Research

Author for correspondence:
Mark M. Wilde
e-mail: mwilde@gmail.com

Sequential decoding of a general classical-quantum channel

Mark M. Wilde

School of Computer Science, McGill University, Montreal, Quebec, Canada H3A 2A7

Because a quantum measurement generally disturbs the state of a quantum system, one might think that it should not be possible for a sender and receiver to communicate reliably when the receiver performs a large number of sequential measurements to determine the message of the sender. We show here that this intuition is not true, by demonstrating that a sequential decoding strategy works well even in the most general 'one-shot' regime, where we are given a single instance of a channel and wish to determine the maximal number of bits that can be communicated up to a small failure probability. This result follows by generalizing a non-commutative union bound to apply for a sequence of general measurements. We also demonstrate two ways in which a receiver can recover a state close to the original state after it has been decoded by a sequence of measurements that each succeed with high probability. The second of these methods will be useful in realizing an efficient decoder for fully quantum polar codes, should a method ever be found to realize an efficient decoder for classical-quantum polar codes.

1. Introduction

The reliable communication of classical data over quantum channels is one of the earliest problems to be considered in quantum information theory. Some of the most important contributions to this problem (to name just a few) are the Holevo upper bound on the accessible information [1], the coding theorem due to [2,3] Holevo–Schumacher–Westmoreland (HSW), and the fact that entangled signalling states can enhance communication rates for certain quantum channels [4].

The main difference between the proofs of the HSW theorem and Shannon's classical channel capacity

theorem [5] is that, in the former case, one has to specify a quantum measurement that recovers the classical data being transmitted (a quantum decoder), as opposed to a classical algorithm that does so. Indeed, in their respective proofs, HSW demonstrated that a quantum measurement known as the 'pretty-good' or 'square-root' measurement [6–8] allows a receiver to decode classical information reliably at a rate equal to the Holevo rate. For a pure-loss bosonic channel modelling free-space communication, for example, this Holevo rate can be significantly higher than data rates that are achievable with more traditional measurement strategies such as homodyne or heterodyne detection [9].

As in the HSW decoding measurement, we typically perform measurements on quantum systems in order to gain information about them, and one well-known feature of quantum mechanics is that a measurement can disturb the state of the system that we are measuring. Thus, it came as a surprise when Lloyd–Giovannetti–Maccone (LGM) [10,11] showed that it is possible to achieve the Holevo rate by performing independent non-commuting sequential measurements, in analogy with classical sequential decoding strategies [12]. A sequential decoding scheme proceeds according to the following simple algorithm:

(1) Let M be the total number of codewords. Initialize a counter $i = 1$.
(2) Perform a quantum measurement to determine whether the transmitted codeword is the ith codeword.
(3) If the measurement result is 'yes', decode as codeword i and conclude. If the measurement result is 'no,' increment i.
(4) If $i \leq M$, go to step 2. Otherwise, declare failure.

After the work of LGM, Sen [13] presented a remarkable simplification of their error analysis, by establishing a non-commutative union bound that holds for a set of projective measurements applied sequentially to a quantum state (where the projective measurements do not necessarily commute). This non-commutative union bound is an extension of the familiar union bound from probability theory, and as such, it should find wide application in settings beyond those considered in quantum communication theory. Sen applied his non-commutative union bound to a variety of problems in [13], including the problem of classical communication over quantum channels, and it has since been applied in designing Holevo-rate-achieving polar codes for classical-quantum channels [14] and in demonstrating how to decode the pure-loss bosonic channel at the Holevo rate [15].

All of the above results apply to a setting in which the channel is memoryless and identically distributed, so that one use of it does not depend on the others and so that each use leads to the same noise at the output as the other uses, respectively. Given that this 'IID' setting is really just an idealization, there has been a strong effort to develop a theory of quantum information that goes beyond the IID setting and applies to channels with no structure whatsoever [16–18]. This regime beyond the IID setting is known as the 'one-shot' regime, where we are concerned with a single instance of a resource and desire to make the best use of it up to some controllable failure probability. In this vein, there have been several contributions characterizing the reliable communication of classical data over quantum channels [19–21], and all of these used the 'pretty-good' measurement as the decoder.

Many of the developments listed above have improved our understanding of classical communication over quantum channels, but there are some important considerations left unanswered:

— We know very well that the most general kind of measurement allowed in quantum mechanics is a positive operator-valued measure (POVM). Does Sen's bound generalize so that it applies for a sequence of general measurements?
— Does sequential decoding work well in the one-shot regime?
— When can one conclude that the state resulting from a sequence of general measurements is close to the state before this sequence of measurements occurs?

This paper resolves the above problems, by showing that

— Sen's non-commutative union bound applies not just for a sequence of projections, but for the more general case of a sequence of positive operators each with spectrum less than one. This result follows simply by applying the well-known Naimark extension theorem.[1] Thus, the non-commutative union bound now applies for a sequence of general measurements (POVMs) and, as such, it should find wide application in other areas of quantum information science.

— Indeed, sequential decoding works well even in the one-shot regime. That is, one can give a meaningful bound on the amount of information that can be transmitted up to a failure probability no larger than ε for some $\varepsilon > 0$ when using a sequential decoding strategy. The information bound we present is very similar to the bound of Wang & Renner [21].

— A sequence of measurements followed by the reverse sequence of these measurements causes only a negligible disturbance to a state if the original sequence of measurements has a high probability of success. This last result generalizes Winter's gentle measurement lemma [22] to the more general setting of a sequence of measurements. One application of this last result is in decoding fully quantum polar codes for arbitrary quantum channels [23].[2]

We structure this paper as follows. Section 2 reviews some background material, including the definition of the hypothesis testing relative entropy [21,24], the Naimark extension theorem and Sen's non-commutative union bound [13]. We then proceed in the order given above.

2. Review

(a) Hypothesis testing relative entropy

The hypothesis testing relative entropy, denoted as $D_H^\varepsilon(\rho\|\sigma)$, is an entropy measure derived from the error probabilities arising from a quantum measurement that attempts to distinguish between the states ρ and σ (a quantum hypothesis test). The most general measurement that one could use in such a test is a two-outcome POVM $\{Q, I - Q\}$, where $0 \leq Q \leq I$. The outcome Q corresponds to deciding that the state is ρ, and the outcome $I - Q$ corresponds to deciding that the state is σ. Thus, the probability of guessing correctly when the state is ρ is equal to $\mathrm{Tr}\{Q\rho\}$, and the probability of guessing incorrectly when the state is σ is equal to $\mathrm{Tr}\{Q\sigma\}$. In an asymmetric quantum hypothesis test, we try to find a POVM that guesses ρ correctly with high probability, so that

$$\mathrm{Tr}\{Q\rho\} \geq 1 - \varepsilon, \qquad (2.1)$$

for some small, fixed $\varepsilon \geq 0$, while minimizing the probability that we guess σ incorrectly. This naturally leads to a semidefinite optimization programme, specified by the following quantity:

$$\beta_\varepsilon(\rho, \sigma) \equiv \min_Q \{\mathrm{Tr}\{Q\sigma\} : 0 \leq Q \leq I, \ \mathrm{Tr}\{Q\rho\} \geq 1 - \varepsilon\}. \qquad (2.2)$$

By taking the negative logarithm of $\beta_\varepsilon(\rho, \sigma)$, we arrive at the hypothesis testing relative entropy defined in Buscemi & Datta [24] and Wang & Renner [21]:

$$D_H^\varepsilon(\rho\|\sigma) \equiv -\log \beta_\varepsilon(\rho, \sigma). \qquad (2.3)$$

One can derive other entropic measures based on the hypothesis testing relative entropy that have various natural properties [25].

---

[1]This observation is due to Andreas Winter and Aram Harrow from a discussion in December 2011 at QIP 2012.

[2]Quantum polar codes are the only known near-explicit quantum-error-correcting codes that achieve the coherent information rate of an arbitrary quantum channel.

(b) Naimark extension theorem

We briefly review the Naimark extension theorem and a straightforward proof of it. The importance of this theorem is that it demonstrates how one can implement a general quantum measurement simply by performing a unitary on the system of interest and a probe system, followed by a von Neumann measurement of the probe.

Theorem 2.1 (Naimark). *For any POVM $\{\Gamma_x\}_{x \in \mathcal{X}}$ acting on a system S, there exists a unitary U_{SP} (acting on the system S and a probe system P) and an orthonormal basis $\{|x\rangle_P\}_{x \in \mathcal{X}}$ such that*

$$\text{Tr}\{U_{SP}^\dagger (I_S \otimes |x\rangle\langle x|_P) U_{SP}(\rho_S \otimes |0\rangle\langle 0|_P)\} = \text{Tr}\{\Gamma_x \rho\}. \tag{2.4}$$

Proof. For every POVM $\{\Gamma_x\}$, we can form the following isometry:

$$V_{SP} \equiv \sum_x (\sqrt{\Gamma_x})_S \otimes |x\rangle\langle 0|_P, \tag{2.5}$$

which can be extended to a unitary operator U_{SP} by appropriately filling out the other $|\mathcal{X}| - 1$ entries of the form:

$$\sum_x (A_{x,x'})_S \otimes |x\rangle\langle x'|, \tag{2.6}$$

for some operators $A_{x,x'}$ and where $x' \in \{1, \dots, |\mathcal{X}| - 1\}$. The statement of the theorem then follows easily from this choice of unitary. ∎

Example 2.2. Let $\{\Gamma, I - \Gamma\}$ be a binary POVM acting on the system S. Consider the following unitary operator U_{SP} acting on the system S and a qubit probe system P:

$$U_{SP} \equiv (\sqrt{\Gamma})_S \otimes |0\rangle\langle 0|_P + (\sqrt{I - \Gamma})_S \otimes |1\rangle\langle 0|_P - (\sqrt{I - \Gamma})_S \otimes |0\rangle\langle 1|_P + (\sqrt{\Gamma})_S \otimes |1\rangle\langle 1|_P. \tag{2.7}$$

The above unitary corresponds to a Naimark extension of the POVM $\{\Gamma, I - \Gamma\}$.

(c) Non-commutative union bound

This section recalls Sen's [13] non-commutative union bound. As we mentioned in §1, this bound should find wide application in settings beyond those considered for communication, because it generalizes the union bound from probability theory.

Theorem 2.3 (Sen). *For a subnormalized state σ such that $\sigma \geq 0$ and $\text{Tr}\{\sigma\} \leq 1$, and a sequence of Hermitian projectors $\Pi_1, \dots, \Pi_M$, the following non-commutative union bound holds:*

$$\text{Tr}\{\sigma\} - \text{Tr}\{\Pi_M \cdots \Pi_1 \sigma \Pi_1 \cdots \Pi_M\} \leq 2\sqrt{\sum_{m=1}^{M} \text{Tr}\{(I - \Pi_m)\sigma\}}. \tag{2.8}$$

3. Non-commutative union bound for positive operator-valued measures

We now give an extension of Sen's non-commutative union bound that applies for general measurements.

Lemma 3.1. *Let σ be a subnormalized state such that $\sigma \geq 0$ and $\text{Tr}\{\sigma\} \leq 1$, and let $\Lambda_1, \dots, \Lambda_M$ denote a set of positive operators such that $0 \leq \Lambda_m \leq I$ for all $m \in \{1, \dots, M\}$. Then, the following non-commutative union bound holds:*

$$\text{Tr}\{\sigma\} - \text{Tr}\{\Pi_{\Lambda_M} \cdots \Pi_{\Lambda_1}(\sigma \otimes |\bar{0}\rangle\langle\bar{0}|_{PM})\Pi_{\Lambda_1} \cdots \Pi_{\Lambda_M}\} \leq 2\sqrt{\sum_{m=1}^{M} \text{Tr}\{(I - \Lambda_m)\sigma\}}, \tag{3.1}$$

where $|\bar{0}\rangle_{PM} \equiv |0\rangle_{P_1} \otimes \cdots \otimes |0\rangle_{P_M}$ is an ancillary state of M probe systems and Π_{Λ_i} is a projector defined as $\Pi_{\Lambda_i} \equiv U_i^\dagger P_i U_i$, for some unitary U_i and projector P_i such that

$$\text{Tr}\{\Pi_{\Lambda_i}(\sigma \otimes |\bar{0}\rangle\langle\bar{0}|_{PM})\} = \text{Tr}\{\Lambda_m \sigma\}. \tag{3.2}$$

Proof. This extension of Sen's bound follows easily by using the Naimark extension theorem and Sen's non-commutative union bound.

To each POVM element Λ_i (as in the statement of the theorem), there exists a unitary $U^{(i)}_{SP_i}$ (acting on the system S and the ith probe system) and a projector $I_S \otimes |i\rangle\langle i|_{P_i}$ such that the following relation holds

$$\mathrm{Tr}\{\Pi_{\Lambda_i}(\rho_S \otimes |\bar{0}\rangle\langle\bar{0}|_{PM})\} = \mathrm{Tr}\{\Lambda_i \rho_S\}, \tag{3.3}$$

where

$$\Pi_{\Lambda_i} \equiv (U^{(i)}_{SP_i})^{\dagger}(I_S \otimes |i\rangle\langle i|_{P_i})U^{(i)}_{SP_i}. \tag{3.4}$$

Observe that the operator Π_{Λ_i} is a Hermitian projector, so that Sen's bound applies to each of these operators. Then, (3.1) follows from theorems 2.3 and 2.1. ∎

Remark 3.2. Because lemma 3.1 applies for general measurements, it can be used in the context of sections 3 and 4 of Sen [13] without the need for constructing a particular kind of 'intersection projector' as is performed there.

4. Sequential decoding in the one-shot regime

This section provides a proof for one of our main results: that a sequential decoding strategy works well even in the one-shot regime. More specifically, the theorem bounds the ε-one-shot classical capacity of a classical-quantum channel, defined operationally as the maximum number of bits that a sender can transmit to a receiver using such a channel with a failure probability no larger than ε. The general idea behind the proof is the same as that in the proof of theorem 1 of Wang & Renner [21], with the exception that we use a sequential decoding strategy and use lemma 3.1 to bound the error probability of this decoding strategy.

Theorem 4.1. *A sequential decoding strategy leads to the following bound on the ε-one-shot classical capacity $C^{\varepsilon}(W)$ of a classical-quantum channel $W : x \to \rho_x$:*

$$C^{\varepsilon}(W) \geq \max_{p_X} D^{\varepsilon'}_H(\rho_{XB}\|\rho_X \otimes \rho_B) - \log_2\left(\frac{1}{\varepsilon^2/4 - \varepsilon'}\right), \tag{4.1}$$

for some ε' such that $\varepsilon^2/4 > \varepsilon'$, where ρ_{XB} is the following classical-quantum state that depends on the distribution $p_X(x)$ and the channel W:

$$\rho_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes (\rho_x)_B. \tag{4.2}$$

Proof. Fix $\varepsilon \geq 0$ and a distribution $p_X(x)$. Let Q_{XB} be an operator such that $0 \leq Q_{XB} \leq I_{XB}$ and

$$\mathrm{Tr}\{Q_{XB}\rho_{XB}\} \geq 1 - \varepsilon', \tag{4.3}$$

where ε' is chosen as in the statement of the theorem. We generate a codebook by choosing its codewords x_j at random, each independently according to $p_X(x)$. Let A_{x_j} denote the following operator:

$$A_{x_j} \equiv \mathrm{Tr}_X\{(|x_j\rangle\langle x_j|_X \otimes I_B)Q_{XB}\}. \tag{4.4}$$

From theorem 2.1, we know that to each A_{x_j} there is associated a qubit probe system P_j, a unitary U_{BP_j} and a projector $I_B \otimes |1\rangle\langle 1|_{P_j}$ such that for every state σ

$$\mathrm{Tr}\{A_{x_j}\sigma\} = \mathrm{Tr}\{(U^{(j)}_{BP_j})^{\dagger}(I_B \otimes |1\rangle\langle 1|_{P_j})U^{(j)}_{BP_j}(\sigma_B \otimes |0\rangle\langle 0|_{P_j})\}. \tag{4.5}$$

Furthermore, it follows that for the complementary operator $I - A_{x_j}$, we have the following relation:

$$\mathrm{Tr}\{(I - A_{x_j})\sigma\} = \mathrm{Tr}\{(U^{(j)}_{BP_j})^{\dagger}(I_B \otimes |0\rangle\langle 0|_{P_j})U^{(j)}_{BP_j}(\sigma_B \otimes |0\rangle\langle 0|_{P_j})\}. \tag{4.6}$$

(Because $\{I - A_{x_j}, A_{x_j}\}$ is a two-outcome POVM, the unitary operator $U^{(j)}_{BP_j}$ can have the form given in example 2.2.)

For a specific codebook $\{x_j\}_{j\in[M]}$, the decoding strategy of the receiver Bob is as follows. Suppose that the sender Alice wishes to transmit message m, so that she transmits codeword x_m over the channel W. Then, the state at the receiver is ρ_{x_m}. The receiver first appends M ancillas, each set to $|0\rangle$, to the state ρ_{x_m} received. Then, the state at the receiving end is as follows:

$$\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M}. \tag{4.7}$$

Bob then checks whether the codeword transmitted by Alice is the first codeword. He does so by performing the unitary $U_{BP_1}^{(1)}$ corresponding to the first POVM $\{I - A_{x_1}, A_{x_1}\}$, and the state becomes

$$U_{BP_1}^{(1)} (\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})(U_{BP_1}^{(1)})^\dagger. \tag{4.8}$$

He then measures the probe system P_1 in the computational basis $\{|0\rangle\langle0|_{P_1}, |1\rangle\langle1|_{P_1}\}$. If he obtains the outcome $|1\rangle$, then he decodes that the first message was sent (in this case, there would be an error if $m \neq 1$). Otherwise, he performs the inverse of $U_{BP_1}^{(1)}$. At this point, if $m \neq 1$ and if there is no error, the subnormalized state becomes

$$(U_{BP_1}^{(1)})^\dagger (I_B \otimes |0\rangle\langle0|_{P_1})U_{BP_1}^{(1)} (\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})(U_{BP_1}^{(1)})^\dagger (I_B \otimes |0\rangle\langle0|_{P_1})(U_{BP_1}^{(1)}). \tag{4.9}$$

Making the abbreviations

$$I - \Pi_{x_1} \equiv (U_{BP_1}^{(1)})^\dagger (I_B \otimes |0\rangle\langle0|_{P_1})U_{BP_1}^{(1)} \tag{4.10}$$

and

$$\Pi_{x_1} \equiv (U_{BP_1}^{(1)})^\dagger (I_B \otimes |1\rangle\langle1|_{P_1})U_{BP_1}^{(1)}, \tag{4.11}$$

we can write the above subnormalized state as

$$(I - \Pi_{x_1})(\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})(I - \Pi_{x_1}). \tag{4.12}$$

The receiver then continues by performing similar actions to determine whether the transmitted codeword was the second one. That is, he performs the unitary $U_{BP_2}^{(2)}$ corresponding to A_{x_2}, measures the probe system P_2 in the computational basis and inverts the unitary $U_{BP_2}^{(2)}$ if he does not receive the outcome $|1\rangle$ from the measurement of P_2.

The success probability of this sequential decoding procedure when the mth codeword is sent is equal to

$$\text{Tr}\{\Pi_{x_m}(I - \Pi_{x_{m-1}})\cdots(I - \Pi_{x_1})(\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})(I - \Pi_{x_1})\cdots(I - \Pi_{x_{m-1}})\Pi_{x_m}\}. \tag{4.13}$$

Thus, the error probability is given by

$$1 - \text{Tr}\{\Pi_{x_m}(I - \Pi_{x_{m-1}})\cdots(I - \Pi_{x_1})(\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})(I - \Pi_{x_1})\cdots(I - \Pi_{x_{m-1}})\Pi_{x_m}\}$$
$$= \text{Tr}\{(\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})\} - \text{Tr}\{\Pi_{x_m}(I - \Pi_{x_{m-1}})\cdots(I - \Pi_{x_1})(\rho_{x_m} \otimes |\overline{0}\rangle\langle\overline{0}|_{P^M})$$
$$\times (I - \Pi_{x_1})\cdots(I - \Pi_{x_{m-1}})\Pi_{x_m}\}. \tag{4.14}$$

We can then upper bound this error probability by using lemma 3.1:

$$\leq 2\sqrt{\text{Tr}\{(I - A_{x_m})\rho_{x_m}\} + \sum_{j=1}^{m-1}\text{Tr}\{A_{x_j}\rho_{x_m}\}} \tag{4.15}$$

and

$$\leq 2\sqrt{1 - \text{Tr}\{A_{x_m}\rho_{x_m}\} + \sum_{j\neq m}\text{Tr}\{A_{x_j}\rho_{x_m}\}}. \tag{4.16}$$

Taking the expectation of the error with respect to all codebooks (but keeping the codeword x_m fixed) and exploiting concavity of the square-root function, this upper bound becomes

$$2\sqrt{1 - \text{Tr}\{A_{x_m}\rho_{x_m}\} + (M - 1)\sum_{x'}p_X(x')\text{Tr}\{A_{x'}\rho_{x_m}\}}. \tag{4.17}$$

Taking the expectation of the error with respect to the codeword x_m itself (and again exploiting concavity), this upper bound becomes

$$2\sqrt{1 - \sum_x p_X(x)\text{Tr}\{A_x\rho_x\} + (M-1)\text{Tr}\left\{\sum_{x'} p_X(x')A_{x'}\sum_x p_X(x)\rho_x\right\}}.$$ (4.18)

Using the facts that

$$\sum_x p_X(x)\text{Tr}\{A_x\rho_x\} = \text{Tr}\{Q_{XB}\rho_{XB}\}$$ (4.19)

and

$$\text{Tr}\left\{\sum_{x'} p_X(x')A_{x'}\sum_x p_X(x)\rho_x\right\} = \text{Tr}\{Q_{XB}(\rho_X \otimes \rho_B)\},$$ (4.20)

we can write the upper bound in (4.18) as

$$2\sqrt{1 - \text{Tr}\{Q_{XB}\rho_{XB}\} + (M-1)\text{Tr}\{Q_{XB}(\rho_X \otimes \rho_B)\}}$$ (4.21)

and

$$\leq 2\sqrt{\varepsilon' + M\text{Tr}\{Q_{XB}(\rho_X \otimes \rho_B)\}}$$ (4.22)

Let $R = \log_2(M)$. By optimizing the choice of the operator Q_{XB} with respect to the hypothesis testing relative entropy defined in (2.2) and (2.3), we find the following upper bound on the error ε:

$$2\sqrt{\varepsilon' + 2^{-[D_H^{\varepsilon'}(\rho_{XB}\|\rho_X\otimes\rho_B)-R]}}.$$ (4.23)

Because we proved an upper bound on the expectation of the average error probability with respect to the codebook choice, we can conclude that there exists at least one code with the above bound on its average error probability. Rewriting this upper bound on ε, we find that the sequential decoding scheme gives the following bound on the ε-one-shot capacity of W:

$$C^\varepsilon(W) \geq \max_{p_X} D_H^{\varepsilon'}(\rho_{XB}\|\rho_X \otimes \rho_B) - \log_2\left(\frac{1}{\varepsilon^2/4 - \varepsilon'}\right).$$ (4.24)

∎

Remark 4.2. We recover the Holevo rate for communication by considering a memoryless classical-quantum channel and evaluating a limit as the number of channel used tends to infinity. We do not discuss this point any further here, because Wang & Renner [21] already discussed it in detail.

Remark 4.3. Of course, it is not actually necessary to use M ancillas when decoding. After performing each measurement, the receiver could store the result in a classical memory and simply refresh a single ancilla to the state $|0\rangle$.

Remark 4.4. The proof of the above theorem and lemma 3.1 make it clear that one can always use Sen's bound in the error analysis for any random coding classical communication scheme of the above form, thus serving as a substitute for the well-known bound in lemma 2 of Hayashi & Nagaoka [19]. However, the performance is slightly worse than that obtained with the Hayashi–Nagaoka bound owing to the square root on the right-hand side of Sen's bound (one can see this explicitly by comparing theorem 4.1 with theorem 1 of [21]).

Remark 4.5. The operation of the sequential decoder is similar in spirit to the conditional pulse nulling receiver introduced in Guha *et al.* [26] and experimentally implemented in Chen *et al.* [27], in the sense that it proceeds by performing a unitary operation, a projection and the inverse of the unitary for every codeword in the codebook.

Remark 4.6. We can also use sequential decoding for a task known as one-shot classical data compression with quantum side information [28–30]. In such a task, the sender and receiver are given a classical-quantum state of the form $\sum_x p_X(x)|x\rangle\langle x|_X \otimes (\rho_x)_B$, where the sender has the

system X and the receiver the system B. The goal is for the sender to transmit as few classical bits as possible to the receiver, such that he can recover the register X up to a failure probability no larger than some $\varepsilon > 0$. In this case, we can show that the number of bits that need to be sent is related to the conditional hypothesis testing entropy [25,30], defined as

$$H_H^\varepsilon(X \mid B)_\rho \equiv \max_{\sigma_B} -D_H^\varepsilon(\rho_{XB} \| I_X \otimes \sigma_B),$$

by exploiting the same kind of proof as given in Renes & Renner [29] and Tomamichel & Hayashi [30] combined with our proof given above. The use of sequential decoding in the IID setting for this task was first performed in §4 of Wilde *et al.* [31].

(a) Performing sequential decoding coherently

We can also consider a fully coherent implementation of the sequential decoding strategy (i.e. with unitary operations alone). For the sake of simplicity, let $|\psi\rangle$ denote the state on which the coherent sequential decoding operations will act. As before, the procedure begins by the receiver appending M probe ancillas, so that the state becomes

$$|\psi\rangle_B \otimes |\bar{0}\rangle_{PM}. \tag{4.25}$$

The receiver first performs the unitary $U_{BP_1}^{(1)}$ corresponding to the first codeword, leading to

$$U_{BP_1}^{(1)} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM}. \tag{4.26}$$

Rather than perform an incoherent projection of the probe, the receiver can perform a controlled-NOT operation from the first probe system to another ancillary system A_1 initialized in the state $|0\rangle_{A_1}$. This leads to the state:

$$(I_B \otimes |0\rangle\langle 0|_{P_1}) U_{BP_1}^{(1)} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |0\rangle_{A_1} + (I_B \otimes |1\rangle\langle 1|_{P_1}) U_{BP_1}^{(1)} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |1\rangle_{A_1}. \tag{4.27}$$

The receiver then performs the inverse unitary $(U_{BP_1}^{(1)})^\dagger$, and by using the shorthand in (4.11) and (4.10), we can write the resulting state as

$$(I - \Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |0\rangle_{A_1} + (\Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |1\rangle_{A_1}. \tag{4.28}$$

Continuing a similar procedure for the second codeword leads to the expansion:

$$(I - \Pi_{x_2})_{BP_2}(I - \Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |00\rangle_{A_1 A_2} + (\Pi_{x_2})_{BP_2}(I - \Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |01\rangle_{A_1 A_2}$$
$$+ (I - \Pi_{x_2})_{BP_2}(\Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |10\rangle_{A_1 A_2} + (\Pi_{x_2})_{BP_2}(\Pi_{x_1})_{BP_1} |\psi\rangle_B \otimes |\bar{0}\rangle_{PM} \otimes |11\rangle_{A_1 A_2}, \tag{4.29}$$

and so forth.

5. Gentle sequential measurements

Winter's gentle operator lemma has found numerous applications in quantum information theory [22,32].[3] It states that if a two-outcome measurement has one outcome that occurs with high probability, then the subnormalized post-measurement state is close to the original state. More formally,

Lemma 5.1 (gentle operator). *Let ρ be a state, and let Λ be an operator such that $0 \leq \Lambda \leq I$. Then*

$$\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\|_1 \leq 2\sqrt{\mathrm{Tr}\{(I - \Lambda)\rho\}}. \tag{5.1}$$

Thus, if $\mathrm{Tr}\{\Lambda\rho\} \geq 1 - \varepsilon$ for some small $\varepsilon \geq 0$, then

$$\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\|_1 \leq 2\sqrt{\varepsilon}. \tag{5.2}$$

---

[3]In quantum complexity theory, there is a similar lemma known as the 'almost as good as new lemma' discovered independently by Aaronson [33].

Of course, lemma 5.1 can be extended with the Naimark extension theorem as well. Suppose that we know that

$$\text{Tr}\{\Lambda\rho\} \geq 1 - \varepsilon, \tag{5.3}$$

for a two-outcome POVM $\{\Lambda, I - \Lambda\}$. By theorem 2.1, we know that there exists a unitary U_{SP} such that for the orthonormal basis $\{|0\rangle_P, |1\rangle_P\}$, we have that

$$\text{Tr}\{\Pi_{SP}(\rho_S \otimes |0\rangle\langle 0|_P)\} = \text{Tr}\{\Lambda\rho\}, \tag{5.4}$$

where

$$\Pi_{SP} \equiv U_{SP}^\dagger(I_S \otimes |0\rangle\langle 0|_P)U_{SP}. \tag{5.5}$$

Thus, if we perform the unitary U_{SP}, the projective measurement $\{|0\rangle\langle 0|_P, |1\rangle\langle 1|_P\}$, followed by the inverse unitary $U_{SP}^\dagger$, we can conclude from lemma 5.1 that the resulting state of both the system and the probe is close to the original state:

$$\|\rho_S \otimes |0\rangle\langle 0|_P - \Pi_{SP}(\rho_S \otimes |0\rangle\langle 0|_P)\Pi_{SP}\|_1 \leq 2\sqrt{\varepsilon}. \tag{5.6}$$

So, for the sake of simplicity, in what follows, we just consider the state ρ to be the state of the combined system and any necessary ancillas so that we can consider projective measurements only (this is due to the above observation and the Naimark extension theorem).

In a sequential decoding scheme, we also might like to conclude that the state after the decoding procedure is close to the original state. This would be pleasing conceptually and would also have applications in constructing decoders for quantum data from decoders for classical data [23,34,35]. However, as noted in [13], we cannot generally make the above conclusion. Here, we show how performing additional operations leads to a state close to the original one.

There are at least two ways that we can perform additional operations in order to guarantee that the sequentially decoded state is close to the original one. The first was mentioned at the end of §4.3 of Wilde *et al.* [31] and relies on the polar decomposition. Given that the post-measurement state is of the following form (omitting normalization):

$$\Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N, \tag{5.7}$$

the receiver could perform a unitary V given by the polar decomposition

$$\sqrt{\Pi_1 \cdots \Pi_N \cdots \Pi_1} = V\Pi_N \cdots \Pi_1 \tag{5.8}$$

so that the post-measurement state becomes

$$\sqrt{\Pi_1 \cdots \Pi_N \cdots \Pi_1}\rho\sqrt{\Pi_1 \cdots \Pi_N \cdots \Pi_1}. \tag{5.9}$$

In this case, we can apply the gentle operator lemma (lemma 5.1) to upper bound the disturbance:

$$\|\rho - \sqrt{\Pi_1 \cdots \Pi_N \cdots \Pi_1}\rho\sqrt{\Pi_1 \cdots \Pi_N \cdots \Pi_1}\|_1 \leq 2\sqrt{\text{Tr}\{(I - \Pi_1 \cdots \Pi_N \cdots \Pi_1)\rho\}} \tag{5.10}$$

and

$$= 2\sqrt{1 - \text{Tr}\{\Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N\}}, \tag{5.11}$$

and then once again apply Sen's non-commutative union bound (theorem 2.3) to upper bound the disturbance as

$$2\sqrt{2} \sqrt[4]{\sum_{i=1}^{N} \text{Tr}\{(I - \Pi_i)\rho\}}. \tag{5.12}$$

(In the context of classical-quantum polar codes, a quantity such as the above will be exponentially small in the number of channel uses because each term is exponentially small while there are only a linear number of terms [14].)

One practical problem with the above approach is as follows. Suppose that we can efficiently implement each of the measurements corresponding to the projections $\Pi_1, \ldots, \Pi_N$ (say, on a quantum computer). Then, we can clearly perform the sequential decoding procedure efficiently if N is not too large. On the other hand, given a particular sequence of measurements, it

is not clear at all that the unitary given by the polar decomposition in (5.8) has an efficient implementation. Thus, the above approach does not realize both desirable requirements of having a small disturbance and an efficient sequential decoding (if each of the measurements can be efficiently implemented to begin with).

There is a simple way to remedy the aforementioned problem if the sequential decoding strategy has a very small error probability. We can simply perform the projections Π_1 through Π_N and then perform them again in the opposite order. This gives the subnormalized post-measurement state

$$\Pi_1 \cdots \Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N \cdots \Pi_1, \tag{5.13}$$

and the error probability is bounded as follows, again by applying Sen's bound (theorem 2.3):

$$\mathrm{Tr}\{\rho\} - \mathrm{Tr}\{\Pi_1 \cdots \Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N \cdots \Pi_1\} \le 2\sqrt{2\sum_{i=1}^{N} \mathrm{Tr}\{(I - \Pi_i)\rho\}}. \tag{5.14}$$

Thus, by performing the measurements again in reverse, we increase only the error probability by a factor of $\sqrt{2}$. Furthermore, the gentle operator lemma (lemma 5.1) gives the following upper bound on the disturbance:

$$\|\rho - \Pi_1 \cdots \Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N \cdots \Pi_1\|_1 \le 2\sqrt{\mathrm{Tr}\{(I - [\Pi_1 \cdots \Pi_N \cdots \Pi_1]^2)\rho\}} \tag{5.15}$$

$$= 2\sqrt{1 - \mathrm{Tr}\{\Pi_1 \cdots \Pi_N \cdots \Pi_1 \rho \Pi_1 \cdots \Pi_N \cdots \Pi_1\}}. \tag{5.16}$$

Applying the bound in (5.14) gives the following upper bound on the disturbance:

$$2\sqrt{2}(\sqrt[4]{2})\sqrt[4]{\sum_{i=1}^{n} \mathrm{Tr}\{(I - \Pi_i)\rho\}}. \tag{5.17}$$

Thus, with this scheme, we can realize both requirements of having an efficient implementation and a small disturbance—the receiver simply has to perform $2N$ measurements (each of which were assumed to have an efficient implementation) while the disturbance increases only by a factor of $\sqrt[4]{2}$. An efficient coherent implementation of these operations follows from the discussion in §4a, if each measurement has an efficient implementation.

By the methods of Wilde & Renes [23], this latter approach will be useful for decoding quantum polar codes, should a method ever be found to realize an efficient decoder for classical-quantum polar codes [14] (see [36] for progress in this direction). At the very least, this latter approach answers an open question from Wilde & Renes [23].

References

1. Holevo AS. 1973 Bounds for the quantity of information transmitted by a quantum communication channel. *Problems Inf. Transm.* **9**, 177–183.
2. Holevo AS. 1998 The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory* **44**, 269–273. (doi:10.1109/18.651037)
3. Schumacher B, Westmoreland MD. 1997 Sending classical information via noisy quantum channels. *Phys. Rev. A* **56**, 131–138. (doi:10.1103/PhysRevA.56.131)
4. Hastings MB. 2009 Superadditivity of communication capacity using entangled inputs. *Nat. Phys.* **5**, 255–257. (doi:10.1038/nphys1224)

5. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423.

6. Belavkin V. 1975 Optimal distinction of non-orthogonal quantum signals. *Radio Eng. Electron. Phys.* **20**, 39–47.

7. Belavkin V. 1975 Optimal multiple quantum statistical hypothesis testing. *Stochastics* **1**, 315–345. (doi:10.1080/17442507508833114)

8. Hausladen P, Jozsa R, Schumacher B, Westmoreland M, Wootters WK. 1996 Classical information capacity of a quantum channel. *Phys. Rev. A* **54**, 1869–1876. (doi:10.1103/PhysRevA.54.1869)

9. Giovannetti V, Guha S, Lloyd S, Maccone L, Shapiro JH, Yuen HP. 2004 Classical capacity of the lossy bosonic channel: the exact solution. *Phys. Rev. Lett.* **92**, 027902. (doi:10.1103/PhysRevLett.92.027902)

10. Lloyd S, Giovannetti V, Maccone L. 2011 Sequential projective measurements for channel decoding. *Phys. Rev. Lett.* **106**, 250501. (doi:10.1103/PhysRevLett.106.250501)

11. Giovannetti V, Lloyd S, Maccone L. 2012 Achieving the Holevo bound via sequential measurements. *Phys. Rev. A* **85**, 012302. (doi:10.1103/PhysRevA.85.012302)

12. Cover TM, Thomas JA. 1991 *Elements of information theory*. London, UK: Wiley-Interscience.

13. Sen P. 2011 Achieving the Han–Kobayashi inner bound for the quantum interference channel by sequential decoding. See http://arxiv.org/abs/1109.0802.

14. Wilde MM, Guha S. 2013 Polar codes for classical-quantum channels. *IEEE Trans. Inf. Theory* **59**, 1175–1187. (doi:10.1109/TIT.2012.2218792)

15. Wilde MM, Guha S, Tan S-H, Lloyd S. 2012 Explicit capacity-achieving receivers for optical communication and quantum reading. In *Proc. the 2012 Int. Symp. on Information Theory*, pp. 551–555. Boston, MA. See http://arxiv.org/abs/1202.0518.

16. Renner R. 2005 Security of quantum key distribution. PhD thesis, ETH Zurich, Switzerland.

17. Datta N, Hsieh M-H. 2011 The apex of the family tree of protocols: optimal rates and resource inequalities. *New J. Phys.* **13**, 093042. (doi:10.1088/1367-2630/13/9/093042)

18. Tomamichel M. 2012 A framework for non-asymptotic quantum information theory. PhD thesis, ETH Zurich, Switzerland.

19. Hayashi M, Nagaoka H. 2003 General formulas for capacity of classical-quantum channels. *IEEE Trans. Inf. Theory* **49**, 1753–1768. (doi:10.1109/TIT.2003.813556)

20. Mosonyi M, Datta N. 2009 Generalized relative entropies and the capacity of classical-quantum channels. *J. Math. Phys.* **50**, 072104. (doi:10.1063/1.3167288)

21. Wang L, Renner R. 2012 One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.* **108**, 200501. (doi:10.1103/PhysRevLett.108.200501)

22. Winter A. 1999 Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory* **45**, 2481–2485. (doi:10.1109/18.796385)

23. Wilde MM, Renes JM. 2012 Quantum polar codes for arbitrary channels. In *Proc. 2012 Int. Symp. on Information Theory*, pp. 334–338. Boston, MA.

24. Buscemi F, Datta N. 2010 The quantum capacity of channels with arbitrarily correlated noise. *IEEE Trans. Inf. Theory* **56**, 1447–1460. (doi:10.1109/TIT.2009.2039166)

25. Dupuis F, Kraemer L, Faist P, Renes JM, Renner R. 2012 Generalized entropies. See http://arxiv.org/abs/1211.3141.

26. Guha S, Habif JL. Takeoka M. 2011 Approaching Helstrom limits to optical pulse-position demodulation using single photon detection and optical feedback. *J. Mod. Opt.* **58**, 257–265. (doi:10.1080/09500340.2010.533204)

27. Chen J, Habif JL, Dutton Z, Lazarus R, Guha S. 2012 Optical codeword demodulation with error rates below the standard quantum limit using a conditional nulling receiver. *Nat. Photonics* **6**, 374–379. (doi:10.1038/nphoton.2012.113)

28. Devetak I, Winter A. 2003 Classical data compression with quantum side information. *Phys. Rev. A* **68**, 042301. (doi:10.1103/PhysRevA.68.042301)

29. Renes JM, Renner R. 2012 One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys. *IEEE Trans. Inf. Theory* **58**, 1985–1991. (doi:10.1109/TIT.2011.2177589)

30. Tomamichel M, Hayashi M. 2012 A hierarchy of information quantities for finite block length analysis of quantum tasks. See http://arxiv.org/abs/1208.1478.

31. Wilde MM, Hayden P, Buscemi F, Hsieh M-H. 2012 The information-theoretic costs of simulating quantum measurements. *J. Phys. Math. Theor.* **45**, 453001. (doi:10.1088/1751-8113/45/45/453001)

32. Ogawa T, Nagaoka H. 2007 Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Trans. Inf. Theory* **53**, 2261–2266. (doi:10.1109/TIT.2007.896874)
33. Aaronson S. 2005 Limitations of quantum advice and one-way communication. *Theory Comp.* **1**, 1–28. (doi:10.4086/toc.2005.v001a001)
34. Devetak I. 2005 The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory* **51**, 44–55. (doi:10.1109/TIT.2004.839515)
35. Renes JM, Boileau J-C. 2008 Physical underpinnings of privacy. *Phys. Rev. A* **78**, 032335. (doi:10.1103/PhysRevA.78.032335)
36. Wilde MM, Landon-Cardinal O, Hayden P. 2013 Towards efficient decoding of classical-quantum polar codes. See http://arxiv.org/abs/1302.0398.