1-24-2013

Polar codes for classical-quantum channels

Mark M. Wilde
Université McGill

Saikat Guha
BBN Technologies

Polar codes for classical-quantum channels

Mark M. Wilde and Saikat Guha

***Abstract*—Holevo, Schumacher, and Westmoreland's coding theorem guarantees the existence of codes that are capacity-achieving for the task of sending classical data over a channel with classical inputs and quantum outputs. Although they demonstrated the existence of such codes, their proof does not provide an explicit construction of codes for this task. The aim of the present paper is to fill this gap by constructing near-explicit "polar" codes that are capacity-achieving. The codes exploit the channel polarization phenomenon observed by Arikan for the case of classical channels. Channel polarization is an effect in which one can synthesize a set of channels, by "channel combining" and "channel splitting," in which a fraction of the synthesized channels are perfect for data transmission while the other fraction are completely useless for data transmission, with the good fraction equal to the capacity of the channel. The channel polarization effect then leads to a simple scheme for data transmission: send the information bits through the perfect channels and "frozen" bits through the useless ones. The main technical contributions of the present paper are threefold. First, we leverage several known results from the quantum information literature to demonstrate that the channel polarization effect occurs for channels with classical inputs and quantum outputs. We then construct linear polar codes based on this effect, and the encoding complexity is $O(N \log N)$, where N is the blocklength of the code. We also demonstrate that a quantum successive cancellation decoder works well, in the sense that the word error rate decays exponentially with the blocklength of the code. For this last result, we exploit Sen's recent "non-commutative union bound" that holds for a sequence of projectors applied to a quantum state.**

I. Introduction

Shannon's fundamental contribution was to establish the capacity of a noisy channel as the highest rate at which a sender can reliably transmit data to a receiver [1]. His method of proof exploited the probabilistic method and was thus non-constructive. Ever since Shannon's contribution, researchers have attempted to construct error-correcting codes that can reach the capacity of a given channel. Some of the most successful schemes for error correction are turbo codes and low-density parity-check codes [2], with numerical results demonstrating that these codes perform well for a variety of channels. In spite of the success of these codes, there is no proof that they are capacity achieving for channels other than the erasure channel [3].

Recently, Arikan constructed polar codes and proved that they are capacity achieving for a wide variety of channels [4]. Polar codes exploit the phenomenon of *channel polarization*, in which a simple, recursive encoding synthesizes a set of channels that polarize, in the sense that a fraction of them become perfect for transmission while the other fraction are completely noisy and thus useless for transmission. The fraction of the channels that become perfect for transmission is equal to the capacity of the channel. In addition, the complexity of both the encoding and decoding scales as $O(N \log N)$, where N is the blocklength of the code. Arikan developed polar codes after studying how the techniques of channel combining and channel splitting affect the rate and reliability of a channel [5]. Arikan and others have now extended the methods of polar coding to many different settings, including arbitrary discrete memoryless channels [6], source coding [7], lossy source coding [3], [8], and the multiple access channel with two senders and one receiver [9].

All of the above results are important for determining both the limits on data transmission and methods for achieving these limits on classical channels. The description of a classical channel $p_{Y|X}$ arises from modeling the signaling alphabet, the physical transmission medium, and the receiver measurement. If we are interested in accurately evaluating and reaching the true data-transmission limits of the physical channels, with an unspecified receiver measurement, and whose information carriers require a quantum-mechanical description, then it becomes necessary to invoke the laws of quantum mechanics. Examples of such channels include deep-space optical channels and ultra-low-temperature quantum-noise-limited RF channels. Achieving the classical communication capacity for such (quantum) channels often requires making collective measurements at the receiver, an action for which no classical description or implementation exists. The quantum-mechanical approach to information theory [10], [11] is not merely a formality or technicality—encoding classical information with quantum states and decoding with collective measurements on the channel outputs [12], [13] can dramatically improve data transmission rates, for example if the sender and receiver are operating in a low-power regime for a pure-loss optical channel (which is a practically relevant regime for long haul free-space terrestrial and deep-space optical communication) [14], [15]. Also, encoding with entangled inputs to the channels can increase capacity for certain channels [16], a superadditive effect which simply does not occur for classical channels.

The proof of one of the most important theorems of quantum information theory is due to Holevo [12], Schumacher, and Westmoreland [13] (HSW). They showed that the Holevo information of a quantum channel is an achievable rate for classical communication over it. Their proof of the HSW theorem bears some similarities with Shannon's technique (including the use of random coding), but their main contribution was the construction of a quantum measurement at the receiving end that allows for reliable decoding at the Holevo information rate. Since the proof of the HSW theorem, several researchers have improved the proof's error analysis [17], and others have demonstrated different techniques for

Mark M. Wilde is with the School of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada. Saikat Guha is with the Quantum Information Processing Group, Raytheon BBN Technologies, Cambridge, Massachusetts, USA 02138. (E-mail: mark.wilde@mcgill.ca; sguha@bbn.com)

achieving the Holevo information [18], [19], [20], [21], [22]. Very recently, Giovannetti *et al*. proved that a sequential decoding approach can achieve the Holevo information [23]. The sequential decoding approach has the receiver ask, through a series of dichotomic quantum measurements, whether the output of the channel was the first codeword, the second codeword, etc. (this approach is similar in spirit to a classical "jointly-typical" decoder [24]). As long as the rate of the code is less than the Holevo information, then this sequential decoder will correctly identify the transmitted codeword with asymptotically negligible error probability. Sen recently simplified the error analysis of this sequential decoding approach (rather significantly) by introducing a "non-commutative union bound" in order to bound the error probability of quantum sequential decoding [25].

In spite of the large amount of effort placed on proving that the Holevo information is achievable, there has been relatively little work on devising explicit codes that approach the Holevo information rate.[1] The aim of the present paper is to fill this gap by generalizing the polar coding approach to quantum channels. In doing so, we construct the first explicit class of linear codes that approach the Holevo information rate with asymptotically small error probability.

The main technical contributions of the present paper are as follows:

1) We characterize *rate* with the symmetric Holevo information [27], [12], [10], [11] and *reliability* with the fidelity [28], [29], [10], [11] between channel outputs corresponding to different classical inputs. These parameters generalize the symmetric Shannon capacity and the Bhattacharya parameter [4], respectively, to the quantum case. We demonstrate that the symmetric Holevo information and the fidelity polarize under a recursive channel transformation similar to Arikan's [4], by exploiting Arikan's proof ideas [4] and several tools from the quantum information literature [30], [31], [32], [10], [33], [11].
2) The second contribution of ours is the generalization of Arikan's successive cancellation decoder [4] to the quantum case. We exploit ideas from quantum hypothesis testing [34], [35], [36], [37], [38] in order to construct the quantum successive cancellation decoder, and we use Sen's recent "non-commutative union bound" [25] in order to demonstrate that the decoder performs reliably in the limit of many channel uses, while achieving the symmetric Holevo information rate.

The complexity of the encoding part of our polar coding scheme is $O(N \log N)$ where N is the blocklength of the code (the argument for this follows directly from Arikan's [4]). However, we have not yet been able to show that the complexity of the decoding part is $O(N \log N)$ (as is the case with Arikan's decoder [4]). Determining how to simplify the complexity of the decoding part is the subject of ongoing research. For now, we should regard our contribution in this paper as a more explicit method for achieving the Holevo information rate (as compared to those from prior work [12], [13], [18], [19], [20], [21]).

One might naively think from a casual glance at our paper that Arikan's results [4] directly apply to our quantum scenario here, but this is not the case. If one were to impose single-symbol detection on the outputs of the quantum channels,[2] such a procedure would induce a classical channel from input to output. In this case, Arikan's results do apply in that they can attain the Shannon capacity of this induced classical channel.

However, the Shannon capacity of the best single-symbol detection strategy may be far below the Holevo limit [14], [15]. Attaining the Holevo information rate generally requires the receiver to perform collective measurements (physical detection of the quantum state of the entire codeword that may not be realizable by detecting single symbols one at a time). We should stress that what we are doing in this paper is different from a naive application of Arikan's results. First, our polar coding rule depends on a quantum parameter, the fidelity, rather than the Bhattacharya distance (a classical parameter). The polar coding rule is then different from Arikan's, and we would thus expect a larger fraction of the channels to be "good" channels than if one were to impose a single-symbol measurement and exploit Arikan's polar coding rule with the Bhattacharya distance. Second, the quantum measurements in our quantum successive cancellation decoder are collective measurements performed on all of the channel outputs. Were it not so, then our polar coding scheme would not achieve the Holevo information rate in general.

We organize the rest of the paper as follows. The next section provides an overview of polar coding for classical-quantum channels (channels with classical inputs and quantum outputs). This overview states the main concepts and the important theorems, while saving their proofs for later in the paper. The main concepts include channel combining, channel splitting, channel polarization, rate of polarization, quantum successive cancellation decoding, and polar code performance. Section III gives more detail on how recursive channel combining and splitting lead to transformation of rate and reliability in the direction of polarization. Section IV proves that channel polarization occurs under the transformations given in Section III (the proofs in Section IV are identical to Arikan's [4] because they merely exploit his martingale approach). We prove in Section V that the performance of the polar coding scheme is good, by analyzing the error probability under quantum successive cancellation decoding. We finally conclude in Section VI with a summary and some open questions.

[1] This is likely due to the large amount of effort that the quantum information community has put towards *quantum* error correction [26], which is important for the task of transmitting quantum bits over a noisy quantum channel or for building a fault-tolerant quantum computer. Also, there might be a general belief that classical coding strategies would extend easily for sending classical information over quantum channels, but this is not the case given that collective measurements on channel outputs are required to achieve the Holevo information rate and the classical strategies do not incorporate these collective measurements.

[2] For instance, all known conventional optical receivers are single-symbol detectors. They detect each modulated pulse individually, followed by classical postprocessing.

II. Overview of Results

Our setting involves a classical-quantum channel W with a classical input x and a quantum output ρ_x:

$$W : x \to \rho_x,$$

where $x \in \{0, 1\}$ and ρ_x is a unit trace, positive operator called a *density operator*. We can associate a probability distribution and a classical label with the states ρ_0 and ρ_1 by writing the following *classical-quantum state* [11]:

$$\rho^{XB} \equiv \frac{1}{2}|0\rangle\langle 0|^X \otimes \rho_0^B + \frac{1}{2}|1\rangle\langle 1|^X \otimes \rho_1^B.$$

Two important parameters for characterizing any classical-quantum channel are its rate and reliability.[3] We define the rate in terms of the channel's symmetric Holevo information $I(W)$ where

$$I(W) \equiv I(X; B)_\rho.$$

$I(X; B)_\rho$ is the quantum mutual information of the state ρ^{XB}, defined as

$$I(X; B)_\rho \equiv H(X)_\rho + H(B)_\rho - H(XB)_\rho,$$

and the von Neumann entropy $H(\sigma)$ of any density operator σ is defined as

$$H(\sigma) \equiv -\text{Tr}\{\sigma \log_2 \sigma\}.$$

(Observe that the von Neumann entropy of σ is equal to the Shannon entropy of its eigenvalues.) It is also straightforward to verify that

$$I(W) = H((\rho_0^B + \rho_1^B)/2) - H(\rho_0^B)/2 - H(\rho_1^B)/2.$$

The symmetric Holevo information is non-negative by concavity of von Neumann entropy, and it can never exceed one if the system X is a classical binary system (as is the case for the classical-quantum state ρ^{XB}). Additionally, the symmetric Holevo information is equal to zero if there is no correlation between X and B. It is equal to the capacity of the channel W for transmitting classical bits over it if the input prior distribution is restricted to be uniform [12], [13]. It also generalizes the symmetric capacity [4] to the quantum setting given above.

We define the reliability of the channel W as the fidelity between the states ρ_0 and ρ_1 [28], [29], [10], [11]:

$$F(\rho_0, \rho_1) \equiv \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1^2,$$

where $\|A\|_1$ is the nuclear norm of the operator A:

$$\|A\|_1 = \text{Tr}\Big\{\sqrt{A^\dagger A}\Big\}.$$

Let $F(W)$ denote the reliability of the channel W:

$$F(W) \equiv F(\rho_0, \rho_1).$$

The fidelity is equal to a number between zero and one, and it characterizes how "close" two quantum states are to one another. It is equal to zero if and only if there exists a quantum measurement that can perfectly distinguish the states, and it is equal to one if the states are indistinguishable by any measurement [10], [11]. The fidelity generalizes the Bhattacharya parameter used in the classical setting [4]. Naturally, we would expect the channel W to be perfectly reliable if $F(W) = 0$ and completely unreliable if $F(W) = 1$. The fidelity also serves as a coarse bound on the probability of error in discriminating the states ρ_0 and ρ_1 [37], [39].

[3]We are using the same terminology as Arikan [4].

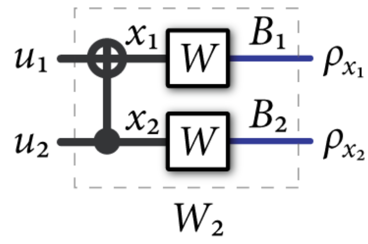

Fig. 1. The channel W_2 synthesized from the first level of recursion. Thick lines denote classical systems while thin lines denote quantum systems (this is our convention for the other figures as well). The depicted gate acting on the channel input is a classical controlled-NOT (CNOT) gate, where the filled-in circle acts on the source bit and the other circle acts on the target bit. Its truth table is $(u_1, u_2) \to (u_1 \oplus u_2, u_2)$.

We would expect the symmetric Holevo information $I(W) \approx 1$ if and only if the channel's fidelity $F(W) \approx 0$ and vice versa: $I(W) \approx 0 \Leftrightarrow F(W) \approx 1$. The following proposition makes this intuition rigorous, and it serves as a generalization of Arikan's first proposition regarding the relationship between rate and reliability. We provide its proof in the appendix.

Proposition 1: For any binary input classical-quantum channel of the above form, the following bounds hold

$$I(W) \geq \log_2\left(\frac{2}{1 + \sqrt{F(W)}}\right), \tag{1}$$

$$I(W) \leq \sqrt{1 - F(W)}. \tag{2}$$

A. Channel Polarization

The channel polarization phenomenon occurs after synthesizing a set of N classical-quantum channels $\{W_N^{(i)} : 1 \leq i \leq N\}$ from N independent copies of the classical-quantum channel W. The effect is known as "polarization" because a fraction of the channels $W_N^{(i)}$ become perfect for data transmission,[4] in the sense that $I(W_N^{(i)}) \approx 1$ for the channels in this fraction, while the channels in the complementary fraction become completely useless in the sense that $I(W_N^{(i)}) \approx 0$ in the limit as N becomes large. Also, the fraction of channels that do not exhibit polarization vanishes as N becomes large. One can induce the polarization effect by means of channel combining and channel splitting.

1) Channel Combining: The channel combining phase takes copies of a classical-quantum channel W and builds from them an N-fold classical-quantum channel W_N in a recursive way, where N is any power of two: $N = 2^n$, $n \geq 0$. The zeroth level of recursion merely sets $W_1 \equiv W$. The first level

[4]One cannot expect to transmit more than one classical bit over a perfect qubit channel due to Holevo's bound [27].

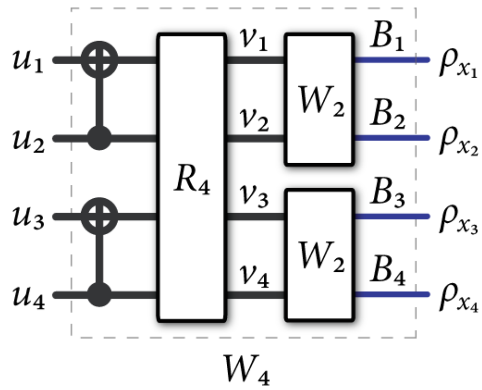

Fig. 2. The second level of recursion in the channel combining phase.

of recursion combines two copies of W_1 and produces the channel W_2, defined as

$$W_2 : u_1 u_2 \to W_2^{B_1 B_2}(u_1, u_2), \tag{3}$$

where

$$W_2^{B_1 B_2}(u_1, u_2) \equiv \rho_{u_1 \oplus u_2}^{B_1} \otimes \rho_{u_2}^{B_2}.$$

Figure 1 depicts this first level of recursion.

The second level of recursion takes two copies of W_2 and produces the channel W_4:

$$W_4 : u_1 u_2 u_3 u_4 \to W_4^{B_1 B_2 B_3 B_4}(u_1, u_2, u_3, u_4), \tag{4}$$

where

$$\begin{aligned} &W_4^{B_1 B_2 B_3 B_4}(u_1, u_2, u_3, u_4) \\ &\qquad \equiv W_2^{B_1 B_2}(u_1 \oplus u_2, u_3 \oplus u_4) \otimes W_2^{B_3 B_4}(u_2, u_4), \end{aligned}$$

so that

$$\begin{aligned} &W_4^{B_1 B_2 B_3 B_4}(u_1, u_2, u_3, u_4) \\ &\qquad = \rho_{u_1 \oplus u_2 \oplus u_3 \oplus u_4}^{B_1} \otimes \rho_{u_3 \oplus u_4}^{B_2} \otimes \rho_{u_2 \oplus u_4}^{B_3} \otimes \rho_{u_4}^{B_4}. \end{aligned}$$

Figure 2 depicts the second level of recursion.

The operation R_4 in Figure 2 is a permutation that takes $(u_1, u_2, u_3, u_4) \to (u_1, u_3, u_2, u_4)$. One can then readily check that the mapping from the row vector u_1^4 to the channel inputs x_1^4 is a linear map given by $x_1^4 = u_1^4 G_4$ with

$$G_4 \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The general recursion at the n^{th} level is to take two copies of $W_{N/2}$ and synthesize a channel W_N from them. The first part is to transform the input sequence u^N according to the following rule for all $i \in \{1, \ldots, N/2\}$:

$$\begin{aligned} s_{2i-1} &= u_{2i-1} \oplus u_{2i}, \\ s_{2i} &= u_{2i}. \end{aligned}$$

The next part of the transformation is a "reverse shuffle" R_N that performs the transformation:

$$\begin{aligned} &(s_1, s_2, s_3, s_4, \ldots, s_{N-1}, s_N) \\ &\qquad \to (s_1, s_3, \ldots, s_{N-1}, s_2, s_4, \ldots, s_N). \end{aligned}$$

The resulting bit sequence is the input to the two copies of $W_{N/2}$.

The overall transformation on the input sequence u^N is a linear transformation given by $x^N = u^N G_N$ where

$$G_N = B_N F^{\otimes n}, \tag{5}$$

where

$$F \equiv \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

and B_N is a permutation matrix known as a "bit-reversal" operation [4].

2) Channel Splitting: The channel splitting phase consists of taking the channels W_N induced by the transformation G_N and defining new channels $W_N^{(i)}$ from them. Let ρ_{u^N} denote the output of the channel W_N when inputting the bit sequence u^N. We define the i^{th} split channel $W_N^{(i)}$ as follows:

$$W_N^{(i)} : u_i \to \rho_{(i),u_i}^{U_1^{i-1} B^N}, \tag{6}$$

where

$$\rho_{(i),u_i}^{U_1^{i-1} B^N} \equiv \sum_{u_1^{i-1}} \frac{1}{2^{i-1}} \left| u_1^{i-1} \right\rangle \left\langle u_1^{i-1} \right|^{U_1^{i-1}} \otimes \overline{\rho}_{u_1^i}^{B^N}, \tag{7}$$

$$\overline{\rho}_{u_1^i}^{B^N} \equiv \sum_{u_{i+1}^N} \frac{1}{2^{N-i}} \rho_{u^N}^{B^N}. \tag{8}$$

We can also write as an alternate notation

$$W_N^{(i)}(u_i) = \rho_{(i),u_i}^{U_1^{i-1} B^N}.$$

These channels have the same interpretation as Arikan's split channels [4]—they are the channels induced by a "genie-aided" quantum successive cancellation decoder, in which the i^{th} decision measurement estimates u_i given that the channel output $\rho_{u^N}^{B^N}$ is available, after observing the previous bits u_1^{i-1} correctly, and if the distribution over u_{i+1}^N is uniform. These split channels arise in our analysis of the error probability for quantum successive cancellation decoding.

3) Channel Polarization: Our channel polarization theorem below is similar to Arikan's Theorem 1 [4], though ours applies for classical-quantum channels with binary inputs and quantum outputs:

Theorem 2 (Channel Polarization): The classical-quantum channels $W_N^{(i)}$ synthesized from the channel $W^{\otimes N}$ polarize, in the sense that the fraction of indices $i \in \{1, \ldots, N\}$ for which $I(W_N^{(i)}) \in (1 - \delta, 1]$ goes to the symmetric Holevo information $I(W)$ and the fraction for which $I(W_N^{(i)}) \in [0, \delta)$ goes to $1 - I(W)$ for any $\delta \in (0, 1)$ as N goes to infinity through powers of two.

The proof of the above theorem is identical to Arikan's proof with a martingale approach [4]. For completeness, we provide a brief proof in Section IV.

4) Rate of Polarization: It is important to characterize the speed with which the polarization phenomenon comes into play for the purpose of proving this paper's polar coding theorem. We exploit the fidelity $F(W^{(i)})$ of the split channels in order to characterize the rate of polarization:

$$F(W_N^{(i)}) \equiv F(\rho_{(i),0}^{U_1^{i-1}B^N}, \rho_{(i),1}^{U_1^{i-1}B^N}). \tag{9}$$

The theorem below exploits the exponential convergence results of Arikan and Telatar [40], which improved upon Arikan's original convergence results [4] (note that we could also use the more general results in Ref. [41]):

Theorem 3 (Rate of Polarization): Given any classical-quantum channel W with $I(W) > 0$, any $R < I(W)$, and any constant $\beta < 1/2$, there exists a sequence of sets $\mathcal{A}_N \subset \{1, \ldots, N\}$ with $|\mathcal{A}_N| \geq NR$ such that

$$\sum_{i \in \mathcal{A}_N} \sqrt{F(W_N^{(i)})} = o(2^{-N^\beta}).$$

Conversely, suppose that $R > 0$ and $\beta > 1/2$. Then for any sequence of sets $\mathcal{A}_N \subset \{1, \ldots, N\}$ with $|\mathcal{A}_N| \geq NR$, the following result holds

$$\max\left\{\sqrt{F(W_N^{(i)})} : i \in \mathcal{A}_N\right\} = \omega(2^{-N^\beta}).$$

The proof of this theorem exploits our results in Section III and Theorem 1 of Ref. [40].

B. Polar Coding

The idea behind polar coding is to exploit the polarization effect for the construction of a capacity-achieving code. The sender should transmit the information bits only through the split channels $W_N^{(i)}$ for which the reliability parameter $F(W_N^{(i)})$ is close to zero. In doing so, the sender and receiver can achieve the symmetric Holevo information $I(W)$ of the channel W.

1) Coset Codes: Polar codes arise from a special class of codes that Arikan calls "G_N-coset codes" [4]. These G_N-coset codes are given by the following mapping from the input sequence u^N to the channel input sequence x^N:

$$x^N = u^N G_N,$$

where G_N is the encoding matrix defined in (5). Suppose that $\mathcal{A}$ is some subset of $\{1, \ldots, N\}$. Then we can write the above transformation as follows:

$$x^N = u_{\mathcal{A}} G_N(\mathcal{A}) \oplus u_{\mathcal{A}^c} G_N(\mathcal{A}^c), \tag{10}$$

where $G_N(\mathcal{A})$ denotes the submatrix of G_N constructed from the rows of G_N with indices in $\mathcal{A}$ and $\oplus$ denotes vector binary addition.

Suppose that we fix the set $\mathcal{A}$ and the bit sequence $u_{\mathcal{A}^c}$. The mapping in (10) then specifies a transformation from the bit sequence $u_{\mathcal{A}}$ to the channel input sequence x^N. This mapping is equivalent to a linear encoding for a code that Arikan calls a G_N-coset code where the sequence $u_{\mathcal{A}^c} G_N(\mathcal{A}^c)$ identifies the coset. We can fully specify a coset code by the parameter vector $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$ where N is the length of the code, $K = |\mathcal{A}|$ is the number of information bits, $\mathcal{A}$ is a set that identifies the indices for the information bits, and $u_{\mathcal{A}^c}$ is the vector of frozen bits. The polar coding rule specifies a way to choose the indices for the information bits based on the channel over which the sender is transmitting data.

2) A Quantum Successive Cancellation Decoder: The specification of the quantum successive cancellation decoder is what mainly distinguishes Arikan's polar codes for classical channels from ours developed here for classical-quantum channels. Let us begin with a G_N-coset code with parameter vector $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$. The sender encodes the information bit vector $u_{\mathcal{A}}$ along with the frozen vector $u_{\mathcal{A}^c}$ according to the transformation in (10). The sender then transmits the encoded sequence x^N through the classical-quantum channel, leading to a state $\rho_{x_1} \otimes \cdots \otimes \rho_{x_N}$, which is equivalent to a state ρ_{u^N} up to the transformation G_N. It is then the goal of the receiver to perform a sequence of quantum measurements on the state ρ_{u^N} in order to determine the bit sequence u^N. We are assuming that the receiver has full knowledge of the frozen vector $u_{\mathcal{A}^c}$ so that he does not make mistakes when decoding these bits.

Corresponding to the split channels $W_N^{(i)}$ in (6) are the following projectors that can attempt to decide whether the input of the i^{th} split channel is zero or one:

$$\begin{aligned}
\Pi_{(i),0}^{U_1^{i-1}B^N} &\equiv \left\{\sqrt{\rho_{(i),0}^{U_1^{i-1}B^N}} - \sqrt{\rho_{(i),1}^{U_1^{i-1}B^N}} \geq 0\right\}, \\
\Pi_{(i),1}^{U_1^{i-1}B^N} &\equiv I - \Pi_{(i),0}^{U_1^{i-1}B^N} \\
&= \left\{\sqrt{\rho_{(i),0}^{U_1^{i-1}B^N}} - \sqrt{\rho_{(i),1}^{U_1^{i-1}B^N}} < 0\right\},
\end{aligned}$$

where $\sqrt{A}$ denotes the square root of a positive operator A, $\{B \geq 0\}$ denotes the projector onto the positive eigenspace of a Hermitian operator B, and $\{B < 0\}$ denotes the projection onto the negative eigenspace of B. After some calculations, we can readily see that

$$\Pi_{(i),0}^{U_1^{i-1}B^N} = \sum_{u_1^{i-1}} |u_1^{i-1}\rangle\langle u_1^{i-1}|^{U_1^{i-1}} \otimes \Pi_{(i),u_1^{i-1}0}^{B^N}, \tag{11}$$

$$\Pi_{(i),1}^{U_1^{i-1}B^N} = \sum_{u_1^{i-1}} |u_1^{i-1}\rangle\langle u_1^{i-1}|^{U_1^{i-1}} \otimes \Pi_{(i),u_1^{i-1}1}^{B^N}, \tag{12}$$

where

$$\begin{aligned}
\Pi_{(i),u_1^{i-1}0}^{B^N} &\equiv \left\{\sqrt{\overline{\rho}_{u_1^{i-1}0}^{B^N}} - \sqrt{\overline{\rho}_{u_1^{i-1}1}^{B^N}} \geq 0\right\}, \\
\Pi_{(i),u_1^{i-1}1}^{B^N} &\equiv \left\{\sqrt{\overline{\rho}_{u_1^{i-1}0}^{B^N}} - \sqrt{\overline{\rho}_{u_1^{i-1}1}^{B^N}} < 0\right\}.
\end{aligned}$$

The above observations lead to a method for a successive cancellation decoder similar to Arikan's [4], with the following decoding rule:

$$\hat{u}_i = \begin{cases} u_i & \text{if } i \in \mathcal{A}^c \\ h(\hat{u}_1^{i-1}) & \text{if } i \in \mathcal{A} \end{cases},$$

where $h(\hat{u}_1^{i-1})$ is the outcome of the following i^{th} measurement on the output of the channel (after $i-1$ measurements have already been performed):

$$\left\{\Pi_{(i),\hat{u}_1^{i-1}0}^{B^N}, \Pi_{(i),\hat{u}_1^{i-1}1}^{B^N}\right\}.$$

We are assuming that the measurement device outputs "0" if the outcome $\Pi^{B^N}_{(i),\hat{u}_1^{i-1}0}$ occurs and it outputs "1" otherwise. (Note that we can set $\Pi^{B^N}_{(i),\hat{u}_1^{i-1}u_i} = I$ if the bit u_i is a frozen bit.) The above sequence of measurements for the whole bit stream u^N corresponds to a positive operator-valued measure (POVM) $\{\Lambda_{u^N}\}$ where

$$\Lambda_{u^N} \equiv \Pi^{B^N}_{(1),u_1} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \\ \cdots \Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1},$$

$$\sum_{u_{\mathcal{A}}} \Lambda_{u^N} = I^{B^N}.$$

The above decoding strategy is suboptimal in two regards. First, the decoder assumes that the future bits are unknown (and random) even if the receiver has full knowledge of the future frozen bits (this suboptimality is similar to the suboptimality of Arikan's decoder [4]). Second, the measurement operators for making a decision are suboptimal as well because we choose them to be projectors onto the positive eigenspace of the difference of the *square roots* of two density operators. The optimal bitwise decision rule is to choose these operators to be the Helstrom-Holevo projector onto the positive eigenspace of the difference of two density operators [34], [35]. Having our quantum successive cancellation decoder operate in these two different suboptimal ways allows for us to analyze its performance easily (though, note that we could just as well have used Helstrom-Holevo measurements to obtain bounds on the error probability). This suboptimality is asymptotically negligible because the symmetric Holevo information is still an achievable rate for data transmission even for the above choice of measurement operators.

3) Polar Code Performance: The probability of error $P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c})$ for code length N, number K of information bits, set $\mathcal{A}$ of information bits, and choice $u_{\mathcal{A}^c}$ for the frozen bits is as follows:

$$P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}) \\ = \frac{1}{2^K} \sum_{u_{\mathcal{A}}} \mathrm{Tr}\{(I - \Lambda_{u^N})\rho_{u^N}\} \\ = 1 - \frac{1}{2^K} \sum_{u_{\mathcal{A}}} \mathrm{Tr}\{\Lambda_{u^N}\rho_{u^N}\} \\ = 1 - \frac{1}{2^K} \sum_{u_{\mathcal{A}}} \mathrm{Tr}\Big\{\Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \\ \cdots \Pi^{B^N}_{(1),u_1}\ \rho_{u^N}\ \Pi^{B^N}_{(1),u_1} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(N),u_1^{N-1}u_N}\Big\},$$

where we are assuming a particular choice of the bits $u_{\mathcal{A}^c}$ in the sequence of projectors $\Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1}$ and the convention mentioned before that $\Pi^{B^N}_{(i),u_1^{i-1}u_i} = I$ if u_i is a frozen bit. We are also assuming that the sender transmits the information sequence $u_{\mathcal{A}}$ with uniform probability 2^{-K}. The probability of error $P_e(N, K, \mathcal{A})$ averaged over all choices of the frozen bits is then

$$P_e(N, K, \mathcal{A}) \\ = \frac{1}{2^{N-K}} \sum_{u_{\mathcal{A}^c}} P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}) \\ = 1 - \frac{1}{2^N} \sum_{u^N} \mathrm{Tr}\Big\{\Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \\ \cdots \Pi^{B^N}_{(1),u_1}\ \rho_{u^N}\ \Pi^{B^N}_{(1),u_1} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(N),u_1^{N-1}u_N}\Big\}. \tag{13}$$

One of the main contributions of this paper is the following proposition regarding the average ensemble performance of polar codes with a quantum successive cancellation decoder:

Proposition 4: For any classical-quantum channel W with binary inputs and quantum outputs and any choice of $(N, K, \mathcal{A})$, the following bound holds

$$P_e(N, K, \mathcal{A}) \leq 2\sqrt{\sum_{i \in \mathcal{A}} \frac{1}{2}\sqrt{F(W_N^{(i)})}}.$$

Thus, there exists a frozen vector $u_{\mathcal{A}^c}$ for each $(N, K, \mathcal{A})$ such that

$$P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}) \leq 2\sqrt{\sum_{i \in \mathcal{A}} \frac{1}{2}\sqrt{F(W_N^{(i)})}}.$$

4) Polar Coding Theorem: Proposition 4 immediately leads to the definition of polar codes for classical-quantum channels:

Definition 5 (Polar Code): A polar code for W is a G_N-coset code with parameters $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$ where the information set $\mathcal{A}$ is such that $|\mathcal{A}| = K$ and

$$F(W_N^{(i)}) \leq F(W_N^{(j)}) \text{ for all } i \in \mathcal{A} \text{ and } j \in \mathcal{A}^c.$$

We can finally state the polar coding theorem for classical-quantum channels. Consider a classical-quantum channel W and a real number $R \geq 0$. Let

$$P_e(N, R) = P_e(N, \lfloor NR \rfloor, \mathcal{A}),$$

with the information bit set chosen according to the polar coding rule in Definition 5. So $P_e(N, \mathcal{A})$ is the block error probability for polar coding over W with blocklength N, rate R, and quantum successive cancellation decoding averaged uniformly over the frozen bits $u_{\mathcal{A}^c}$.

Theorem 6 (Polar Coding Theorem): For any classical-quantum channel W with binary inputs and quantum outputs, a fixed $R < I(W)$, and $\beta < 1/2$, the block error probability $P_e(N, R)$ satisfies the following bound:

$$P_e(N, R) = o(2^{-\frac{1}{2}N^\beta}).$$

The polar coding theorem above follows as a straightforward corollary of Theorem 3 and Proposition 4.

III. Recursive Channel Transformations

This section delves into more detail regarding recursive channel combining and channel splitting. Recall the channel combining in (3-5) and the channel splitting in (6). These allowed for us to take N independent copies of a classical-quantum channel $W^{\otimes N}$ and transform them into the N split

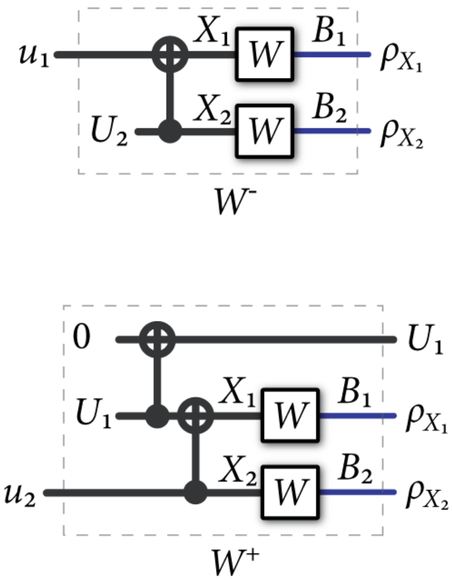

Fig. 3. The channels W^- and W^+ induced from channel combining and channel splitting. The channel W^- with input u_1 is induced by selecting the bit U_2 uniformly at random, passing both u_1 and U_2 through the encoder, and then through the two channel uses. The channel W^+ with input u_2 is induced by selecting U_1 uniformly at random, copying it to another bit (via the classical CNOT gate), sending both U_1 and u_2 through the encoder, and the outputs are the quantum outputs and the bit U_1.

channels $W_N^{(1)}, \ldots, W_N^{(N)}$. We show here how to break the channel transformation into a series of single-step transformations. Much of the discussion here parallels Arikan's discussion in Sections II and III of Ref. [4].

We obtain a pair of channels W^- and W^+ from two independent copies of a channel $W : x \rightarrow \rho_x$ by a single-step transformation if it holds that

$$W^- : u_1 \rightarrow \rho^-_{u_1},$$

where

$$\rho^-_{u_1} \equiv \sum_{u_2} \frac{1}{2} \rho^{B_1}_{u_2 \oplus u_1} \otimes \rho^{B_2}_{u_2}. \tag{14}$$

Also, it should hold that

$$W^+ : u_2 \rightarrow \rho^+_{u_2},$$

where

$$\begin{aligned}\rho^+_{u_2} &\equiv \sum_{u_1} \frac{1}{2} |u_1\rangle\langle u_1|^{U_1} \otimes \rho^{B_1}_{u_2 \oplus u_1} \otimes \rho^{B_2}_{u_2} \qquad (15)\\ &= \left(\sum_{u_1} \frac{1}{2} |u_1\rangle\langle u_1|^{U_1} \otimes \rho^{B_1}_{u_2 \oplus u_1} \right) \otimes \rho^{B_2}_{u_2}.\end{aligned}$$

We use the following notation to denote such a transformation:

$$(W, W) \rightarrow \left(W^-, W^+\right).$$

Additionally, we choose the notation W^- and W^+ so that W^- denotes the worse channel and W^+ denotes the better channel. Figure 3 depicts the channels W^- and W^+.

Thus, from the above, we can write $(W, W) \rightarrow (W_2^{(1)}, W_2^{(2)})$ because, by the definition in (6), we have

$$\begin{aligned}W_2^{(1)}(u_1) &= \sum_{u_2} \frac{1}{2} \rho^{B_1}_{u_1 \oplus u_2} \otimes \rho^{B_2}_{u_2},\\ W_2^{(2)}(u_2) &= \sum_{u_1} \frac{1}{2} |u_1\rangle\langle u_1|^{U_1} \otimes \rho^{B_1}_{u_1 \oplus u_2} \otimes \rho^{B_2}_{u_2}.\end{aligned}$$

We can actually write more generally

$$(W_N^{(i)}, W_N^{(i)}) \rightarrow (W_{2N}^{(2i-1)}, W_{2N}^{(2i)}), \tag{16}$$

which follows as a corollary to

Proposition 7: For any $n \geq 0$, $N = 2^n$, and $1 \leq i \leq N$, it holds that

$$W_{2N}^{(2i-1)}(u_{2i-1}) = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(u_{2i-1} \oplus u_{2i}) \otimes W_N^{(i)}(u_{2i}), \tag{17}$$

$$W_{2N}^{(2i)}(u_{2i}) = W_N^{(i)}(u_{2i-1} \oplus u_{2i}) \otimes W_N^{(i)}(u_{2i}), \tag{18}$$

with $W_N^{(i)}$ defined in (6).

Proof: The proof of the above proposition is similar to the proof of Arikan's Proposition 3 [4]. ■

We can justify the relationship in (16) by observing that (17) and (18) have the same form as (14) and (15) with the following substitutions:

$$\begin{aligned}W &\leftarrow W_N^{(i)},\\ W^+ &\leftarrow W_{2N}^{(2i)},\\ W^- &\leftarrow W_{2N}^{(2i-1)},\\ u_1 &\leftarrow u_{2i-1},\\ u_2 &\leftarrow u_{2i}.\end{aligned}$$

A. Transformation of Rate and Reliability

This section considers how both the rate $I(W_N^{(i)})$ and reliability $F(W_N^{(i)})$ evolve under the general transformation in (16). All proofs of the results in this section appear in the appendix.

Proposition 8: Suppose that $(W, W) \rightarrow (W^-, W^+)$ for some channels satisfying (14-15). Then the following rate conservation and polarizing relations hold

$$I\left(W^-\right) + I\left(W^+\right) = 2I(W), \tag{19}$$

$$I\left(W^-\right) \leq I\left(W^+\right). \tag{20}$$

We can conclude from the above two relations that

$$I\left(W^-\right) \leq I(W) \leq I\left(W^+\right).$$

The following proposition states how the reliability evolves under the channel transformation:

Proposition 9: Suppose $(W, W) \rightarrow (W^-, W^+)$ for some channels satisfying (14-15). Then

$$\sqrt{F(W^+)} = F(W), \tag{21}$$

$$\sqrt{F(W^-)} \leq 2\sqrt{F(W)} - F(W), \tag{22}$$

$$F\left(W^-\right) \geq F(W) \geq F\left(W^+\right). \tag{23}$$

By combining (21) with (22), we observe that the reliability only improves under a single-step transformation:

$$\sqrt{F(W^-)} + \sqrt{F(W^+)} \le 2\sqrt{F(W)}.$$

The above propositions for the single-step transformation lead us to the following proposition in the general case:

Proposition 10: For any classical-quantum channel W, $N = 2^n$, $n \ge 0$, and $1 \le i \le N$, the local transformation in (16) preserves rate and improves reliability in the following sense:

$$I(W_{2N}^{(2i-1)}) + I(W_{2N}^{(2i)}) = 2I(W_N^{(i)}), \tag{24}$$

$$\sqrt{F(W_{2N}^{(2i-1)})} + \sqrt{F(W_{2N}^{(2i)})} \le 2\sqrt{F(W_N^{(i)})}. \tag{25}$$

Channel splitting moves rate and reliability "away from the center":

$$I(W_{2N}^{(2i-1)}) \le I(W_N^{(i)}) \le I(W_{2N}^{(2i)}),$$

$$\sqrt{F(W_{2N}^{(2i-1)})} \ge \sqrt{F(W_N^{(i)})} \ge \sqrt{F(W_{2N}^{(2i)})}.$$

The reliability terms satisfy

$$\sqrt{F(W_{2N}^{(2i)})} = F(W_N^{(i)}), \tag{26}$$

$$\sqrt{F(W_{2N}^{(2i-1)})} \le 2\sqrt{F(W_N^{(i)})} - F(W_N^{(i)}), \tag{27}$$

and the cumulative rate and reliability satisfy

$$\sum_{i=1}^{N} I(W_N^{(i)}) = N\ I(W), \tag{28}$$

$$\sum_{i=1}^{N} \sqrt{F(W_N^{(i)})} \le N\ \sqrt{F(W)}. \tag{29}$$

The above proposition follows directly from Propositions 7, 8, and 9. The relations in (28) and (29) follow from applying (24) and (25) repeatedly.

IV. Channel Polarization

We are now in a position to prove Theorem 2 on channel polarization. The idea behind the proof of this theorem is identical to Arikan's proof of his Theorem 1 in Ref. [4]—with the relationships in Propositions 8 and 9 already established, we can readily exploit the martingale proof technique. Thus, we only provide a brief summary of the proof of Theorem 2 by following the presentation in Chapter 2 of Ref. [3].

Consider the channel $W_N^{(i)}$. Let $b_1 \cdots b_n$ denote an n-bit binary expansion of the channel index i and let $W_{(b_1 \cdots b_n)} = W_N^{(i)}$. Then we can construct the channel $W_{(b_1 \cdots b_k)}$ by combining two copies of $W_{(b_1 \cdots b_{k-1})}$ according to (17) if $b_k = 0$ or by combining two copies of $W_{(b_1 \cdots b_{k-1})}$ according to (18) if $b_k = 1$. We repeatedly construct all the way from b_1 until b_n with the above rule.

Arikan's idea was to represent the channel construction as a random birth process in order to analyze its limiting behavior. In order to do so, we let $\{B_n : n \ge 1\}$ be a sequence of IID uniform Bernoulli random variables, where we define each over a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_0$ denote the trivial σ-field. Also, let $\{\mathcal{F}_n : n \ge 1\}$ denote the σ-fields that the random variables $(B_1, \ldots, B_n)$ generate. We also assume that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. Let $W_0 = W$ and let $\{W_n : n \ge 0\}$ denote a sequence of operator-valued random variables that forms a tree process where W_{n+1} is constructed from two copies of W_n according to (17) if $B_n = 0$ and according to (18) if $B_n = 1$. The output space of the operator-valued random variable W_n is equal to $\{W_{2^n}^{(i)}\}_{i=1}^{2^n}$. We are not really concerned with the channel process $\{W_n : n \ge 0\}$ but more so with the fidelities $\{F(W_N^{(i)})\}$ and Holevo informations $\{I(W_N^{(i)})\}$. Thus, we can simply analyze the limiting behavior of the two random processes $\{F_n : n \ge 0\} \equiv \{\sqrt{F(W_n)} : n \ge 0\}$ and $\{I_n : n \ge 0\} \equiv \{I(W_n) : n \ge 0\}$. By the definitions of the random variables F_n and I_n, it follows that

$$\Pr\{I_n \in (a,b)\} = \frac{1}{2^n}|\{i : I(W_{2^n}^{(i)}) \in (a,b)\}|,$$

$$\Pr\{F_n \in (a,b)\} = \frac{1}{2^n}|\{i : F(W_{2^n}^{(i)}) \in (a,b)\}|.$$

We then have the following lemma.

Lemma 11: The sequence $\{(F_n, \mathcal{F}_n) : n \ge 0\}$ is a bounded super-martingale, and the sequence $\{(I_n, \mathcal{F}_n) : n \ge 0\}$ is a bounded martingale.

Proof: Let $b_1 \cdots b_n$ be a particular realization of the random sequence $B_1 \cdots B_n$. Then the conditional expectation satisfies

$$\begin{aligned} &\mathbb{E}\{I_{n+1} \mid B_1 = b_1, \ldots, B_n = b_n\} \\ &= \frac{1}{2} I\big(W_{(b_1,\ldots,b_n,0)}\big) + \frac{1}{2} I\big(W_{(b_1,\ldots,b_n,1)}\big) \\ &= I\big(W_{(b_1,\ldots,b_n)}\big) \\ &= I_n, \end{aligned}$$

where the second equality follows from the definition of $W_{(b_1,\ldots,b_n,0)}$ and $W_{(b_1,\ldots,b_n,1)}$ and Proposition 10. The proof for $\{F_n\}$ similarly follows from the definitions and Proposition 10. The boundedness condition follows because $0 \le I(W), F(W) \le 1$ for any classical-quantum channel W with binary inputs and quantum outputs. ■

We can now finally prove Theorem 2 regarding channel polarization. Given that $\{I_n\}$ is a bounded martingale and $\{F_n\}$ is a bounded super-martingale, the limits $\lim_{n\to\infty} I_n$ and $\lim_{n\to\infty} F_n$ converge almost surely and in $\mathcal{L}_1$ to the random variables I_∞ and F_∞. The convergence implies that $\mathbb{E}\{|F_{n+1} - F_n|\} \to 0$ as $n \to \infty$. By the definition of the process $\{F_n\}$, it holds that $F_{n+1} = F_n^2$ with probability $\frac{1}{2}$, so that

$$\mathbb{E}\{|F_{n+1} - F_n|\} \ge \frac{1}{2}\mathbb{E}\{|F_n(1 - F_n)|\} \ge 0.$$

It then follows that $\mathbb{E}\{|F_n(1 - F_n)|\} \to 0$ as $n \to \infty$, which in turn implies that $\mathbb{E}\{|F_\infty(1 - F_\infty)|\} = 0$. We conclude that $F_\infty \in \{0, 1\}$ almost surely. Combining this result with Proposition 1 proves that $I_\infty \in \{0, 1\}$ almost surely. Finally, we have that $\Pr\{I_\infty = 1\} = \mathbb{E}\{I_\infty\} = \mathbb{E}\{I_0\} = I(W)$ because I_n is a martingale.

V. Performance of Polar Coding

We can now analyze the performance under the above successive cancellation decoding scheme and provide a proof of Proposition 4. The proof of Theorem 6 readily follows by applying Proposition 4 and Theorem 3.

First recall the following "non-commutative union bound" of Sen (Lemma 3 in Ref. [25]):

$$1-\mathrm{Tr}\{\Pi_N\cdots\Pi_1\rho\Pi_1\cdots\Pi_N\}\leq 2\sqrt{\sum_{i=1}^{N}\mathrm{Tr}\{(I-\Pi_i)\rho\}}, \tag{30}$$

which holds for projectors $\Pi_1,\ldots,\Pi_N$ and a density operator ρ.[5] We begin by applying the above inequality to $P_e(N,K,\mathcal{A})$ (defined in (13)):

$$\begin{aligned}
&P_e(N,K,\mathcal{A})\\
&=\frac{1}{2^N}\sum_{u^N}\Bigg(1-\mathrm{Tr}\Big\{\Pi^{B^N}_{(N),u_1^{N-1}u_N}\cdots\Pi^{B^N}_{(i),u_1^{i-1}u_i}\cdots\\
&\qquad\cdots\Pi^{B^N}_{(1),u_1}\ \rho_{u^N}\ \Pi^{B^N}_{(1),u_1}\cdots\Pi^{B^N}_{(i),u_1^{i-1}u_i}\cdots\Pi^{B^N}_{(N),u_1^{N-1}u_N}\Big\}\Bigg)\\
&\leq\frac{1}{2^N}\sum_{u^N}2\sqrt{\sum_{i=1}^{N}\mathrm{Tr}\Big\{\Big(I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}\Big)\rho_{u^N}\Big\}}\\
&=\frac{1}{2^N}\sum_{u^N}2\sqrt{\sum_{i\in\mathcal{A}}\mathrm{Tr}\Big\{\Big(I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}\Big)\rho_{u^N}\Big\}}\\
&\leq 2\sqrt{\frac{1}{2^N}\sum_{u^N}\sum_{i\in\mathcal{A}}\mathrm{Tr}\Big\{\Big(I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}\Big)\rho_{u^N}\Big\}}
\end{aligned}$$

where the second equality follows from our convention that $\Pi^{B^N}_{(i),u_1^{i-1}u_i}=I$ if u_i is a frozen bit and the second inequality follows from concavity of the square root. Continuing, we have

$$\begin{aligned}
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u^N}\frac{1}{2^N}\mathrm{Tr}\Big\{\hat{\Pi}^{B^N}_{(i),u_1^{i-1}u_i}\rho_{u^N}\Big\}}\\
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u_1^{i-1}}\frac{1}{2^{i-1}}\sum_{u_i}\frac{1}{2}\sum_{u_{i+1}^N}\frac{1}{2^{N-i}}\mathrm{Tr}\Big\{\hat{\Pi}^{B^N}_{(i),u_1^{i-1}u_i}\rho_{u^N}\Big\}}\\
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u_1^{i-1}}\frac{1}{2^{i-1}}\sum_{u_i}\frac{1}{2}\mathrm{Tr}\Bigg\{\hat{\Pi}^{B^N}_{(i),u_1^{i-1}u_i}\sum_{u_{i+1}^N}\frac{1}{2^{N-i}}\rho_{u^N}\Bigg\}},
\end{aligned}$$

where we define

$$\hat{\Pi}^{B^N}_{(i),u_1^{i-1}u_i}=I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}.$$

The first equality follows from exchanging the sums. The second equality follows from expanding the sum and normalization $\sum_{u^N}\frac{1}{2^N}$. The third equality follows from bringing the sum $\sum_{u_{i+1}^N}\frac{1}{2^{N-i}}$ inside the trace. Continuing,

$$\begin{aligned}
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u_1^{i-1}}\frac{1}{2^{i-1}}\sum_{u_i}\frac{1}{2}\mathrm{Tr}\Big\{\Big(I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}\Big)\overline{\rho}^{B^N}_{u_1^i}\Big\}}\\
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u_i}\frac{1}{2}\sum_{u_1^{i-1}}\frac{1}{2^{i-1}}\mathrm{Tr}\Big\{\Big(I-\Pi^{B^N}_{(i),u_1^{i-1}u_i}\Big)\overline{\rho}^{B^N}_{u_1^i}\Big\}}\\
&=2\Bigg(\sum_{i\in\mathcal{A}}\sum_{u_i}\frac{1}{2}\mathrm{Tr}\Big\{(I-\sum_{u_1^{i-1}}\big|u_1^{i-1}\big\rangle\big\langle u_1^{i-1}\big|^{U_1^{i-1}}\otimes\Pi^{B^N}_{(i),u_1^{i-1}u_i})\\
&\qquad\sum_{u_1^{i-1}}\frac{1}{2^{i-1}}\big|u_1^{i-1}\big\rangle\big\langle u_1^{i-1}\big|^{U_1^{i-1}}\otimes\overline{\rho}^{B^N}_{u_1^i}\Big\}\Bigg)^{-\frac{1}{2}}
\end{aligned}$$

The first equality is from the definition in (8). The second equality is from exchanging sums. The third equality is from the fact that

$$\sum_x p(x)\mathrm{Tr}\{A_x\rho_x\}=\mathrm{Tr}\Bigg\{\Bigg(\sum_x|x\rangle\langle x|\otimes A_x\Bigg)\Bigg(\sum_{x'}p(x')|x'\rangle\langle x'|\otimes\rho_{x'}\Bigg)\Bigg\}.$$

Continuing,

$$\begin{aligned}
&=2\sqrt{\sum_{i\in\mathcal{A}}\sum_{u_i}\frac{1}{2}\mathrm{Tr}\Big\{\Big(I-\Pi^{U_1^{i-1}B^N}_{(i),u_i}\Big)\rho^{U_1^{i-1}B^N}_{(i),u_i}\Big\}}\\
&\leq 2\sqrt{\sum_{i\in\mathcal{A}}\frac{1}{2}\sqrt{F\big(W^{(i)}\big)}}
\end{aligned}$$

The first equality is from the observations in (11-12) and the definition in (6). The final inequality follows from Lemma 3.2 of Ref. [37] and the definition in (9). This completes the proof of Proposition 4.

We state the proof of Theorem 6 for completeness. Invoking Theorem 3, there exists a sequence of sets $\mathcal{A}_N$ with size $|\mathcal{A}_N|\geq NR$ for any $R<I(W)$ and $\beta<1/2$ such that

$$\sum_{i\in\mathcal{A}_N}\sqrt{F\big(W^{(i)}\big)}=o(2^{-N^\beta}),$$

and thus

$$2\sqrt{\sum_{i\in\mathcal{A}_N}\frac{1}{2}\sqrt{F\big(W^{(i)}\big)}}=o(2^{-\frac{1}{2}N^\beta}).$$

This bound holds if we choose the set $\mathcal{A}_N$ according to the polar coding rule because this rule minimizes the above sum by definition. Theorem 6 follows by combining Proposition 4 with this fact about the polar coding rule.

VI. Conclusion

We have shown how to construct polar codes for channels with classical binary inputs and quantum outputs, and we showed that they can achieve the symmetric Holevo information rate for classical communication. In fact, for a quantum channel with binary pure state outputs, such as a binary-phase-shift-keyed (BPSK) coherent-state optical communication alphabet, the symmetric Holevo information rate is the

[5] We say that Sen's bound is a "non-commutative union bound" because it is analogous to the following union bound from probability theory: $\Pr\{(A_1\cap\cdots\cap A_N)^c\}=\Pr\{A_1^c\cup\cdots\cup A_N^c\}\leq\sum_{i=1}^N\Pr\{A_i^c\}$, where $A_1,\ldots,A_N$ are events. The analogous bound for projector logic would be $\mathrm{Tr}\{(I-\Pi_1\cdots\Pi_N\cdots\Pi_1)\rho\}\leq\sum_{i=1}^N\mathrm{Tr}\{(I-\Pi_i)\rho\}$, if we think of $\Pi_1\cdots\Pi_N$ as a projector onto the intersection of subspaces. Though, the above bound only holds if the projectors $\Pi_1,\ldots,\Pi_N$ are commuting (choosing $\Pi_1=|+\rangle\langle+|$, $\Pi_2=|0\rangle\langle 0|$, and $\rho=|0\rangle\langle 0|$ gives a counterexample). If the projectors are non-commuting, then Sen's bound in (30) is the next best thing and suffices for our purposes here.

ultimate channel capacity [15], which is therefore achieved by our polar code [42]. The general idea behind the construction is similar to Arikan's [4], but we required several technical advances in order to demonstrate both channel polarization at the symmetric Holevo information rate and the operation of the quantum successive cancellation decoder. To prove that channel polarization takes hold, we could exploit several results in the quantum information literature [30], [31], [32], [10], [33], [11] and some of Arikan's tools. To prove that the quantum successive cancellation decoder works well, we exploited some ideas from quantum hypothesis testing [34], [35], [36], [37], [38] and Sen's recent "non-commutative union bound" [25]. The result is a near-explicit code construction that achieves the symmetric Holevo information rate for channels with classical inputs and quantum outputs. (When we say "near-explicit," we mean that it still remains open in the quantum case to determine which synthesized channels are good or bad.) Also, several works have now appeared on polar coding for private classical communication and quantum communication [43], [44], [43], [45], [46], [47], most of which use the results developed in this paper.

One of the main open problems going forward from here is to simplify the quantum successive cancellation decoder. Arikan could show how to calculate later estimates by exploiting the results of earlier estimates in an "FFT-like" fashion, and this observation reduced the complexity of the decoding to $O(N \log N)$. It is not clear to us yet how to reduce the complexity of the quantum successive cancellation decoder because it is not merely a matter of computing formulas, but rather a sequence of physical operations (measurements) that the receiver needs to perform on the channel output systems. If there were some way to perform the measurements on smaller systems and then adaptively perform other measurements based on earlier results, then this would be helpful in demonstrating a reduced complexity.

Another important open question is to devise an efficient construction of the polar codes, something that remains an open problem even for classical polar codes. However, there has been recent work on efficient suboptimal classical polar code constructions [48], which one might try to extend to polar codes for the classical-quantum channel. Finally, extending our code and decoder construction to a classical-quantum channel with a non-binary (M-ary) alphabet remains a good open line for investigation.

VII. Acknowledgments

MMW acknowledges financial support from the MDEIE (Québec) PSR-SIIRI international collaboration grant. SG was supported by the DARPA Information in a Photon (InPho) program under contract number HR0011-10-C-0159. We thank David Forney, MIT for suggesting us to try polar codes for the quantum channel. We also thank Emre Telatar, EPFL for an intuitive tutorial on channel polarization at ISIT 2011.

Appendix

Proof of Proposition 1: The first bound in (1) follows from Holevo's characterization of the quantum cutoff rate (Proposition 1 of Ref. [32]). In particular, Holevo proved that the following inequality holds for all $s \in [0,1]$:

$$I(X;B)_\omega \geq -\log \operatorname{Tr}\left\{ \left(\sum_{x\in\mathcal{X}} p_X(x) (\omega_x)^{\frac{1}{1+s}} \right)^{1+s} \right\},$$

where the entropy on the LHS is with respect to a classical-quantum state

$$\omega^{XB} \equiv \sum_{x\in\mathcal{X}} p_X(x) |x\rangle\langle x|^X \otimes \omega_x^B .$$

By setting $s = 1$, the alphabet $\mathcal{X} = \{0,1\}$, and the distribution $p_X(x)$ to be uniform, we obtain the bound

$$\begin{aligned} I(W) &\geq -\log\left(\operatorname{Tr}\left\{ \left(\frac{1}{2}(\sqrt{\rho_0} + \sqrt{\rho_1}) \right)^2 \right\} \right) \\ &= -\log\left(\frac{1}{4} \operatorname{Tr}\{\rho_0 + \sqrt{\rho_1}\sqrt{\rho_0} + \sqrt{\rho_0}\sqrt{\rho_1} + \rho_1\} \right) \\ &= -\log\left(\frac{1}{2}(1 + \operatorname{Tr}\{\sqrt{\rho_0}\sqrt{\rho_1}\}) \right) \\ &= \log\left(\frac{2}{1 + \operatorname{Tr}\{\sqrt{\rho_0}\sqrt{\rho_1}\}} \right) \\ &\geq \log\left(\frac{2}{1 + \sqrt{F(W)}} \right), \end{aligned}$$

where the last line follows from

$$\begin{aligned} \operatorname{Tr}\{\sqrt{\rho_0}\sqrt{\rho_1}\} &\leq \operatorname{Tr}\{|\sqrt{\rho_0}\sqrt{\rho_1}|\} \\ &= \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1 \\ &= \sqrt{F(W)}. \end{aligned}$$

The other inequality in (2) follows from (21) in Ref. [33]. In particular, they showed that

$$I(W) \leq H_2\left(\frac{1}{2}\left(1 - \sqrt{F(W)}\right) \right),$$

where the binary entropy $H_2(x) \equiv -x \log_2 x - (1-x)\log_2(1-x)$. Combining this with the following observation that holds for all $0 \leq F(W) \leq 1$ gives the second inequality:

$$H_2\left(\frac{1}{2}\left(1 - \sqrt{F(W)}\right) \right) \leq \sqrt{1 - F(W)}.$$

■

Proof of Proposition 8: These follow from the same line of reasoning as in the proof of Arikan's Proposition 4 [4]. We prove the first equality. Consider the mutual information

$$\begin{aligned} I(U_1U_2; B_1B_2) &= I(X_1X_2; B_1B_2) \\ &= I(X_1; B_1) + I(X_2; B_2) \\ &= 2I(W). \end{aligned}$$

By the chain rule for quantum mutual information [11], we have

$$\begin{aligned} I(U_1U_2; B_1B_2) &= I(U_1; B_1B_2) + I(U_2; B_1B_2U_1) \\ &= I\big(W^-\big) + I\big(W^+\big). \end{aligned}$$

The inequality follows because

$$\begin{aligned}I\big(W^{+}\big) &= I(U_2; B_1 B_2 U_1) \\ &= I(U_2; B_2) + I(U_2; B_1 U_1 | B_2) \\ &= I(W) + I(U_2; B_1 U_1 | B_2).\end{aligned}$$

Thus,

$$I\big(W^{+}\big) \geq I(W)$$

because $I(U_2; B_1 U_1 | B_2) \geq 0$ [30], [10], [11]. We then have

$$\begin{aligned}2I\big(W^{+}\big) &\geq 2I(W) \\ &= I\big(W^{-}\big) + I\big(W^{+}\big),\end{aligned}$$

and the inequality follows. ■

Proof of Proposition 9: We begin with the first equality. Consider that

$$\begin{aligned}&\sqrt{F(W^{+})} \\ &= \sqrt{F\big(\rho_0^{+}, \rho_1^{+}\big)} \\ &= \sqrt{F\left(\frac{1}{2}\sum_{u_1} |u_1\rangle\langle u_1| \otimes \rho_{u_1}^{B_1}, \frac{1}{2}\sum_{u_1} |u_1 \oplus 1\rangle\langle u_1 \oplus 1| \otimes \rho_{u_1}^{B_1}\right)} \\ &\qquad \times \sqrt{F\left(\rho_0^{B_2}, \rho_1^{B_2}\right)} \\ &= \frac{1}{2}\left(\sqrt{F\left(\rho_0^{B_1}, \rho_1^{B_1}\right)} + \sqrt{F\left(\rho_1^{B_1}, \rho_0^{B_1}\right)}\right)\sqrt{F\left(\rho_0^{B_2}, \rho_1^{B_2}\right)} \\ &= F(\rho_0, \rho_1) \\ &= F(W)\end{aligned}$$

The first two equalities follow by definition. The third equality follows from the multiplicativity of fidelity under tensor product states [10], [11]:

$$F(\rho \otimes \sigma, \tau \otimes \omega) = F(\rho, \tau) F(\sigma, \omega).$$

The fourth equality follows from the following formula that holds for the fidelity of classical-quantum states:

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

We now consider the second inequality. The fidelity also has the following characterization as the minimum Bhattacharya overlap between distributions induced by a POVM on the states [31], [10], [11]:

$$F(\rho_0, \rho_1) = \min_{\{\Lambda_m\}}\left(\sum_m \sqrt{\mathrm{Tr}\{\Lambda_m \rho_0\}\mathrm{Tr}\{\Lambda_m \rho_1\}}\right)^2.$$

So

$$\sqrt{F(W^{-})} = \min_{\left\{\Gamma_m^{B_1 B_2}\right\}} \sum_m \sqrt{\mathrm{Tr}\left\{\Gamma_m^{B_1 B_2}\rho_0^{-}\right\}\mathrm{Tr}\left\{\Gamma_m^{B_1 B_2}\rho_1^{-}\right\}}$$

Let Λ_m denote the POVM that achieves the minimum for $\sqrt{F(W)}$:

$$\begin{aligned}\sqrt{F(W)} &= \sqrt{F(\rho_0, \rho_1)} \\ &= \min_{\{\Lambda_m\}} \sum_m \sqrt{\mathrm{Tr}\{\Lambda_m \rho_0\}\mathrm{Tr}\{\Lambda_m \rho_1\}}.\end{aligned}$$

Then the POVM $\{\Lambda_l \otimes \Lambda_m\}$ is a particular POVM that can try to distinguish the states ρ_0^{-} and ρ_1^{-}. We then have

$$\begin{aligned}&\sqrt{F(W^{-})} \\ &\leq \sum_{l,m} \sqrt{\mathrm{Tr}\big\{(\Lambda_l \otimes \Lambda_m)\big(\rho_0^{-}\big)\big\}\mathrm{Tr}\big\{(\Lambda_l \otimes \Lambda_m)\big(\rho_1^{-}\big)\big\}} \\ &= \sum_{l,m} \sqrt{\mathrm{Tr}\left\{(\Lambda_l \otimes \Lambda_m)\frac{1}{2}\left(\rho_0^{B_1} \otimes \rho_0^{B_2} + \rho_1^{B_1} \otimes \rho_1^{B_2}\right)\right\}} \\ &\qquad \times \sqrt{\mathrm{Tr}\left\{(\Lambda_l \otimes \Lambda_m)\frac{1}{2}\left(\rho_1^{B_1} \otimes \rho_0^{B_2} + \rho_0^{B_1} \otimes \rho_1^{B_2}\right)\right\}}\end{aligned}$$

$$\begin{aligned}&= \frac{1}{2}\sum_{l,m}\Bigg[\Bigg(\mathrm{Tr}\left\{\Lambda_l \rho_0^{B_1}\right\}\mathrm{Tr}\left\{\Lambda_m \rho_0^{B_2}\right\} + \\ &\quad \mathrm{Tr}\left\{\Lambda_l \rho_1^{B_1}\right\}\mathrm{Tr}\left\{\Lambda_m \rho_1^{B_2}\right\}\Bigg)\Bigg(\mathrm{Tr}\left\{\Lambda_l \rho_1^{B_1}\right\}\mathrm{Tr}\left\{\Lambda_m \rho_0^{B_2}\right\} + \\ &\qquad \mathrm{Tr}\left\{\Lambda_l \rho_0^{B_1}\right\}\mathrm{Tr}\left\{\Lambda_m \rho_1^{B_2}\right\}\Bigg)\Bigg]^{1/2}\end{aligned}$$

Making the assignments

$$\begin{aligned}\alpha_m &\equiv \mathrm{Tr}\left\{\Lambda_m \rho_0^{B_2}\right\}, \\ \beta_l &\equiv \mathrm{Tr}\left\{\Lambda_l \rho_0^{B_1}\right\}, \\ \gamma_l &\equiv \mathrm{Tr}\left\{\Lambda_l \rho_1^{B_1}\right\}, \\ \delta_m &\equiv \mathrm{Tr}\left\{\Lambda_m \rho_1^{B_2}\right\},\end{aligned}$$

the above expression is equal to

$$\sum_{l,m} \frac{1}{2}\sqrt{\beta_l \alpha_m + \gamma_l \delta_m}\sqrt{\gamma_l \alpha_m + \beta_l \delta_m}$$

We can then exploit Arikan's inequality in Appendix D of Ref. [4] to have

$$\begin{aligned}&\sum_{l,m} \frac{1}{2}\sqrt{\beta_l \alpha_m + \gamma_l \delta_m}\sqrt{\gamma_l \alpha_m + \beta_l \delta_m} \\ &\leq \sum_{l,m} \frac{1}{2}\left(\sqrt{\beta_l \alpha_m} + \sqrt{\gamma_l \delta_m}\right)\left(\sqrt{\gamma_l \alpha_m} + \sqrt{\beta_l \delta_m}\right) \\ &\qquad - \sum_{l,m} \sqrt{\beta_l \alpha_m \gamma_l \delta_m} \\ &= \sum_{l,m} \frac{1}{2}\left(\alpha_m \sqrt{\beta_l \gamma_l} + \gamma_l \sqrt{\delta_m \alpha_m} + \beta_l \sqrt{\alpha_m \delta_m} + \delta_m \sqrt{\gamma_l \beta_l}\right) \\ &\qquad - \sum_l \sqrt{\beta_l \gamma_l} \sum_m \sqrt{\alpha_m \delta_m} \\ &= \sum_l \sqrt{\beta_l \gamma_l} + \sum_m \sqrt{\delta_m \alpha_m} - \sum_l \sqrt{\beta_l \gamma_l} \sum_m \sqrt{\alpha_m \delta_m} \\ &= 2\sqrt{F(W)} - F(W).\end{aligned}$$

The inequality $F(W^-) \geq F(W)$ follows from concavity of fidelity and its multiplicativity under tensor products [10], [11]:

$$\begin{aligned}
F(W^-) &= F(\rho_0^-, \rho_1^-) \\
&\geq \frac{1}{2}F\Big(\rho_0^{B_1}\otimes\rho_0^{B_2}, \rho_1^{B_1}\otimes\rho_0^{B_2}\Big) \\
&\quad + \frac{1}{2}F\Big(\rho_1^{B_1}\otimes\rho_1^{B_2}, \rho_0^{B_1}\otimes\rho_1^{B_2}\Big) \\
&= \frac{1}{2}F\Big(\rho_0^{B_1}, \rho_1^{B_1}\Big)F\Big(\rho_0^{B_2}, \rho_0^{B_2}\Big) \\
&\quad + \frac{1}{2}F\Big(\rho_1^{B_1}, \rho_0^{B_1}\Big)F\Big(\rho_1^{B_2}, \rho_1^{B_2}\Big) \\
&= \frac{1}{2}F\Big(\rho_0^{B_1}, \rho_1^{B_1}\Big) + \frac{1}{2}F\Big(\rho_1^{B_1}, \rho_0^{B_1}\Big) \\
&= F(W)
\end{aligned}$$

The inequality $F(W) \geq F(W^+)$ follows from the relation $\sqrt{F(W^+)} = F(W)$ and the fact that $0 \leq F \leq 1$. ■

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

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

[3] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, July 2009.

[4] E. Arikan, "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, July 2009, arXiv:0807.3917.

[5] ——, "Channel combining and splitting for cutoff rate improvement," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 628–639, February 2006, arXiv:cs/0508034.

[6] E. Sasoglu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Proceedings of the 2009 Information Theory Workshop*, Taormina, Sicily, Italy, October 2009, pp. 144–148, arXiv:0908.0302.

[7] E. Arikan, "Source polarization," in *Proceedings of the 2010 IEEE International Symposium on Information Theory*, Austin, Texas, USA, June 2010, pp. 899–903, arXiv:1001.3087.

[8] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010, arXiv:0903.0307.

[9] E. Sasoglu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," June 2010, arXiv:1006.4255.

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

[11] M. M. Wilde, *From Classical to Quantum Shannon Theory*, June 2011, arXiv:1106.1445.

[12] A. S. Holevo, "The capacity of the quantum channel with general signal states." *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, January 1998, arXiv:quant-ph/9611023.

[13] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, pp. 131–138, July 1997.

[14] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, "Classical capacity of the lossy bosonic channel: The exact solution," *Physical Review Letters*, vol. 92, no. 2, p. 027902, January 2004.

[15] S. Guha, "Structured optical receivers to attain superadditive capacity and the Holevo limit," *Physical Review Letters*, vol. 106, p. 240502, June 2011, arXiv:1101.1550.

[16] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Physics*, vol. 5, pp. 255–257, April 2009, arXiv:0809.3972.

[17] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1753–1768, 2003, arXiv:quant-ph/0206186.

[18] A. S. Holevo, "Coding theorems for quantum channels," Tamagawa University Research Review, Tech. Rep. 4, 1998, arXiv:quant-ph/9809023.

[19] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481–2485, 1999.

[20] N. Datta and T. Dorlas, "A quantum version of Feinstein's theorem and its application to channel coding," in *Proceedings of the IEEE International Symposium on Information Theory*, Seattle, Washington, USA, 2006, pp. 441–445.

[21] T. Ogawa and H. Nagaoka, "Making good codes for classical-quantum channel coding via quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2261–2266, June 2007.

[22] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Physical Review Letters*, vol. 108, p. 200501, May 2012.

[23] V. Giovannetti, S. Lloyd, and L. Maccone, "Achieving the Holevo bound via sequential measurements," *Physical Review A*, vol. 85, p. 012302, January 2012.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[25] P. Sen, "Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding," September 2011, arXiv:1109.0802.

[26] S. J. Devitt, K. Nemoto, and W. J. Munro, "Quantum error correction for beginners," May 2009, arXiv:0905.2794.

[27] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problems of Information Transmission*, vol. 9, pp. 177–183, 1973.

[28] A. Uhlmann, "The "transition probability" in the state space of a *-algebra," *Reports on Mathematical Physics*, vol. 9, no. 2, pp. 273–279, 1976.

[29] R. Jozsa, "Fidelity for mixed quantum states," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2315–2323, 1994.

[30] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *Journal of Mathematical Physics*, vol. 14, pp. 1938–1941, 1973.

[31] 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, arXiv:quant-ph/9712042.

[32] A. S. Holevo, "Reliability function of general classical-quantum channel," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2256–2261, September 2000, arXiv:quant-ph/9907087.

[33] W. Roga, M. Fannes, and K. Życzkowski, "Universal bounds for the Holevo quantity, coherent information, and the Jensen-Shannon divergence," *Physical Review Letters*, vol. 105, p. 040505, July 2010, arXiv:1004.4782.

[34] C. W. Helstrom, "Quantum detection and estimation theory," *Journal of Statistical Physics*, vol. 1, pp. 231–252, 1969. [Online]. Available: http://dx.doi.org/10.1007/BF01007479

[35] A. S. Holevo, "An analog of the theory of statistical decisions in noncommutative theory of probability," *Trudy Moscov Mat. Obsc.*, vol. 26, pp. 133–149, 1972, english translation: Trans. Moscow Math Soc. 26, 133–149 (1972).

[36] C. W. Helstrom, *Quantum Detection and Estimation Theory*. New York: Academic, 1976.

[37] M. Hayashi, *Quantum Information: An Introduction*. Berlin Heidelberg: Springer-Verlag, 2006.

[38] H. Nagaoka and M. Hayashi, "An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 534–549, February 2007, arXiv:quant-ph/0206185.

[39] J. Calsamiglia, R. Muñoz Tapia, L. Masanes, A. Acin, and E. Bagan, "Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states," *Physical Review A*, vol. 77, p. 032311, March 2008, arXiv:0708.2343.

[40] E. Arikan and E. Telatar, "On the rate of channel polarization," in *Proceedings of the 2009 International Symposium on Information Theory*, Seoul, Korea, June 2009, pp. 1493–1495, arXiv:0807.3806.

[41] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6253–6264, December 2010, arXiv:0901.0536.

[42] S. Guha and M. M. Wilde, "Polar coding to achieve the holevo capacity of a pure-loss optical channel," *Proceedings of the 2012 International Symposium on Information Theory*, pp. 551–555, July 2012, arXiv:1202.0533.

[43] M. M. Wilde and S. Guha, "Polar codes for degradable quantum channels," September 2011, arXiv:1109.5346.

[44] J. M. Renes, F. Dupuis, and R. Renner, "Efficient quantum polar coding," September 2011, arXiv:1109.3195.

[45] M. M. Wilde and J. M. Renes, "Quantum polar codes for arbitrary channels," *Proceedings of the 2012 International Symposium on Information Theory*, pp. 339–343, July 2012, arXiv:1201.2906.

[46] ——, "Polar codes for private classical communication," 2012, arXiv:1203.5794.

[47] Z. Dutton, S. Guha, and M. M. Wilde, "Performance of polar codes for quantum and private classical communication," *Submitted to the Allerton Conference on Control, Communication, and Computing*, 2012, arXiv:1205.5980.

[48] I. Tal and A. Vardy, "How to construct polar codes," May 2011, arXiv:1105.6164.