

# Efficient Quantum Polar Codes Requiring No Preshared Entanglement

Joseph M. Renes *Member, IEEE*, David Sutter *Student Member, IEEE*,  
Frédéric Dupuis *Member, IEEE*, and Renato Renner *Member, IEEE*

**Abstract**—We construct an *explicit* quantum coding scheme which achieves a communication rate not less than the coherent information when used to transmit quantum information over a noisy quantum channel. For Pauli and erasure channels we also present *efficient* encoding and decoding algorithms for this communication scheme based on polar codes (essentially linear in the blocklength), but which do not require the sender and receiver to share any entanglement before the protocol begins. Due to the existence of degeneracies in the involved error-correcting codes it is indeed possible that the rate of the scheme exceeds the coherent information. We provide a simple criterion which indicates such performance. Finally we discuss how the scheme can be used for secret key distillation as well as private channel coding.

**Index Terms**—Quantum polar codes, coherent information, entanglement distillation, privacy amplification, information reconciliation, secret key distillation, private channel coding

## I. INTRODUCTION

SHANNON'S channel coding theorem determines the capacity of a classical discrete memoryless channel  $W$  by random coding arguments and finds that it is given by

$$C(W) = \max_{P_X} I(X : W(X)), \quad (1)$$

where the random variable  $X$  describes the input to the channel, and  $P_X$  is its probability distribution [1]. Analogous random coding arguments for the problem of transmitting quantum information over a memoryless quantum channel  $\mathcal{N}^{A' \rightarrow B}$  lead to a communication rate given by

$$Q_1(\mathcal{N}) := \max_{\phi} I(A)B_{\sigma}, \quad (2)$$

where the optimization is over all pure, bipartite states  $\phi^{AA'}$ ,  $\sigma^{AB} := \mathcal{N}^{A' \rightarrow B}(\phi^{AA'})$  and  $I(A)B_{\rho} := H(B)_{\rho} - H(AB)_{\rho}$  is the *coherent information* and  $H$  the von Neumann entropy [2], [3], [4], [5], [6], [7]. It has been shown that  $Q_1(\mathcal{N})$  is not generally optimal [8] and that the quantum capacity is given by its *regularization*

$$Q(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} Q_1(\mathcal{N}^{\otimes k}). \quad (3)$$

J.M. Renes, D. Sutter and R. Renner are with the Institute for Theoretical Physics, ETH Zurich, Zürich 8092, Switzerland (e-mail: renes@phys.ethz.ch; sutedav@phys.ethz.ch; renner@phys.ethz.ch).

F. Dupuis is with Faculty of Informatics, Masaryk University, Brno, Czech Republic (e-mail: dupuis@fi.muni.cz).

This work was supported by the Swiss National Science Foundation (through the National Centre of Competence in Research ‘Quantum Science and Technology’ and grant No. 200020-135048) and by the European Research Council (grant No. 258932).

This paper was presented in part at the IEEE Information Theory Workshop (2012) and the IEEE International Symposium on Information Theory (2013).

Notwithstanding the complications surrounding the regularized expression, it is already difficult to construct *explicit* coding schemes that achieve the coherent information of an arbitrary quantum channel, and the task becomes that much harder if we also ask for *efficient* encoding and decoding. Here, a task is termed efficient if it can be completed in a number of steps scaling essentially linearly in the input size, not just polynomially as in the complexity-theoretic setting. Until very recently, almost nothing was known about explicit, efficient, *provably* capacity-achieving classical error-correcting codes, to say nothing of the quantum case.

Polar codes, introduced in 2008 by Arkan [9], are the first family of classical error-correcting codes which both provably achieve the *symmetric* (i.e.,  $X \sim$  uniform in (1)) classical capacity for any discrete memoryless channel and have an essentially linear encoding and decoding complexity. These codes have been generalized to the quantum setup. Wilde and Guha adapted polar codes to transmit classical information over quantum channels [10] and gave a scheme for transmitting quantum information over degradable quantum channels [11], at the cost of an unknown decoding efficiency. Three of us showed in [12] how to achieve the *symmetric* coherent information ( $\phi^{AA'}$  a Bell state in (2)) of any Pauli or erasure channel with efficient encoding and decoding operations. In [13], Wilde and Renes extended this method to arbitrary quantum channels and showed that the quantum decoder can be constructed by combining sequential decoders for suitable classical-quantum channels (see also [14]), but without providing an efficient decoder.

However, all of these quantum channel coding schemes suffer from two important drawbacks. The first is the need for noiseless entanglement to be shared by the sender and receiver prior to the start of the protocol. Second, the aforementioned protocols only achieve rates given by the symmetric coherent information. More details about entanglement-assisted quantum coding can be found in [15]; a precursor in the field of quantum key distribution is [16]. In this contribution we present an explicit coding scheme that provably achieves the (true) coherent information for an arbitrary quantum channel without using any entanglement assistance. Our protocol is a concatenated two level scheme in which the method of polar coding is employed separately at each level. For Pauli or erasure channels we show how to perform efficient encoding and decoding.

Recently, an extension to quantum polar codes that is based on branching MERA codes has been proposed [17], [18]. It is claimed (however not yet proven) that these codes do not

require entanglement assistance while achieving high rates and an efficient encoding and decoding.

The paper is structured as follows. Section II-A introduces basic notation and definitions. As our protocol is based on the use of information reconciliation (IR) and stabilizer-based quantum error-correcting codes, Section II-B provides background on these topics. Section II-C recalls two general uncertainty relations needed to establish the rate of the coding scheme. Section III introduces the classical polarization phenomenon and proves that it also holds for general classical-quantum (cq) states.<sup>1</sup> Section IV then describes a protocol for the problem of entanglement distillation and Section V shows that it is computationally efficient for Pauli and erasure channels when using quantum polar codes. It is well known that an entanglement distillation featuring only forward classical communication can be turned into a channel coding scheme [19], and Section VI explains the channel coding view of the scheme in detail. We prove that the communication rate of the channel coding scheme is at least as great as  $Q_1(\mathcal{N})$ , as defined in (2), and present an efficient encoder and decoder based on polar codes for Pauli and erasure channels. It is possible, but still unproven, that the scheme can achieve rates beyond the coherent information. Section VII shows, however, that the rate expression reduces to the coherent information when the channel is degradable. Sections VII and VIII state two different open problems addressing the question of whether our scheme can achieve rates beyond the coherent information or even achieve the quantum capacity for channels where the coherent information is not optimal. In Section IX it is discussed how the scheme can be used for secret key distillation.

## II. BACKGROUND

### A. Notation and Definitions

Let  $\mathcal{N} : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$  be a quantum channel, where  $\mathfrak{T}(\mathcal{H})$  denotes the space of all operators in some Hilbert space  $\mathcal{H}$  that are equipped with the trace norm. We assume that the underlying state spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$  for systems  $A$  and  $B$  are finite dimensional. Using the Stinespring dilation [20] there exists an “environment” system  $\mathcal{H}_E$  and a partial isometry  $U_{BE} : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ , such that

$$\mathcal{N}(\rho) = \text{Tr}_E \left( U_{BE} \rho U_{BE}^\dagger \right) \quad (4)$$

for every  $\rho \in \mathcal{D}(\mathcal{H}_A)$ , where  $\mathcal{D}(\mathcal{H}_A)$  denotes the space of all density operators on  $\mathcal{H}_A$ . Then we may define the *complementary channel* to  $\mathcal{N}$  as

$$\mathcal{N}_c(\rho) = \text{Tr}_B \left( U_{BE} \rho U_{BE}^\dagger \right). \quad (5)$$

A quantum channel  $\mathcal{N}$  is *more capable* if the quantum capacity of its complementary channel is zero, i.e.,  $Q(\mathcal{N}_c) = 0$ . Equivalently, the notion of more capable may be formulated as follows. For all states  $\rho^{A^N}$  on  $\mathcal{D}(\mathcal{H}_A^{\otimes N})$  let the classical random variable  $X^N$  denote its eigenvalues and  $B^N$  and  $E^N$  the quantum outputs under the action of  $\mathcal{N}^{\otimes N}$ . Then  $\mathcal{N}$  is

more capable when  $I(X^N : B^N) \geq I(X^N : E^N)$  for all  $N \geq 1$  [21].

If the private capacity of the complementary channel is zero, i.e.,  $C_p(\mathcal{N}_c) = 0$ , then  $\mathcal{N}$  is said to be *less noisy*.<sup>2</sup> Equivalently,  $\mathcal{N}$  is less noisy if  $I(T : B^N) \geq I(T : E^N)$ , for every  $N \geq 1$  and for all distributions on  $(T, X^N)$ , where  $T$  has finite support and  $T \text{---} X^N \text{---} (B^N, E^N)$  form a Markov chain [21].

Finally, a quantum channel  $\mathcal{N}$  is called *degradable* if there exists a trace-preserving and completely positive (TPCP) map  $\mathcal{T}$  such that  $\mathcal{T} \circ \mathcal{N} = \mathcal{N}_c$ . The concepts of a channel being more capable, less noisy, or degradable were introduced in a classical framework in [23] and have recently been generalized to the quantum mechanical setup in [21]. For classical channels, it has been shown that being more capable is a strictly weaker condition than being less noisy, which is again a strictly weaker condition than being degradable [23]. For quantum channels it is unknown whether these relations are strict [21]. However a degradable (quantum) channel  $\mathcal{N}$  is less noisy and *a fortiori* more capable. In [21] it is shown that both the private and quantum capacities of a less noisy channel  $\mathcal{N}$  are equal to the coherent information:

$$C_p(\mathcal{N}) = Q(\mathcal{N}) = Q_1(\mathcal{N}), \quad (7)$$

a relation that was first shown for degradable channels [24], [25].

### B. Information Reconciliation & Stabilizer Codes

In this subsection we recall some details about the information reconciliation protocols which form the basis of our coding schemes. The form of information reconciliation needed here is an instance of data compression in which the decoder has access to quantum side information. Consider a classical-quantum state of a random variable  $Z$  and a quantum system  $B$ , which may be written  $\psi^{ZB} = \sum_z p_z |z\rangle \langle z|^Z \otimes \varphi_z^B$ , where  $p_z$  is the probability distribution of  $Z$  and  $\varphi_z$  are arbitrary states. The goal of information reconciliation is to compress  $Z$ , i.e. apply some function to  $Z$ , so that  $Z$  itself can be reconstructed (with high probability) by a decoder with access to  $B$  and the compressed output.

In quantum language, the decoding step may be regarded as a measurement of the systems available to the decoder. The probability of error in determining the value of a random variable  $Z$  given measurement of some quantum system  $B$  can be expressed as

$$p_{\text{err}}(Z^A|B)_\psi := 1 - \max_{\mathcal{M}_Z} \sum_z p_z \text{Tr}[\Lambda_z^B \varphi_z^B], \quad (8)$$

where the maximum is taken over all measurements  $\mathcal{M}_Z$  with elements  $\Lambda_z^B$ .

Devetak and Winter showed that in the case of  $N \rightarrow \infty$  copies of  $\psi^{ZB}$ , there exist compression functions outputting

<sup>2</sup>Recall that according to [4], [22] the private capacity of a quantum channel  $\mathcal{N}$  is given by

$$C_p(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{T, X^k} I(T : B^k) - I(T : E^k), \quad (6)$$

where  $T \text{---} X^k \text{---} (B^k, E^k)$  form a Markov chain.

<sup>1</sup>Since, to the best of our knowledge, the polarization phenomenon has thus far only been generalized to symmetric cq-states [10].

roughly  $NH(Z|B)_\psi + o(N)$  bits which suffice to reliably determine  $Z^N$  [26]. Specifically, calling  $B_C$  the classical compressor output, they constructed a sequence of measurements such that  $p_{\text{err}}(Z^N|B^N B_C)_{\psi^{\otimes N}} \rightarrow 0$  as  $N \rightarrow \infty$ . Moreover, this compression rate was shown to be optimal.

We shall be interested in information reconciliation of classical information that represents the outcomes of measurements on a quantum system in one of two complementary bases, the ‘‘amplitude’’ basis and the ‘‘phase’’ basis. Any basis  $\{|z\rangle\}_{z=0}^{d-1}$  of a  $d$ -dimensional quantum system may be chosen to be the amplitude basis, and for concreteness we then take the phase basis to be comprised of elements  $|\tilde{x}\rangle = \frac{1}{\sqrt{d}} \sum_{z=0}^{d-1} \omega^{xz} |z\rangle$ , where  $\omega = e^{2\pi i/d}$ . Although the coding scheme we describe below can be made to work for any finite  $d$ , there is no real loss of generality in setting  $d = 2$ , which we do henceforth.

For the purposes of information reconciliation in this context, it is convenient that the random coding argument can be specialized to random linear functions (as, for instance, in [27]). Then, since the compressed output is a linear function of the input, we may regard it as resulting from the measurement of a corresponding set of stabilizer observables. Here we recall the basic facts of the stabilizer formalism; for more details see [28]. For instance, the result of computing the parity of the outcome of measuring 3 qubit systems in the amplitude basis can just as well be thought of as the result of measuring the observable  $Z \otimes Z \otimes Z$ , where  $Z = \text{diag}(1, -1)$ . In general, any function of the form  $f(z^N) = v^N \cdot z^N$  corresponds to the stabilizer operator  $Z^{v_1} \otimes Z^{v_2} \otimes \dots \otimes Z^{v_N}$ . An arbitrary linear function is just a sequence of functions of the form just considered, and therefore any linear function corresponds to a sequence of stabilizer operators.

Linear functions of phase basis measurements can be similarly regarded as measurements of stabilizer operators using the operator  $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . Since the operators  $X$  and  $Z$  anticommute, stabilizers corresponding to vectors  $u^N$  and  $v^N$  commute when  $u^N \cdot v^N = 0$ . Thus, it is possible to consider simultaneously measuring a compressed output for both the amplitude and phase bases. Such a set of commuting stabilizer operators constitutes a CSS code.

The measurement of any set of stabilizer operators can be accomplished by applying a unitary transformation and then performing appropriate measurements on individual qubits. Each stabilizer measurement necessitates the measurement of one qubit, so that not every qubit is necessarily measured after the transformation. Due to the linear structure of stabilizer observables, this unitary transformation is simultaneously a mapping of amplitude basis states to themselves, as well as a mapping of phase basis states to themselves.

Note that, although we will consider using CSS codes for information reconciliation in the following sections, the code structure is only used by the encoder. Stabilizer measurements for recovery operations are not generally used by the decoder. Instead, the coding techniques for information reconciliation make use of more general measurements, such as the pretty-good measurement.

## C. Generalized Uncertainty Relations

The coding scheme introduced in this paper consists of two layers, both performing an information reconciliation operation in complementary bases.<sup>3</sup> In the derivation of the exact rate expression (cf. Theorem 9) as well as in the formulation of the two open problems (cf. Sections VII and VIII) we will use entropic uncertainty relations that arise when measuring a system in complementary bases. This subsection gives an overview about the uncertainty relations we need to prove several properties of the scheme.

**Lemma 1** (Renes & Boileau [29]). *Suppose  $\psi^{AB}$  is a bipartite state for which  $H(Z^A|B)_\psi = 0$ . Then*

$$H(X^A|B)_\psi = \log \dim(A) + H(A|B)_\psi. \quad (9)$$

*Proof:* Let  $R$  be a purification of  $\psi^{AB}$ ; expressing the  $A$  system in the amplitude basis, we may write

$$|\psi\rangle^{ABR} = \sum_z \sqrt{p_z} |z\rangle^A |\varphi_z\rangle^{BR} \quad (10)$$

for some probability distribution  $p_z$  and normalized states  $|\varphi_z\rangle^{BR}$ . Since  $H(Z^A|B) = 0$ , the states  $\varphi_z^B$  are completely distinguishable. Thus, there exists a measurement which precisely determines the value of  $z$  given  $B$ . Let  $U^{B \rightarrow BC}$  be a coherent implementation (Stinespring dilation) of this measurement, in which the measurement result is stored in system  $C$ . Then define

$$|\psi'\rangle^{ABCR} := U^{B \rightarrow BC} |\psi\rangle^{ABR} \quad (11)$$

$$= \sum_z \sqrt{p_z} |z\rangle^A |z\rangle^C |\varphi_z\rangle^{BR}. \quad (12)$$

As  $U$  is a partial isometry,  $H(X^A|B)_\psi = H(X^A|BC)_{\psi'}$  and  $H(A|B)_\psi = H(A|BC)_{\psi'}$ . Expressing  $A$  in  $|\psi'\rangle$  in the phase basis, we have

$$|\psi'\rangle^{ABCR} = \frac{1}{\sqrt{d}} \sum_{x,z} \sqrt{p_z} \omega^{-xz} |\tilde{x}\rangle^A |z\rangle^C |\varphi_z\rangle^{BR} \quad (13)$$

$$= \frac{1}{\sqrt{d}} \sum_{x,z} |\tilde{x}\rangle^A (Z^{-x})^C |\psi\rangle^{CBR}, \quad (14)$$

where  $d = \dim(A)$ . Using the chain rule we can write

$$H(X^A|BC)_{\psi'} = H(X^A BC)_{\psi'} - H(BC)_{\psi'} \quad (15)$$

$$= H(BC|X^A)_{\psi'} + H(X^A)_{\psi'} - H(BC)_{\psi'}. \quad (16)$$

Clearly  $H(X^A)_{\psi'} = \log \dim(A)$ , while  $H(BC|X^A)_{\psi'} = H(AB)_\psi$  since all the  $BC$  marginals conditioned on  $X^A$  are related by the unitary  $Z^{-x}$  and the  $BC$  marginal for  $X^A = 0$  is  $|\psi\rangle^{CBR} \simeq |\psi\rangle^{ABR}$ . Finally,  $H(BC)_{\psi'} = H(B)_\psi$  again by the isometry property of  $U$ . Therefore,

$$H(X^A|BC)_{\psi'} = H(BC)_\psi + \log \dim(A) - H(B)_\psi, \quad (17)$$

which by the chain rule is the expression we set out to prove. ■

<sup>3</sup>Recall that we denote by  $X^A$  and  $Z^A$  an operator  $X$  and  $Z$  acting on a system  $A$ .

**Lemma 2** (Renes & Boileau [29]). *Suppose  $\psi^{ABR}$  is a pure state for which  $H(Z^A|B)_\psi = 0$ . Then*

$$H(Z^A|R)_\psi + H(X^A|B)_\psi = \log \dim A. \quad (18)$$

*Proof:* The statement follows from Lemma 1 and the fact that  $H(A|B)_\psi = H(Z^A|B)_\psi - H(Z^A|R)_\psi$  for any state  $\psi$ . To see the latter expression, use the chain rule to find

$$\begin{aligned} H(Z^A|B)_\psi - H(Z^A|R)_\psi &= H(Z^A B)_\psi - H(B)_\psi - H(Z^A R)_\psi + H(R)_\psi \\ &= H(B|Z^A)_\psi - H(R|Z^A)_\psi + H(A|B)_\psi. \end{aligned} \quad (19)$$

$$= H(B|Z^A)_\psi - H(R|Z^A)_\psi + H(A|B)_\psi. \quad (20)$$

Since  $ABR$  is pure, the  $BR$  marginals conditioned on a projective measurement of  $A$  are, too, and therefore  $H(B|Z^A)_\psi = H(R|Z^A)_\psi$ . ■

### III. POLARIZATION PHENOMENON

In this section, we introduce the *polarization phenomenon* discovered by Arıkan, which has been used to construct codes (called *polar codes*) that can be used for channel [9] as well as for source coding [30]. We show how it can be generalized to the setup of classical-quantum (cq) states.

#### A. Classical Polarization Phenomenon

Polar codes have several desirable attributes [9], [31], [32], [33]: they achieve the capacity when used for transmitting information over a discrete memoryless channel (DMC); they can be encoded and decoded efficiently (with a complexity that is essentially linear in the blocklength); the error probability of the efficient decoder decays exponentially in the square root of the blocklength.

Let  $X^N$  be a vector whose entries are i.i.d. Bernoulli( $p$ ) distributed for some  $p \in [0, 1]$  and  $N = 2^n$  for  $n \in \mathbb{Z}^+$ . Furthermore, let  $U^N = G_N X^N$ , where  $G_N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\otimes \log N}$  denotes the polarization transform and  $W : \mathcal{X} \rightarrow \mathcal{Y}$  be a DMC with a binary input alphabet  $\mathcal{X} = \{0, 1\}$ , an arbitrary output alphabet  $\mathcal{Y}$  and transition probabilities  $W(y|x)$  for  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ .  $W^N$  denotes the channel corresponding to  $N$  uses of  $W$ . For  $Y^N = W^N X^N$  and  $\epsilon \in (0, 1)$ , we define the two sets

$$\mathcal{R}_\epsilon^N(X|Y) := \{i \in [N] : H(U_i|U^{i-1}Y^N) \geq 1 - \epsilon\} \quad (21)$$

and

$$\mathcal{D}_\epsilon^N(X|Y) := \{i \in [N] : H(U_i|U^{i-1}Y^N) \leq \epsilon\}. \quad (22)$$

The former consists of outputs  $U_j$  which are essentially uniformly random, even given all previous outputs  $U^{j-1}$  as well as  $Y^N$ , while the latter set consists of the essentially deterministic outputs. The polarization phenomenon is that essentially all outputs are in one of these two subsets, and their sizes are given by the conditional entropy of the input  $X$  given  $Y$ .

**Theorem 3** (Arıkan [9], [30]). *For any  $\epsilon \in (0, 1)$*

$$|\mathcal{R}_\epsilon^N(X|Y)| = NH(X|Y) - o(N) \quad (23)$$

and

$$|\mathcal{D}_\epsilon^N(X|Y)| = N(1 - H(X|Y)) - o(N). \quad (24)$$

Based on this result, Arıkan showed in [9] how to construct polar codes having the desirable properties mentioned above.

Non-binary random variables can be represented by a sequence of correlated binary random variables, which are then encoded separately. Correlated sequences of binary random variables may be polarized using a multilevel construction, as shown in [31].<sup>4</sup> Given  $M$  i.i.d. instances of a sequence  $X = (X_{(1)}, X_{(2)}, \dots, X_{(K)})$  and possibly a correlated random variable  $Y$ , the basic idea is to first polarize  $X_{(1)}^M$  relative to  $Y^M$ , then treat  $X_{(1)}^M Y^M$  as side information in polarizing  $X_{(2)}^M$ , and so on. More precisely, defining  $U_{(j)}^M = G_M X_{(j)}^M$  for  $j = 1, \dots, K$ , we may define the random and deterministic sets for each  $j$  as

$$\begin{aligned} \mathcal{R}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y) := \\ \{i \in [M] : H(U_{(j), i}^M | U_{(j)}^{i-1}, X_{(j-1)}^M, \dots, X_{(1)}^M, Y^M) \geq 1 - \epsilon\} \end{aligned} \quad (25)$$

and

$$\begin{aligned} \mathcal{D}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y) := \\ \{i \in [M] : H(U_{(j), i}^M | U_{(j)}^{i-1}, X_{(j-1)}^M, \dots, X_{(1)}^M, Y^M) \leq \epsilon\}. \end{aligned} \quad (26)$$

In principle we could choose different  $\epsilon$  parameters for each  $j$ , but this will not be necessary here. Now, Theorem 3 applies to the random and deterministic sets for every  $j$ . The sets  $\mathcal{R}_{\epsilon, (j)}^M(X|Y) = \{\mathcal{R}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y)\}_{j=1}^K$  and  $\mathcal{D}_{\epsilon, (j)}^M(X|Y) = \{\mathcal{D}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y)\}_{j=1}^K$  have sizes given by

$$\begin{aligned} |\mathcal{R}_\epsilon^M(X|Y)| &= \sum_{j=1}^K |\mathcal{R}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y)| \\ &= \sum_{j=1}^K MH(X_{(j)}|X_{(1)}, \dots, X_{(j-1)}, Y) - o(M) \end{aligned} \quad (27)$$

$$= \sum_{j=1}^K MH(X_{(j)}|X_{(1)}, \dots, X_{(j-1)}, Y) - o(M) \quad (28)$$

$$= MH(X|Y) - o(KM), \quad (29)$$

and

$$\begin{aligned} |\mathcal{D}_\epsilon^M(X|Y)| &= \sum_{j=1}^K |\mathcal{D}_{\epsilon, (j)}^M(X_{(j)}|X_{(j-1)}, \dots, X_{(1)}, Y)| \\ &= \sum_{j=1}^K M(1 - H(X_{(j)}|X_{(1)}, \dots, X_{(j-1)}, Y)) - o(M) \end{aligned} \quad (30)$$

$$= \sum_{j=1}^K M(1 - H(X_{(j)}|X_{(1)}, \dots, X_{(j-1)}, Y)) - o(M) \quad (31)$$

$$= M(K - H(X|Y)) - o(KM). \quad (32)$$

In the following we will make use of both the polarization phenomenon in its original form, Theorem 3, and the multilevel extension. To simplify the presentation, we denote by  $\tilde{G}_M^K$  the  $K$  parallel applications of  $G_M$  to the  $K$  random variables  $X_{(j)}^M$ .

<sup>4</sup>An alternative approach is given in [34], [35], where the polarization phenomenon has been generalized for arbitrary finite fields. We will however focus on the multilevel construction in this paper.



### B. Polarization Phenomenon for General cq-States

In this subsection, we generalize the polarization phenomenon to the setup of classical-quantum (cq) states of the form

$$\rho^{XB} = \sum_{x \in \{0,1\}} p_x |x\rangle \langle x|^X \otimes \rho_x^B, \quad (33)$$

where  $p_0, p_1 \geq 0$  such that  $p_0 + p_1 = 1$ . We note that the special case where  $p_0 = p_1 = \frac{1}{2}$  has been discussed in [10]. We denote the fidelity between the states  $\rho_0$  and  $\rho_1$  by [36], [37]

$$F(\rho_0, \rho_1) := \|\sqrt{\rho_0} \sqrt{\rho_1}\|_1, \quad (34)$$

where  $\|A\|_1$  denotes the *trace norm* of the operator  $A$ , defined as  $\|A\|_1 := \text{Tr}[\sqrt{A^\dagger A}]$ . Furthermore, for a cq state  $\rho^{XB}$  as in (33), we define the quantity

$$Z(X|B)_\rho := 2\sqrt{p_0 p_1} F(\rho_0, \rho_1) = 2\sqrt{p_0 p_1} \|\sqrt{\rho_0} \sqrt{\rho_1}\|_1. \quad (35)$$

*Fidelity Polarization.*— Applying Arıkan's transformation [9, Section I.B] to the  $X$  systems of two independent cq-states  $\rho^{XB}$  as defined in (33) gives

$$\tilde{\rho} = \sum_{u_1 u_2} p_{u_1 \oplus u_2} p_{u_2} |u_1\rangle \langle u_1|^{U_1} \otimes |u_2\rangle \langle u_2|^{U_2} \otimes \rho_{u_1 \oplus u_2}^{B_1} \otimes \rho_{u_2}^{B_2}. \quad (36)$$

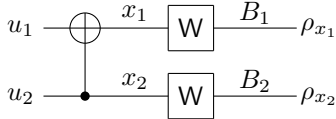


Fig. 1. Notation used to derived the fidelity polarization phenomenon (cf. Proposition 4), including the classical-quantum channel  $W : x \rightarrow \rho_x$ .

#### Proposition 4.

$$Z(U_2|U_1 B_1 B_2)_{\tilde{\rho}} = Z(X|B)_\rho^2 \quad \text{and} \quad (37)$$

$$Z(U_1|B_1 B_2)_{\tilde{\rho}} \leq 2Z(X|B)_\rho - Z(X|B)_\rho^2. \quad (38)$$

*Proof:* Note that this proof is a generalization of the proof for the symmetric case (i.e.,  $p_0 = p_1 = \frac{1}{2}$ ) given by Wilde and Guha [10, Proof of Proposition 9]. We first prove (37).

$$\begin{aligned} & Z(U_2|U_1 B_1 B_2)_{\tilde{\rho}} \\ &= 2\sqrt{p_0 p_1} F \left( \sum_{u_1} p_{u_1} |u_1\rangle \langle u_1| \otimes \rho_{u_1} \otimes \rho_0, \right. \\ & \quad \left. \sum_{u_1} p_{u_1 \oplus 1} |u_1\rangle \langle u_1| \otimes \rho_{u_1 \oplus 1} \otimes \rho_1 \right) \quad (39) \end{aligned}$$

$$\begin{aligned} &= 2\sqrt{p_0 p_1} F \left( \sum_{u_1} p_{u_1} |u_1\rangle \langle u_1| \otimes \rho_{u_1}, \right. \\ & \quad \left. \sum_{u_1} p_{u_1 \oplus 1} |u_1\rangle \langle u_1| \otimes \rho_{u_1 \oplus 1} \right) F(\rho_0, \rho_1) \quad (40) \end{aligned}$$

$$= 4\sqrt{p_0 p_1} F(p_0 \rho_0, p_1 \rho_1) F(\rho_0, \rho_1) \quad (41)$$

$$= 4p_0 p_1 F(\rho_0, \rho_1)^2 \quad (42)$$

$$= Z(X|B)_\rho^2, \quad (43)$$

where (40) uses the multiplicativity of the fidelity under tensor product states (i.e.,  $F(\rho \otimes \sigma, \tau \otimes \nu) = F(\rho, \tau)F(\sigma, \nu)$ ) and (41) uses the following relation for cq-states

$$\begin{aligned} & F \left( \sum_x p(x) |x\rangle \langle x| \otimes \rho_x, \sum_x p(x) |x\rangle \langle x| \otimes \sigma_x \right) \\ &= \sum_x p(x) F(\rho_x, \sigma_x). \quad (44) \end{aligned}$$

We next prove (38). Note that the fidelity can be expressed as the minimum Bhattacharya overlap between distributions induced by a POVM on the states [38]

$$F(\rho_0, \rho_1) = \min_{\{\Lambda_m\}} \sum_m \sqrt{\text{Tr}[\Lambda_m \rho_0] \text{Tr}[\Lambda_m \rho_1]}. \quad (45)$$

Let  $\Lambda_m$  denote a POVM that achieves the minimum for  $Z(X|B)_\rho$ , we obtain

$$Z(X|B)_\rho = 2\sqrt{p_0 p_1} F(\rho_0, \rho_1) \quad (46)$$

$$= 2\sqrt{p_0 p_1} \sum_m \sqrt{\text{Tr}[\Lambda_m \rho_0] \text{Tr}[\Lambda_m \rho_1]}. \quad (47)$$

We can use the POVM  $\{\Lambda_l \otimes \Lambda_m\}$  to bound  $Z(U_1|B_1 B_2)_{\tilde{\rho}}$ :

$$\begin{aligned} & Z(U_1|B_1 B_2)_{\tilde{\rho}} \\ &= 2F \left( \sum_{u_2} p_{u_2}^2 \rho_{u_2} \otimes \rho_{u_2}, \sum_{u_2} p_{u_2 \oplus 1} p_{u_2} \rho_{u_2 \oplus 1} \otimes \rho_{u_2} \right) \quad (48) \end{aligned}$$

$$\begin{aligned} &\leq 2 \sum_{l,m} \left( \sum_u p_u^2 \text{Tr}[(\Lambda_l \otimes \Lambda_m)(\rho_u \otimes \rho_u)] \right. \\ & \quad \left. \sum_{u'} p_{u'} p_{u' \oplus 1} \text{Tr}[(\Lambda_l \otimes \Lambda_m)(\rho_{u' \oplus 1} \otimes \rho_{u'})] \right)^{\frac{1}{2}} \quad (49) \end{aligned}$$

$$\begin{aligned} &= 2 \sum_{l,m} \left( \sum_u p_u \text{Tr}[\Lambda_l \rho_u] p_u \text{Tr}[\Lambda_m \rho_u] \right. \\ & \quad \left. \sum_{u'} p_{u' \oplus 1} \text{Tr}[\Lambda_l \rho_{u' \oplus 1}] p_{u'} \text{Tr}[\Lambda_m \rho_{u'}] \right)^{\frac{1}{2}}. \quad (50) \end{aligned}$$

Now introduce the notation  $\alpha_m := p_0 \text{Tr}[\Lambda_m \rho_0]$ ,  $\beta_l := p_0 \text{Tr}[\Lambda_l \rho_0]$ ,  $\gamma_l := p_1 \text{Tr}[\Lambda_l \rho_1]$ , and  $\delta_m := p_1 \text{Tr}[\Lambda_m \rho_1]$  and notice that  $Z(X|B)_\rho = 2 \sum_l \sqrt{\beta_l \gamma_l} = 2 \sum_m \sqrt{\alpha_m \delta_m}$ . Then we can write (50) as

$$\begin{aligned} & Z(U_1|B_1 B_2)_{\tilde{\rho}} \\ &\leq 2 \sum_{l,m} \sqrt{\alpha_m \beta_l + \gamma_l \delta_m} \sqrt{\alpha_m \gamma_l + \beta_l \delta_m} \quad (51) \end{aligned}$$

$$\begin{aligned} &\leq 2 \left( \sum_{l,m} \left( \sqrt{\alpha_m \beta_l} + \sqrt{\gamma_l \delta_m} \right) \left( \sqrt{\alpha_m \gamma_l} + \sqrt{\beta_l \delta_m} \right) \right. \\ & \quad \left. - 2 \sum_{l,m} \sqrt{\alpha_m \beta_l \gamma_l \delta_m} \right) \quad (52) \end{aligned}$$

$$= 2 \left( \sum_{l,m} \left( (\alpha_m + \delta_m) \sqrt{\beta_l \gamma_l} + (\beta_l + \gamma_l) \sqrt{\alpha_m \delta_m} \right) \right)$$

$$-2 \sum_l \sqrt{\beta_l \gamma_l} \sum_m \sqrt{\alpha_m \delta_m} \quad (53)$$

$$= 2 \left( \sum_l \sqrt{\beta_l \gamma_l} + \sum_m \sqrt{\alpha_m \delta_m} - 2 \sum_l \sqrt{\beta_l \gamma_l} \sum_m \sqrt{\alpha_m \delta_m} \right) \quad (54)$$

$$= 2Z(X|B)_\rho - Z(X|B)_\rho^2, \quad (55)$$

where inequality (52) is due to Arıkan [9, Appendix D]. ■

*Entropy Polarization.*— We begin by bounding the entropy  $H(X|B)_\rho$  in terms of the probability of error  $P_e(X|B)_\rho$  when using  $B$  to determine  $X$ . This is formally defined by

$$P_e(X|B)_\rho := \min_{0 \leq \Lambda \leq 1} p_0 \text{Tr}[\Lambda \rho_0] + p_1 \text{Tr}[(\mathbb{1} - \Lambda) \rho_1]. \quad (56)$$

The error probability can be expressed in terms of a trace distance, as follows. From the definition it follows immediately that

$$P_e(X|B)_\rho = \min_{0 \leq \Lambda \leq 1} p_1 + \text{Tr}[\Lambda(p_0 \rho_0 - p_1 \rho_1)] \quad (57)$$

$$= \min_{0 \leq \Lambda \leq 1} p_0 - \text{Tr}[\Lambda(p_0 \rho_0 - p_1 \rho_1)]. \quad (58)$$

Letting  $\Gamma := p_0 \rho_0 - p_1 \rho_1$ , the minimum of (57) is achieved for  $\Lambda = \{\Gamma\}_-$ , the projector onto the negative part of  $\Gamma$ . We thus have  $P_e(X|B)_\rho = p_1 + \text{Tr}[\{\Gamma\}_- \Gamma] = p_0 - \text{Tr}[\{\Gamma\}_+ \Gamma]$ . Averaging the two different expressions leads to

$$P_e = \frac{1}{2} (p_1 + \text{Tr}[\{\Gamma\}_- \Gamma] + p_0 - \text{Tr}[\{\Gamma\}_+ \Gamma]) \quad (59)$$

$$= \frac{1}{2} - \frac{1}{2} \|\Gamma\|_1 \quad (60)$$

$$= \frac{1}{2} - \frac{1}{2} \|p_0 \rho_0 - p_1 \rho_1\|_1. \quad (61)$$

Now we have

**Lemma 5.**  $-\log(1 - P_e(X|B)_\rho) \leq H(X|B)_\rho \leq H_b(P_e(X|B)_\rho)$ .

*Proof:* Let  $\hat{X}$  denote a guess of  $X$  generated by an optimal measurement, the data processing inequality ensures that  $H(X|B)_\rho \leq H(X|\hat{X})_\rho$ . Using Fano's inequality, i.e.,  $H(X|\hat{X})_\rho \leq H_b(P_e) + P_e \log(|\mathcal{X}| - 1) = H_b(P_e)$  proves the upper bound. To prove the lower bound we use  $H_{\min}(X|B)_\rho = -\log(1 - P_e)$  [39, Section IC]. According to [40, Proposition 4.3],  $H(X|B)_\rho \geq H_{\min}(X|B)_\rho$  which proves the assertion. ■

We next bound  $Z(X|B)_\rho$  in terms of  $P_e$ .

**Lemma 6.** Let  $\rho_0$  and  $\rho_1$  be two arbitrary density operators and  $p_0, p_1 \geq 0$  such that  $p_0 + p_1 = 1$ . The parameter  $Z(X|B)_\rho := 2\sqrt{p_0 p_1} \|\sqrt{\rho_0} \sqrt{\rho_1}\|_1$  can be bounded in terms of error probability  $P_e := \frac{1}{2} - \frac{1}{2} \|p_0 \rho_0 - p_1 \rho_1\|_1$  as

$$2P_e \leq Z(X|B)_\rho \leq \sqrt{1 - (1 - 2P_e)^2}. \quad (62)$$

*Proof:* According to [41, Lemma A.2.4] we have  $\|p_0 \rho_0 - p_1 \rho_1\|_1^2 \leq 1 - Z(X|B)_\rho^2$ . Lemma A.2.6 of [41] ensures that  $1 - Z(X|B)_\rho \leq \|p_0 \rho_0 - p_1 \rho_1\|_1$ . By definition of  $P_e$  we have  $\|p_0 \rho_0 - p_1 \rho_1\|_1 = 1 - 2P_e$  which completes the proof. ■

Using Lemma 5 and 6 we can bound  $H(X|B)_\rho$  in terms of  $Z(X|B)_\rho$ .

**Proposition 7.** For  $\rho^{XB} = \sum_{x \in \{0,1\}} p_x |x\rangle \langle x|^X \otimes \rho_x^B$  and  $Z(X|B)_\rho := 2\sqrt{p_0 p_1} \|\sqrt{\rho_0} \sqrt{\rho_1}\|_1$ , we have

$$1 - \log \left( 1 + \sqrt{1 - Z(X|B)_\rho^2} \right) \leq H(X|B)_\rho \leq H_b \left( \frac{1}{2} Z(X|B)_\rho \right). \quad (63)$$

*Proof:* Follows immediately from Lemma 5 and 6. ■

Proposition 7 serves the purpose of showing that  $H(X|B)_\rho$  is near 0 or 1 if and only if  $Z(X|B)_\rho$  is near 0 or 1, respectively, i.e.,  $H(X|B)_\rho$  and  $Z(X|B)_\rho$  polarize simultaneously. This is visualized in Figure 2.

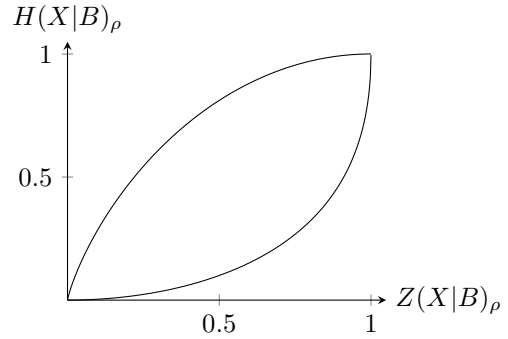


Fig. 2. The lower and upper bound for  $H(X|B)_\rho$  given in (63). It shows that  $H(X|B)_\rho$  and  $Z(X|B)_\rho$  polarize simultaneously.

#### IV. ENTANGLEMENT DISTILLATION

Inspired by previous work in a purely classical scenario [42], we consider a *concatenated* entanglement distillation scheme based on CSS codes. The explicitly concatenated structure differentiates our approach from that of Devetak and Winter [43], based on Devetak's CSS-like approach for channel coding [4].

##### A. Protocol

The scheme consists of an inner layer which performs information reconciliation (IR) in the amplitude, or  $Z$ -basis and an outer layer which performs information reconciliation in the phase, or  $X$ -basis. Each layer utilizes a quantum stabilizer code, and together the amplitude and phase codes form a CSS quantum error-correcting code. Information reconciliation at the inner layer is performed on  $M$  independent blocks, each consisting of  $L$  input systems. Due to this two-level structure the scheme has a blocklength  $N = LM$ . Letting  $K$  denote the number of unmeasured outputs per amplitude block, Figure 3 depicts the case  $L = 4$ ,  $M = 2$ , and  $K = 2$ . The unmeasured qubits after the amplitude information reconciliation are forwarded to the phase information reconciliation block.

To explain the scheme in more detail, we start with a single bipartite system  $\psi^{AB}$  shared by Alice and Bob, with  $A$  a qubit. Purifying  $\psi^{AB}$  and expressing  $A$  in the amplitude basis gives

$$|\psi\rangle^{ABE} = \sum_{z \in \{0,1\}} \sqrt{p_z} |z\rangle^A |\varphi_z\rangle^{BE}, \quad (64)$$

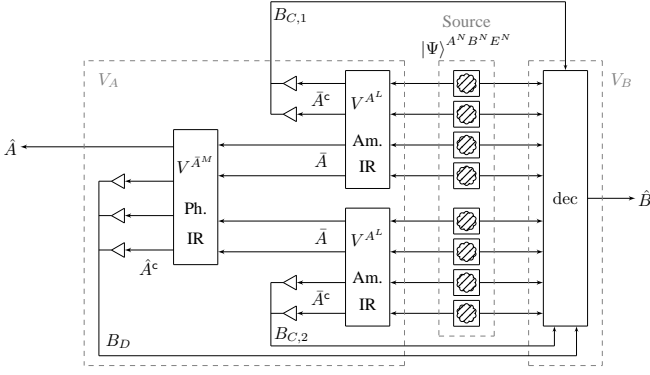


Fig. 3. The entanglement distillation scheme for  $L = 4$ ,  $M = 2$  and  $K = 2$ . A source (middle) produces states  $|\Psi\rangle$  and distributes the  $A$  subsystems to Alice (left), the  $B$  subsystems to Bob (right), and retains the  $E$  subsystems. Alice performs the amplitude information reconciliation (IR) transformation  $M$  times, measures part of the output with respect to the amplitude basis and sends the outcomes to Bob over a classical channel. The non-measured qubits are fed to an IR operation in the complementary phase basis and part of the outcome is again measured and sent to Bob. Using Alice's classical information, Bob runs a decoder such that his and Alice's outcome—described by the systems  $\hat{A}$  and  $\hat{B}$ —are a good approximation to maximally entangled qubits.

where  $p_z$  is some probability distribution and  $\{|\varphi_z\rangle^{BE}\}$  some set of normalized states, not necessarily orthogonal. The input to each block of the first layer will be the state

$$|\Psi\rangle^{A^L B^L E^L} = \left(|\psi\rangle^{ABE}\right)^{\otimes L}. \quad (65)$$

The amplitude stabilizer code is chosen so that Bob can determine  $z^L$ , with probability exceeding  $1 - \epsilon_1$  using his systems  $B^L$  if he is also supplied with the syndromes of the code. These are determined by Alice and transmitted to him over a public classical channel. Given a particular  $z^L$ , denote by  $\bar{z}^c$  the syndrome and  $\bar{z}$  the encoded information. Since stabilizer codes are linear codes,  $z^L$  determines  $(\bar{z}, \bar{z}^c)$  and *vice versa*. Moreover, for every stabilizer code on  $L$  systems there exists a unitary operation which maps the stabilizer and encoded operators to physical qubits. Call this unitary  $V^{A^L}$ ; after applying it, Alice need only measure certain subsystems to generate the syndrome. Let  $\bar{A}^c$  be the systems which are then measured to yield the syndromes and  $\bar{A}$  the remaining systems (corresponding to encoded qubits). After applying the unitary, the joint state becomes

$$\begin{aligned} |\Psi_1\rangle^{A^L B^L E^L} &= V^{A^L} |\Psi\rangle^{A^L B^L E^L} \\ &= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sqrt{p_{z^L(\bar{z}, \bar{z}^c)}} |\bar{z}\rangle^{\bar{A}} |\bar{z}^c\rangle^{\bar{A}^c} |\varphi_{z^L(\bar{z}, \bar{z}^c)}\rangle^{B^L E^L}. \end{aligned} \quad (66)$$

$$= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sqrt{p_{z^L(\bar{z}, \bar{z}^c)}} |\bar{z}\rangle^{\bar{A}} |\bar{z}^c\rangle^{\bar{A}^c} |\varphi_{z^L(\bar{z}, \bar{z}^c)}\rangle^{B^L E^L}. \quad (67)$$

Sending  $\bar{z}^c$  to Bob can be modeled as copying  $\bar{z}^c$  to a register  $B_C$  he controls, plus another one for the environment,  $E_C$  (for notational simplicity, we suppress the dependence of  $z^L$  on  $(\bar{z}, \bar{z}^c)$ ):

$$|\Psi_2\rangle^{A^L B^L B_C E^L E_C}$$

$$= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sqrt{p_{z^L}} |\bar{z}\rangle^{\bar{A}} |\bar{z}^c\rangle^{\bar{A}^c} |\bar{z}^c\rangle^{B_C} |\bar{z}^c\rangle^{E_C} |\varphi_{z^L}\rangle^{B^L E^L}. \quad (68)$$

Bob's decoding operation attempts to determine  $\bar{z}$  using the information contained in systems  $B^L$  and  $B_C$ . It can be thought of as a measurement on  $B^L$  conditioned on the value in the register  $B_C$ ; call its elements  $\Lambda_{\bar{z}; \bar{z}^c}^{B^L}$ . Performed coherently, it stores the result  $z^L$  in an extra ancillary system, say  $B'_C$ . The post-measurement state is then

$$\begin{aligned} &|\Psi_3\rangle^{A^L B^L B_C B'_C E^L E_C} \\ &= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sum_{z'} \sqrt{p_{z^L}} |\bar{z}\rangle^{\bar{A}} |\bar{z}^c\rangle^{\bar{A}^c} |\bar{z}^c\rangle^{B_C} |\bar{z}'\rangle^{B'_C} \\ &\quad \sqrt{\Lambda_{\bar{z}; \bar{z}^c}^{B^L}} |\varphi_{z^L}\rangle^{B^L E^L} |\bar{z}^c\rangle^{E_C}. \end{aligned} \quad (69)$$

Regarding the pair  $(B_C B'_C)$  as the system  $C^L$ , a simple fidelity calculation shows that  $\langle \Psi_3 | \hat{\Psi}_3 \rangle \geq 1 - \epsilon_1$ , where

$$\begin{aligned} &|\hat{\Psi}_3\rangle^{A^L B^L C^L E^L E_C} \\ &= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sqrt{p_{z^L}} |(\bar{z}, \bar{z}^c)\rangle^{A^L} |\varphi_{z^L}\rangle^{B^L E^L} |(\bar{z}, \bar{z}^c)\rangle^{C^L} |\bar{z}^c\rangle^{E_C}. \end{aligned} \quad (70)$$

The outer layer performs phase information reconciliation on  $M$  instances of the  $\hat{A}$  systems of the state  $|\Psi_3\rangle$ , where Bob's side information is given by  $B^L C^L$  in each instance. In contrast to the inner layer, here the information to be reconciled is not a bit, but a sequence of bits. Therefore, to use the formalism of stabilizer codes, we either need to consider codes over larger dimension or multilevel coding schemes. Either would work for our purposes, but for concreteness let us opt for the latter. Here, Alice and Bob assemble a block of  $M$  systems  $\hat{A}$  and sequentially run blocksize- $M$  phase IR protocols on the first qubits in each of the  $\hat{A}$ , then the second, and so forth. At each step they treat already reconciled systems as side information for the current step.

Ultimately the effect of this procedure can be regarded, as at the inner layer, as applying a unitary  $V^{\hat{A}^M}$  and measuring a subset of the output qubits to obtain the syndromes. These measurement results are sent to Bob, which is modeled as copying them to a register  $B_D$  he controls, plus another one for the environment  $E_D$ . Remaining at the end of this process are a set of unmeasured qubits, the encoded qubits  $\hat{A}$  of the error-correcting code used in phase information reconciliation.

## B. Reliability & Rate

Now let us examine the scheme more quantitatively. Associated with any set of qubits are a set of  $X$  and  $Z$  operators acting on these qubits; abusing notation, let us refer to the entire collection of these by, for instance,  $X^{\hat{A}}$  and  $Z^{\hat{A}}$  for the set  $\hat{A}$ . The amplitude IR protocol is chosen to be  $\epsilon_1$ -good, i.e.,  $p_{\text{err}}(Z^{A^L} | B^L B_C)_{\Psi_2} \leq \epsilon_1$ . Since the scheme uses  $M$  independent amplitude information reconciliation blocks, we can use the union bound to write

$$p_{\text{err}}\left(Z^{A^N} | B^N B_C^M\right)_{\Psi_2^{\otimes M}} \leq M\epsilon_1. \quad (71)$$

Sidestepping the details of the multilevel coding for the moment, the phase IR protocol is chosen to have

$$p_{\text{err}} \left( X^{\hat{A}^M} \middle| B^N C^N B_D \right) \leq \epsilon_2. \quad (72)$$

Clearly  $X^{\hat{A}}$  (cf. Figure 3) is a deterministic function of  $X^{\hat{A}^M}$  due to the action of  $V^{\hat{A}^M}$ . However, since this unitary implements a linear function in the basis conjugate to the amplitude observable  $Z^{\hat{A}^M}$ , it also implements a linear function in the amplitude basis itself. (This fact was used to show that Arıkan's polar encoding circuit is directly useful in the quantum setting in [12].) Therefore,  $Z^{\hat{A}}$  is a deterministic function of  $Z^{\hat{A}^M}$  and hence also of  $Z^{\hat{A}^N}$ .

$$p_{\text{err}} \left( X^{\hat{A}} \middle| B^N C^N B_D \right) \leq \epsilon_2 \quad \text{and} \quad (73)$$

$$p_{\text{err}} \left( Z^{\hat{A}} \middle| B^N B_C^M \right) \leq M\epsilon_1. \quad (74)$$

These conditions ensure that Alice and Bob share a good approximation to  $|\hat{A}\rangle$  maximally entangled qubit pairs. Alice's part of the distillation process (summarized in the left hand side of Figure 3) can be described by a unitary  $U^{A^N \rightarrow \hat{A} B_C^M B_D E_C^M E_D}$ . Bob's part is to decode the state  $|\Psi\rangle^{A^N \hat{B}^N E^N}$  using  $B^N$  and the side information  $B_C^M B_D$  he receives from Alice. Inequalities (73) and (74) together with [44, Theorem 1] ensure that there exists a decoding unitary  $U_B^{B^N B_C^M B_D \rightarrow \hat{B}}$ . It is constructed directly from the two IR decoders.

To make the reliability statement precise, define  $V_A := U_A^{A^N \rightarrow \hat{A} B_C^M B_D E_C^M E_D}$ ,  $V_B := U_B^{B^N B_C^M B_D \rightarrow \hat{B}}$ , introduce

$$\mathcal{E}(\cdot) := \text{Tr}_{E_C^M E_D E^N} [V_B (V_A(\cdot) V_A^\dagger) V_B^\dagger] \quad (75)$$

and define  $\delta(\psi, \phi) = \frac{1}{2} \|\psi - \phi\|_1$ .

**Proposition 8.** *Let  $|\phi\rangle_d^{\hat{A}\hat{B}}$  be a maximally entangled state of dimension  $d$ , where  $d = \dim \hat{A}$ . Then*

$$\delta \left( \phi_d^{\hat{A}\hat{B}}, \mathcal{E}(\Psi^{A^N B^N E^N}) \right) \leq \sqrt{2\epsilon_2} + \sqrt{2M\epsilon_1}. \quad (76)$$

*Proof:* This proposition follows immediately from (73), (74) and [44, Theorem 1]. ■

The rate of the scheme is defined as the number of output qubits divided by the number of input qubits (cf. Equation (79)).

**Theorem 9.** *The rate of the scheme is*

$$R = \frac{1}{L} I(\bar{A})_{B^L C^L} + \frac{o(L)}{L}, \quad (77)$$

where

$$\begin{aligned} & |\Psi'_3\rangle \\ &= \sum_{(\bar{z}, \bar{z}^c) \in \{0,1\}^L} \sqrt{p_{\bar{z}^L(\bar{z}, \bar{z}^c)}} |\bar{z}\rangle^{\bar{A}} |\bar{z}^c\rangle^{\bar{A}^c} |\varphi_{\bar{z}^L(\bar{z}, \bar{z}^c)}\rangle^{B^L E^L} |z^L(\bar{z}, \bar{z}^c)\rangle^{C^L}. \end{aligned} \quad (78)$$

*Proof:* The rate of the scheme is

$$R := \frac{|\hat{A}|}{N} \quad (79)$$

$$= \frac{N - M|\bar{A}^c| - |\hat{A}^c|}{N} \quad (80)$$

$$= \frac{N - MLH(Z^A|B)_\psi - MH(X^{\bar{A}}|B^L C^L)_{\Psi'_3}}{N}. \quad (81)$$

Since we only care about the  $\bar{A}$  register, the copy of  $\bar{z}^c$  in  $E_C$  does not contribute to the entropy in the final term. Thus,

$$H(X^{\bar{A}}|B^L C^L)_{\Psi'_3} = H(X^{\bar{A}}|B^L C^L)_{\Psi'_3}, \quad (82)$$

where  $|\Psi'_3\rangle$  is as defined in (78). Using Lemma 1 we obtain

$$\begin{aligned} & H(X^{\bar{A}}|B^L C^L)_{\Psi'_3} \\ &= \log |\bar{A}| + H(\bar{A}|B^L C^L)_{\Psi'_3} \end{aligned} \quad (83)$$

$$= L \left( 1 - H(Z^A|B)_\psi \right) + H(\bar{A}|B^L C^L)_{\Psi'_3} + o(L), \quad (84)$$

where the final step is ensured by the fact that the code is capacity-achieving. We thus can write (81) as

$$R = - \frac{H(\bar{A}|B^L C^L)_{\Psi'_3}}{L} + \frac{o(L)}{L} \quad (85)$$

$$= \frac{I(\bar{A})_{B^L C^L}}{L} + \frac{o(L)}{L}, \quad (86)$$

which proves the assertion. ■

**Corollary 10.** *For  $\psi$  as given in (64), the rate is larger than or equal to the coherent information, i.e.,*

$$R \geq \max \left\{ 0, I(A)_B \right\}. \quad (87)$$

*Proof:* Recall that the states  $|\psi\rangle$  and  $|\Psi'_3\rangle$  are defined in (64) and (78). Let

$$|\psi'\rangle^{ABCE} = \sum_{z \in \{0,1\}} \sqrt{p_z} |z\rangle^A |z\rangle^C |\varphi_z\rangle^{BE}. \quad (88)$$

By definition (cf. (79)) the rate is non-negative. As explained in the proof of Theorem 9 we can write

$$R = \frac{N - MLH(Z^A|B)_\psi - MH(X^{\bar{A}}|B^L C^L)_{\Psi'_3}}{N}. \quad (89)$$

Using the chain rule we obtain

$$\begin{aligned} & H(X^{\bar{A}}|B^L C^L)_{\Psi'_3} \\ &= H(X^{A^L}|B^L C^L)_{\Psi'_3} - H(X^{\bar{A}^c}|B^L C^L X^{\bar{A}})_{\Psi'_3} \end{aligned} \quad (90)$$

$$= LH(X^A|BC)_{\psi'} - H(X^{\bar{A}^c}|B^L C^L X^{\bar{A}})_{\Psi'_3}. \quad (91)$$

We thus have

$$\begin{aligned} R &= 1 - H(Z^A|B)_\psi - H(X^A|BC)_{\psi'} \\ &\quad + \frac{H(X^{\bar{A}^c}|B^L C^L X^{\bar{A}})_{\Psi'_3}}{L} \end{aligned} \quad (92)$$

$$\geq \max \left\{ 0, 1 - H(Z^A|B)_\psi - H(X^A|BC)_{\psi'} \right\} \quad (93)$$

$$= \max \left\{ 0, -H(A)_B \right\} \quad (94)$$

$$= \max \left\{ 0, I(A)_B \right\}. \quad (95)$$



Equality (94) holds since

$$1 - H(Z^A|B)_{\psi} - H(X^A|BC)_{\psi'} = 1 - H(A|B)_{\psi'} - H(X^A|BC)_{\psi'} \quad (96)$$

$$= -H(A|B)_{\psi'} - H(A|BC)_{\psi'} \quad (97)$$

$$= -H(A|B)_{\psi'} - H(C|AB)_{\psi'} \quad (98)$$

$$= -H(AC|B)_{\psi'} \quad (99)$$

$$= -H(A|B)_{\psi}, \quad (100)$$

where (97) uses Lemma 1 and that  $H(Z^A|BC)_{\psi'} = 0$ . ■

Note that the scheme presented above works for any CSS code meeting the two reliability conditions (73) and (74). A particularly favorable family of codes are the quantum polar codes. Using these codes for Pauli or erasure channels, we know how to build an efficient encoder and decoder having essentially linear complexity and being reliable for large enough blocklength. This will be explained next.

## V. USING QUANTUM POLAR CODES FOR PAULI OR ERASURE CHANNELS

By using polar codes, Alice and Bob can perform the operations explained in Section IV in a computationally efficient manner for states  $|\psi\rangle^{ABE}$  that arise from sending half of an entangled pair through a Pauli or erasure channel.

### A. Code Construction

Before the protocol starts one must construct the code, i.e., determine the qubits comprising the systems  $\bar{A}^c$  at the inner layer and  $\hat{A}^c$  at the outer layer. (Recall that these are the qubits that are measured by Alice and whose measurement outcomes are sent to Bob.) Constructing the system  $\bar{A}^c$  can be approximately done in linear time using Tal and Vardy's algorithm [45] and its adaptation to an asymmetric setup as explained in [33], or alternatively using the more recent algorithm by Tal *et al.* [46].

To determine the system  $\hat{A}^c$  requires more effort. Applying the above algorithm for a "super-source" seen by the outer layer will not be efficient in the overall blocklength  $N$  since its alphabet size is exponential in  $L$ . Nonetheless, due to the structure of the inner layer, it is perhaps possible that the method of approximation by limiting the alphabet size [45], [46] can be extended to this case.

### B. Encoding

As described in Section IV-A and Figure 3, starting with a state  $|\Psi\rangle^{ANB^NE^N}$ , Alice first applies  $M$  times a unitary  $V^{A^L}$  to perform amplitude information reconciliation, which is in the specific case of using quantum polar codes  $V^{A^L} = \sum_{z^L \in \{0,1\}^L} |G_L z^L\rangle \langle z^L|$ , where  $G_L = G^{\otimes \log L}$  and  $G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  denotes the Arkan polar transform [9]. Alice measures the frozen qubits with respect to the amplitude basis and sends the outcome to Bob.

Alice next applies another polar transform  $V^{\bar{A}^M}$ —this time with respect to the phase basis—to the  $M$  systems  $\bar{A}$ . Here we use a multilevel coding scheme, as mentioned above and

described in more detail in Section III-A. The unitary  $V^{\bar{A}^M}$  applied to elements in the phase basis acts as  $(\tilde{G}_M^K)^T$ , i.e.,

$$V^{\bar{A}^M} = \frac{1}{2^{KM}} \sum_{x,x',z \in \{0,1\}^{KM}} (-1)^{x' \cdot \tilde{G}_M^K z + x \cdot z} |\tilde{x}'\rangle \langle \tilde{x}| \quad (101)$$

$$= \sum_{x \in \{0,1\}^{KM}} |(\tilde{G}_M^K)^T \tilde{x}\rangle \langle \tilde{x}|, \quad (102)$$

where we have used  $(\tilde{G}_M^K)^{-1} = \tilde{G}_M^K$ . The frozen qubits are measured with respect to the phase basis and its outcomes are sent again to Bob. The remaining qubits form the system  $\hat{A}$ .

**Theorem 11.** *The encoding of the distillation scheme can be done with  $O(N \log N)$  steps.*

*Proof:* The  $M$  polar transformations with respect to the amplitude basis can be performed in  $O(ML \log L)$  complexity. The second polar transform, this time with respect to the phase basis, has  $O(MK \log(MK))$  complexity, where  $K = |\bar{A}|$ . Hence it follows that all the amplitude measurements require  $O(M(L - K))$  operations and all phase measurements can be done with  $O(MK)$  complexity. From the polarization phenomenon (cf. Theorem 3) we obtain that  $K = O(L)$  which proves the assertion. ■

### C. Decoding

Decoding of the  $B^N$  system, with the additional information stored in the  $B_C^M B_D$  registers, can be done by combining ideas from [44] and [12]. As shown schematically in Figure 4, Bob's operation is constructed by using the classical polar decoders for amplitude and phase IR in sequence. Note however that these two decoding tasks are not independent as this would neglect possible correlations between amplitude and phase.

In the first step Bob performs the amplitude IR decoding operation  $\mathcal{D}_A$  ( $M$  times), which corresponds to the standard classical polar decoder as introduced in [9], and stores the result in an auxiliary system  $F_i$ ,  $i \in \{1, \dots, M\}$ . Each instance of  $\mathcal{D}_A$  requires the corresponding frozen information, the values  $\bar{z}^c$ , which is provided in  $B_C^M$ .

Bob next performs the phase IR decoding operation  $\mathcal{D}_P$ , using the information gained from decoding the first layer. Following [42], we next show how the standard polar decoding procedure can be adapted for the outer layer of a two-stage polar scheme.

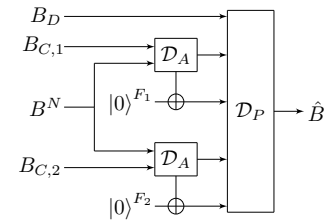


Fig. 4. Bob's task in the entanglement distillation process for  $M = 2$ . With the help of the ancilla systems  $F_1$  and  $F_2$ , the classical decoders  $\mathcal{D}_A$  and  $\mathcal{D}_P$  are utilized to distill entanglement.

*Efficient Concatenated Classical Source Coding.*— We start by presenting a concatenated classical coding scheme based

on polar codes that can be used for efficient source coding. The scheme has been introduced in [42] for efficient channel coding at the optimal rate. Let  $\mathcal{I} \subseteq [L]$  denote the indices of the frozen bits at the inner layer and let  $K = L - |\mathcal{I}|$ . The set  $\mathcal{O} \subseteq [KM]$  denotes the indices of the frozen bits at the outer layer.<sup>5</sup> Figure 5, depicts the scheme schematically for the setup of  $L = 4$ ,  $M = 2$ ,  $\mathcal{I} = \{2, 3\}$ ,  $K = 2$  and  $\mathcal{O} = \{2\}$ . In the following we prove that there exists an encoder and decoder that require  $O(N \log N)$  steps having an error probability not greater than  $O(L2^{-M^\beta})$  for any  $\beta < \frac{1}{2}$ .

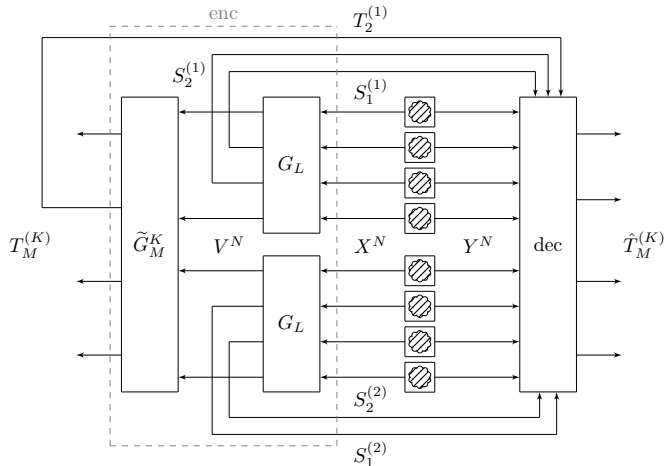


Fig. 5. The source coding scheme for  $L = 4$ ,  $M = 2$ ,  $\mathcal{I} = \{2, 3\}$ ,  $K = 2$  and  $\mathcal{O} = \{2\}$ . A source produces  $N$  i.i.d. copies of correlated random variables  $(X, Y)^N$ . The encoder first applies  $M$  polarization transforms  $G_L$  to  $X^L$ , obtaining a vector  $V^N$ . The frozen bits that are determined by  $\mathcal{I}$  and denoted by  $S_{L-K}^{(M)}$  are sent to the decoder whereas the remaining bits are encoded by another polarization transform—performed in a multilevel construction as explained in Section III-A—resulting in a vector  $T_M^{(K)}$ . The frozen bits of  $T_M^{(K)}$  (determined by  $\mathcal{O}$ ) are sent to the decoder again. Using the bits received from the encoder the decoder outputs  $\hat{T}_M^{(K)}$ ; a guess for  $T_M^{(K)}$ .

As depicted in Figure 5 the encoder consists of two parts. It first applies  $M$  identical inner encoding transforms—it performs  $M$  times the polar transform  $G_L$ . The outcome  $V^N$  can be classified into two systems determined by the code construction. The frozen bits, denoted by  $S_{L-K}^{(M)}$ , are sent to the decoder whereas the others are encoded a second time by the outer encoder. The outer encoder applies another polarization transform in the multilevel technique explained in Section III-A which outputs  $T_M^{(K)}$ .<sup>6</sup> The frozen bits of  $T_M^{(K)}$ , determined by the code construction, are sent to the decoder again.

**Lemma 12.** *The classical encoder explained above has complexity  $O(N \log N)$ .*

*Proof:* The inner encoder performs  $M$  times a standard polar transform  $G_L$  which has been shown to require at

<sup>5</sup>The bits at the outer layer are numbered with respect to the multilevel structure, i.e., each binary decompressor is numbered sequentially.

<sup>6</sup> $T_j^{(i)}$  denotes the  $j$ th output of the  $i$ th binary outer encoder in the sequential multilevel structure.

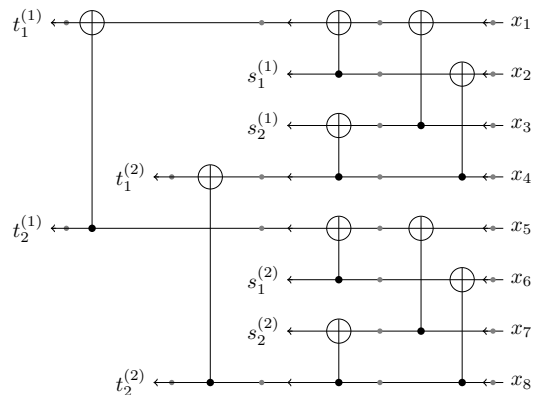


Fig. 6. Encoding circuit for the setup  $L = 4$ ,  $M = 2$ ,  $\mathcal{I} = \{2, 3\}$ ,  $K = 2$  and  $\mathcal{O} = \{2\}$ . Here  $s_j^{(i)}$  denote the frozen bits that are sent to the decoder at the inner layer. We have a single frozen bit at the outer layer,  $t_2^{(1)}$ , that is sent to the decoder as well. The small gray dots represent variables in the network and correspond to nodes in Fig. 7.

most  $O(L \log L)$  steps each [9]. The outer encoder consists of  $K$  multiplications with the matrix  $G_M$ , each requiring  $O(M \log M)$  operations [9]. As justified in Section V and ensured by the polarization phenomenon (cf. Theorem 3),  $K = O(L)$ . We thus conclude that the total encoding requires  $O(ML \log L) + O(LM \log M) = O(N \log N)$  steps. ■

The decoding is more challenging. An important feature of the decoder is that the inner layer decompressors must be interleaved with the outer layer decompressors in order to ensure that all required variables are known at the appropriate steps. To illustrate, we explain in detail how the decoding is done for the setup  $L = 4$ ,  $M = 2$ ,  $\mathcal{I} = \{2, 3\}$ ,  $K = 2$  and  $\mathcal{O} = \{2\}$ . The logical structure of the successive cancellation decoder is shown in Figure 7. Figure 10 of [9] depicts a similar representation of the original successive cancellation decoder. To see the close affinity between the encoding and decoding process, Figure 6 visualizes the encoder for the setup defined above.

Each node in Figure 7 is responsible for computing a likelihood ratio (LR) arising during the algorithm; the parameters below each node represent the variables involved in the associated LR computation. Starting from the left we traverse the diagram to the right at whose border we can compute the LRs. Then we transmit the results back to the left. Here  $\hat{t}_j^{(i)}$  denotes the  $j$ th output of the  $i$ th decompressor at the outer layer and  $s_j^{(i)}$  denotes the  $j$ th frozen input for the  $i$ th inner encoding block which has been sent to the decoder as explained above.

The decoding begins by activating node 1, which would like to compute the LR for  $T_1^{(1)}$  given  $Y_1^8$ . For this it needs the LRs for the first inputs to the two super-channels, and so node 1 activates node 2, which is responsible for computing the LR for the first input to the first super-channel. This computation proceeds exactly as the usual successive cancellation decoder, recursively combining the LRs of the physical channels by calling node 3 and then 6. Assembling their results, node 2 can compute its LR and transmits the result to nodes 1 and 16.

Meanwhile, node 1 has also requested the LR of node 9, which performs the same calculation as node 2 for the second super-channel, again forwards the result to nodes 1 and 16. Now node 1 is able to compute the final desired LR and can therefore guess  $\hat{t}_1^{(1)}$ . Node 16 next guesses  $\hat{t}_2^{(1)}$ , which is easy since this is a frozen bit and therefore available at the decoder (i.e., the decoder sets  $\hat{t}_2^{(1)} = t_2^{(1)}$ ), completing first decompressor of the outer layer.

Node 16 passes control to node 17 in order to compute the LR for  $T_1^{(2)}$ . This requires the LR for second inputs to the two super-channels, so nodes 18 (and later 21) are called. Node 18 finishes the decompression of the first super-channel in the usual way, while node 21 completes the decompression of the second super-channel. *Neither of these can occur until the first outer layer decompressor is finished.* After the inner layer decompression is complete, node 17 can guess  $\hat{t}_1^{(2)}$  and node 24 can finally guess  $\hat{t}_2^{(2)}$ , completing the second decompressor of the outer layer. In general, decompression of the  $M$  different  $k$ th inputs at the inner layer has to wait for the  $(k-1)$ th decompressor to finish at the outer layer.

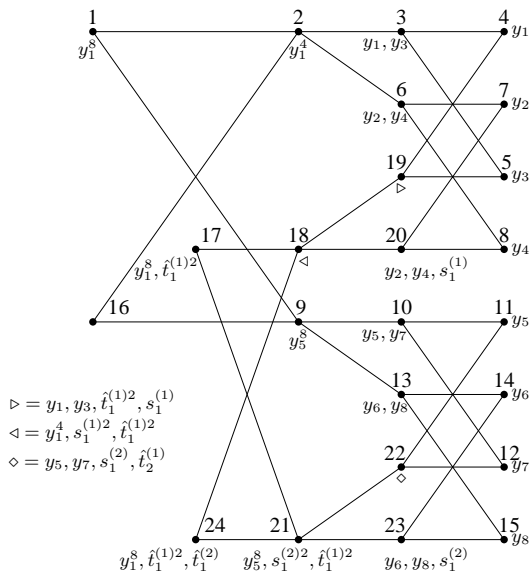


Fig. 7. Logical structure of the successive cancellation decoder for the setup  $L = 4$ ,  $M = 2$ ,  $\mathcal{I} = \{2, 3\}$ ,  $K = 2$  and  $\mathcal{O} = \{2\}$  (compare with [9, Fig. 10]). Note that  $\hat{t}_j^{(i)}$  denotes the  $j$ th output of the  $i$ th decompressor at the outer layer and  $s_j^{(i)}$  denotes the  $j$ th frozen bit at the  $i$ th inner encoding block. The numbering of the nodes represents the order in which they get activated in the decoding process.

**Lemma 13.** *The classical decoder explained above has complexity  $O(N \log N)$ .*

*Proof:* Let  $V_i$  denote the  $i$ th non-frozen output of an inner encoding block. The decoder proceeds by employing, in sequence, the  $K$  decompressors for blocklength- $M$  compression of  $V_i$  given  $Y^L V^{i-1}$ . This ensures that at all times the decoder has all the required previous inputs  $V^{i-1}$ . Each decompressor can be executed using  $O(M \log M)$  operations, given the corresponding likelihood ratio of  $V_i | Y^L V^{i-1}$ . All such likelihoods can be computed in  $O(L \log L)$  steps, and

each of the  $M$  inner encoding blocks requires its own likelihood calculation, as the values taken by  $V^{i-1}$  can differ in each case. Using  $K = O(L)$  which follows from the polarization phenomenon (cf. Theorem 3), we find that the decoder has complexity  $O(N \log N)$ . ■

We next analyze the reliability of the multilevel encoder and decoder explained above. Suppose we would like to compress ( $L$  instances of)  $(V_1, \dots, V_n)$  relative to side information  $Y$ , by sequentially compressing  $V_i$  relative to  $V^{i-1} Y$ . Define  $\hat{V}_i$  to be the output of the decompressor, let  $\mathcal{A}_i$  be the event that  $\hat{V}_i \neq V_i$  (i.e., that the decompressor makes a mistake at position  $i$ ), and let  $\mathcal{B}_i := \cup_{k=1}^i \mathcal{A}_k$ . Note that  $\Pr[\mathcal{B}_n]$  is the probability of incorrectly decoding at least one  $V_i$  for  $i \in [n]$ . Let  $r$  be a bound on the probability of that we decode incorrectly at any step and that the previous steps are all correct:  $\Pr[\mathcal{A}_j \cap \mathcal{B}_{j-1}^c] \leq r$  for all  $j \in [n]$ . Then

**Lemma 14.** *For  $n \in \mathbb{Z}^+$  and  $r$  as defined above, we have*

$$\Pr[\mathcal{B}_n] \leq nr. \quad (103)$$

*Proof:* The proof proceeds by induction over  $n$ ; the case  $n = 1$  holds by assumption. The induction step is as follows:

$$\Pr[\mathcal{B}_{n+1}] = \Pr[\mathcal{B}_n \cup \mathcal{A}_{n+1}] \quad (104)$$

$$= \Pr[\mathcal{B}_n] + \Pr[\mathcal{A}_{n+1} \cap \mathcal{B}_n^c] \quad (105)$$

$$\leq \Pr[\mathcal{B}_n] + r \quad (106)$$

$$\leq (n+1)r. \quad (107)$$

where (106) follows by assumption and (107) uses the induction hypothesis. ■

Now the reliability statement follows easily.

**Lemma 15.** *The error probability of the encoder and decoder introduced above satisfies  $P_{\text{err}} = O(L 2^{-M^\beta})$  for any  $\beta < \frac{1}{2}$ .*

*Proof:* For the polar source coding scheme, note that  $\Pr[\mathcal{A}_i \cap \mathcal{B}_{i-1}^c] + x \in O(2^{-M^\beta})$ , where  $x$  is the probability that  $\hat{V}_i \neq V_i$  given that a mistake previously occurred, but where we still give the correct  $V^{i-1}$  to the decompressor. We can therefore upper bound  $r$  in Lemma 14 by  $O(2^{-M^\beta})$  [30]. Thus, the probability of incorrectly decoding any of the  $K$   $V_i$  is  $O(L 2^{-M^\beta})$ . ■

We note that for the phase reconciliation task we cannot directly use that decoder since we have measured some qubits at the inner layer with respect to the amplitude basis (the qubits belonging to the system  $\hat{A}^c$ ), which implies that we do not have knowledge about the phase-basis-measurements of these qubits. In the following we present two different approaches to resolve this problem. We first show that the phase-basis measurements we do not know can be chosen at random without affecting the decoder's reliability. Alternatively, we show how to adapt the classical decoding algorithm introduced in [42] such that it does not require the above mentioned phase-basis measurements and still remains computationally efficient.

*Choosing Measured Qubits at Inner Layer at Random.—*

We show that randomly choosing the qubits that have been measured with respect to the amplitude basis at the inner layer does not affect the reliability of the outer IR decoder

for symmetric sources. This implies that the computationally efficient decoder introduced in [42] and explained above can be used to reliably decode the phase IR.

Recall that the state

$$|\psi\rangle^{ABCR} = \sum_z \sqrt{p_z} |z\rangle^A |z\rangle^C |\varphi_z\rangle^{BR} \quad (108)$$

produced by the source is relevant for the phase IR task. Taking  $L$  copies and applying the polarization transform  $G_L$  to  $A^L$  and  $C^L$  gives

$$\begin{aligned} & |\Psi\rangle^{A^L B^L C^L R^L} \\ &= \sum_{z^L \in \{0,1\}^L} \sqrt{p_{z^L}} |G_L z^L\rangle^{A^L} |G_L z^L\rangle^{C^L} |\varphi_{z^L}\rangle^{B^L R^L}. \end{aligned} \quad (109)$$

Expressing  $A^L$  in the conjugate phase basis gives

$$\begin{aligned} & |\Psi\rangle^{A^L B^L C^L R^L} \\ &= \frac{1}{\sqrt{2^L}} \sum_{x^L, z^L \in \{0,1\}^L} \sqrt{p_{z^L}} (-1)^{x^L \cdot G_L z^L} |\tilde{x}^L\rangle^{A^L} \\ & \quad |G_L z^L\rangle^{C^L} |\varphi_{z^L}\rangle^{B^L R^L} \end{aligned} \quad (110)$$

$$= \frac{1}{\sqrt{2^L}} \sum_{x^L \in \{0,1\}^L} |\tilde{x}^L\rangle^{A^L} (Z^{x^L})^{C^L} |\xi\rangle^{C^L B^L R^L}, \quad (111)$$

where

$$|\xi\rangle^{C^L B^L R^L} = \sum_{z^L \in \{0,1\}^L} \sqrt{p_{z^L}} |G_L z^L\rangle^{C^L} |\varphi_{z^L}\rangle^{B^L R^L}. \quad (112)$$

The state of  $B^L C^L$  conditioned on the value  $x^L$  on  $A^L$  is just  $(Z^{x^L})^{C^L} \xi^{C^L B^L} (Z^{x^L})^{C^L}$ . The source is symmetric because the conditional states are all related by unitary action, here on the  $C^L$  systems.

Dividing  $A^L$  and  $C^L$  into two systems  $\bar{A}$  and  $\bar{A}^c$  respectively  $\bar{C}$  and  $\bar{C}^c$  enables us to write the marginal state of  $B^L C^L$  conditioned on only  $\bar{x}$  as

$$\Theta_{\bar{x}}^{B^L C^L} = \frac{1}{2^{|\bar{A}^c|}} \sum_{\bar{x}^c \in \{0,1\}^{|\bar{A}^c|}} (Z^{\bar{x}})^{\bar{C}} (Z^{\bar{x}^c})^{\bar{C}^c} \xi^{C^L} (Z^{\bar{x}})^{\bar{C}} (Z^{\bar{x}^c})^{\bar{C}^c}. \quad (113)$$

Recall that we would like to use the recursive likelihood formulas in the successive cancellation decoder to determine the likelihoods of  $\bar{x}_i$  given  $\bar{x}^{i-1} B^L C^L$ . Note that for this likelihoods only the state  $|\Psi\rangle^{\bar{A} B^L C^L}$  is relevant— $\bar{A}^c$  has been discarded.

We can mimic this having this state in the standard polar coding setup as follows. Given  $|\Psi\rangle^{A^L B^L C^L R^L}$ , applying a CPTP map to  $C^L$  that randomly performs a  $Z$  operation on each of the qubits in  $\bar{A}^c$  yields a state whose  $B^L C^L$  marginal given  $X^{\bar{A}}$  is the same as  $\Theta_{\bar{x}}^{B^L C^L}$ . Moreover, the bits  $\bar{x}^c$  are now uncorrelated with  $B^L C^L$  and  $\bar{x}$ , which can be seen from direct calculation:  $Z^{\bar{x}^c}$  acting on  $\bar{C}^c$  goes to  $Z^{u+\bar{x}^c}$  for a random  $u$ , so the  $\bar{x}^c$  dependence is eliminated.

Now proceed with the usual recursive likelihood calculation. Nominally, these likelihood (ratios) are functions of the bits  $\bar{x}^c$ . However, since these are independent of everything else, the ratios computed by the recursion formulas using  $\bar{x}^c$  have the

same values as the ratios we are looking for. The dependence on  $\bar{x}^c$  is irrelevant. Moreover, the decoder may as well choose  $\bar{x}^c$  himself.

*Neglecting Measured Qubits at Inner Layer.*— We show how to efficiently compute the likelihood ratios at the outer layer without using the qubits measured with respect to the amplitude basis at the inner layer. Let  $\mathcal{F} \subseteq [N]$  denote the indices of the qubits measured at the inner layer, i.e., the qubits belonging to the  $M$  systems  $\bar{A}^c$  and let  $\mathcal{F}^c := [N] \setminus \mathcal{F}$ . For each  $i \subseteq [N]$  we define  $\mathcal{F}_{(i)} := [i] \cap \mathcal{F}$  and  $\mathcal{F}_{(i)}^c := [i] \cap \mathcal{F}^c$ .

In order to execute Arkan's classical SC-decoder at the outer layer, we need to compute for each  $i \in [N]$  such that  $X_i^{A^N} \in \mathcal{F}^c$  the following likelihood ratio

$$L^{(i)}(b^N, \hat{x}[\mathcal{F}_{(i)}^c]) := \frac{\Pr \left[ X_i^{A^N} = 0 \mid B^N = b^N, \bigcap_{j \in \mathcal{F}_{(i)}^c} \{X_j^{A^N} = \hat{x}_j\} \right]}{\Pr \left[ X_i^{A^N} = 1 \mid B^N = b^N, \bigcap_{j \in \mathcal{F}_{(i)}^c} \{X_j^{A^N} = \hat{x}_j\} \right]}. \quad (114)$$

However, since in (114) the qubits measured with respect to the amplitude basis at the inner layer are missing in the conditioning, it is not straightforward how to compute this likelihood ratio efficiently.<sup>7</sup> Using Bayes' theorem we can write

$$\begin{aligned} & \Pr \left[ X_i^{A^N} = 0 \mid B^N = b^N, \bigcap_{j \in \mathcal{F}_{(i)}^c} \{X_j^{A^N} = \hat{x}_j\} \right] \\ &= \sum_{\hat{x}_k \in \{0,1\} \forall k \in \mathcal{F}_{(i)}} \Pr \left[ X_i^{A^N} = 0, \bigcap_{k \in \mathcal{F}_{(i)}} \{X_k^{A^N} = \hat{x}_k\} \mid \right. \\ & \quad \left. B^N = b^N, \bigcap_{j \in \mathcal{F}_{(i)}^c} \{X_j^{A^N} = \hat{x}_j\} \right] \quad (115) \\ &= \left( \frac{1}{2} \right)^{|\mathcal{F}_{(i)}|} \sum_{\hat{x}_k \in \{0,1\} \forall k \in \mathcal{F}_{(i)}} \Pr \left[ X_i^{A^N} = 0, \bigcap_{k \in \mathcal{F}_{(i)}} \{X_k^{A^N} = \hat{x}_k\} \mid \right. \\ & \quad \left. B^N = b^N, \bigcap_{j \in \mathcal{F}_{(i)}^c} \{X_j^{A^N} = \hat{x}_j\} \right] / \\ & \quad \Pr \left[ \bigcap_{k \in \mathcal{F}_{(i)}} \{X_k^{A^N} = \hat{x}_k\} \right] \quad (116) \\ &= \left( \frac{1}{2} \right)^{|\mathcal{F}_{(i)}|} \sum_{\hat{x}_k \in \{0,1\} \forall k \in \mathcal{F}_{(i)}} \Pr \left[ X_i^{A^N} = 0 \mid B^N = b^N, \right. \\ & \quad \left. \bigcap_{j=1}^{i-1} \{X_j^{A^N} = \hat{x}_j\} \right], \quad (117) \end{aligned}$$

where (116) uses that the random variables  $X_j^{A^N}$  for  $j \in \mathcal{F}_j$  are independent uniformly distributed, since they have been measured in the complementary amplitude basis. The elements

<sup>7</sup>Arkan's recursive formula cannot be applied directly.



of the final sum (117) can be computed efficiently using Arıkan’s recursive formula [9]. Since the elements of the sum in (117) are bounded, we can sample the sum obtaining a (reasonably) good approximation of the true value which can be done efficiently.

We are now ready to state the main result of this section—that the distillation scheme introduced in Section IV for Pauli or erasure channels is efficiently decodable when using quantum polar codes.

**Theorem 16.** *The decoding of the distillation scheme can be done with  $O(N \log N)$  steps.*

*Proof:* According to [9], the  $M$  amplitude recovery blocks together have a complexity of  $O(ML \log L)$ . Using Lemma 13 it follows that the phase correction can be done in  $O(N \log N)$ . Since  $N = ML$ , this proves the assertion. ■

When using quantum polar codes for Pauli or erasure channels we can derive explicit expressions for  $\epsilon_1$  and  $\epsilon_2$  and hence make a precise statement about the reliability of the distillation scheme.

**Corollary 17.** *The reliability of the distillation scheme is as given in Proposition 8 with  $\epsilon_1 = O(2^{-L^\beta})$  and  $\epsilon_2 = O(L2^{-M^{\beta'}})$  for any  $\beta, \beta' < \frac{1}{2}$ .*

*Proof:* Let  $p_{\text{err}}(\mathcal{D}_A)$  denote the error probability of the decoding operation  $\mathcal{D}_A$  and  $p_{\text{err}}(\mathcal{D}_P)$  the error probability of the decoding process  $\mathcal{D}_P$ . The error probability for all decoding operations with respect to the amplitude basis is denoted by  $p_{\text{err}}(\mathcal{D}_{A^M})$ . According to Proposition 8, the trace distance between the scheme’s output and a maximally entangled state of appropriate dimension is less than  $\sqrt{2p_{\text{err}}(\mathcal{D}_{A^M})} + \sqrt{2p_{\text{err}}(\mathcal{D}_P)}$ . Using the union bound we obtain  $p_{\text{err}}(\mathcal{D}_{A^M}) \leq Mp_{\text{err}}(\mathcal{D}_A)$ . Since we use the standard polar decoder [9] for the amplitude error-correction, we have

$$p_{\text{err}}(\mathcal{D}_A) = O\left(2^{-L^\beta}\right) \quad \text{for } \beta < \frac{1}{2}. \quad (118)$$

Furthermore, according to Lemma 15,

$$p_{\text{err}}(\mathcal{D}_P) = O\left(L2^{-M^{\beta'}}\right) \quad \text{for any } \beta' < \frac{1}{2}, \quad (119)$$

which proves the assertion. ■

## VI. CHANNEL CODING

Bennett *et al.* [19] showed that any entanglement distillation scheme can be turned into a channel coding scheme, which however is not known to be computationally efficient—even if the entanglement distillation protocol we started with is efficient. In this section, we show how to modify the entanglement distillation scheme introduced in Section IV such that it can be used for efficient channel coding. The resulting coding scheme is depicted schematically in Figure 8. Before applying the actual encoding transformation, the outer encoder adds redundancy in form of random qubits which are sent to the decoder. As explained in the entanglement distillation scheme in the previous section, we know that after the inner layer the state is perfectly known with respect to the amplitude basis. Therefore, we can choose the additional qubits at random in the complementary phase basis.

The inner encoder also adds redundancy. The additional qubits are generated as explained in [42, Section II] and sent to the decoder, before applying the actual encoding transform. The encoded data is then transmitted over  $N$  identical channels  $\mathcal{N}$ . The decoding is identical to Bob’s task in the entanglement distillation scenario, explained in Section IV.

For the code construction, the set of frozen qubits (the indices which determine at which position the redundant qubits are added) at the outer and inner layer have to be determined. This can be done efficiently for the inner layer as explained in Section V. The existence of an efficient algorithm for the outer layer remains an open question.

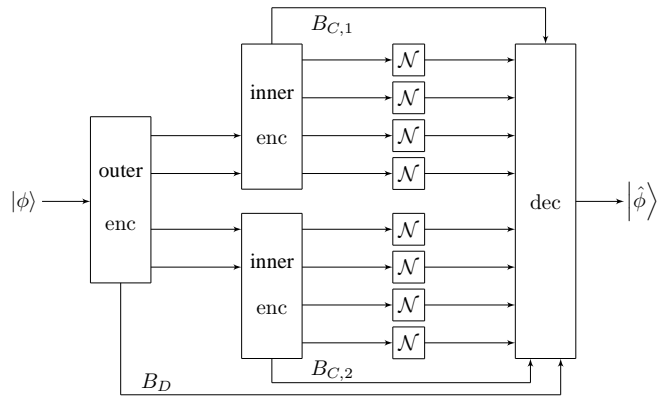


Fig. 8. The channel coding view of the scheme for  $L = 4$  and  $M = 2$ . The outer encoder adds randomly generated qubits in the phase basis, the identities of which are forwarded to the decoder (over a classical channel), before applying the actual encoding transform. At the inner layer the encoder mimics the extra amplitude basis qubits as explained in Section VI-B (see also [42, Section II]) and sends them to the decoder as well. Decoding is the same as for the entanglement distillation setup explained in Section IV.

### A. Encoding

We show that for Pauli and erasure channels together with the use of quantum polar codes an efficient encoder and decoder can be obtained.

■ **Corollary 18.** *For Pauli channels and the use of polar codes, there exists an encoder for the scheme described above that has  $O(N \log N)$  complexity.*

*Proof:* The encoder is equivalent to the one introduced in Section V. Note that at the outer encoder the frozen qubits are chosen at random as justified above. The inner encoder chooses its frozen qubits as explained in [42, Section II]. ■

In order to explain the decoding strategy we first present an efficient concatenated classical channel coding scheme based on polar codes

### B. Efficient Concatenated Classical Channel Coding

In order to explain the decoding strategy we first present an efficient concatenated classical channel coding scheme based on polar codes. The scheme achieves the capacity and has been introduced in [42]. It will serve as a building block to prove the reliability and efficiency of the quantum channel coding



scheme introduced in Section VI when using quantum polar codes for Pauli and erasure channels.

We consider a discrete memoryless channel  $W : \mathcal{X} \rightarrow \mathcal{Y}$  with a binary input alphabet  $\mathcal{X} = \{0, 1\}$  and an arbitrary output alphabet  $\mathcal{Y}$ . The idea for the efficient classical channel coding scheme is to run the encoder of the source coding scheme—introduced in Section V—in reverse and to use the same decoder. One main difficulty that occurs is that we need to simulate the frozen bits at the inner and outer layer. As we show next, by cleverly choosing the frozen bits at the inner layer (we call this in the following *shaping*), we can approximate the source coding case, i.e., the sequence of  $N$  i.i.d. correlated random variables  $(X, Y)^N$  arbitrarily well for large  $N$  (cf. Lemma 20). This enables us to prove that the efficient decoder introduced in Section V for source coding, can also be used for reliable channel coding. More details about this approach can be found in [42].

*Shaping.*— The idea of how to approximate  $(X, Y)^N$  and hence how to choose the frozen bits at the inner encoding block is to run an extractor for the distribution  $(X, Y)^N$  we want to approximate in reverse.<sup>8</sup> A priori it is not clear this process can be done efficiently, however luckily we show that this is the case for extractors based on the source polarization phenomenon.

A  $K$ -bit polarization extractor  $E_{L,K}$  for  $X^L$  simply outputs the  $K$  bits of  $U^L = X^L G_L$  for which  $H(U_i|U^{i-1})$  are greatest. We denote this (ordered) set of indices by  $\mathcal{E}_K$  and the output of the extractor by  $U^L[\mathcal{E}_K]$ .

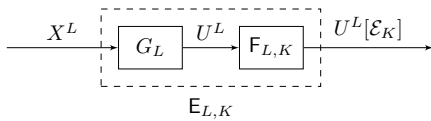


Fig. 9. Polarization-based randomness extractor  $E_{L,K}$ . The input  $X^L$  is first transformed to  $U^L$  via the polarization transformation  $G_L$ , and subsequently  $F_{L,K}$  filters out the  $K$  bits of  $U^L$  for which  $H(U_i|U^{i-1})$  are greatest.

The aim of randomness extraction is to output  $K$  approximately uniform bits, where the approximation is quantified using the variational distance. Recall that for distributions  $P$  and  $Q$  over the same alphabet  $\mathcal{X}$ , the variational distance is defined by  $\delta(P, Q) := \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$ . We will often abuse notation slightly and write a random variable instead of its distribution in  $\delta$ .

Using  $\mathcal{E}_K$  we define the shaper for  $X^L$  as follows

**Definition 1.** For  $\mathcal{U} = \{0, 1\}$ , the shaper  $S_{K,L}$  for  $X^L$  is the map  $S_{K,L} : \mathcal{U}^K \rightarrow \mathcal{X}^L$  taking input  $U^K$  to  $\hat{X}^L = \hat{U}^L G_L$ , with

$$\hat{U}_i = \begin{cases} U_{\text{pos}_{\mathcal{E}_K}(i)} & i \in \mathcal{E}_K \\ Z_i & \text{else} \end{cases}, \quad (120)$$

where  $\text{pos}_{\mathcal{A}}(a)$  denotes the position of the entry  $a$  in  $\mathcal{A}$ . Here  $Z_i$  is a random variable generated from the distribution of  $U_i|U^{i-1}$ , using  $U^L = X^L G_L$ .

<sup>8</sup>We only need to show how to approximate  $(X, Y)^L$  since taking  $M$  identical blocks then leads to an approximation of  $(X, Y)^N$ .

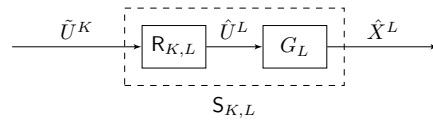


Fig. 10. Generation of an approximation to  $X^L$  from a uniform input  $\tilde{U}^K$  using the shaper  $S_{K,L}$ .  $\hat{U}^L$  is first constructed by  $R_{K,L}$  from the uniform input according to (120). Applying  $G_L$  gives  $\hat{X}^L$ , which has nearly the same distribution as  $X^L$ .

Using the shaper with uniform input  $\tilde{U}^K$  (a  $K$ -bit vector whose entries are i.i.d. Bernoulli  $(\frac{1}{2})$ ) generates an approximation  $\hat{X}^L := S_{K,L}(\tilde{U}^K)$  to  $X^L$  (see also [47, Lemma 11]).

**Lemma 19.** For  $\epsilon \geq 0$  and  $K$  such that  $H(U_i|U^{i-1}) \geq 1 - \epsilon$  for all  $i \in \mathcal{E}_K$ ,

$$\delta(\hat{X}^L, X^L) \leq K \sqrt{\frac{\ln 2}{2}} \epsilon. \quad (121)$$

*Proof:* Let  $\hat{U}^L$  be the  $L$ -bit string obtained when using the shaper with uniform input  $\tilde{U}^K$  (cf. Equation 120). We have  $X^L = U^L G_L$  and  $\hat{X}^L = \hat{U}^L G_L$  and, hence,

$$\delta(\hat{X}^L, X^L) = \delta(\hat{U}^L, U^L). \quad (122)$$

We will bound the distance on the right hand side. For this, we introduce a family of intermediate distributions  $P_{U_1 \dots U_i \hat{U}_{i+1} \dots \hat{U}_L}^{(i)}$ , for  $i = 0, \dots, L$ , defined by

$$P_{U_1 \dots U_i \hat{U}_{i+1} \dots \hat{U}_L}^{(i)} := P_{U_1 \dots U_i} P_{\hat{U}_{i+1} \dots \hat{U}_L | \hat{U}_1 \dots \hat{U}_i}, \quad (123)$$

so that  $P_{\hat{U}_1 \dots \hat{U}_L}^{(0)} = P_{\hat{U}_1 \dots \hat{U}_L}$  and  $P_{U_1 \dots U_L}^{(L)} = P_{U_1 \dots U_L}$ . By the triangle inequality,

$$\begin{aligned} \delta(\hat{U}^L, U^L) &\leq \sum_{i=1}^L \delta\left(P_{U_1 \dots U_{i-1} \hat{U}_i \dots \hat{U}_L}^{(i-1)}, P_{U_1 \dots U_i \hat{U}_{i+1} \dots \hat{U}_L}^{(i)}\right) \quad (124) \\ &\leq \sum_{i=1}^L \delta\left(P_{U_1 \dots U_{i-1} \hat{U}_i}^{(i-1)}, P_{U_1 \dots U_{i-1} U_i}^{(i)}\right), \quad (125) \end{aligned}$$

where the last line follows from the fact that the variational distance is non-increasing under stochastic maps [48] (we apply this to the map that generates  $\hat{U}_{i+1} \dots \hat{U}_L$  according to the distribution  $P_{\hat{U}_{i+1} \dots \hat{U}_L | \hat{U}_1 \dots \hat{U}_i}$ ). Each term of the sum can be written as  $\delta(P_{U^{i-1}} P_{\hat{U}_i | \hat{U}^{i-1}}, P_{U^{i-1}} P_{U_i | U^{i-1}})$  or, equivalently,  $E_{U^{i-1}} \left[ \delta(P_{\hat{U}_i | \hat{U}^{i-1}}, P_{U_i | U^{i-1}}) \right]$ . To bound this, we use Pinsker's inequality [49, p.58] as well as the concavity of the square root,

$$\begin{aligned} E_{U^{i-1}} \left[ \delta\left(P_{\hat{U}_i | \hat{U}^{i-1}}, P_{U_i | U^{i-1}}\right) \right] &\leq E_{U^{i-1}} \left[ \sqrt{\frac{\ln 2}{2} D(P_{U_i | U^{i-1}} \| P_{\hat{U}_i | \hat{U}^{i-1}})} \right] \quad (126) \\ &\leq \sqrt{\frac{\ln 2}{2}} E_{U^{i-1}} \left[ \sqrt{D(P_{U_i | U^{i-1}} \| P_{\hat{U}_i | \hat{U}^{i-1}})} \right]. \quad (127) \end{aligned}$$

By construction, the conditional distribution of  $\hat{U}_i$  for all  $i \in \mathcal{E}_K$  is the uniform distribution, so that

$$E_{U^{i-1}} \left[ D(P_{U_i | U^{i-1}} \| P_{\hat{U}_i | \hat{U}^{i-1}}) \right] = 1 - H(U_i | U^{i-1}) \quad (128)$$

$$\leq \epsilon. \quad (129)$$

Furthermore, for all  $i \notin \mathcal{E}_K$ , the conditional distribution of  $\hat{U}_i$  equals  $P_{U_i|U^{i-1}}$ , so that the corresponding term in the sum (125) vanishes. The sum can thus be rewritten as

$$\delta(\hat{U}^L, U^L) \leq \sum_{i \in \mathcal{E}_K} \sqrt{\frac{\ln 2}{2}} \epsilon, \quad (130)$$

from which the assertion follows.  $\blacksquare$

Concatenating the shaper with the channel gives the super-channel  $W'_{K,L} := W^L \circ S_{K,L}$ . Monotonicity of the variational distance gives the following lemma, which is the basis of our coding scheme. Letting  $\hat{Y}^L := W^L(\hat{X}^L)$  and  $Y^L = W^L(X^L)$ , we have

**Lemma 20.** *For  $\epsilon \geq 0$  and  $K$  such that  $H(U_i|U^{i-1}) \geq 1 - \epsilon$  for all  $i \in \mathcal{E}_K$ ,*

$$\delta((\hat{U}^K, \hat{Y}^L), (U^L[\mathcal{E}_K], Y^L)) \leq K \sqrt{\frac{\ln 2}{2}} \epsilon. \quad (131)$$

*Proof:* Let  $\epsilon' = K \sqrt{\frac{\ln 2}{2}} \epsilon$ , then Lemma 19 implies  $\delta((\hat{X}^L, \hat{Y}^L), (X^L, Y^L)) \leq \epsilon'$  by the monotonicity of the variational distance under stochastic maps. Applying  $G_L$  to  $X^L$  or  $\hat{X}^L$  and marginalizing over the elements not in  $\mathcal{E}_K$  is also a stochastic map, so  $\delta((\hat{U}^L[\mathcal{E}_K], \hat{Y}^L), (U^L[\mathcal{E}_K], Y^L)) \leq \epsilon'$ . Observing that  $\hat{U}^L[\mathcal{E}_K] = \hat{U}^K$  completes the proof.  $\blacksquare$

*Efficient Classical Encoding and Decoding.*— The encoding for channel coding is the reverse operation of the encoding in the source polarization setup explained in Section V and Figure 6. The frozen bits which are sent to the decoder in the source polarization scenario are simulated by the encoder in this setup. The frozen bits at the outer layer are generated as explained in [33, Section IV]. The frozen bits at the inner layer are chosen according to our shaper construction explained above. Note that the frozen bits at the inner and outer layer are forwarded to the decoder.

For the decoding we use the source coding decoder introduced in Section V. It is therefore clear that it is efficient, however it remains to be shown that it is reliable for channel coding which is done next.

**Corollary 21.** *The encoder and decoder explained above have  $O(N \log N)$  complexity.*

*Proof:* This corollary is an immediate consequence of Lemma 12 and Lemma 13.  $\blacksquare$

*Reliability.*— To analyze the reliability of the decoder introduced above we start with a general lemma on the reliability of using the “wrong” compressor / decompressor pair in the problem of source coding.<sup>9</sup>

**Lemma 22.** *Let  $X$  and  $X'$  be arbitrary random variables such that  $\delta(X', X) \leq \eta$  and let  $W$  denote an arbitrary stochastic map. If  $C$  and  $D$  are a compressor / decompressor pair for  $(X, W(X))$ , such that  $\Pr[\hat{X} \neq X] \leq \eta'$  where*

<sup>9</sup>Recall that we are using the decoder built for source coding, i.e., for pairs of random variables  $(X, Y)^N$ , but we actually have only a (good) approximation of those.

$\hat{X} = D(W(X), C(X))$ , then, for  $\hat{X}' = D(W(X'), C(X'))$ ,

$$\Pr[\hat{X}' \neq X'] \leq \eta + \eta'. \quad (132)$$

*Proof:* Note that the pairs  $(X, \hat{X})$  and  $(X', \hat{X}')$  are obtained from  $X$  and  $X'$  by applying the stochastic map that takes  $x$  to  $(x, D(W(x), C(x)))$ . Because the variational distance is non-increasing under such maps, we have

$$\delta((X, \hat{X}), (X', \hat{X}')) \leq \delta(X, X') \leq \eta. \quad (133)$$

Furthermore, defining  $(X, X)$  to be the random variable  $(X, \bar{X})$  with distribution  $P_{X\bar{X}} = P_X \delta_{X\bar{X}}$ , we have

$$\delta((X, X), (X, \hat{X})) = \Pr[\hat{X} \neq X] \leq \eta'. \quad (134)$$

Hence, applying the triangle inequality, we obtain

$$\delta((X, X), (X', \hat{X}')) \leq \eta + \eta'. \quad (135)$$

Now note that the variational distance can also be written as

$$\delta(A, A') = \sum_{a: P_A(a) \leq P_{A'}(a)} P_{A'}(a) - P_A(a). \quad (136)$$

Applied to  $A = (X, X)$  and  $A' = (X', \hat{X}')$ , and using that  $P_{X\bar{X}}(x, \hat{x}) = 0$  for  $x \neq \hat{x}$ , we immediately obtain

$$\delta((X, X), (X', \hat{X}')) \geq \sum_{x \neq \hat{x}} P_{X', \hat{X}'}(x, \hat{x}), \quad (137)$$

which implies that  $\Pr[\hat{X}' \neq X'] \leq \eta + \eta'$ .  $\blacksquare$

Using Lemmas 14 and 22 the statement of reliability follows easily.

**Lemma 23.** *The error probability using the source coding decoder in the classical setup of channel coding as explained above is  $P_{\text{err}} = O(L 2^{-M^\beta} + L 2^{-\frac{1}{2}L^{\beta'}})$  for any  $\beta, \beta' < \frac{1}{2}$ .*

*Proof:* For the polar source coding scheme, note that  $\Pr[\mathcal{A}_i \cap \mathcal{B}_{i-1}^c] + x \in O(2^{-M^\beta})$ , where  $x$  is the probability that  $\hat{V}_i \neq V_i$  given that a mistake previously occurred, but where we still give the correct  $V^{i-1}$  to the decompressor. We can therefore upper bound  $r$  in Lemma 14 by  $O(2^{-M^\beta})$  [30]. Thus, the probability of incorrectly decoding any of the  $K$   $V_i$  is  $O(L 2^{-M^\beta})$ ; this is  $\eta'$  in Lemma 22. Lemma 20 and the properties of  $K$  give  $\eta = O(L 2^{-\frac{1}{2}L^{\beta'}})$  for  $\beta' < \frac{1}{2}$ , establishing the theorem.  $\blacksquare$

### C. Decoding

The decoding is equivalent to Bob’s task in the entanglement distillation scheme. Note however that it has a slightly worse reliability since the coding scheme approximates the entanglement distillation setup by generating the frozen qubits at the inner and outer layer as explained above. Recall that the inner and outer encoder forward the values of the frozen qubits to the decoder which is necessary to decode reliably and efficiently. An immediate corollary of Theorem 16 states that

**Corollary 24.** *For Pauli channels and the use of polar codes, the scheme introduced above has a  $O(N \log N)$  complexity decoder.*

It remains to be shown that the efficient encoding and decoding introduced above is reliable.

**Theorem 25.** *The trace distance between the state produced by the decoder and the ideal, maximally entangled state is less than  $\sqrt{2\epsilon_2} + \sqrt{2M\epsilon_1}$  where  $\epsilon_1 = O(2^{-L^\beta})$ ,  $\epsilon_2 = O(L2^{-M^{\beta'}} + L2^{-\frac{1}{2}L^{\beta''}})$  for  $\beta, \beta', \beta'' < \frac{1}{2}$ .*

*Proof:* Similar as in the proof of Corollary 17 let  $p_{\text{err}}(\mathcal{D}_A)$  denote the error probability of the amplitude-decoding operation  $\mathcal{D}_A$  and  $p_{\text{err}}(\mathcal{D}_P)$  the error probability of the phase-decoding process  $\mathcal{D}_P$ . The error probability for all decoding operations with respect to the amplitude basis is denoted by  $p_{\text{err}}(\mathcal{D}_{A^M})$ . According to Proposition 8, the trace distance between the scheme's output and a maximally entangled state of appropriate dimension is less than  $\sqrt{2p_{\text{err}}(\mathcal{D}_{A^M})} + \sqrt{2p_{\text{err}}(\mathcal{D}_P)}$ . Using the union bound we obtain  $p_{\text{err}}(\mathcal{D}_{A^M}) \leq Mp_{\text{err}}(\mathcal{D}_A)$ . Since we use the standard polar decoder [9] for the amplitude error-correction, we have

$$p_{\text{err}}(\mathcal{D}_A) = O\left(2^{-L^\beta}\right) \quad \text{for } \beta < \frac{1}{2}. \quad (138)$$

Furthermore, according to Lemma 23,

$$p_{\text{err}}(\mathcal{D}_P) = O(L2^{-M^{\beta'}} + L2^{-\frac{1}{2}L^{\beta''}}) \quad (139)$$

for any  $\beta', \beta'' < \frac{1}{2}$ , which proves the assertion. ■

## VII. ACHIEVING RATES BEYOND THE COHERENT INFORMATION?

In Corollary 10 it is shown that the scheme achieves the coherent information. However if (93) were not tight the scheme could achieve a higher rate. We formulate a series of related open questions addressing this point and provide possible approaches to answer them.

**Open Question 1.** Is it possible that  $R > I(A)B_\psi$  for  $R > 0$ ?

There are several ways to phrase the question above differently. Using (92), an equivalent formulation of Open Question 1 is the following:

**Open Question 1'.** Is it possible that  $\lim_{L \rightarrow \infty} \frac{1}{L} H\left(X^{\bar{A}^c} | B^L C^L X^{\bar{A}}\right)_{\Psi'_3} > 0$  for  $R > 0$ ?

The following Proposition leads to a different, particularly clean reformulation of Open Questions 1 and 1'.

**Proposition 26.** *The rate of the scheme introduced above can be written as*

$$R = -H(A|B)_\psi + H(Z^A|B)_\psi - \frac{1}{L} H\left(Z^{\bar{A}^c} | E^L\right)_{\Psi'_3}. \quad (140)$$

*Proof:* Recall that a possible rate expression is given in Theorem 9. Using the chain rule we can write

$$\begin{aligned} & -H(\bar{A} | B^L C^L)_{\Psi'_3} \\ &= -H(A^L | B^L C^L)_{\Psi'_3} + H(\bar{A}^c | B^L C^L \bar{A})_{\Psi'_3} \end{aligned} \quad (141)$$

$$= -LH(A|BC)_{\psi'} + H\left(X^{\bar{A}^c} | B^L C^L \bar{A}\right)_{\Psi'_3} - |\bar{A}^c| \quad (142)$$

$$= -LH(A|BC)_{\psi'} - H\left(Z^{\bar{A}^c} | E^L\right)_{\Psi'_3} \quad (143)$$

$$= -LH(ABC)_{\psi'} - LH(BC)_{\psi'} - H\left(Z^{\bar{A}^c} | E^L\right)_{\Psi'_3} \quad (144)$$

$$= -LH(AC|B)_{\psi'} + LH(Z^A|B)_\psi - H\left(Z^{\bar{A}^c} | E^L\right)_{\Psi'_3} \quad (145)$$

$$= -LH(A|B)_\psi + LH(Z^A|B)_\psi - H\left(Z^{\bar{A}^c} | E^L\right)_{\Psi'_3}. \quad (146)$$

Equality (142) uses Lemma 1 and that  $H(Z^{\bar{A}^c} | B^L C^L \bar{A}) = 0$ . Equation (143) uses Lemma 2. The remaining steps using the chain rule and the form of  $|\psi\rangle$  and  $|\psi'\rangle$ . ■

Note that from the polarization phenomenon (cf. Theorem 3) we know that  $|\bar{A}^c| = LH(Z^A|B)_\psi - o(L)$  and hence we immediately see that we can bound the rate term as  $R \geq I(A)B_\psi$ . Proposition 26 implies that we have an equivalent formulation of Open Question 1, which might be easier to answer as it is a purely classical problem.<sup>10</sup>

**Open Question 1''.** Is it possible that  $\lim_{L \rightarrow \infty} \frac{1}{L} H(Z^{\bar{A}^c} | E^L)_{\Psi'_3} < \lim_{L \rightarrow \infty} \frac{1}{L} |\bar{A}^c|$  for  $R > 0$ ?

Equation (7) states that for *less noisy* channels the coherent information is optimal. Therefore we must be able to show that  $R = I(A)B_\psi$ . Using the less noisy characterization (cf. Section II-A), we can write  $H(Z^{\bar{A}^c} | E^L)_{\Psi'_3} \geq H(Z^{\bar{A}^c} | B^L)_{\Psi'_3} = LH(Z^A|B)_\psi + o(L)$ , where the last step follows from the polarization phenomenon as stated in Theorem 3. Using (143), we obtain for sufficiently large  $L$ , that  $R \leq I(A)B_\psi$  which together with Corollary 10 proves that for less noisy channels the rate of our scheme is equal to the coherent information.

Recent advances in understanding the polarization phenomenon [51] may be useful to resolve this open question, since Questions 1' and 1'' when applied to quantum polar codes involve a statement about the structure of the polarized sets (see Section III). A possible indication why our scheme could indeed achieve rates beyond the coherent information is the observation that the outer layer can introduce degeneracies for the code at the inner layer. In addition, the states at the inner layer are not product states, which would rule out having a rate beyond the coherent information.<sup>11</sup>

## VIII. ACHIEVING THE QUANTUM CAPACITY?

In this section we state a second open problem which addresses the question of whether it is possible that our scheme achieves the quantum capacity for channels where the coherent information is not optimal. This question is related to the one introduced in Section VII before. Whereas Open Questions 1, 1' and 1'' ask if it is possible to achieve a rate beyond the coherent information for certain quantum channels, Open Question 2 in this section raises the more specific question if for certain quantum channels our scheme achieves the quantum

<sup>10</sup>Note that Open Question 1'' is formulated in a purely classical framework in [50, Section V.A].

<sup>11</sup>Note that for product states the von Neumann entropy is additive.

capacity.<sup>12</sup> As depicted in Figure 8, we can define a super-channel  $\mathcal{N}' : \mathfrak{T}(\mathcal{H}_{\bar{A}}) \rightarrow \mathfrak{T}(\mathcal{H}_{B^{\otimes L}} \otimes \mathcal{H}_{B_C})$  which consists of an inner encoding block and  $L$  basic channels  $\mathcal{N}$ . Then we have

**Proposition 27.** *For sufficiently large  $L$ , the channel  $\mathcal{N}'$  is approximately less noisy, irrespective of  $\mathcal{N}$ .*

*Proof:* Recall that as explained in Section IV-A, we have  $p_{\text{err}}(Z^{\bar{A}}|B^L B_C)_{\Psi_2} \leq \epsilon_1$ . Using Fano's inequality we obtain

$$H\left(Z^{\bar{A}}|B^L B_C\right)_{\Psi_2} \leq H_b(\epsilon_1) + \epsilon_1 \log \dim \bar{A} =: \xi. \quad (147)$$

Recall that we choose the CSS code used in our scheme such that  $\xi \rightarrow 0$  for  $L \rightarrow \infty$ . For example using quantum polar codes, we have  $\epsilon_1 = O\left(2^{-L^\beta}\right)$  for  $\beta < \frac{1}{2}$  and  $\log \dim \bar{A} = O(L)$ . The assertion then follows from [50, Lemma 12]. ■

Proposition 27 and (7) imply that

$$\begin{aligned} & \lim_{L \rightarrow \infty} \frac{1}{L} Q(\mathcal{N}') \\ &= \lim_{L \rightarrow \infty} \frac{1}{L} Q_1(\mathcal{N}') \end{aligned} \quad (148)$$

$$= \lim_{L \rightarrow \infty} \frac{1}{L} \max_{\phi^{\bar{A}\bar{A}'}} I(\bar{A})^{B^L C^L}_{\Psi_3, \mathcal{N}'^{\bar{A}' \rightarrow B^L C^L}(\phi^{\bar{A}\bar{A}'})} \quad (149)$$

$$= R, \quad (150)$$

where the last step follows from Theorem 9. Our scheme hence achieves the capacity of the super-channel. This raises the question of when the super-channel  $\mathcal{N}'$  has the same capacity than the original channel  $\mathcal{N}$ , i.e., how much is lost in the first layer of our scheme.

**Open Question 2.** Under which conditions does  $\frac{1}{L}Q(\mathcal{N}') = Q(\mathcal{N})$  hold?

Equation 7 and Theorem 9 imply that  $\frac{1}{L}Q(\mathcal{N}') = Q(\mathcal{N})$  holds if  $\mathcal{N}$  is less noisy. However, it is not yet known if this is necessary for the condition to hold.

## IX. SECRET KEY DISTILLATION & PRIVATE CHANNEL CODING

### A. Scheme, Reliability, Secrecy and Rate

With minor changes, the above entanglement distillation scheme also works for secret key distillation in a setup where we have two quantum parties Alice and Bob as well as a quantum adversary Eve. Consider the scenario in which Alice and Bob share an additional “shield” system  $S$  [52], [27]. A shield is any system not held by the eavesdropper Eve but nevertheless cannot be used for amplitude IR by Alice and Bob, where the amplitude information is used to create the secret key. One can show that

$$p_{\text{err}}\left(X^{\bar{A}}|B^N C^N B_D S\right) \leq \epsilon_2 \quad \text{and} \quad (151)$$

$$p_{\text{err}}\left(Z^{\bar{A}}|B^N B_C^M\right) \leq M\epsilon_1, \quad (152)$$

characterizes a state where  $Z^{\bar{A}}$  can be used as a secret key.

<sup>12</sup>The existence of a channel fulfilling the condition in Question 2 which is not less noisy would imply a positive answer to Questions 1, 1' and 1''.

Due to the uncertainty principle, the secrecy of the amplitude information from Eve is ensured if Alice and Bob could implement phase IR. Moreover, it is only necessary that phase IR *could* be performed; it is not necessary to actually do it, and thus  $S$  may be used as side information in the procedure, no matter who holds which parts of  $S$  [53].

An observable  $Z^{\bar{A}}$  is approximately secure if the trace distance to the ideal case is small. Therefore we introduce for  $\psi^S = \text{Tr}_A[\psi^{AS}]$ ,

$$p_{\text{secure}}(Z^{\bar{A}}|S) := \frac{1}{2} \left\| \psi^{AS} - \frac{1}{d} \mathbb{1} \otimes \psi^S \right\|_1, \quad (153)$$

where  $\|M\|_1 := \text{Tr}[\sqrt{M^\dagger M}]$ . Note that  $p_{\text{secure}}(Z^{\bar{A}}|S)_\psi \leq \epsilon$  means that the actual  $\epsilon$ -secure  $Z^{\bar{A}}$  can only be distinguished from the ideal, a perfect key, with probability at most  $\epsilon$ .

**Corollary 28.** *For the scheme explained above we have*

$$p_{\text{secure}}\left(X^{\hat{A}}|E^N E_C^M S\right) \leq \sqrt{2\epsilon_2} \quad \text{and} \quad (154)$$

$$p_{\text{secure}}\left(Z^{\hat{A}}|E^N E_C^M E_D\right) \leq \sqrt{2M\epsilon_1}. \quad (155)$$

*Proof:* We show that (151) and (152) are satisfied. The assertion then immediately follows from [54], [53, Theorem 4.1.]. Due to the i.i.d. structure of the scheme and the union bound, it is sufficient to show that  $p_{\text{err}}(Z^{\bar{A}}|B^L B_C) \leq \epsilon_1$  in order to prove (152). This however is a direct consequence of Bob's decoding task for the entanglement distillation setup, as introduced in Section IV-A.

Inequality (151) is equivalent to the error probability for the phase decoding Bob performs (as explained in Section IV-A) with the difference that there is extra side information  $S$  in this case. Fortunately it is straightforward to modify the decoder to the setup with extra side information [30]. ■

The rate for this scenario can be computed analogously as in Section IV-A, which leads to

$$R \geq 1 - H(Z^{\bar{A}}|B)_\psi - H(X^{\bar{A}}|BCS)_\psi \quad (156)$$

$$= H(Z^{\bar{A}}|E)_\psi - H(Z^{\bar{A}}|B)_\psi. \quad (157)$$

The final step uses the exact-uncertainty relation given in [29], which ensures that  $H(Z^{\bar{A}}|E)_\psi + H(X^{\bar{A}}|BCS)_\psi = 1$ . Note that we no longer obtain the coherent information  $-H(A|B)_\psi$ , since  $E$  no longer purifies  $AB$ . The reliability of the secret key distillation scheme is analogous to that of the entanglement distillation scheme (cf. Proposition 8).

### B. Using Quantum Polar Codes for Pauli or Erasure Channels

As mentioned above the secret key distillation scheme is very similar to the entanglement distillation scheme introduced in Section IV. More precisely, Alice's first task, i.e., the  $M$  amplitude information reconciliation blocks are identical as in the entanglement distillation. The phase IR step is slightly different as one has to consider side information  $S$ , which leads to a different set of frozen qubits. Furthermore, Alice does not send the outcomes from measuring  $\hat{A}^c$  to Bob, but simply keeps them secret from Eve. Alice's task thus can be done with  $O(N \log N)$  complexity as proven in Theorems 11 and 16.



Bob's task is also similar to the decoding he performs in the entanglement distillation setup. He first decodes the amplitude  $Z^{A^N}$ , which can be done with a standard classical polar decoder. He next computes the value of  $Z^{\hat{A}}$  using the details of the phase IR code. Hence Bob's decoding operation has  $O(N \log N)$  complexity.

The reliability and secrecy are as given in Proposition 8 and Corollary 28 with  $\epsilon_1 = O(2^{-L^{\beta'}})$  and  $\epsilon_2 = O(L2^{-M^{\beta'}})$  for any  $\beta, \beta' < \frac{1}{2}$ .

The secret key distillation scheme described above also works in a purely classical setup, since the phase IR protocol can be turned into a privacy amplification protocol needing only classical operations as shown in [50]. Our scheme for classical one-way secret key agreement improves practically efficient protocols where the eavesdropper has no prior knowledge [55] or/and degradability assumptions are required [56].

### C. Private Communication

The quantum coding scheme can be used for efficient private channel coding at a high rate (as given in (156) and (157)). As in Section VI, the idea is to run key distillation in reverse, simulating the measurement outputs with appropriately-chosen random inputs. These are then the frozen bits. The frozen bits of the inner and outer encoder can be sent over an insecure channel to Bob, since privacy is ensured by the outer layer whose frozen bits are uncorrelated to the message bits since the corresponding protocol produces entanglement.

In [50], it is shown how to use a similar scheme in a purely classical scenario for private channel coding at the secrecy capacity such that there exists an encoding and decoding operation that have an essentially linear complexity. Our scheme improves previous work on efficient wiretap coding, where either only weak secrecy could be proven [57] or/and degradability assumptions are required [58].

## X. CONCLUSION

We have constructed a protocol that can be used to perform reliable entanglement distillation or quantum communication at a rate equal to (or possibly larger than) the coherent information. Compared to previous work in this area, our scheme does not require any preshared entanglement and achieves the coherent information also for asymmetric channels (where  $\phi^{AA'}$  in (2) is not necessarily a Bell state). When communicating over a Pauli or erasure channel using polar codes, encoding and decoding can be performed with a number of operations essentially linear in the blocklength. We have also shown how the protocol can be modified for efficient, high-rate secret key distillation and private channel coding.

## REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948. [Online]. Available: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>

[2] S. Lloyd, "Capacity of the noisy quantum channel," *Physical Review A*, vol. 55, no. 3, pp. 1613–1622, 1997. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.55.1613>

[3] P. W. Shor, "The quantum channel capacity and coherent information," Presented at the MSRI Workshop on Quantum Computation, San Francisco, 2002. [Online]. Available: <http://www.msri.org/publications/In/msri/2002/quantumcrypto/shor/1/>

[4] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 44–55, 2005. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2004.839515>

[5] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, p. 131, Jul 1997. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.56.131>

[6] H. Barnum, M. A. Nielsen, and B. Schumacher, "Information transmission through a noisy quantum channel," *Physical Review A*, vol. 57, no. 6, p. 4153, June 1998.

[7] H. Barnum, E. Knill, and M. A. Nielsen, "On quantum fidelities and channel capacities," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1317–1329, 2000.

[8] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, "Quantum-channel capacity of very noisy channels," *Physical Review A*, vol. 57, pp. 830–839, Feb 1998. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.57.830>

[9] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jul 2009.

[10] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 1175–1187, Feb 2013.

[11] M. Wilde and S. Guha, "Polar codes for degradable quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4718–4729, 2013.

[12] J. M. Renes, F. Dupuis, and R. Renner, "Efficient polar coding of quantum information," *Physical Review Letters*, vol. 109, p. 050504, Aug 2012. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.109.050504>

[13] J. Renes and M. Wilde, "Polar codes for private and quantum communication over arbitrary channels," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3090–3103, June 2014.

[14] M. M. Wilde, "Sequential decoding of a general classical-quantum channel," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 469, no. 2157, pp. 20130259–20130259, July 2013. [Online]. Available: <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.2013.0259>

[15] T. Brun, I. Devetak, and M.-H. Hsieh, "Correcting quantum errors with entanglement," *Science*, vol. 314, no. 5798, pp. 436–439, 2006. [Online]. Available: <http://dx.doi.org/10.1126/science.1131563>

[16] H.-K. Lo, "Method for decoupling error correction from privacy amplification," *New Journal of Physics*, vol. 5, pp. 36–36, April 2003. [Online]. Available: <http://iopscience.iop.org/1367-2630/5/1/336>

[17] A. J. Ferris and D. Poulin, "Branching MERA codes: a natural extension of polar codes," available at [arXiv:1312.4575](https://arxiv.org/abs/1312.4575).

[18] —, "Tensor networks and quantum error correction," *Phys. Rev. Lett.*, vol. 113, p. 030501, Jul 2014. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.113.030501>

[19] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Physical Review A*, vol. 54, no. 5, pp. 3824–3851, 1996. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.54.3824>

[20] W. F. Stinespring, "Positive functions on  $c^*$ -algebras," *Proceedings of the American Mathematical Society*, vol. 6, no. 211, pp. pp. 211–216, April 1955. [Online]. Available: <http://dx.doi.org/10.1090/S0002-9939-1955-0069403-4>

[21] S. Watanabe, "Private and quantum capacities of more capable and less noisy quantum channels," *Physical Review A*, vol. 85, p. 012326, Jan 2012. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.85.012326>

[22] N. Cai, A. Winter, and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *Problems of Information Transmission*, vol. 40, no. 4, pp. 318–336, 2004. [Online]. Available: <http://dx.doi.org/10.1007/s11122-004-0002-2>

[23] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory*, ser. Colloquia Mathematica Societatis, J. Bolyai, Ed. The Netherlands: North-Holland, 1977, pp. 411–424.

[24] I. Devetak and P. W. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information,"



- Communications in Mathematical Physics*, vol. 256, pp. 287–303, 2005. [Online]. Available: <http://dx.doi.org/10.1007/s00220-005-1317-6>
- [25] G. Smith, “Private classical capacity with a symmetric side channel and its application to quantum cryptography,” *Phys. Rev. A*, vol. 78, p. 022306, Aug 2008. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.78.022306>
- [26] I. Devetak and A. Winter, “Classical data compression with quantum side information,” *Physical Review A*, vol. 68, p. 042301, Oct 2003. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.68.042301>
- [27] J. M. Renes and J.-C. Boileau, “Physical underpinnings of privacy,” *Physical Review A*, vol. 78, p. 032335, Sep 2008. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevA.78.032335>
- [28] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [29] J. M. Renes and J.-C. Boileau, “Conjectured strong complementary information tradeoff,” *Physical Review Letters*, vol. 103, p. 020402, Jul 2009. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.103.020402>
- [30] E. Arıkan, “Source polarization,” *Proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 899–903, Jun 2010.
- [31] E. Şaşıoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” *Proceedings IEEE Information Theory Workshop (ITW)*, pp. 144–148, Oct 2009.
- [32] E. Arıkan and E. Telatar, “On the rate of channel polarization,” *Proceedings IEEE International Symposium on Information Theory (ISIT)*, June 28 – July 3, 2009.
- [33] J. Honda and H. Yamamoto, “Polar coding without alphabet extension for asymmetric channels,” *Proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 2147–2151, Jul 2012.
- [34] E. Abbe, “Randomness and dependencies extraction via polarization,” *Information Theory and Applications Workshop (ITA)*, pp. 1–7, Feb 2011.
- [35] A. G. Sahebi and S. S. Pradhan, “Multilevel polarization of polar codes over arbitrary discrete memoryless channels,” *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1718–1725, 2011.
- [36] A. Uhlmann, “The “transition probability” in the state space of a  $*$ -algebra,” *Reports on Mathematical Physics*, vol. 9, no. 2, pp. 273 – 279, 1976. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/003448776900604>
- [37] R. Jozsa, “Fidelity for mixed quantum states,” *Journal of Modern Optics*, vol. 41, no. 12, pp. 2315–2323, 1994. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/09500349414552171>
- [38] C. A. Fuchs and J. van de Graaf, “Cryptographic distinguishability measures for quantum-mechanical states,” *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1216–1227, May 1999.
- [39] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4337–4347, Sep 2009.
- [40] M. Tomamichel, “A framework for non-asymptotic quantum information theory,” *PhD thesis, ETH Zurich*, 2012, available at [arXiv:1203.2142](https://arxiv.org/abs/1203.2142).
- [41] R. Renner, “Security of quantum key distribution,” *PhD thesis, ETH Zurich*, 2005, available at [arXiv:quant-ph/0512258](https://arxiv.org/abs/quant-ph/0512258).
- [42] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner, “Achieving the capacity of any DMC using only polar codes,” *Proceedings IEEE Information Theory Workshop (ITW)*, pp. 114–118, Sep 2012, extended version available at [arXiv:1205.3756](https://arxiv.org/abs/1205.3756).
- [43] I. Devetak and A. Winter, “Distillation of secret key and entanglement from quantum states,” *Proceedings of the Royal Society A*, vol. 461, no. 2053, pp. 207–235, January 2005. [Online]. Available: <http://www.journals.royalsoc.ac.uk/content/klkmux3dp34he1qm/?p=ee42141c2a1d4c279367945283db73a2&pi=0>
- [44] J. M. Renes, “Approximate quantum error correction via complementary observables,” 2012, available at [arXiv:1003.1150](https://arxiv.org/abs/1003.1150).
- [45] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.
- [46] I. Tal, A. Sharov, and A. Vardy, “Constructing polar codes for non-binary alphabets and MACs,” *Proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 2132–2136, July 2012.
- [47] S. B. Korada and R. L. Urbanke, “Polar codes are optimal for lossy source coding,” *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.
- [48] F. Liese and I. Vajda, *Convex Statistical Distances*. Leipzig, Germany: Teubner, 1987.
- [49] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [50] J. M. Renes, R. Renner, and D. Sutter, “Efficient one-way secret-key agreement and private channel coding via polarization,” in *Advances in Cryptology - ASIACRYPT 2013*, ser. Lecture Notes in Computer Science 8269, K. Sako and P. Sarkar, Eds. Springer Berlin Heidelberg, 2013, pp. 194–213. [Online]. Available: [http://link.springer.com/chapter/10.1007/978-3-642-42033-7\\_11](http://link.springer.com/chapter/10.1007/978-3-642-42033-7_11)
- [51] J. M. Renes, D. Sutter, and S. H. Hassani, “Alignment of polarized sets,” available at [arXiv:1411.7925](https://arxiv.org/abs/1411.7925).
- [52] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, “Secure key from bound entanglement,” *Physical Review Letters*, vol. 94, no. 16, pp. 160502–4, Apr 2005. [Online]. Available: <http://link.aps.org/abstract/PRL/v94/e160502>
- [53] J. M. Renes, “Duality of privacy amplification against quantum adversaries and data compression with quantum side information,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 2010. [Online]. Available: <http://rspa.royalsocietypublishing.org/content/early/2010/12/10/rspa.2010.0445.abstract>
- [54] —, “The physics of quantum information: Complementarity, uncertainty, and entanglement,” *Habilitation thesis, TU Darmstadt*, 2011, available at [arXiv:1212.2379](https://arxiv.org/abs/1212.2379).
- [55] E. Abbe, “Low complexity constructions of secret keys using polar coding,” *Proceedings IEEE Information Theory Workshop (ITW)*, Sep 2012.
- [56] R. A. Chou, M. R. Bloch, and E. Abbe, “Polar coding for secret-key generation,” available at [arXiv:1305.4746](https://arxiv.org/abs/1305.4746).
- [57] H. Mahdavi and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6428–6443, Oct 2011.
- [58] E. Şaşıoğlu and A. Vardy, “A new polar coding scheme for strong security on wiretap channels,” in *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2013, pp. 1117–1121.