equivalent channels W and W' have equivalent duals:

Proposition 1. *For any two CQ channels W and W' such that $W \simeq W'$, it holds that $W^\perp \simeq W'^\perp$.*

Proof. Let $\mathcal{N}_{B|A}$ and $\mathcal{N}'_{B'|A}$ be the quantum channels associated to W and W', respectively. By equivalence, there exist channels $\mathcal{E}_{B'|B}$ and $\mathcal{E}'_{B|B'}$ such that $\mathcal{E}_{B'|B} \circ \mathcal{N}_{B|A} = \mathcal{N}'_{B'|A}$ and $\mathcal{E}'_{B|B'} \circ \mathcal{N}'_{B'|A} = \mathcal{N}_{B|A}$. Now suppose $V_{BE|A}$ and $V'_{B'E'|A}$ are Stinespring dilations of $\mathcal{N}_{B|A}$ and $\mathcal{N}'_{B'|A}$, while $U_{B'D'|B}$ and $U'_{BD|B'}$ are dilations of $\mathcal{E}_{B'|B}$ and $\mathcal{E}'_{B|B'}$. By

the first equivance statement, there must exist a partial isometry $T'_{D'E|E'}$ such that $U_{B'D'|B}V_{BE|A} = T'_{D'E|E'}V'_{B'E'|A}$. Similarly, by the second, there exists a partial isometry $T_{DE'|E}$ such that $U'_{BD|B'}V'_{B'E'|A} = T_{DE'|E}V_{BE|A}$. Hence we can define $\mathcal{F}_{E'|E}(\cdot) = \mathrm{Tr}_D[T(\cdot)T^*]$ and $\mathcal{F}'_{E|E'}(\cdot) = \mathrm{Tr}_{D'}[T'(\cdot)T'^*]$ to satisfy $W'^\perp = \mathcal{F}\circ W^\perp$ and $W^\perp = \mathcal{F}'\circ W'^\perp$, where V and V' are used to define the complements in the duals $W^\perp$ and $W'^\perp$. $\qquad\square$

Because the channel $\mathcal{N}_{B|A}$ measures the input in the $|z\rangle$ basis, the outcome $|z\rangle$ shows up in the Stinespring isometry. Therefore it is in some sense copied to the output of $W^\perp$. This leads to symmetry of the dual channel, which is present even if the original channel W is not symmetric. Specifically, an easy calculation shows that $|\theta_x\rangle_{BCD} = Z_C^x|\theta_0\rangle_{BCD}$, and therefore the value of x modulates system C with the unitary operator Z and doesn't involve B or D. Thus, the dual channel has a simple group covariance structure $W^\perp(x) = Z_C^x W^\perp(0)Z_C^{-x}$, irrespective of the properties of W.

This also immediately implies that $(W^\perp)^\perp \not\simeq W$ in general. However the dual of the dual is the symmetrized version W_{sym} of W, in the sense of [21, Definition 1.3]. More specifically, for a general CQ channel W, let W_{sym} be defined by $W_{\mathrm{sym}}(z) = \frac{1}{|\mathcal{Z}|}\sum_{z'}|z+z'\rangle\langle z+z'|\otimes W(z')$. Then we have

Proposition 2. *For any CQ channel W, $(W^\perp)^\perp \simeq W_{\mathrm{sym}}$. If W is symmetric, then $(W^\perp)^\perp \simeq W$.*

Proof. Iterating the above construction of the dual, it follows that the output of $(W^\perp)^\perp$ for input y is the $B'B$ marginal of the state

$$|\xi_y\rangle_{BB'CD} = \frac{1}{\sqrt{d}}\sum_x \omega^{xy}|x\rangle_{B'}|\theta_x\rangle_{BCD} \tag{20a}$$

$$= \frac{1}{d}\sum_{xz}\omega^{x(y+z)}|x\rangle_{B'}|z\rangle_C|\varphi_z\rangle_{BD}. \tag{20b}$$

Direct calculation gives $(\xi_y)_{BB'} = \frac{1}{d}\sum_z |\widetilde{y+z}\rangle\langle\widetilde{y+z}|_{B'}\otimes\varphi_z$, which is the output of $W_{\mathrm{sym}}(y)$, up to Fourier transform on B'.

For symmetric W, suppose we perform a controlled-unitary operation on $B'B$ which applies U^*_{y+z} to B when B' is in the state $|\widetilde{y+z}\rangle\langle\widetilde{y+z}|$. This produces $(\xi'_y)_{B'B} = \mu_{B'}\otimes(\varphi_{-y})_B$. Since we can invert the input to W and append $\mu_{B'}$ to obtain this state, $(W^\perp)^\perp$ is equivalent to W. $\qquad\square$

3.2 Duals of classical channels

The dual of a classical channel has a particular form. Suppose that the φ_z are determined by a conditional probability distribution $P_{Y|Z}$ in the sense that $\varphi_z = \sum_y P_{Y|Z=z}(y)|y\rangle\langle y|$, and define the unnormalized states $|\eta_y\rangle = \frac{1}{\sqrt{d}}\sum_z \sqrt{P_{Y|Z=z}(y)}|z\rangle$. Computing $V|\tilde{0}\rangle$, we find

$$V|\tilde{0}\rangle = \sum_y |\eta_y\rangle_C|y\rangle_B|y\rangle_D. \tag{21}$$

Observe that the norm $\langle\eta_y|\eta_y\rangle$ is just $\frac{1}{d}\sum_z P_{Y|Z=z}(y)$, i.e. $P_Y(y)$ assuming that Z is uniformly distributed. Now we can write a useful form for θ_x:

$$\theta_x = \sum_y Z_C^x|\eta_y\rangle\langle\eta_y|_C Z_C^{-x}\otimes|y\rangle\langle y|_D. \tag{22}$$

Thus, the dual of a classical channel outputs two systems, one classical and one quantum. The former (system D) records the classical value of y, while the latter (system C) is a pure state $|\eta_y\rangle$ which has been modulated by Z according to the value of x.

In case the quantum outputs of the channel are commuting states, we can regard the channel as a classical channel. This can only happen in two ways: either $P_{Z|Y=y}$ is uniform, in which case the outputs are orthogonal states, or $P_{Z|Y=y}$ is concentrated on only one value of Z, and the output is the same for all inputs. In the former case, the input is completely recoverable from the output in C, while for the latter recovery better than blind guessing is completely impossible. Thus the only classical channels which have classical duals are erasure-like channels in which the output y in D indicates whether the input is perfectly recoverable from C or has been essentially erased.

For binary-input channels, since the quantum outputs $Z^x|\eta_y\rangle$ are pure states, they are completely characterized by their overlap $\cos\vartheta_y = |\langle\eta_y|Z|\eta_y\rangle|/\langle\eta_y|\eta_y\rangle$. Working this out explicitly, one finds

$$\cos\vartheta_y = |P_{Z|Y=y}(0) - P_{Z|Y=y}(1)|, \tag{23}$$

which is just the $|D|$-*density* of the original binary-input classical channel [22, Eq. 4.12]. The duals of binary symmetric and erasure channels are computed in [4], and the results can be immediately understood using (23). The BEC is its own dual, in the sense that $\mathrm{BEC}(p)^{\perp} = \mathrm{BEC}(1-p)$. This can be seen because the overlaps are either zero ($y =?$) or one ($y = 0, 1$), corresponding to quantum outputs $\frac{1}{\sqrt{2}}(|0\rangle + (-1)^x|1\rangle)$ or $|0\rangle$ respectively. Thus, the value of x is perfectly recoverable when $y =?$ but not at all when $y = 0, 1$. The dual of the BSC, meanwhile, has pure state outputs (up to equivalence), taking x to $\sqrt{p}|0\rangle + (-1)^x\sqrt{1-p}|1\rangle$. Again using (23), it is clear that the overlap $|1 - 2p|$ is the same for both values of y, so we can dispense with this part of the output of the dual. Comparing to the form of (22), the fact that that any symmetric binary-input classical channel can be thought of as a heralded mixture of BSCs (see, e.g. [23]) is reflected in the fact that its dual is a heralded mixture of BSCs.

Constructing the dual of the binary-input additive white Gaussian noise channel, $W : z \to y = z + N$ with N normally-distributed, is more subtle, because strictly speaking the framework above does not apply. Since the outputs are continuous, we would like to use $|y\rangle$ for $y \in \mathbb{R}$, but this is not a proper basis set. Put differently, $\varphi_z = \int \mathrm{d}y \, P_{Y|Z=z}(y)|y\rangle\langle y|$ is not a proper density operator. Nonetheless, the ultimate result will look essentially the same: The dual $W^{\perp}$ will take x to a joint classical-quantum system, a classical random variable Y governed by $P_Y(y)$ and an associated qubit C in the state $Z^x|\eta_y\rangle$. The precise details of the construction will be reported elsewhere.

3.3 Extremality of the BSC and its dual

The BSC and its dual are extremal binary-input channels in the following sense. First, any such channel can be degraded to a BSC, simply by performing the optimal measurement for distinguishing the outputs. This operation preserves the trace distance (the quantum analog of the variational distance) of the two outputs, $\delta(W) := \frac{1}{2}\|\varphi_0 - \varphi_1\|_1$, since the optimal measurement is known to have an error probability of $\frac{1}{2}(1 - \delta(W))$ [24, 25]. Let us denote this channel by W_{BSC}. Similarly, any binary-input channel can be upgraded to a channel with pure state outputs, simply by finding pure states $|\Psi_j\rangle_{BB'}$ such that $\varphi_j = \mathrm{Tr}_B[(\Psi_j)_{BB'}]$ for $j = 0, 1$. By Uhlmann's theorem [26] (see also [27, Theorem 9.4]), it is possible to find purifications such that $F(\varphi_0, \varphi_1) = |\langle\Psi_0|\Psi_1\rangle|$, so this construction preserves the fidelity. Let us denote the resulting pure state channel by W_{pure}, and the fidelity of the outputs of any binary-input channel symmetric W as $F(W)$.

Duality relates these two constructions in an elegant way. To see this, we first state a result shown in the proof of [4, Proposition 3.6], and we include the proof here for completeness.

Proposition 3. *For any binary-input symmetric channel W, $\delta(W) = F(W^{\perp})$.*

Proof. Using Uhlmann's theorem and (19), for unitary U we have,

$$F(W^{\perp}) = \max_U |\langle\theta_0|U_B|\theta_1\rangle_{BCD}| \tag{24a}$$

$$= \max_U \frac{1}{2}\Big| \sum_{z\in\{0,1\}} (-1)^z \langle\varphi_z|U_B|\varphi_z\rangle_{BD}\Big| \tag{24b}$$

$$= \max_U \frac{1}{2}|\mathrm{Tr}[U(\varphi_0 - \varphi_1)]| \tag{24c}$$

Since $\|A\|_1 = \max_U |\mathrm{Tr}[UA]|$, the desired result follows. $\qquad\square$

Now degrade $W^{\perp}$ to a BSC. We have $\delta(W^{\perp}) = \delta((W^{\perp})_{\mathrm{BSC}})$ by the properties of the degrading map, which together with Proposition 3 implies $F(W) = F(((W^{\perp})_{\mathrm{BSC}})^{\perp})$. Since $((W^{\perp})_{\mathrm{BSC}})^{\perp}$ is a pure state channel with the same fidelity as W, it is necessarily equivalent to W_{pure}. Therefore, we have shown

Corollary 1. *For any binary-input symmetric CQ channel W, $W_{pure} \simeq ((W^{\perp})_{BSC})^{\perp}$.*

3.4 Channel convolution

As discussed in [4], the dual is compatible with the notion of channel convolution appearing in the setting of polar codes. Essentially the same notion also appears in belief propagation decoding of general binary linear codes, as the update rules for messages at check and variable nodes [22]. The check ($\boxast$) and variable ($\circledast$) convolutions are defined by

$$[W \circledast W'](z) := W(z) \otimes W'(z), \tag{25}$$

$$[W \boxast W'](z) := \frac{1}{2}W(z) \otimes W'(0) + \frac{1}{2}W(z + 1) \otimes W'(1). \tag{26}$$

In the context of polar coding, the check convolution is precisely the "worse" channel synthesized from W and W', call it $W \boxminus W'$, while for symmetric W and W', the variable convolution is equivalent to the "better" synthesized channel $W \boxplus W'$. Formally,

$$W \boxminus W' = W \circledast W' \qquad \text{and} \tag{27}$$

$$W \boxplus W' \simeq W \circledast W'. \tag{28}$$

To see the latter, first observe that the better channel has outputs

$$[W \boxplus W')](z) = \tfrac{1}{2}|0\rangle\langle 0| \otimes W(z) \otimes W'(z) + \tfrac{1}{2}|1\rangle\langle 1| \otimes W(z+1) \otimes W'(z). \tag{29}$$

Symmetry of W amounts to the existence of a unitary operator U such that $W(z+1) = UW(z)U^*$ for $z = 0, 1$. Therefore, applying U if the first system is in the state $|1\rangle$ and doing nothing otherwise results in the state $\tfrac{1}{2}\sum_{z'}|z'\rangle\langle z'| \otimes W(z) \otimes W'(z)$ for input z. Since the first system is independent of the second two, we have $W \boxplus W' \simeq W \circledast W'$ for symmetric W, as intended.

The compatibility of convolution with the dual is the following theorem.

Theorem 1. *For any two binary-input CQ channels W and W',*

$$(W \circledast W')^{\perp} \simeq W^{\perp} \circledast W'^{\perp} \qquad \text{and} \tag{30}$$

$$(W \circledast W')^{\perp} \simeq W^{\perp} \circledast W'^{\perp}. \tag{31}$$

Proof. Let φ_z and φ'_z be the outputs of W and W', and similarly θ_x and θ'_x the outputs of $W^{\perp}$ and $W'^{\perp}$, respectively. Now consider the states $|\psi\rangle$ and $|\psi'\rangle$ from (19) associated with W and W', and denote the respective systems involved by $ABCD$ and $A'B'C'D'$. Applying a CNOT operation from A' to A yields

$$|\eta\rangle = U_{A' \to A}^{\text{CNOT}}|\psi\rangle_{ABCD}|\psi'\rangle_{A'B'C'D'} \tag{32}$$

$$= \tfrac{1}{2}\sum_{z,z'}|z+z'\rangle_A|z'\rangle_{A'}|z\rangle_C|z'\rangle_{C'}|\varphi_z\rangle_{BD}|\varphi'_{z'}\rangle_{B'D'}. \tag{33}$$

In the conjugate basis the CNOT gate has the same action as in the standard basis, but with control and target reversed. Therefore we may write

$$|\eta\rangle = \tfrac{1}{2}\sum_{x,x'}|x\rangle_A|x+x'\rangle_{A'}|\theta_x\rangle_{BCD}|\theta'_{x'}\rangle_{B'C'D'}, \tag{34}$$

where we have abused notation by omitting tildes on the A and A' basis states to denote use of the conjugate basis.

The outputs of $W \circledast W'$ can be generated from this state by measuring system A' in the $|z\rangle$ basis, discarding the C and D systems, and making use of channel symmetry. In the binary-input setting, symmetry amounts to the existence of a unitary operator U such that $W(z+1) = UW(z)U^*$ for $z = 0, 1$. Applying U to B conditional on the value of $z + z'$ in A therefore gives the state

$$|\eta'\rangle = \tfrac{1}{2}\sum_{z,z'}|z+z'\rangle_A|z'\rangle_{A'}|z\rangle_C|z'\rangle_{C'}|\varphi_{z'}\rangle_{BD}|\varphi'_{z'}\rangle_{B'D'}. \tag{35}$$

Measuring A' and discarding CD gives output states $\tfrac{1}{2}\sum_z|z\rangle\langle z|_A \otimes (\varphi_{z'})_B \otimes (\varphi'_{z'})_{B'}$ for measurement result z'. Since the A part of the state is independent of the rest, the outputs are equivalent to $[W \circledast W'](z')$. By Proposition 1, the outputs of the dual can therefore be obtained by measuring system A' of $|\eta'\rangle$ in the Fourier-conjugate basis and discarding systems ABB'. But since $|\eta'\rangle$ differs from $|\eta\rangle$ only by a unitary action on AB, which will anyway be discarded, the outputs of the dual can just as well be obtained from $|\eta\rangle$. Using (34), these are easily seen to be just $[W^{\perp} \circledast W'^{\perp}](x)$.

For the second statement, return to (33) and note that measuring A in the $|z\rangle$ basis and discarding $A'CC'DD'$ gives the states $[W \circledast W'](z)$. Thus, the outputs of $(W \circledast W')^{\perp}$ can be generated by measuring A in the $|\tilde{x}\rangle$ basis and discarding BB', for which it is convenient to use (34). Again using channel symmetry to shift the index x' to x in $|\theta'_{x'}\rangle_{B'C'D'}$, the outputs are easily seen to be equivalent to those of $(W^{\perp} \circledast W'^{\perp})$. $\square$

In the context of polar coding over memoryless channels, one considers repeated convolution of a channel with itself, with a random choice of which convolution to use at each step. Theorem 1 immediately gives a duality relation, a weaker version of which was recently used by the author and collaborators to study the capability of polar codes constructed for a given channel to be used for another [4]. Suppose $y^n \in \{0,1\}^n$ for integer $n > 0$ and let $\bar{y}^n = 1^n + y^n$ (understood modulo 2), where 1^n is the length-n string of 1s. Then define W_{y^n} recursively as $W_{y^{n-1}} \circledast W_{y^{n-1}}$ if $b_n = 0$ and $W_{y^{n-1}} \circledast W_{y^{n-1}}$ if $b_n = 1$. Repeatedly applying Theorem 1 gives the following:

Corollary 2. *For any symmetric CQ channel W, $(W_{y^n})^\perp \simeq (W^\perp)_{\bar{y}^n}$.*

This is an improvement over [4], which showed that $(W^\perp)_{\bar{y}^n}$ is a degraded version of $(W_{y^n})^\perp$.

3.5 Entropic relations between a channel and its dual

Entropic uncertainty relations constrain the behavior of a channel by that of its dual. In fact, due to the use of conjugate bases and the form the dual, the entropic uncertainty relations hold with equality, not just as inequalities, as is generally the case. Thus, the behavior of a channel is in fact completely characterized by that of its dual.

For symmetric channels, we are often interested in the conditional entropy of the input given the output, assuming uniform inputs; for the von Neumann or Shannon entropy this leads to the formula for capacity. Let us define $H(W) := H(Z|W(Z))$ for any of the entropy functions considered in §2. Then we have

Theorem 2. *For any CQ channel W with input Z and any conditional entropy H,*

$$H(W) + H^\perp(W^\perp) = \log|Z|. \tag{36}$$

The proof is based entirely on the following uncertainty equality for the kinds of tripartite states that are found in the state-based definition of the dual channel. Indeed, using (19), both entropy terms can be computed from the state $|\psi\rangle_{ABCD}$: $H(W) = H(Z_A|B)_\psi$ while $H^\perp(W^\perp) = H(X_A|CD)_\psi$. Invoking the following lemma with $E = B$ and $F = CD$ gives (36). Its proof is given in Appendix A.

Lemma 1. *For any tripartite pure state $|\psi\rangle_{AEF}$ in which the unnormalized conditional states $(\sigma_z)_F := \mathrm{Tr}_{AE}[|z\rangle\langle z|_A \psi_{AEF}]$ are pairwise disjoint, i.e. $\sigma_z \sigma_{z'} = 0$ for $z \neq z'$, we have*

$$H(Z_A|E)_\psi + H^\perp(X_A|F)_\psi = \log|A|, \tag{37}$$

for H any entropy defined as in (1), (2), (13), or (14).

Theorem 2 has several important implications. First, from duality of min- and max-entropy, the guessing probability of the channel is directly related to the decoupling of the dual channel. For W taking Z to B, let $P(W) = P(Z|W(Z))$ and $Q(W) = Q(Z|W(Z))$, with Z uniformly distributed. Then, taking H to be $H_{\min}$ and $H_{\max}$, respectively, leads to

Corollary 3. *For any CQ channel W,*

$$P(W) = Q(W^\perp) \quad and \tag{38}$$
$$Q(W) = P(W^\perp). \tag{39}$$

Second, from self-duality of the von Neumann entropy, the capacity of the channel is determined by the capacity of the dual, and vice versa. For $I(W) := \max_{P_Z}(H(Z) - H(Z|W(Z)))$, since the optimal input distribution for symmetric channels is the uniform distribution, we have

Corollary 4. *For any symmetric CQ channel W with input Z,*

$$I(W) + I(W^\perp) = \log|Z|. \tag{40}$$

Moreover, the duality of Rényi entropies implies that the dispersions of a channel and its dual are identical. The channel dispersion determines the second order asymptotic behavior of the maximal achievable communication rate as a function of blocklength for large blocklength [28–30], just as the capacity determines the first order behavior. To define the dispersion, let $V(Z|B)_\psi := V(\psi_{ZB}, \mathbb{I}_Z \otimes \psi_B)$ and $V(W) := V(Z|W(Z))$. Then we have

Corollary 5. *For any symmetric CQ channel W,*

$$V(W) = V(W^\perp). \tag{41}$$

This follows by using $H = \bar{H}_\alpha^\downarrow$, which leads to

$$\bar{D}_\alpha(\psi_{Z_A B}, \mathbb{I}_{Z_A} \otimes \psi_B) + \bar{D}_{2-\alpha}(\psi_{X_A CD}, \mathbb{I}_{X_A} \otimes \psi_{CD}) = \log|A|. \tag{42}$$

Then making use of the following, Proposition 4 in [31], gives the desired result.

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\bar{D}_\alpha(\varrho,\sigma)\big|_{\alpha=1} = \tfrac{1}{2}V(\varrho,\sigma).\tag{43}$$

Entropic duality also implies an interesting result on the rate of polarization of a CQ channel under repeated convolution, choosing among the two choices uniformly at random. Suppose Y^n is a random variable with values in $\{0,1\}^n$, each with the same probability. Then W_{Y^n} is a random convolution of W with itself according to the particular sequence Y^n. Depending on the application, one is interested in the probability that the resulting channel W_{Y^n} is either essentially deterministic, in that $H_{\min}(W_{Y^n}) \approx 0$, or essentially random, in that $H_{\max}(W_{Y^n}) \approx 1$. Here, and in the remainder of this section, we take the base of the logarithm to be 2. The former case is useful in constructing codes for noisy channel communication [32] or information reconciliation [33], the latter for lossy compression [34] or wiretap coding [35]. The rate of polarization refers to how fast the min-entropy approaches 0 with increasing n or how fast the max-entropy approaches 1, and in principle the rate of polarization to determinstic channels could be distinct from the rate of polarization to random channels. However, combining Corollary 2 with Theorem 2 implies that the rates must be identical. Thus, it is only necessary to establish the precise rate for only one of them. This is formalized in the following corollary.

Corollary 6. *Let W be any symmetric binary-input CQ channel. For any function f, the following are equivalent:*

$$\lim_{n\to\infty} P[H_{\min}(W_{Y^n}) \leqslant f(n)] = I(W) \qquad and \tag{44}$$

$$\lim_{n\to\infty} P[H_{\max}(W_{Y^n}) \geqslant 1 - f(n)] = 1 - I(W).\tag{45}$$

Similarly, for $I(W) < 1$ and any function g, the following are equivalent:

$$\lim_{n\to\infty} P[H_{\min}(W_{Y^n}) \geqslant g(n)] = 1 \qquad and \tag{46}$$

$$\lim_{n\to\infty} P[H_{\max}(W_{Y^n}) \leqslant 1 - g(n)] = 1.\tag{47}$$

To see this, first apply (44) to the dual channel and use (40) to obtain $\lim_{n\to\infty} P[H_{\min}((W^\perp)_{\bar{Y}^n}) \leqslant f(n)] = 1 - I(W)$. By Theorem 2, the y^n such that $H_{\min}((W^\perp)_{\bar{y}^n}) \leqslant f(n)$ are precisely those for which $H_{\max}(W_{y^n}) \geqslant 1 - f(n)$. This implies (45). The other implications proceed similarly

Note that polarization statements are not typically made in terms of the min- or max-entropies, but in terms of the Bhattacharyya parameter, which in the quantum case is the fidelity of the output states $B(W) := F(W(0), W(1))$. Following the approach of [36], in [37] it is shown that $\lim_{n\to\infty} P[B(W_{Y^n}) \leqslant 2^{-2^{n\beta}}] = I(W)$ for any $\beta < \tfrac{1}{2}$ and any CQ channel W. Conversely, for any $\beta > \tfrac{1}{2}$ we have $\lim_{n\to\infty} P[B(W_{Y^n}) \geqslant 2^{-2^{n\beta}}] = 1$.

We can relate this fidelity to the min-entropy using Lemma 6 of [38]. This gives $P(W) \geqslant 1 - \tfrac{1}{2}B(W)$ and therefore $H_{\min}(W) \leqslant -\log(1 - \tfrac{1}{2}B(W))$, which we can further bound by $B(W)$ itself, since it takes values in $[0,1]$. Hence $f(n) = 2^{-2^{n\beta}}$ for $\beta < \tfrac{1}{2}$ is feasible in Corollary 6. On the other hand, the other bound in Lemma 6 yields $P(W) \leqslant 1 - \tfrac{1}{2}(1 - \sqrt{1 - B(W)^2})$, from which the bound $H_{\min}(W) \geqslant \tfrac{1}{4}B(W)^2$ follows. Thus, $g(n) = 2^{-2(2^{n\beta}-1)}$ for any $\beta > \tfrac{1}{2}$ is feasible in the converse statement. A more refined analysis would presumably show that $g(n)$ has the same form as $f(n)$, but $\beta > \tfrac{1}{2}$, but this is left for future work.

Using an uncertainty relation developed for channel fidelities $B(W)$ leads to a version of Corollary 6 directly in terms of the Bhattacharyya parameter. Proposition 3.6 of [4] shows that $B(W) + B(W^\perp) \geqslant 1$. Thus, $P[B(W_{Y^n}) \geqslant 1 - f(n)] \geqslant P[B((W^\perp)_{\bar{Y}^n}) \leqslant f(n)]$, which implies

$$\lim_{n\to\infty} P[B(W_{Y^n}) \geqslant 1 - f(n)] \geqslant 1 - I(W).\tag{48}$$

A corresponding upper bound follows from the converse bounds on randomness extraction, since exceeding $1 - I(W)$ would give a means of extracting random bits from Z^n in $\psi_{ZB}^{\otimes n}$ which are uncorrelated from B^n at a rate greater than $H(Z|B)_\psi$. Since the fidelity uncertainty relation is an inequality, it can only be used to show that the rate of polarization to deterministic channels implies a polarization rate to random channels, not vice versa, as in Corollary 6.

Lemma 1 can be directly applied to source scenarios of data compression and randomness extraction, as well as to channel scenarios. Importantly, here we are freed from the constraint of symmetry channels and the four corollaries above can be applied to a general state $|\psi\rangle_{ABCD} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|z\rangle_C|\varphi_z\rangle_{BD}$ for arbitrary probability distribution P_Z and states $|\varphi_z\rangle$. That is, we need not take P_Z to be uniform, as in (19), though note that X_A is uniform no matter the choice of P_Z. Here we have $P(Z_A|B)_\psi = Q(X_A|CD)_\psi$, $Q(Z_A|B)_\psi =$

$P(X_A|CD)_\psi$ from choosing $H_{\min}$ and $H_{\max}$, $H(Z_A|B)_\psi + H(X_A|CD)_\psi = \log|A|$ from the von Neumann entropy. The dispersion argument goes through as above, so that $V(Z_A|B)_\psi = V(X_A|CD)_\psi$. Finally, the connection between the rates of polarization also goes through, now using $B(W) = 2\sqrt{p_0 p_1} F(W(0), W(1))$, and the bound $f(n) = 2^{-2^{n\beta}}$ for the general CQ scenario can be obtained from [38] following [33, 36, 37].

4 Codes and channels

4.1 Codes and complementarity

The notion of duality extends to include linear codes, because a linear code C and its dual $C^\perp$ can be combined into a single quantum code. Here we elucidate this combination by taking a somewhat nonstandard approach to describing a code and its dual. First consider a reversible linear transformation M from $\mathbb{F}_q^n$ to itself, with prime q. We can regard the first $n-k$ outputs of M as defining the parity checks of a linear code C and the remaining k outputs as specifying its encoded information. That is, if we define the $(n-k) \times n$ matrix $\hat{M}$ as the first $n-k$ rows of M and similarly $\bar{M}$ as the last k rows, then $\hat{M}$ is the parity check matrix of C and $\bar{M}$ correspond to the logical bits (message bits). Here we regard the matrix as implementing the linear transformation by acting to the right. Now let $M' = (M^{-1})^T$ and define $\hat{M}'$ and $\bar{M}'$ to be its first $n-k$ and last k rows, respectively. Since $M(M')^T = \mathbb{I}$, $\hat{M}\bar{M}'^T = 0$, and therefore $\bar{M}'$ is the parity check matrix of the dual code $C^\perp$. We will also have occasion to make use of the code $C^\top$, whose parity check matrix is $\bar{M}$. This is the complement of C in $\mathbb{F}_q^n$ in the sense that $\mathbb{F}_q^n = C \oplus C^\top$. The dual also has a complement, call it $C^\perp$, with parity check matrix $\hat{M}'$. Invertibility of M also implies $\bar{M}\hat{M}'^T = 0$, meaning $(C^\top)^\perp = C^\perp$.

We can promote M to a unitary operator U by using M on the standard basis: $U = \sum_{z^n} |Mz^n\rangle\langle z^n|$. The resulting U then necessarily has the action M' in the conjugate basis: $U|x^n\rangle = |(M^{-1})^T x^n\rangle$ (here and subsequently we drop the tilde and always use z refer to the standard basis or x to the conjugate basis). To see this, just use the Fourier transform:

$$U|x^n\rangle = \frac{1}{\sqrt{q^n}} \sum_{z^n} \omega^{x^n \cdot z^n} |Mz^n\rangle \tag{49a}$$

$$= \frac{1}{\sqrt{q^n}} \sum_{z^n} \omega^{x^n \cdot M^{-1} z^n} |z^n\rangle \tag{49b}$$

$$= \frac{1}{\sqrt{q^n}} \sum_{z^n} \omega^{(M^{-1})^T x^n \cdot z^n} |z^n\rangle \tag{49c}$$

$$= |(M^{-1})^T x^n\rangle. \tag{49d}$$

4.2 Encoded channel outputs by measurement

Just as the outputs of a CQ channel W and its dual can be generated by measuring an appropriate state, the same is true for the encoded outputs of W^n. However, there are additional subtleties in the encoded case that are worth exploring first before examining duality.

Suppose we apply U to the A systems in $|\psi\rangle^{\otimes n}$, using $|\psi\rangle$ from (19). Denoting these collectively as A^n, the decomposition of M is mirrored in a similar decomposition of A^n into the first $n-k$ and last k systems, call them $\hat{A}$ and $\bar{A}$, respectively. If we call E_C the encoder of the code C, then from the state

$$|\Psi\rangle_{\hat{A}\bar{A}B^n C^n D^n} := U_{\hat{A}\bar{A}|A^n} |\psi\rangle_{ABCD}^{\otimes n} \tag{50}$$

we can generate both the outputs of $W^n \circ E_C$ as well as $(W^\perp)^n \circ E_{C^\perp}$. For $\bar{z} \in \mathbb{F}_q^k$, we have

$$[W^n \circ E_C](\bar{z}) = \operatorname{Tr}_{C^n D^n}\left[(\Pi_0)_{\hat{A}} \otimes |\bar{z}\rangle\langle\bar{z}|_{\bar{A}} \Psi_{\hat{A}\bar{A}B^n C^n D^n}\right]. \tag{51}$$

The projection Π_0 on $\hat{A}$ is just $|0\rangle\langle 0|^{\otimes n-k}$ and ensures that all the parity checks are satisfied and the projection onto $|\bar{z}\rangle_{\bar{A}}$ picks out the term in the superposition corresponding to the input $\bar{z}$. Similarly, for $\hat{x} \in \mathbb{F}_q^{n-k}$,

$$[(W^\perp)^n \circ E_{C^\perp}](\hat{x}) = \operatorname{Tr}_{B^n}\left[|-\hat{x}\rangle\langle-\hat{x}|_{\hat{A}} \otimes (\tilde{\Pi}_0)_{\bar{A}} \Psi_{\hat{A}\bar{A}B^n C^n D^n}\right]. \tag{52}$$

Now the projection $\tilde{\Pi}_0 = |\tilde{0}\rangle\langle\tilde{0}|^{\otimes k}$ on $\bar{A}$ ensures that the parity checks of the dual code are satisfied. By swapping the roles of the $\hat{A}$ and $\bar{A}$ systems, we can equally-well generate the outputs of $W^n \circ E_{C^\top}$ as well as $(W^\perp)^n \circ E_{C^\perp}$.

For symmetric channels, as we are considering here, we need not insist on using setting all the parity checks to zero as opposed to some other value. Put differently, any parity check matrix specifies an entire

family of codes, one for each choice of the parity check values (syndromes), and for symmetric channels all codes lead to equivalent decoding tasks. To see this more formally, define $E_C(\hat{z}, \bar{z}) = M^{-1}(\hat{z} \oplus \bar{z})$, so that the usual encoder is $E_C(\bar{z}) = E_C(0, \bar{z})$. Now let $s = M^{-1}(\hat{z} \oplus 0)$. By linearity $E_C(\hat{z}, \bar{z}) = s + E_C(\bar{z})$, and therefore by channel symmetry there exists an appropriate unitary operator V_s such that

$$W^n(s + E_C(\bar{z})) = V_s W^n(E_C(\bar{z})) V_s^*. \tag{53}$$

Hence the two channels $W^n \circ E_C(0, \cdot)$ and $W^n \circ E_C(\hat{z}, \cdot)$ are equivalent. Moreover, we can allow the syndrome to be chosen randomly, provided that it is also delivered as part of the channel output. That is, $W^n \circ E_C$ is equivalent to the channel W' which takes $\bar{z}$ to the pair $(W^n \circ E_C(\hat{Z}, \bar{z}), \hat{Z})$ for random $\hat{Z}$. This implies that $\mathsf{H}(W^n \circ E_C) = \mathsf{H}(\bar{Z}|B^n\hat{Z})_\Psi$ for all $\hat{z}$. The latter conditional entropy is relevant in the setting of data compression of Z^n, where the decompressor will have access to B^n as well as the compressed output $\hat{Z}$. In particular, $H_{\min}(\bar{Z}|B^n\hat{Z})_\Psi$ characterizes the error probability of the compression task, just as it does for the coding task. In this sense the coding and compression tasks are equivalent for symmetric channels when using linear codes.

Besides deterministic encoding, it is sometimes useful to employ randomized encoding in which the message is fixed but the syndrome is chosen uniformly at random. This is particularly relevant in coding for the wiretap channel, i.e. private classical communication. Randomized encoding will also play an important role in duality. More formally, let $R_C(\bar{z}) = \frac{1}{q^{n-k}} \sum_{\hat{z}} E_C(\hat{z}, \bar{z})$, where the summation is to be understood as the probabilistic mixture of the outputs of $E_C(\hat{z}, \bar{z})$. Since the syndrome is unknown at the channel output, the relevant conditional entropy is $\mathsf{H}(W^n \circ R_C) = \mathsf{H}(\bar{Z}|B^n)_\Psi$. This conditional entropy is also relevant in the setting of randomness extraction from Z^n relative to side information B^n, where now $\bar{Z}$ is the output of the extraction scheme.[2] In particular, $H_{\max}(\bar{Z}|B^n)_\Psi$ directly characterizes the closeness of the output to the ideal output, a uniformly random string uncorrelated with system B^n. As above, randomized encoding and randomness extraction are in this sense equivalent for symmetric channels when using linear codes.

4.3 Duality of deterministic and randomized encoding

Observe that $(W^n \circ E_C)^\perp \not\simeq (W^\perp)^n \circ E_{C^\perp}$ since the input spaces do not match. Nor is it the case that $(W^n \circ E_C)^\perp$ is equal or equivalent to $(W^\perp)^n \circ E_{C^\intercal}$. The latter would require projecting $\hat{A}$ in $|\Psi\rangle$ onto the conjugate basis state $|\tilde{0}\rangle^{\otimes n-k}$, but the construction of $W^n \circ E_C$ uses the projection onto the standard basis state $|0\rangle^{\otimes n-k}$. Instead, the dual converts deterministic encoding to randomized encoding, and vice versa, as formalized in the following.

Proposition 4. *For any CQ channel W and linear code C,*

$$(W^n \circ E_C)^\perp \simeq (W^\perp)^n \circ R_{C^\intercal} \quad and \tag{54}$$

$$(W^n \circ R_C)^\perp \simeq (W^\perp)^n \circ E_{C^\intercal}. \tag{55}$$

Proof. Consider the state $|\Psi'\rangle_{\hat{A}\bar{A}B^n\hat{C}\bar{C}D^n} = U_{\hat{C}\bar{C}|C^n}|\Psi\rangle_{\hat{A}\bar{A}B^nC^nD^n}$, which can be used to compute the two duals in question. More explicitly, we have

$$|\Psi'\rangle_{\hat{A}\bar{A}B^n\hat{C}\bar{C}D^n} = \frac{1}{\sqrt{q^n}} \sum_{\hat{z}, \bar{z}} |\hat{z}\rangle_{\hat{A}}|\bar{z}\rangle_{\bar{A}}|\hat{z}\rangle_{\hat{C}}|\bar{z}\rangle_{\bar{C}}|\varphi_{M^{-1}(\hat{z} \oplus \bar{z})}\rangle_{B^nD^n}. \tag{56}$$

We first prove (54). Nominally, the outputs of $W^n \circ E_C$ are obtained by projecting onto $|0^{n-k}\rangle_{\hat{A}}|\bar{z}\rangle_{\bar{A}}$ and keeping the B^n system. As described above, we can just as well consider the equivalent scenario in which the output is given by projecting onto $|\bar{z}\rangle_{\bar{A}}$ and keeping the $\hat{A}$ and B^n systems. There is no need to measure $\hat{A}$ to remove superpositions between different syndrome values, as these are wiped out when tracing out C^n. Thus, the outputs of the dual $(W^n \circ E_C)^\perp$ are obtained by projecting onto $|-\bar{x}\rangle_{\bar{A}}$ and keeping the C^nD^n systems. The projection gives

$$\langle -\bar{x}|\Psi'\rangle_{\hat{A}\bar{A}B^n\hat{C}\bar{C}D^n} = \frac{1}{\sqrt{q^n}} \sum_{\hat{z}, \bar{z}} \langle -\bar{x}|\bar{z}\rangle|\hat{z}\rangle_{\hat{A}}|\hat{z}\rangle_{\hat{C}}|\bar{z}\rangle_{\bar{C}}|\varphi_{M^{-1}(\hat{z} \oplus \bar{z})}\rangle_{B^nD^n}. \tag{57}$$

Defining $|\sigma_{\hat{z}}\rangle_{\bar{C}B^nD^n} = \frac{1}{\sqrt{q^k}} \sum_{\bar{z}} |\bar{z}\rangle_{\bar{C}}|\varphi_{M^{-1}(\hat{z} \oplus \bar{z})}\rangle_{B^nD^n}$, the dual channel outputs are then given by

$$[(W^n \circ E_C)^\perp](\bar{x}) = Z_{\bar{C}}^{\bar{x}} \left(\frac{1}{q^{n-k}} \sum_{\hat{z}} |\hat{z}\rangle\langle\hat{z}|_{\hat{C}} \otimes (\sigma_{\hat{z}})_{\bar{C}D^n} \right) Z_{\bar{C}}^{-\bar{x}}, \tag{58}$$

---

since tracing out $\hat{A}$ dephases the $\hat{C}$ system.

Meanwhile, the outputs of $(W^\perp)^n \circ R_{C^\intercal}$ are by definition $\varrho_{\bar{x}} = \frac{1}{q^{n-k}} \sum_{\hat{x}} (\theta_{M'^{-1}(\hat{x} \oplus \bar{x})})_{C^n D^n}$, where $\theta_{x^n} = \theta_{x_1} \otimes \cdots \otimes \theta_{x_n}$. By symmetry of the θ_{x_k},

$$\theta_{M'^{-1}(\hat{x} \oplus \bar{x})} = (Z^{M'^{-1}(\hat{x} \oplus \bar{x})})_{C^n} (\theta_{0^n})_{C^n D^n} (Z^{M'^{-1}(\hat{x} \oplus \bar{x})})^*_{C^n}. \tag{59}$$

Observe that applying $U_{\hat{C}\bar{C}|C^n}$ to $Z^{M'^{-1}(\hat{x} \oplus \bar{x})}_{C^n}$ results in $Z^{\hat{x}}_{\hat{C}} \otimes Z^{\bar{x}}_{\bar{C}}$, as might be expected:

$$UZ^{M'^{-1}(\hat{x} \oplus \bar{x})}U^* = \sum_{z^n} \omega^{z^n \cdot M'^{-1}(\hat{x} \oplus \bar{x})} |Mz^n\rangle\langle Mz^n| \tag{60a}$$

$$= \sum_{\hat{z},\bar{z}} \omega^{M^{-1}(\hat{z} \oplus \bar{z}) \cdot M'^{-1}(\hat{x} \oplus \bar{x})} |\hat{z} \oplus \bar{z}\rangle\langle \hat{z} \oplus \bar{z}| \tag{60b}$$

$$= \sum_{\hat{z},\bar{z}} \omega^{\hat{z} \cdot \hat{x} + \bar{z} \cdot \bar{x}} |\hat{z}\rangle\langle \hat{z}| \otimes |\bar{z}\rangle\langle \bar{z}| \tag{60c}$$

$$= Z^{\hat{x}} \otimes Z^{\bar{x}} \tag{60d}$$

Applying $U_{\hat{C}\bar{C}|C^n}$ to the $\varrho_{\bar{x}}$ yields an equivalent set of outputs, namely

$$U_{\hat{C}\bar{C}|C^n}(\varrho_{\bar{x}})_{C^n D^n} U^*_{\hat{C}\bar{C}|C^n} = \frac{1}{q^{n-k}} \sum_{\hat{x}} Z^{\hat{x}}_{\hat{C}} \otimes Z^{\bar{x}}_{\bar{C}} (\theta'_{0^n})_{\hat{C}\bar{C}D^n} Z^{-\hat{x}}_{\hat{C}} \otimes Z^{-\bar{x}}_{\bar{C}}, \tag{61}$$

where we have used $|\theta'_{0^n}\rangle = U_{\hat{C}\bar{C}|C^n}|\theta_0\rangle^{\otimes n}$. More explicitly,

$$|\theta'_{0^n}\rangle_{\hat{C}\bar{C}B^n D^n} = \frac{1}{\sqrt{q^n}} \sum_{\hat{z},\bar{z}} |\hat{z}\rangle_{\hat{C}} |\bar{z}\rangle_{\bar{C}} |\varphi_{M^{-1}(\hat{z} \oplus \bar{z})}\rangle_{B^n D^n} \tag{62a}$$

$$= \frac{1}{\sqrt{q^{n-k}}} \sum_{\hat{z}} |\hat{z}\rangle_{\hat{C}} |\sigma_{\hat{z}}\rangle_{\bar{C}B^n D^n}. \tag{62b}$$

The average over $\hat{x}$ in (61) will dephase the $\hat{C}$ system, leading to equivalent output states identical to those in (58).

To establish (55), first observe that the outputs of $(W^n \circ R_C)$ can be generated by measuring $\bar{A}$ of $|\Psi'\rangle$ in the standard basis and keeping just the B^n systems. Therefore the dual outputs are obtained by measuring $\bar{A}$ in the conjugate basis and keeping the $\hat{A}\hat{C}\bar{C}D^n$ systems. The projection is precisely that of (57), but now since the dual output also includes $\hat{A}$, it can clearly be absorbed into $\hat{C}$. The dual outputs are then just

$$[(W^n \circ R_C)^\perp](\bar{x}) \simeq Z^{\bar{x}}_{\bar{C}} (\theta'_{0^n})_{\hat{C}\bar{C}D^n} Z^{-\bar{x}}_{\bar{C}}. \tag{63}$$

Note that this differs from (58) in that $\hat{C}$ is not dephased. The outputs of $(W^n)^\perp \circ E_{C^\intercal}$ are just $\theta_{M'^{-1}(0^{n-k} \oplus \bar{x})}$. By the calculations above for $(W^\perp)^n \circ R_{C^\intercal}$, these are plainly equivalent to $[(W^n \circ R_C)^\perp](\bar{x})$. $\qquad \square$

4.4 Entropic relations for codes and channels

There are entropic relationships between channels and codes just as there are for bare channels as in Theorem 2. In particular, taking the base of the logarithm to be q, we have

Theorem 3. *For $|\Psi\rangle$ as in (50),*

$$H(\bar{Z}|B^n\hat{Z})_\Psi + H^\perp(\bar{X}|C^n D^n)_\Psi = k \qquad and \tag{64}$$

$$H(\hat{Z}|B^n)_\Psi + H^\perp(\hat{X}|C^n D^n \bar{X})_\Psi = n - k. \tag{65}$$

Then, for any CQ channel W and linear code C,

$$H(W^n \circ E_C) + H^\perp((W^\perp)^n \circ R_{C^\intercal}) = k \qquad and \tag{66}$$

$$H(W^n \circ R_{C^\top}) + H^\perp((W^\perp)^n \circ E_{C^\perp}) = n - k. \tag{67}$$

Proof. The latter two follow from the former by the discussion in §4.2. To establish the former, first observe that the following two statements follow from Lemma 1:

$$H(\bar{Z}|B^n\hat{Z})_\Psi + H^\perp(\bar{X}|C^n D^n \hat{Z})_\Psi = k \qquad and \tag{68}$$

$$H(\hat{Z}|B^n\bar{X})_\Psi + H^\perp(\hat{X}|C^n D^n \bar{X})_\Psi = n - k, \tag{69}$$

Compared to the statements we are trying to prove, here the entropies in the second and first terms are additionally conditioned on $\hat{Z}$ in the first equation and $\bar{X}$ in the second, respectively. This conditioning can be obtained by extending $\hat{A}$ to two copies and conditioning on the first copy in the first term and the second copy in the second. That is, if we define $|\Psi'\rangle_{\hat{A}_1\hat{A}_2\bar{A}B^nC^nD^n}$ by $|\Psi'\rangle = U_{\hat{A}_1\hat{A}_2|\hat{A}}|\Psi\rangle$ for $U_{\hat{A}_1\hat{A}_2|\hat{A}} = \sum_{\hat{z}}|\hat{z}\rangle_{\hat{A}_1}|\hat{z}\rangle_{\hat{A}_2}\langle\hat{z}|_{\hat{A}}$, then $\mathsf{H}(\bar{Z}|B^n\hat{Z})_\Psi = \mathsf{H}(\bar{Z}|B^n\hat{A}_1)_{\Psi'}$ and $\mathsf{H}^\perp(\bar{X}|C^nD^n\hat{Z})_\Psi = \mathsf{H}^\perp(\bar{X}|C^nD^n\hat{A}_2)_{\Psi'}$. Applying Lemma 1 to $|\Psi'\rangle$ with A therein equal to $\bar{A}$ here, $E = \hat{A}_1B^n$ and $F = \hat{A}_2C^nD^n$ gives the first equality, and an entirely similar argument gives the second.

It then remains to show that $\hat{Z}$ is irrelevant in the second term of the first equation and $\bar{X}$ is irrelevant in the first term of the second. We can dispense with $\hat{Z}$ in $\mathsf{H}^\perp(\bar{X}|C^nD^n\hat{Z})_\Psi$ since it can be obtained from C^n anyway; it is redundant. On the other hand, we can dispense with $\bar{X}$ in $\mathsf{H}(\hat{Z}|B^n\bar{X})_\Psi$ because tracing out C^nD^n leaves the $\bar{A}$ system of $\Psi_{\hat{A}\bar{A}B^n}$ in a random $|\bar{z}\rangle\langle\bar{z}|$ state. Thus measurement of $\bar{X}$ results in a random outcome, completely independent of the remaining parts of Ψ. $\qquad\square$

Due to the connections with the source tasks of data compression and randomness extraction, Theorem 3 allows us to convert randomness extractors for symmetric sources into error-correcting codes for symmetric channels and vice versa. This was first suggested by the author in [2], but here we can draw much tighter conclusions. Note that this is a different relation between codes and extractors than that of e.g. Ta-Shma and Zuckerman [39]. Even the setting therein is different; while here we consider extraction from known sources, whereas randomness extraction in the cryptographic literature usually refers to functions which produce randomness from sources that are only guaranteed to have a certain min-entropy. The resulting extractor codes have codewords which are sequences running through the different seed values. This has no analog in the present setting, as there is no seed. Instead, for a length-n code C encoding length-k messages, the corresponding randomness extractor function is given by $f(x^n) = \bar{M}'x^n$. Observe that $\bar{M}'$ is the generator matrix of the code, but is used in the opposite sense by the extractor; messages m^k are encoded as $z^n = m^k\bar{M}'$.

As a simple example of the use of Theorem 3, consider the recent result that Reed-Muller codes achieve the capacity of the binary erasure channel [5]. By duality, this implies that Reed-Muller codes can also extract randomness at the optimal rate from the source describing the joint input and output to the BEC. If $\hat{M}$ is the parity-check matrix of a Reed-Muller code used for error-correction, then the associated extractor function is given by the matrix $\bar{M}'$ acting to the right. This is the parity check matrix of the dual code, which is also a Reed-Muller code. The error of the optimal decoder will translate directly into the quality of the extracted randomness, which follows by choosing $\mathsf{H} = H_{\min}$ in Theorem 3 just as in Corollary 3. The rates of the two procedures are of course also linked by the relationship between $\hat{M}$ and $\bar{M}'$; since the extractor uses the generator matrix of the code, the size of the extractor output is just the size of the code $|C|$. By self-duality of the BEC, a code of size $|C|$ with error ε for $\mathrm{BEC}(p)$ also functions as a randomness extractor of output length $|C|$ and quality $Q = 1 - \varepsilon$ from the source describing $\mathrm{BEC}(1-p)$. In the limit of large blocklength, $|C|$ will tend to $1 - p$ while ε tends to zero.

This argument could just as well be run in reverse to convert an optimal-rate extractor into a capacity-achieving channel code. It would be interesting to further investigate this possibility, for instance to construct an optimal extractor for the dual of the BSC and thereby find a capacity-achieving code for the binary symmetric channel. Indeed, by the arguments in Appendix 7 of [40], this would provide a capacity-achieving code for *all* symmetric binary-input channels.

Just as remarked at the end of §3.5, the first part of Theorem 3 also holds for $|\Psi\rangle = U_{\hat{A}\bar{A}|A}|\psi\rangle^{\otimes n}$ with more general $|\psi\rangle_{ABCD} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|z\rangle_C|\varphi_z\rangle_{BD}$. Thus, not only can we relate entropic properties of the sources involved in data compression and randomness extraction, but also protocols for the two. It turns out that the sizes of the optimal data compression and randomness extraction procedures (using linear codes) satisfy a simple relationship for any blocklength n. As in [41], let $m_\varepsilon^L(Z_A|B)_\psi$ be the minimal compression length of Z^n relative to B^n using a linear code with error ε. By the discussion in §4.2, this is the smallest $|\hat{Z}|$ such that $P(\bar{Z}|B^n\hat{Z})_\Psi \geqslant 1 - \varepsilon$, for $\bar{Z}$ and $\hat{Z}$ obtained from code C. Similarly, following [41], let $\ell_\varepsilon^L(X_A|CD)_\psi$ be the maximal randomness extractable by a linear function from X^n, which is independent of C^nD^n up to quality $1 - \varepsilon^2$. Again by §4.2, this is the largest $|\bar{X}|$ such that $Q(\bar{X}|C^nD^n)_\Psi \geqslant 1 - \varepsilon^2$. This formulation ensures that the purification distance between the actual and ideal outputs is less than ε. Then we have

Corollary 7. *For any state $|\psi\rangle_{ABCD} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|z\rangle_C|\varphi_z\rangle_{BD}$ and any $\varepsilon \in [0,1]$,*

$$m_{\varepsilon^2}^L(Z_A|B)_\psi + \ell_\varepsilon^L(X_A|CD)_\psi = n. \tag{70}$$

Proof. First note that by picking $\mathsf{H} = H_{\min}$ in Theorem 3, we have $P(\bar{Z}|B^n\hat{Z})_\Psi = Q(\bar{X}|C^nD^n)_\Psi$. Now suppose the optimal linear compression procedure with error ε^2 is based on the linear transformation of Z^n to compressed output $\hat{Z}$, so that $P(\bar{Z}|B^n\hat{Z})_\Psi \geqslant 1 - \varepsilon^2$. Thus the transformation of X^n to $\bar{X}$ has $Q(\bar{X}|C^nD^n)_\Psi \geqslant 1 - \varepsilon^2$,

Figure 1: Comparison of finite blocklength bounds on randomness extraction from the CQ state in (71). The ℓ_ε upper and lower bounds are based on information-spectrum quantities. By duality, specifically Corollary 7, tighter bounds are available by appealing to bounds on coding rates for the binary symmetric channel. The BSC metaconverse and lower bound are both based on the hypothesis testing quantity β_ε. The Poltyrev achievability bound, based on weight spectra of linear codes, already matches the metaconverse extremely closely for blocklengths in the hundreds: At 500 the bounds differ by just four bits! However, it becomes time-consuming to compute for blocklengths in the thousands.

meaning $\ell_\varepsilon^L(X_A|CD)_\psi \geqslant n - \log|\hat{Z}| = n - m_{\varepsilon^2}(Z_A|B)_\psi$. For the opposite bound, suppose the optimal linear extraction procedure with parameter ε uses the transformation of X^n to $\bar{X}$, meaning $Q(\bar{X}|C^nD^n)_\Psi \geqslant 1 - \varepsilon^2$. The transformation from Z^n to $\hat{Z}$ satisfies $P(\bar{Z}|B^n\hat{Z})_\Psi \geqslant 1 - \varepsilon^2$, implying $m_\varepsilon^L(Z_A|B)_\psi \leqslant n - \log|\bar{X}| = n - \ell_{\varepsilon^2}^L(X_A|CD)_\psi$. $\qquad\square$

Hence bounds on randomness extraction can be applied to data compression, and vice versa, at least for compression and extraction based on linear codes. For example, this gives a unified derivation of the second-order asymptotic analysis of these tasks in [41]. Starting from Corollary 15 therein, $m_{\varepsilon^2}^L(Z_A|B)_\psi = nH(Z_A|B)_\psi + \sqrt{n}V(Z_A|B)_\psi\Phi^{-1}(1-\varepsilon^2) + O(\log n)$ we immediately have $\ell_\varepsilon^L(X_A|CD)_\psi = n(\log|A| - H(Z_A|B)_\psi) - \sqrt{n}V(Z_A|B)_\psi\Phi^{-1}(1-\varepsilon^2) + O(\log n)$, which upon using the relations $H(Z_A|B)_\psi + H(X_A|CD)_\psi = \log|A|$ and $V(Z_A|B)_\psi = V(X_A|CD)_\psi$ is Corollary 16. Similarly, we could start from Corollary 16 and infer Corollary 15. Note that the restriction to linear compression and extraction schemes does not affect this argument, since the converse of each applies to linear schemes and the achievability statements are established using two-universal hashing, which includes linear schemes.

An interesting question is how the bounds on compression and extraction compare for finite blocklength. For example, we can significantly tighten the bounds on randomness extraction from the CQ ensemble considered in [41]. They consider the state

$$\psi_{XC} = \tfrac{1}{2}\sum_x |x\rangle\langle x|_X \otimes (Z^x|\eta\rangle\langle\eta|Z^x)_C, \tag{71}$$

where $|\eta\rangle = \sqrt{p}|0\rangle + \sqrt{1-p}|1\rangle$. This is precisely the output of the dual to the binary symmetric channel, and since compression is directly related to coding for symmetric channels, we can use bounds on the coding problem to infer bounds on randomness extraction from $\psi^{\otimes n}$. In particular, the metaconverse involving the hypothesis-testing quantity β_ε applies to linear codes [29, 42][43, Lemma 4.7], as does the Poltyrev achievability bound [44]. For versions of these bounds specifically formulated for the BSC, see Theorems 34 and Theorem 35 of [29]. Alternately, the achievability bound of Theorem 9 in [41] also involves β_ε, making it easier to compute though significantly worse than the Poltyrev bound. A comparison of the bounds appears in Figure 1. One could also investigate the comparison in the other direction, using extraction bounds for classical side information generated by the BSC to give bounds on the coding problem for the channel with pure state outputs. A converse for extraction in classical scenarios is formulated in [45, Lemma 19], for instance.

14

Not every extraction problem will be the dual of such a simple classical channel, particularly not one for which very tight bounds can be readily computed. Tomamichel and Hayashi refer to bounds involving the hypothesis-testing quantity, as we have used in the example, as giving a "microscopic" analysis of the coding problem, and hence very tight bounds. But in general such a microscopic analysis cannot be easily performed, and instead one has to rely on a more "macroscopic" approach, their term for employing information-spectrum quantities (whose definition we shall not give here). Indeed, the bounds on extraction in their example are computed using this approach. It would be interesting to compare the performance of their macroscopic bounds on compression and extraction, in particular Theorem 17, in light of Corollary 7. We leave this question to future work.

4.5 EXIT functions

Duality also implies that the EXIT function of a channel and code combination and that of the dual channel and dual code combination sum to a fixed constant, the logarithm of the alphabet size. The EXIT function for a code and channel is defined as follows. Let Z^n be a random codeword in C and denote by Z_i the ith bit of Z^n. For $B^n = W^n(Z^n)$, denote by $B^n_{\sim i}$ everything but the ith B system. Then the EXIT function using entropy H is

$$\Xi_{\mathsf{H}}(W, C) := \frac{1}{n} \sum_{i=1}^{n} \mathsf{H}(Z_i | B^n_{\sim i}). \tag{72}$$

Nominally the EXIT function is defined in terms of the von Neumann or Shannon conditional entropy, but here will consider more general H or $\mathsf{H}^{\perp}$ entropies. For simplicity, we omit the smooth min- and max-entropies and show

Theorem 4. *For any symmetric CQ channel W with input alphabet of size q and linear code C,*

$$\Xi_{\mathsf{H}}(W, C) + \Xi_{\mathsf{H}^{\perp}}(W^{\perp}, C^{\perp}) = \log q, \tag{73}$$

where H is any entropy in (1) or (2).

Proof. By symmetry and the discussion in §4.2, it is sufficient to show

$$\mathsf{H}(Z_i | B^n_{\sim i} \hat{Z})_{\psi^{\otimes n}} + \mathsf{H}^{\perp}(X_i | C^n_{\sim i} D^n_{\sim i} \bar{X})_{\psi^{\otimes n}} = \log q, \tag{74}$$

for ψ from (19) and where Z_i refers to the result of measuring the Z observable of the ith bit of A^n, $\hat{Z}$ to the value of the syndrome measurement, and similarly for X_i and $\bar{X}$. Again the goal is to make use of Lemma 1, though doing so requires a little work.

First note that $\hat{Z}$ can be regarded as a sequence of Z-type operators, usually called stabilizers, one for each of the rows of $\hat{M}$. That is, $\hat{Z} = (\hat{Z}_1, \ldots, \hat{Z}_{n-k})$, where $\hat{Z}_j = Z^{\hat{M}_j}$ and $Z^{\nu} = Z^{\nu_1} \otimes Z^{\nu_2} \otimes \cdots \otimes Z^{\nu_n}$. By employing row reduction, we can assume without loss of generality that $\hat{M}$ has only one 1 in the ith column. This implies that only one of the stabilizers involves the ith qubit, and we can also assume without loss of generality that it is the first. Then $\hat{Z}_1 = Z_i \cdot Z'_1$, for Z'_1 a Z-type operator on the remaining $n-1$ qubits. Let us denote the set of remaining $n-k-1$ stabilizers $\hat{Z}_{\sim 1}$. By a similar procedure we can define $\bar{X}_1 = X_i \cdot X'_1$ and $\bar{X}_{\sim 1}$. The two row reduction procedures are independent, since row reduction does not affect orthogonality.

Since the stabilizers all commute, but X_i and Z_i anticommute, so too do X'_1 and Z'_1. Now use $\mathsf{H}(Z_i | B^n_{\sim i} \hat{Z})_{\psi^{\otimes n}} = \mathsf{H}(Z'_1 | B^{n-1} \hat{Z}_{\sim 1})_{\psi^{\otimes n-1}}$ and $\mathsf{H}^{\perp}(X_i | C^n_{\sim i} D^n_{\sim i} \bar{X})_{\psi^{\otimes n}} = \mathsf{H}^{\perp}(X'_1 | C^{n-1} D^{n-1} \bar{X}_{\sim 1})_{\psi^{\otimes n-1}}$ from the following Lemma 2. Projecting onto fixed values for $\hat{Z}_{\sim 1}$ and $\bar{X}_{\sim 1}$ yields a pure state, and certainly Z'_1 can be obtained by measuring the C^{n-1} appropriately. Thus, we may apply Lemma 1 to complete the proof. $\qquad\square$

Lemma 2. *For a CQ state of the form $\psi_{XYB} = \frac{1}{|X|} \sum_{xz} P_Y(z) |x\rangle\langle x|_X \otimes |x+z\rangle\langle x+z|_Y \otimes (\sigma_z)_B$, let $\varrho_{YB} = \sum_y P_Y(y) |y\rangle\langle y|_Y \otimes (\sigma_y)_B$. Then, for any conditional entropy measure from (1) or (2),*

$$\mathsf{H}(X | YB)_{\psi} = \mathsf{H}(Y | B)_{\varrho} \quad and \tag{75}$$

$$\mathsf{H}^{\perp}(X | YB)_{\psi} = \mathsf{H}^{\perp}(Y | B)_{\varrho}. \tag{76}$$

The proof is given in Appendix B. EXIT functions figure prominently in the study of belief propagation decoding [22], as well as in the recent proof by Kudekar *et al.* that Reed-Muller codes achieve capacity on erasure channels [5]. Let us briefly recall their proof; we will then be able to see how Theorem 4 offers a potential route to generalizing the argument for other channels. The proof is based on the fact that the ith

Figure 2: EXIT function transition and capacity. The figure depicts the capacity of $W(p)$ the BSC with crossover probability p, the capacity of its dual $W(p)^\perp$, as well as a putative EXIT function for a rate $R = 1/2$ code over $W(p)$ and EXIT function of the dual code over $W(p)^\perp$. Here the EXIT function $h(p)$ displays a sharp transition at $p^\star = 0.8$ such that the capacity $I(W(p^\star)) \approx 0.6$ exceeds the rate R. By duality, this implies that the dual code, which also has rate $1/2$, is reliably decodable for values of p (say 0.1) such that the rate exceeds the capacity of $W(p)^\perp$ (≈ 0.47). As this cannot be the case by the converse to the noisy channel coding theorem, it must be that the transition satisfies $I(W(p^\star)) = R$, i.e. $p^\star \approx 0.11$ in this example.

EXIT function (the ith term in (72) using the Shannon entropy) is the error probability of the optimal bitwise decoder for the BEC, so that if the EXIT function is essentially zero, then decoding is reliable. For doubly transitive codes like Reed-Muller codes, the EXIT function is the same for each codeword bit; let us define $h(p) = \Xi_H(\text{BEC}(p), C)$ as the EXIT function (using the Shannon entropy) for a given code C. Kudekar *et al.* show that for doubly-transitive codes $h(p)$ exhibits a sharp transition as p increases, jumping from zero to one in an interval that decreases with the blocklength. The location of the transition depends on the chosen code C, and in particular it must not be so high as to imply that the code is reliably decodable above the capacity of the channel. For $\text{BEC}(p)$ the capacity is $1 - p$, and therefore for given code of rate R the transition $p^\star$ must satisfy $R \leqslant 1 - p^\star$. The area theorem implies that in fact $p^\star = 1 - R$, so Reed-Muller codes achieve capacity.

Theorem 4 offers two means of potentially extending this argument to more general chanels. First, one can shift the problem of showing a transition in the EXIT function to that of the dual. To study the BSC for instance, one could instead look at the EXIT function associated with the state in (71), which may be easier to study with existing tools. Moreover, one can examine EXIT functions for different entropies, for example $H_{\min}$, and still appeal to duality.

Secondly, Theorems 4 and 2 imply that, for any channel, if a sharp transition exists, it should be located at capacity. Here we give a rough sketch of the argument, which is also illustrated in Figure 2. Let us simplify to the case of binary input channels, $q = 2$, and fix a binary code C for use on the family of channels $W(p) = \text{BSC}(p)$, with p the crossover probability. Note that, by Theorem 2, the family $W^\perp(p)$ is decreasingly noisy with increasing p. Defining $h(p) = \Xi_H(W(p), C)$ and $h^\perp(p) = \Xi_H(W(p)^\perp, C^\perp)$, we immediately have $h(p) + h^\perp(p) = 1$ by Theorem 4. Thus, $h(p)$ has a sharp transition if and only if $h^\perp(p)$ does. As before, the transition value $p^\star$ must be constrained by the capacity, else codes with rates exceeding the capacity could still be reliably decoded. (For the strong converse to CQ channel coding, see [46].) This means we must have $R \leqslant I(W(p^\star))$ for the family $W(p)$, as well as $R^\perp \leqslant I(W(p^\star)^\perp)$ for $W(p)^\perp$, where $R^\perp$ is the rate of $C^\perp$. But $R + R^\perp = 1$ by construction, and $I(W(p^\star)) + I(W(p^\star)^\perp) = 1$ by Theorem 2, which implies that $p^\star$ satisfies $R = I(W(p^\star))$.

5 Discussion

We have shown that a channel and its dual are very tightly related by uncertainty relations for a general class of entropies, and that this duality is compatible both with channel convolution as in polar coding and with the use of linear codes and their duals. We have also investigated several consequences of duality, finding applications to the phenomenon of polarization, the relationship between randomness extraction and coding, as well as possible means of showing a code family achieves capacity of a given channel.

We have confined much of our analysis to symmetric channels, though this restriction was seen to be unnecessary in applications to source problems such as compression or randomness extraction. It would be interesting to extend the results to nonsymmetric channels, but there appear to be some obstacles to doing so. The chief difficulty is that we are confined to considering uniform inputs to W in order to employ the definition of the dual using (19). For example, suppose we take W to be the classical Z channel, for which the capacity-achieving distribution is not uniform. In anticipation of using the entropic uncertainty relation, we could consider the state $|\psi\rangle_{ABCD} = \sum_z \sqrt{p_z} |z\rangle_A |z\rangle_C |\varphi_z\rangle_{BD}$ where p_0 and p_1 are the capacity-achieving distribution and φ_z are the Z channel outputs. The capacity itself is then $I(W) = H(Z_A)_\psi - H(Z_A|B)_\psi$, and $H(Z_A|B)_\psi + H(X_A|CD)_\psi = \log 2$ certainly still holds. However, $H(X_A|CD)_\psi$ is no longer immediately related to $W^\perp$. Nevertheless, for questions involving average coding error of W, where a uniform input is appropriate, Theorems 2 and 3 continue to give meaningful relations to the dual.

The proof of Theorem 4 raises an interesting question regarding entropic uncertainty relations that, to the author's knowledge, has somehow eluded previous investigation. Therein it is crucial that the ith output system be excluded from the conditioning system in order to be able to appeal to Lemma 1, though ultimately the quantity $\mathsf{H}(Z_i|B^n\hat{Z})$ is perhaps more relevant. While the lemma is a novel relation in that it holds with equality, it is still in the usual "tripartite" framework of inequality uncertainty relations as discovered in [6, 47–49]. That is, one of two (conjugate) measurements is performed on a system A and we are interested in the conditional entropies $\mathsf{H}(X_A|E)$ and $\mathsf{H}^\perp(Z_A|F)$ for distinct E and F. Using chain rules for the von Neumann entropy, one can convert between the tripartite and "bipartite" version which is a relation between $H(X_A|E)$ and $H(Z_A|E)$. Considering $\mathsf{H}(Z_i|B^n\hat{Z})$ suggests a different scenario intermediate between bipartite and tripartite, where one is interested in a relation involving $\mathsf{H}(X_A|EC)$ and $\mathsf{H}^\perp(Z_A|FC)$. Whether a useful relation exists is an open question.

References

[1] J. M. Renes and J.-C. Boileau, "Physical underpinnings of privacy", Physical Review A **78**, 032335 (2008), arXiv:0803.3096 [quant-ph].

[2] J. M. Renes, "Duality of privacy amplification against quantum adversaries and data compression with quantum side information", Proceedings of the Royal Society A **467**, 1604–1623 (2011), arXiv:1003.0703 [quant-ph].

[3] J. M. Renes and M. M. Wilde, "Polar Codes for Private and Quantum Communication Over Arbitrary Channels", IEEE Transactions on Information Theory **60**, 3090–3103 (2014), arXiv:1201.2906 [quant-ph].

[4] J. M. Renes, D. Sutter, and S. H. Hassani, "Alignment of Polarized Sets", IEEE Journal on Selected Areas in Communications **34**, 224–238 (2016), arXiv:1411.7925 [quant-ph].

[5] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller Codes Achieve Capacity on Erasure Channels", in Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing, STOC '16 (2016), pp. 658–669, arXiv:1601.04689 [cs.IT].

[6] P. J. Coles, R. Colbeck, L. Yu, and M. Zwolak, "Uncertainty Relations from Simple Entropic Properties", Physical Review Letters **108**, 210405 (2012), arXiv:1112.0543 [quant-ph].

[7] D. Petz, "Quasi-entropies for finite quantum systems", Reports on Mathematical Physics **23**, 57–65 (1986).

[8] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties", Journal of Mathematical Physics **54**, 122203 (2013), arXiv:1306.3142 [quant-ph].

[9] M. M. Wilde, A. Winter, and D. Yang, "Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy", Communications in Mathematical Physics **331**, 593–622 (2014), arXiv:1306.1586 [quant-ph].

[10] M. Tomamichel, *Quantum Information Processing with Finite Resources*, Vol. 5, SpringerBriefs in Mathematical Physics (Springer International Publishing, Cham, 2016), arXiv:1504.00233 [quant-ph].

[11] M. Tomamichel, R. Colbeck, and R. Renner, "A Fully Quantum Asymptotic Equipartition Property", IEEE Transactions on Information Theory **55**, 5840–5847 (2009), arXiv:0811.1221 [quant-ph].

[12] R. König, R. Renner, and C. Schaffner, "The Operational Meaning of Min- and Max-Entropy", IEEE Transactions on Information Theory **55**, 4337–4347 (2009), arXiv:0807.1338 [quant-ph].

[13] S. Beigi, "Sandwiched Rényi divergence satisfies data processing inequality", Journal of Mathematical Physics **54**, 122202 (2013), arXiv:1306.5920 [quant-ph].

[14] M. Tomamichel, M. Berta, and M. Hayashi, "Relating different quantum generalizations of the conditional Rényi entropy", Journal of Mathematical Physics **55**, 082206 (2014), arXiv:1311.3887 [quant-ph].

[15] K. Matsumoto, "A new quantum version of f-divergence", (2013), arXiv:1311.4722 [quant-ph].

[16] K. M. R. Audenaert and N. Datta, "α-z-Rényi relative entropies", Journal of Mathematical Physics **56**, 022202 (2015), arXiv:1310.7178 [quant-ph].

[17] L. Wang and R. Renner, "One-Shot Classical-Quantum Capacity and Hypothesis Testing", Physical Review Letters **108**, 200501 (2012), arXiv:1007.5456 [quant-ph].

[18] F. Dupuis, L. Krämer, P. Faist, J. M. Renes, and R. Renner, "Generalized Entropies", in *XVIIth International Congress on Mathematical Physics*, edited by A. Jensen (World Scientific, 2013), pp. 134–153, arXiv:1211.3141 [quant-ph].

[19] M. Tomamichel, R. Colbeck, and R. Renner, "Duality Between Smooth Min- and Max-Entropies", IEEE Transactions on Information Theory **56**, 4674–4681 (2010), arXiv:0907.5238 [quant-ph].

[20] J.-C. Boileau and J. M. Renes, "Optimal State Merging Without Decoupling", in Fourth Workshop on Theory of Quantum Computation, Communication, and Cryptography, Vol. 5906, edited by A. M. Childs and M. Mosca, Lecture Notes in Computer Science (May 8, 2009), p. 76, arXiv:0905.1324 [quant-ph].

[21] S. B. Korada, "Polar codes for channel and source coding", PhD Thesis (EPFL, Lausanne, Switzerland, 2009).

[22] T. Richardson and R. Urbanke, *Modern Coding Theory* (Cambridge University Press, 2008).

[23] I. Land, S. Huettinger, P. A. Hoeher, and J. B. Huber, "Bounds on information combining", IEEE Transactions on Information Theory **51**, 612–619 (2005).

[24] C. W. Helstrom, "Detection theory and quantum mechanics", Information and Control **10**, 254–291 (1967).

[25] C. W. Helstrom, *Quantum detection and estimation theory*, Vol. 123, Mathematics in Science and Engineering (Academic, London, 1976).

[26] A. Uhlmann, "The "transition probability" in the state space of a *-algebra", Reports on Mathematical Physics **9**, 273–279 (1976).

[27] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[28] M. Hayashi, "Information Spectrum Approach to Second-Order Coding Rate in Channel Coding", IEEE Transactions on Information Theory **55**, 4947–4966 (2009), arXiv:0801.2242 [cs.IT].

[29] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime", IEEE Transactions on Information Theory **56**, 2307–2359 (2010).

[30] M. Tomamichel and V. Y. F. Tan, "Second-Order Asymptotics for the Classical Capacity of Image-Additive Quantum Channels", Communications in Mathematical Physics **338**, 103–137 (2015), arXiv:1308.6503 [quant-ph].

[31] S. M. Lin and M. Tomamichel, "Investigating properties of a family of quantum Rényi divergences", Quantum Information Processing **14**, 1501–1512 (2015), arXiv:1408.6897 [quant-ph].

[32] E. Arıkan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels", IEEE Transactions on Information Theory **55**, 3051–3073 (2009), arXiv:0807.3917 [cs.IT].

[33] E. Arıkan, "Source polarization", in Proceedings of the 2010 IEEE International Symposium on Information Theory (2010), pp. 899–903, arXiv:1001.3087 [cs.IT].

[34] S. B. Korada and R. L. Urbanke, "Polar Codes are Optimal for Lossy Source Coding", IEEE Transactions on Information Theory **56**, 1751–1768 (2010), arXiv:0903.0307 [cs.IT].

[35] H. Mahdavifar and A. Vardy, "Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes", IEEE Transactions on Information Theory **57**, 6428–6443 (2011), arXiv:1007.3568 [cs.IT].

[36] E. Arıkan and E. Telatar, "On the rate of channel polarization", in Proceedings of the 2009 IEEE International Symposium on Information Theory (2009), pp. 1493–1495, arXiv:0807.3806 [cs.IT].

[37] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels", IEEE Transactions on Information Theory **59**, 1175–1187 (2013), arXiv:1109.2591 [quant-ph].

[38] J. M. Renes, D. Sutter, F. Dupuis, and R. Renner, "Efficient Quantum Polar Codes Requiring No Preshared Entanglement", IEEE Transactions on Information Theory **61**, 6395–6414 (2015), arXiv:1307.1136 [quant-ph].

[39] A. Ta-Shma and D. Zuckerman, "Extractor codes", IEEE Transactions on Information Theory **50**, 3015–3025 (2004).

[40] E. Şaşoğlu, "Polar Coding Theorems for Discrete Systems", PhD Thesis (EPFL, Lausanne, Switzerland, 2011).

[41] M. Tomamichel and M. Hayashi, "A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks", IEEE Transactions on Information Theory **59**, 7693–7710 (2013), arXiv:1208.1478 [quant-ph].

[42] H. Nagaoka, "Strong converse theorems in quantum information theory", in Proceedings of the ERATO Conference on Quantum Information Science (EQIS), Vol. 33 (2001).

[43] M. Hayashi, *Quantum Information Theory*, Graduate Texts in Physics (Springer, Berlin, Heidelberg, 2017).

[44] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra", IEEE Transactions on Information Theory **40**, 1284–1292 (1994).

[45] M. Hayashi, H. Tyagi, and S. Watanabe, "Secret Key Agreement: General Capacity and Second-Order Asymptotics", IEEE Transactions on Information Theory **62**, 3796–3810 (2016), arXiv:1411.0735 [cs.IT].

[46] A. Winter, "Coding theorem and strong converse for quantum channels", Information Theory, IEEE Transactions on **45**, 2481–2485 (1999), arXiv:1409.2536 [quant-ph].

[47] J. M. Renes and J.-C. Boileau, "Conjectured Strong Complementary Information Tradeoff", Physical Review Letters **103**, 020402 (2009), arXiv:0806.3984 [quant-ph].

[48] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty principle in the presence of quantum memory", Nature Physics **6**, 659–662 (2010), arXiv:0909.0950 [quant-ph].

[49] M. Tomamichel and R. Renner, "Uncertainty Relation for Smooth Entropies", Physical Review Letters **106**, 110506 (2011), arXiv:1009.2015 [quant-ph].

[50] M. Tomamichel, "A Framework for Non-Asymptotic Quantum Information Theory", PhD Thesis (ETH Zürich, Zürich, Switzerland, 2012), arXiv:1203.2142 [quant-ph].

A   Entropic uncertainty relations

In the context of entropy duality, it is more convenient to describe CQ conditional entropies $H(Z_A|E)_\psi$ by using isometries to generate the required state, not directly by measuring ψ_{AE}. Abusing notation somewhat, let $U_{XA|A} = \sum_x |x\rangle_X |\tilde{x}\rangle\langle\tilde{x}|_A$ and $U_{ZA|A} = \sum_z |z\rangle_Z |z\rangle\langle z|_A$. Then defining $\xi_{XAE} = U_{XA|A}\psi_{AE}U_{XA|A}^*$ and $\eta_{ZAE} = U_{ZA|A}\psi_{AE}U_{ZA|A}^*$ we have $H(X_A|E)_\psi = H(X|E)_\xi$ and $H(Z_A|E)_\psi = H(Z|E)_\eta$. It will also be useful to define $\Pi_{XA} = \sum_x |x\rangle\langle x|_X \otimes |\tilde{x}\rangle\langle\tilde{x}|_A$, which is the projector onto the image of $U_{XA|A}$.

Proof of Lemma 1. That the sum of entropies is not smaller than $\log|A|$ is the entropic uncertainty relation of conjugate observables, first shown for the von Neumann entropy in [47], then for smooth min- and max-entropies in [49], and generally for entropies based on divergence in [6]. Hence we need only establish the upper bound.

To do so, first write $|\psi\rangle_{AEF} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|\sigma_z\rangle_{EF}$ for $\sqrt{P_Z(z)}|\sigma\rangle_{EF} = {}_A\langle z|\psi\rangle_{AEF}$. The conditional marginals of F are then the reduced states of $|\sigma_z\rangle_{EF}$. Since these are disjoint, the state $|\psi'\rangle_{AEFZ} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|z\rangle_Z|\sigma_z\rangle_{EF}$ can be created locally by measuring F. Hence $H^\perp(X_A|F)_\psi = H^\perp(X_A|FZ)_{\psi'}$. Furthermore, $\psi_{AE} = \psi'_{AE} = \sum_z P_Z(z)|z\rangle\langle z|_A \otimes (\sigma_z)_E$, and thus $H(Z_A|E)_\psi = H(A|E)_{\psi'}$. Therefore, we need only show that $H(A|E)_{\psi'} + H^\perp(X_A|F)_{\psi'} \leqslant \log_2|A|$ for states of this form. This is done for the various cases in the following Lemmas 3, 4, 5, and 6. $\qquad\square$

In fact, only the case of $H_\downarrow$ requires the state $|\psi\rangle$ to have this precise form.

Lemma 3. *For $H_\downarrow$ any conditional entropy measure as in* (1) *and any normalized pure state $|\psi\rangle_{AEFZ}$ of the form $|\psi\rangle_{AEFZ} = \sum_z \sqrt{P_Z(z)}|z\rangle_A|z\rangle_Z|\sigma_z\rangle_{EF}$ with arbitrary $|\sigma_z\rangle$,*

$$H_\downarrow(A|E)_\psi + H_\downarrow^\perp(X_A|FZ)_\psi \leqslant \log|A|. \tag{77}$$

Proof. Define $|\xi\rangle_{XAEFZ} = \sum_{xz}\langle\tilde{x}|z\rangle\sqrt{P_Z(z)}|x\rangle_X|\tilde{x}\rangle_A|z\rangle_Z|\sigma_z\rangle_{EF}$ and observe that $\xi_{AE} = \mu_A \otimes \psi_E$ and $\psi_E = \sum_z P_Z(z)(\sigma_z)_E$ By entropy duality we have $H_\downarrow^\perp(X_A|FZ)_\psi = H_\downarrow^\perp(X|FZ)_\xi = -H_\downarrow(X|AE)_\xi$. Then

$$-H_\downarrow(A|E)_\psi = D(\psi_{AE}, \mathbb{I}_A \otimes \psi_E) \tag{78a}$$
$$= D(U_{XA|A}\psi_{AE}U_{XA|A}^*, U_{XA|A}U_{XA|A}^* \otimes \psi_E) \tag{78b}$$
$$= D(\xi_{XAE}, \Pi_{XA} \otimes \psi_E) \tag{78c}$$
$$\geqslant D(\xi_{XAE}, \mathbb{I}_{XA} \otimes \psi_E) \tag{78d}$$
$$= D(\xi_{XAE}, \mathbb{I}_X \otimes \mu_A \otimes \psi_E) - \log_2|A| \tag{78e}$$
$$= D(\xi_{XAE}, \mathbb{I}_X \otimes \xi_{AE}) - \log_2|A| \tag{78f}$$
$$= -H_\downarrow^\perp(X_A|FZ)_\psi - \log_2|A|. \tag{78g}$$

The first equality is the definition of $H_\downarrow(A|E)_\psi$, the second invariance of D under isometries. In the third we use the fact that $UU^* = \Pi$. The inequality uses the dominance property of D, since, $\Pi \leqslant \mathbb{I}$, while the following equality uses normalization. The penultimate equality uses the specific form of ξ_{XAE}, and the final equality the entropy duality relation above. $\qquad\square$

Lemma 4. *For $H_\uparrow$ any conditional entropy measure as in* (2) *and any normalized pure state $|\psi\rangle_{AEF}$,*

$$H_\uparrow(A|E)_\psi + H_\uparrow^\perp(X_A|F)_\psi \leqslant \log|A|. \tag{79}$$

Proof. We have the same entropy duality $H_{\uparrow}^{\perp}(X_A|F)_{\psi} = -H_{\uparrow}(X|AE)_{\xi}$, with $|\xi\rangle_{XAEF} = U_{XA|A}|\psi\rangle_{AEF}$. Then

$$-H_{\uparrow}(A|E)_{\psi} = \min_{\tau} D(\psi_{AE}, \mathbb{I}_A \otimes \tau_E) \tag{80a}$$

$$\geqslant \min_{\tau} D(\xi_{XAE}, \mathbb{I}_X \otimes \mu_A \otimes \tau_E) - \log|A| \tag{80b}$$

$$\geqslant \min_{\sigma} D(\xi_{XAE}, \mathbb{I}_X \otimes \sigma_{AE}) - \log|A| \tag{80c}$$

$$= -H^{\perp}(X_A|F)_{\psi} - \log_2|A|. \tag{80d}$$

The first inequality encapsulates (78b)-(78e) from the proof of Lemma 3, since these steps hold for an arbitrary ψ_E. $\qquad\square$

Lemma 5. *For any pure state $|\psi\rangle_{AEF}$ and $0 \leqslant \varepsilon \leqslant 1$,*

$$H_{\min}^{\varepsilon}(A|E)_{\psi} + H_{\max}^{\varepsilon}(X_A|F)_{\psi} \leqslant \log|A|. \tag{81}$$

Proof. Define $\lambda = H_{\min}^{\varepsilon}(A|E)_{\psi}$ and let $\widetilde{\psi}_{AE} \in \mathcal{B}_{\varepsilon}(\psi_{AE})$ and σ_E be such that $\widetilde{\psi}_{AE} \leqslant 2^{-\lambda}\mathbb{I}_A \otimes \sigma_E$. Applying the isometry $U_{XA|A}$ yields

$$\widetilde{\xi}_{XAE} = U_{XA|A}\widetilde{\psi}_{AE}U_{XA|A}^{*} \tag{82a}$$

$$\leqslant 2^{-\lambda}U_{XA|A}(\mathbb{I}_A \otimes \sigma_E)U_{XA|A}^{*} \tag{82b}$$

$$= 2^{-\lambda}\Pi_{XA} \otimes \sigma_E \tag{82c}$$

$$\leqslant 2^{-\lambda}\mathbb{I}_{XA} \otimes \sigma_E \tag{82d}$$

$$= 2^{-(\lambda-\log|A|)}\mathbb{I}_X \otimes \mu_A \otimes \sigma_E. \tag{82e}$$

Note that the first step implies $\widetilde{\xi}_{XAE} \in \mathcal{B}_{\varepsilon}(\xi_{XAE})$. Thus, $\nu = \lambda - \log|A|$ and $\mu_A \otimes \sigma_E$ are feasible for $H_{\min}^{\varepsilon}(X|AE)_{\xi}$, meaning

$$H_{\min}^{\varepsilon}(X|AE)_{\xi} \geqslant H_{\min}^{\varepsilon}(A|E)_{\psi} - \log|A|. \tag{83}$$

Therefore the claim follows, since $H_{\min}^{\varepsilon}(X|AE)_{\xi} = -H_{\max}^{\varepsilon}(X|F)_{\xi} = -H_{\max}^{\varepsilon}(X_A|F)_{\psi}.$ $\qquad\square$

Lemma 6. *For any pure state $|\psi\rangle_{AEF}$ and $0 \leqslant \varepsilon \leqslant 1$,*

$$H_{\max}^{\varepsilon}(A|E)_{\psi} + H_{\min}^{\varepsilon}(X_A|F)_{\psi} \leqslant \log|A|. \tag{84}$$

Proof. Here we make use of the formulation of the smooth max-entropy as a semidefinite program and appeal to the dual problem as given in [12]. This avoids the minimax formulation inherent in (14). For ψ_{ABR} an arbitrary purification of ψ_{AB}, we have

$$2^{H_{\max}(A|B)_{\psi}} = \min\{\nu : \nu\mathbb{I}_B \geqslant Y_B, Y_{AB} \otimes \mathbb{I}_R \geqslant \psi_{ABR}, Y_{AB} \geqslant 0\}. \tag{85}$$

Now consider $H_{\max}^{\varepsilon}(X|AE)_{\xi}$ and let $\widetilde{\xi}_{XAE} \in \mathcal{B}_{\varepsilon}(\xi_{XAE})$ be such that $H_{\max}^{\varepsilon}(X|AE)_{\psi} = H_{\max}(X|E)_{\widetilde{\xi}}$. Next, define ν and $\widetilde{Y}_{XAEF}$ to be the optimal variables in (85) so that $\nu = 2^{H_{\max}^{\varepsilon}(X|AE)_{\xi}}$, while $\widetilde{Y}_{AE} \leqslant \nu\mathbb{I}_{AE}$ and $\widetilde{Y}_{XAE} \otimes \mathbb{I}_F \geqslant \widetilde{\xi}_{XAEF}$. Our goal is to construct a feasible set of variables for $H_{\max}^{\varepsilon}(A|E)_{\psi}$ from ν and $\widetilde{Y}_{XAEF}$.

For notational simplicity, let U be the isometry $U_{XA|A}$. Defining $Y_{AE} = U^{*}\widetilde{Y}_{XAE}U$ and $\psi'_{AEF} = U^{*}\widetilde{\xi}_{XAEF}U$, we have $\psi'_{AEF} \geqslant Y_{AE} \otimes \mathbb{I}_F$. The partial isometry U^{*} acts as a projection Π_{XA} and then an isometry on its support, and therefore ψ'_{AEF} is a possibly subnormalized pure state and $\psi'_{AEF} \in \mathcal{B}_{\varepsilon}(\psi_{AEF})$, since the purification distance only decreases under projections [50, Theorem 3.4]. Furthermore, $UY_{AE}U^{*} = \Pi_{XA}\widetilde{Y}_{XAE}\Pi_{XA}$, which implies

$$Y_E = \mathrm{Tr}_{XA}[UY_{AE}U^{*}] \tag{86a}$$

$$= \mathrm{Tr}_{XA}[\Pi_{XA}\widetilde{Y}_{XAE}\Pi_{XA}] \tag{86b}$$

$$\leqslant \widetilde{Y}_E \tag{86c}$$

$$\leqslant \nu|A|\mathbb{I}_E. \tag{86d}$$

Altogether, Y_{AE}, ψ'_{AEF}, and $\lambda = \nu|A|$ are feasible for $H_{\max}^{\varepsilon}(A|E)_{\psi}$, meaning

$$2^{H_{\max}^{\varepsilon}(A|E)_{\psi}} \leqslant |A|2^{H_{\max}^{\varepsilon}(X|AE)_{\xi}}. \tag{87}$$

The claim then follows because $H_{\max}^{\varepsilon}(X|AE)_{\xi} = -H_{\min}^{\varepsilon}(X_A|F)_{\psi}.$ $\qquad\square$

B Entropy simplification lemma

Proof of Lemma 2. We first prove (75). Letting $\bar{\sigma} = \sum_y P_Y(y)\sigma_y$, note that $\psi_{YB} = \mu_Y \otimes \bar{\sigma}_B$, and $\psi_{XYB} = T_{XY}(\mu_X \otimes \varrho_Y)T_{XY}^*$, where $T_{XY} = \sum_{xy}|x\rangle\langle x|_X \otimes |x+y\rangle\langle y|_Y$. Then, for entropies as in (1), we have

$$H(X|YB)_\psi = -D(\psi_{XYB}, \mathbb{I}_X \otimes \mu_Y \otimes \bar{\sigma}_B) \tag{88a}$$

$$= -D(\mu_X \otimes \varrho_{YB}, \mu_X \otimes \mathbb{I}_Y \otimes \bar{\sigma}_B) \tag{88b}$$

$$= H(Y|B)_\varrho. \tag{88c}$$

The first equality is the definition of H, while the second follows from unitary invariance under T^*. The third equality holds by monotonicity of D under creating and removing μ_X.

The argument is slightly more complicated for entropies as in (2), where we are interested in $\min_{\tau_{YB}} D(\psi_{XYB}, \mathbb{I}_X \otimes \tau_{YB})$. Using monotonicity, we can show that without loss of generality the optimal τ_{YB} has the form $\mu_Y \otimes \tau_B$, and so the above argument can be applied to reach the desired conclusion. Observe that ψ_{XYB} is invariant under both the operation $V_{XY} = \sum_{xy}|x+1\rangle\langle x|_X \otimes |y+1\rangle\langle y|_Y$ as well as $U_{XY} = \sum_{xy} \omega^{x+y}|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y$. Letting G be the group generated by these Letting $\mathcal{E}(\varrho_{XY}) = \frac{1}{d^2}\sum_{j,k=0}^{d-1} U^j V^k \varrho (U^j V^k)^*$, we have $\mathcal{E}(\psi_{XYB}) = \psi_{XYB}$ and $\mathcal{E}(\mathbb{I}_X \otimes \tau_{YB}) = \mathbb{I}_X \otimes \mu_Y \otimes \tau_B$. By monotonicity, then,

$$D(\psi_{XYB}, \mathbb{I}_X \otimes \mu_Y \otimes \tau_B) \leqslant D(\psi_{XYB}, \mathbb{I}_X \otimes \tau_{YB}), \tag{89}$$

and therefore the optimal τ_{YB} indeed has the desired form.

Finally, for the smooth min- and max-entropies we can again appeal to monotonicity to ensure that the optimal state in the ε-ball is also classical on X and Y (see also [50, Proposition 5.8]) and is uniform on X by making use of the fact that ψ_{XYB} is invariant under the map which applies T^*, traces out X, creates μ_X in its place, and finally reapplies T.

Now for (76). Define the purifications

$$|\psi\rangle_{XX'YY'BR} = \frac{1}{\sqrt{|X|}}\sum_{xz}\sqrt{P_Y(z)}|x\rangle_X|x\rangle_{X'}|x+z\rangle_Y|x+z\rangle_{Y'}|\sigma_z\rangle_{BR} \quad \text{and} \tag{90}$$

$$|\varrho\rangle_{YY'BR} = \sum_y \sqrt{P_Y(y)}|y\rangle_Y|y\rangle_{Y'}|\sigma_y\rangle_{BR}, \tag{91}$$

where $|\sigma_z\rangle_{BR}$ is a purification of σ_z for each z. By duality, we want to show $H(X|X'Y'R)_\psi = H(Y|Y'R)_\varrho$. First rewrite $|\psi\rangle$ as

$$|\psi\rangle_{XX'YY'BR} = \frac{1}{\sqrt{|X|}}\sum_{xz}\sqrt{P_Y(z)}|y-z\rangle_X|y-z\rangle_{X'}|y\rangle_Y|y\rangle_{Y'}|\sigma_z\rangle_{BR}, \tag{92}$$

and note that the $X'Y'R$ marginal is

$$\psi_{X'Y'R} = \frac{1}{|X|}\sum_{yz}P_Y(x)|y-z\rangle\langle y-z|_{X'} \otimes |y\rangle\langle y|_{Y'} \otimes (\sigma_z)_R. \tag{93}$$

Let $U_{XX'Y'}$ be the unitary which subtracts the value in the Y' register from the X and X' registers, and then negates the latter two. For $|\psi'\rangle = U|\psi\rangle$, we have

$$|\psi'\rangle_{XX'YY'BR} = \frac{1}{\sqrt{|X|}}\sum_{xz}\sqrt{P_Y(z)}|z\rangle_X|z\rangle_{X'}|y\rangle_Y|y\rangle_{Y'}|\sigma_z\rangle_{BR} \tag{94a}$$

$$= |\Phi\rangle_{YY'}|\varrho\rangle_{XX'BR}, \tag{94b}$$

where $|\varrho\rangle_{XX'BR}$ is just $|\varrho\rangle_{YY'BR}$ with Y and Y' relabelled X and X', respectively. From this expression we immediate see that $\psi_{XX'Y'R} = \mu_Y \otimes \varrho_{XX'R}$. Moreover, $U(\mathbb{I}_X \otimes \psi_{X'Y'R})U^* = \mathbb{I}_X \otimes \mu_{Y'} \otimes \varrho_{X'R}$.

Therefore, for entropies H as in (1),

$$H(X|X'Y'R)_\psi = -D(\psi_{XX'Y'R}, \mathbb{I}_X \otimes \psi_{X'Y'R}) \tag{95a}$$

$$= -D(\psi'_{XX'Y'R}, U(\mathbb{I}_X \otimes \psi_{X'Y'R})U^*) \tag{95b}$$

$$= -D(\mu_{Y'} \otimes \varrho_{XX'R}, \mu_{Y'} \otimes \mathbb{I}_X \otimes \varrho_{X'R}) \tag{95c}$$

$$= H(Y|Y'R)_\varrho. \tag{95d}$$

For entropies H involving a marginal optimization as in (2), we have

$$\mathsf{H}(X|X'Y'R)_\psi = \max_\tau [-\mathsf{D}(\psi'_{XX'Y'R}, U(\mathbb{I}_X \otimes \tau_{X'Y'R})U^*)] \tag{96}$$

as above. Due to the form of U, $U(\mathbb{I}_X \otimes \tau_{X'Y'R})U^* = \mathbb{I}_X \otimes U'\tau_{X'Y'R}U'^*$, where U' has the same action as U, just not applied to X. Therefore

$$\mathsf{H}(X|X'Y'R)_\psi = \max_\tau [-\mathsf{D}(\psi'_{XX'Y'R}, \mathbb{I}_X \otimes \tau_{X'Y'R})]. \tag{97}$$

By monotonicity we can assume the optimal $\tau_{X'Y'R}$ has the form $\mu_{Y'} \otimes \tau_{X'R}$, at which point we can follow the above derivation to complete the proof. $\qquad\square$