Achieving the Holevo Capacity of a Pure State Classical-Quantum Channel via Unambiguous State Discrimination

Masahiro Takeoka*†, Hari Krovi†, and Saikat Guha†

* *Quanutm ICT Laboratory, National Institute of Information and Communications Technology, Koganei, Tokyo, Japan*
† *Quantum Information Processing Group, Raytheon BBN Technologies, Cambridge, Massachusetts, USA*
Email: takeoka@nict.go.jp, hkrovi@bbn.com, sguha@bbn.com

***Abstract*—We prove that the ultimate channel capacity, the Holevo bound, for sending classical data on a quantum channel (the so-called *classical-quantum*, or cq channel) can be achieved for a pure-state cq channel by decoding codewords via a collective quantum measurement based on *unambiguous state discrimination* (USD). In cq communication theory, the channel decoder acts directly on the modulated codeword waveform in the quantum (viz., electromagnetic or optical) domain, and it is known that collective measurements on long codeword blocks are needed to attain the Holevo capacity, which is strictly larger than the Shannon capacity of the classical channel induced by any specific measurement choice on each channel use. The USD measurement based channel decoder we propose, can distinguish finite blocklength codeword quantum states unambiguously (i.e., an incorrect codeword is never chosen) provided one allows for a finite probability of obtaining an inconclusive (erasure) outcome. We show that the probability of the inconclusive outcome goes to zero for asymptotically long codewords whenever the code rate is below the Holevo bound. The USD channel decoder is an addition to a small list of other collective measurements known to achieve the Holevo capacity (such as, the *square root measurement*, the *minimum probability of error* measurement, the *sequential decoding* measurement, and the quantum *successive cancellation decoder* for the cq polar code). A structured optical receiver design is not known yet for any of these decoders. What makes the USD decoder special is that there is no classical analogue to truly unambiguous discrimination (say, of samples drawn from a set of probability distributions). Secondly, the erasures-only decoding of USD is likely to result in a better channel reliability function. Finally, the USD measurement seems more likely to lead naturally to a structured optical receiver design and implementation.**

I. Introduction

The Shannon capacity gives the maximum rate of reliable information transmission over a noisy channel, where the channel is described by a transition probability map $P_{Y|X}$. For a *classical-quantum, or cq* channel, i.e., when classical information is sent over a quantum channel (the modulation symbols are quantum states), the maximum rate of reliable information transmission is upper bounded by the *Holevo bound*. The achievability of the Holevo bound was first proven by Holevo, Schumacher and Westmoreland (HSW) [1], [2], [3]. Thus, the Holevo bound is synonymous to *Holevo capacity*. In the HSW proof, Shannon's channel-capacity proof was generalized to the classical-quantum channel by introducing a notion of the typical subspace [4] and an appropriate quantum measurement for the decoder, known as the *square-root measurement* (SRM) or "pretty good measurement" [5] to discriminate the quantum states of codewords of a randomly generated codebook.

Quantum state discrimination is one of the fundamental topics in quantum information theory. Since the nonorthogonality of quantum states fundamentally prohibits perfect discrimination, much attention has been paid to investigate measurements to discriminate optimally among non-orthogonal states, modulo a given criteria, such as the average probability of error [6]. Two different strategies have been investigated so far. The first strategy is to maximize the probability of successful detection (or, equivalently minimizing the probability of error) [7]. The SRM achieves the minimum probability of error, if the states being discriminated have certain symmetry properties [8]. The decoding the cq channel, even though the SRM does not achieve minimum error discrimination of the codewords of a random codebook, it is sufficiently good—hence the name "pretty good"—to achieve the Holevo capacity [1], [2], [3]. It is therefore easy to see that the actual minimum probability of error measurement on a random code achieves the Holevo capacity. In recent years, two more measurements for the channel decoder were shown to meet the Holevo capacity. The first one, the *sequential decoding measurement* is somewhat analogous to jointly-typical decoding in classical information theory, which 'checks' each codeword one at a time via a sequence of 2^{nR} 'yes-no' binary-outcome quantum non-demolition measurements [9], [10], [11]. Secondly, *quantum polar codes*, were shown to achieve the Holevo capacity of any cq channel, which are the first near-explicit (and, linear) Holevo capacity achieving codes [12]. The *quantum successive cancellation* decoder of the polar code decodes one bit at a time via a sequence of nR binary-outcome quantum non-demolition collective measurements on the received codeword.

In this paper, we demonstrate yet another measurement strategy, known as *unambiguous state discrimination* (USD), which achieves the Holevo capacity for any pure state cq

This material is based upon work supported by the Defense Advanced Research Projects Agency's (DARPA) Information in a Photon (InPho) program, under Contract No. HR0011-10-C-0159.

channel. The notion of USD was first addressed in [13], [14], [15]. They showed that it is possible to discriminate two nonorthogonal states unambiguously (i.e., without even making a wrong guess), provided one allows for a finite probability of obtaining an inconclusive measurement result. Approximately unambiguous discrimination of overlapping probability distributions is commonplace in classical detection theory, where a guard-band is imposed in the overlap region, and the detector makes a guess only when the data falls within a certain confidence interval. However, true USD for non-orthogonal signals is a truly quantum phenomenon grounded in the quantum mechanical description of the signal states. USD has hence been extensively studied over the last decade from a viewpoint of fundamentals of quantum mechanics [6], [16], [17], [18], [19], [20], [21], [22], as well as for its applications to quantum information protocols such as quantum cryptography [23], [24], [25].

Our main contribution in this paper is a proof that USD measurement achieves the Holevo capacity of any pure-state cq channel. Our proof may lead to new insights in quantum Shannon theory, since unlike the other measurement strategies known to achieve the Holevo capacity, USD does not have any classical analogue. From the practical point of view, our proof may provide an alternative way to tackle the experimental implementation of channel decoding for optical communication at rates approaching the Holevo capacity, since optical implementations of USD measurements have been investigated theoretically for some special cases [23], and have also been demonstrated experimentally [6].

II. Unambiguous state discrimination

Consider the measurement discriminating a set of pure quantum states $\{|\phi_j\rangle\}_{1\le j\le J}$ occurring with a prior probability distribution $\{p(j)\}$. A quantum measurement is generally described by a positive operator-valued measure (POVM) consisting of a set of measurement operators $\{\Pi_k\}$ where $\Pi_k \ge 0$ and $\sum_k \Pi_k = I$. The POVM for unambiguous state discrimination is defined by a set of $J+1$ measurements where for Π_k with $1 \le k \le J$, one can identify $|\phi_j\rangle$ without errors, i.e. $\langle\phi_j|\Pi_k|\phi_j\rangle = p_j\delta_{jk}$. On the other hand, if one obtains the outcome $\Pi_{J+1} = I - \sum_{j=1}^{J}\Pi_j$, the information about the state completely disappears (an *erasure* outcome). Here δ_{jk} is a Kronecker delta function and p_j corresponds to the success probability of detecting $|\phi_j\rangle$.

The equal probability measurement (EPM) is a special case of the USD measurement with equal success probabilities for each state, which was proved to be optimal if the states have certain symmetry properties [17], [18], [19]. The measurement operators of the EPM are given by:

$$\Pi_j = p|\tilde{\phi}_j\rangle\langle\tilde{\phi}_j| \quad (1 \le j \le J), \tag{1}$$

and $\Pi_{J+1} = I - \sum_{j=1}^{J}\Pi_j$, where

$$|\tilde{\phi}_j\rangle = \left(\sum_i |\phi_i\rangle\langle\phi_i|\right)^{-1} |\phi_j\rangle, \tag{2}$$

$0 \le p \le 1$, and the superscript -1 is the Moore-Penrose inverse [26]. The success probability of unambiguously detecting $|\phi_j\rangle$ is p for all j where the maximum p under the condition $\sum_{j=1}^{J+1}\Pi_j = I$ is given by the minimum eigenvalue of $\sum_i |\phi_i\rangle\langle\phi_i|$, or equivalently that of the Gram matrix $[\langle\phi_i|\phi_j\rangle]$. Note that generally USD is possible only if the states are linearly independent [17]. Otherwise, we always have $p = 0$. In the next section, we will use the EPM based USD for cq channel decoding.

III. Channel decoding via the USD measurement

For a set of quantum states $\{p(x), \rho_x\}$, $x \in X$, with $\rho = \sum_x p(x)\rho_x$, the Holevo bound is given by

$$I(X;Y) \le S(\rho) - \sum_x p(x)S(\rho_x), \tag{3}$$

where $I(X;Y)$ is the classical mutual information between the input X and the classical outcome Y upon measuring the output state ρ_x by a measurement with POVM $\{\Pi_y\}$, which induces a classical DMC with transition probabilities $p_{Y|X}(y|x) = \mathrm{Tr}(\rho_x\Pi_y)$. $S(\rho) = -\mathrm{Tr}[\rho\log\rho]$ is the von Neumann entropy of the quantum state ρ. When ρ_x are pure states, the right hand of Eq. (3) simply reduces to $S(\rho)$, since the entropy of pure states is zero.

We consider here a pure-state cq channel $x \to |\psi_x\rangle$, with the state ensemble $\{p(x), |\psi_x\rangle\}$ at the receiver. The average single-symbol density operator at the receiver is thus given by

$$\rho = \sum_x p(x)|\psi_x\rangle\langle\psi_x|, \tag{4}$$

with spectral decomposition, $\rho = \sum_z q_Z(z)|z\rangle\langle z|$. We then define the δ-typical projector for $\rho^{\otimes n}$ as

$$\Pi_\delta \equiv \sum_{z^n \in T_\delta^{Z^n}} |z^n\rangle\langle z^n|, \tag{5}$$

where $|z^n\rangle = |z_1\rangle|z_2\rangle\cdots$ and $T_\delta^{Z^n}$ is a δ-typical set of z^n. See Appendix for the properties of typical subspaces. We will now state the main result of this paper.

Theorem 1: A USD decoding measurement can achieve the Holevo bound for a pure state point-to-point cq channel.

The proof is split into three steps: encoding, decoding, and error analysis.

Encoding. Our codebook is constructed by random coding. Each messeage $m \in \mathcal{M} \equiv \{1, ..., M\}$ where $M = 2^{nR}$, is encoded in a codeword $x^n(m) \equiv x_1(m)\cdots x_n(m)$ which is randomly generated according to the probability distribution $p(x)$, and transmitted through the (pure state) channel. The quantum state of the received codeword is $|\psi_{x^n(m)}\rangle = |\psi_{x_1(m)}\rangle\cdots|\psi_{x_n(m)}\rangle$.

Decoding. Our decoding is based on the EPM. The EPM for codeword states $\{|\psi_{x^n(m)}\rangle\}$ $(m = 1, \cdots, M)$ is described by a POVM consisting of $M+1$ elements defined as

$$\Lambda_m = p\Lambda^{-1}|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|\Lambda^{-1} \quad (1 \le m \le M), \tag{6}$$

and $\Lambda_{M+1} = I - \sum_{m=1}^{M}\Lambda_m$, where $\Lambda = \sum_{m'=1}^{M}|\psi_{x^n(m')}\rangle\langle\psi_{x^n(m')}|$, and p is the minimum eigenvalue of the Gram matrix $G = [\langle\psi_{x^n(k)}|\psi_{x^n(l)}\rangle]$. The success probability of detecting $|\psi_{x^n(m)}\rangle$ is $\mathrm{Tr}[|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|\Lambda_m] = p$ for all $m \in \{1, \ldots, M\}$.

Error Analysis. As mentioned above, the success probability of the EPM decoding is given by the minimum eigenvalue of the Gram matrix. We will first analyze the eigenvalue distribution of G. That analysis will reveal that we need to slightly modify the EPM. We will then show that the modified EPM can achieve the Holevo bound.

First of all, note that,

$$\begin{aligned} G &= [\langle\psi_{x^n(k)}|\psi_{x^n(l)}\rangle] \\ &= [\langle\psi_{x^n(k)}|\Pi_\delta|\psi_{x^n(l)}\rangle] + [\langle\psi_{x^n(k)}|\Pi_\delta^\perp|\psi_{x^n(l)}\rangle] \\ &\equiv \tilde{G} + \tilde{G}^\perp, \end{aligned} \tag{7}$$

where $\Pi_\delta^\perp$ is a projector onto an atypical subspace. We can thus bound the eigenvalues of G as,

$$\begin{aligned} \lambda_i(G) &= \lambda_i(\tilde{G} + \tilde{G}^\perp) \geq \lambda_i(\tilde{G}) + \lambda_{\min}(\tilde{G}^\perp) \\ &\geq \lambda_i(\tilde{G}), \end{aligned} \tag{8}$$

where $i = 1, \cdots, M$, and we used Lemma 3 (see Appendix) and the fact that $\lambda_{\min}(\tilde{G}^\perp) \geq 0$. This implies that for the minimum eigenvalue analysis, we only need to investigate the distribution of $\lambda_i(\tilde{G})$. The mean of $\lambda_i(\tilde{G})$ over the choice of the random codebook $\mathcal{C}$ can be calculated as follows:

$$\begin{aligned} \bar{\mu} &= \mathbf{E}_{\mathcal{C}} \frac{1}{M} \sum_{i=1}^{M} \lambda_i(\tilde{G}) = \frac{1}{M} \mathbf{E}_{\mathcal{C}} \mathrm{Tr}\left[\tilde{G}\right] \\ &= \frac{1}{M} \mathbf{E}_{\mathcal{C}} \mathrm{Tr}\left[\sum_{m=1}^{M} \Pi_\delta |\psi_{x(m)}\rangle\langle\psi_{x(m)}| \Pi_\delta\right] \\ &= \mathrm{Tr}\left[\Pi_\delta \rho^n \Pi_\delta\right] \geq 1 - \epsilon_1, \end{aligned} \tag{9}$$

where the inequality follows from the property of the typical projector (see Appendix). Similarly, the variance is given by

$$\begin{aligned} \bar{\sigma}^2 &= \mathbf{E}_{\mathcal{C}} \frac{1}{M} \sum_{i=1}^{M} (\lambda_i - \bar{\mu})^2 \\ &= \mathbf{E}_{\mathcal{C}} \frac{1}{M} \mathrm{Tr}\left[\left(\tilde{G} - \bar{\mu} I\right)^2\right] \\ &= \frac{1}{M} \mathbf{E}_{\mathcal{C}} \Bigg\{ \sum_{m=1}^{M} \left(\langle\psi_{x^n(m)}|\Pi_\delta|\psi_{x^n(m)}\rangle - \bar{\mu}\right)^2 \\ &\qquad + \sum_{k=1}^{M} \sum_{l \neq k}^{M} \left|\langle\psi_{x^n(k)}|\Pi_\delta|\psi_{x^n(l)}\rangle\right|^2 \Bigg\}, \end{aligned} \tag{10}$$

where the first term above can be upper bounded as follows:

$$\begin{aligned} &= \frac{1}{M} \mathbf{E}_{\mathcal{C}} \sum_{m=1}^{M} \left(\mathrm{Tr}\left[|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|\Pi_\delta\right] - \bar{\mu}\right)^2 \\ &= \frac{1}{M} \sum_{m=1}^{M} \left(\mathbf{E}_{\mathcal{C}} \mathrm{Tr}\left[|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|\Pi_\delta\right]^2 - \bar{\mu}^2\right) \\ &\leq \frac{1}{M} \sum_{m=1}^{M} (1 - (1 - \epsilon_1)^2) \leq 2\epsilon_1, \end{aligned} \tag{11}$$

where we used Eq. (9) and $\mathrm{Tr}[|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|\Pi_\delta] \leq 1$. The second term can be upper bounded as follows:

$$\begin{aligned} &= \frac{1}{M} \mathbf{E}_{\mathcal{C}} \sum_{k=1}^{M} \sum_{l \neq k}^{M} \mathrm{Tr}\left[\Pi_\delta |\psi_{x^n(k)}\rangle\langle\psi_{x^n(k)}| \Pi_\delta |\psi_{x^n(l)}\rangle\langle\psi_{x^n(l)}|\right] \\ &= \frac{1}{M} \sum_{k=1}^{M} \sum_{l \neq k}^{M} \mathrm{Tr}\left[\Pi_\delta \rho^n \Pi_\delta \mathbf{E}_l |\psi_{x^n(l)}\rangle\langle\psi_{x^n(l)}|\right] \\ &\leq \frac{1}{M} 2^{-n(S(\rho)-\delta_1)} \sum_{k=1}^{M} \sum_{l \neq k}^{M} \mathrm{Tr}\left[\Pi_\delta \mathbf{E}_l |\psi_{x^n(l)}\rangle\langle\psi_{x^n(l)}|\right] \\ &\leq (M-1) 2^{-n(S(\rho)-\delta_1)} \\ &< 2^{nR}\, 2^{-n(S(\rho)-\delta_1)} \\ &= 2^{-n(S(\rho)-R-\delta_1)}, \end{aligned} \tag{12}$$

where the first inequality follows from the property of the typical projector (see Appendix). Consequently, we have

$$\bar{\sigma}^2 \leq 2\epsilon_1 + 2^{-n(S(\rho)-R-\delta_1)}. \tag{13}$$

Equations (9) and (13) suggest that on an average, most of the eigenvalues are distributed around 1 and the variance can be made arbitrarily small for large n, if $R < S(\rho) - \delta_1$. Let us now choose a particular codebook and consider its eigenvalue distribution. Let $\mu(\mathcal{C}) = M^{-1} \sum_{i=1}^{M} \lambda_i(\mathcal{C})$ and $\sigma^2(\mathcal{C}) = M^{-1} \sum_{i=1}^{M} (\lambda_i(\mathcal{C}) - \mu(\mathcal{C}))^2$ be the mean and the variance for codebook $\mathcal{C}$, respectively. ¿From Eqs. (9) and (13), it is clear that there must exist a codebook $\mathcal{C}^*$ such that

$$\mu(\mathcal{C}^*) - \sigma^2(\mathcal{C}^*) \geq \bar{\mu} - \bar{\sigma}^2 \geq 1 - 3\epsilon_1 - 2^{-n(S(\rho)-R-\delta_1)}. \tag{14}$$

Due to the fact that $\mu(\mathcal{C}^*) \leq 1$ and $\sigma^2(\mathcal{C}^*) \geq 0$, we have

$$\mu(\mathcal{C}^*) \geq 1 - 3\epsilon_1 - 2^{-n(S(\rho)-R-\delta_1)}, \tag{15}$$

$$\sigma^2(\mathcal{C}^*) \leq 3\epsilon_1 + 2^{-n(S(\rho)-R-\delta_1)}. \tag{16}$$

In what follows, we will consider this particular codebook $\mathcal{C}^*$. For notational simplicity, $\mu(\mathcal{C}^*)$ and $\sigma^2(\mathcal{C}^*)$ will henceforth be denoted as μ and σ^2. The number of eigenvalues distributed far from μ is estimated by the Chebyshev's inequality:

$$\Pr\left(|\lambda_i - \mu| \geq \delta_2\right) \leq \frac{\sigma^2}{\delta_2^2}, \tag{17}$$

where without loss of generality, we can choose δ_2 as a function of σ^2 such that $\delta_2 \to 0$ and $\sigma^2/\delta_2^2 \to 0$ as $\sigma^2 \to 0$. Chebyshev's inequality implies that for sufficiently large n, the number of eigenvalues smaller than $\mu - \delta_2$ is bounded above by $\sigma^2 M/\delta_2^2$. Note that due to Eq. (8), the same bound holds for the eigenvalues of the original Gram matrix G.

These smaller eigenvalues contribute to decreasing the success probability. To circumvent that, we slightly modify the EPM in Eq. (6) as follows. The decoder first applies a (nondestructive) projective measurement $\{P, P^\perp\}$, where $P + P^\perp = I$, onto the received codeword state. P and $P^\perp$ act on the space spanned by $\sum_m |\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|$, with P defined to be the projection onto the subspace corresponding to the λ_i's with $\lambda_i \geq \mu - \delta_2$ and $P^\perp$ its orthogonal complement.

If the initial projective measurement gives the $P^{\perp}$ outcome, it is regarded as an inconclusive result. The probability of this outcome is estimated as follows. Since the number of λ_i's such that $\lambda_i \leq \mu - \delta_2$ is less than $\sigma^2 M/\delta_2^2$, the rank of $P^{\perp}$,

$$\operatorname{rank}\left(P^{\perp}\right) \leq \frac{\sigma^2 M}{\delta_2^2}. \tag{18}$$

Therefore, the probability of the inconclusive outcome of the projective measurement, averaged over all codewords, is:

$$\begin{aligned} P_I^P &= \frac{1}{M}\sum_{m=1}^{M} \operatorname{Tr}\left[|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|P^{\perp}\right] \\ &= \frac{1}{M}\operatorname{Tr}\left[\langle\psi_{x^n(k)}|P^{\perp}|\psi_{x^n(l)}\rangle\right]_{kl} \\ &\leq \frac{1}{M}(\mu-\delta_2)\frac{\sigma^2 M}{\delta_2^2} \leq \frac{\sigma^2}{\delta_2^2} = \epsilon_2. \end{aligned} \tag{19}$$

Therefore, the average probability of the outcome P is given by $1-\epsilon_2$. Now at most half of the codewords can have the probability of this outcome less than $1-2\epsilon_2$. We throw away this half from the codebook and for each codeword in the remaining half, we have that the probability of outcome P is at least $1-2\epsilon_2$.

If the outcome of this projective measurement is P, then the output state $P|\psi_{x^n(m)}\rangle$ is further measured by the EPM (6). The modified EPM is defined by using a set of (unnormalized) states $\{P|\psi_{x^n(m)}\rangle\}$ $(i = 1, \cdots, M)$, whose POVM elements are given by:

$$\tilde{\Lambda}_m = p(P\Lambda P)^{-1}P|\psi_{x^n(m)}\rangle\langle\psi_{x^n(m)}|P(P\Lambda P)^{-1}, \tag{20}$$

for $1 \leq m \leq M$ and $\tilde{\Lambda}_{M+1} = I - \sum_{m=1}^{M}\tilde{\Lambda}_m$. Now, the minimum eigenvalue of the new Gram matrix $\Gamma = [\langle\psi_{x^n(k)}|P|\psi_{x^n(l)}\rangle]$ is bounded by $\lambda_{\min} \geq \mu - \delta_2$. Notice that the Gram matrix of the normalized states is simply a diagonal matrix times Γ. The diagonal entries contain the inverse of the outcome probabilities which satisfy $1 \geq \operatorname{Tr}[P|\psi_{x^n(k)}\rangle\langle\psi_{x^n(k)}|] \geq 1 - 2\epsilon_2$. Using the fact that for any product AB, we have $\lambda_{\min}(AB) \geq \lambda_{\min}(A)\lambda_{\max}(B)$ and that the maximum value of the diagonal matrix is ≥ 1, we obtain the lower bound on the EPM's success probability as $\lambda_{\min} \geq (\mu - \delta_2)$. Therefore, the inconclusive probability is upper bounded as, $P_I^{EPM} \leq 1 - (\mu - \delta_2)$.

Note that the overall (two-stage) measurement we described above is a USD measurement, since if a codeword outcome results (i.e., by the projective resulting in the P outcome followed by the EPM succeeding in detecting the received codeword correctly), it results in an unambiguous detection. However, there are two ways an inconclusive outcome may result—either due to the initial projective measurement resulting in the $P^{\perp}$ outcome, or it resulting in the P outcome followed by the EPM measurement resulting in its inconclusive outcome. Therefore, the overall probability of the inconclusive outcome, averaged over all codewords, is given by:

$$\bar{P}_I = P_I^P + (1 - P_I^P)P_I^{EPM} \leq \frac{\sigma^2}{\delta_2^2} + 1 - \mu - \delta_2 = \epsilon \tag{21}$$

where $\epsilon \to 0$ as $n \to \infty$ if $R < S(\rho) - \delta_1$, which follows from the inequalities in equations (15) and (16).

Finally, the maximum inconclusive probability over all codebooks can be improved by the usual codeword-expurgation argument [27]. By throwing away the worst half of the codewords from the codebook $\mathcal{C}^*$, we obtain a new codebook for which the inconclusive probability is less than 2ϵ for all 2^{nR-2} codewords (note that we already throw away the half before this expurgation), which completes the proof.

IV. Conclusion

We proved that channel decoding based on an unambiguous state discrimination (USD) measurement attains the Holevo capacity for a general pure state classical-quantum channel. Our result provides an alternative way of achieving the cq channel's capacity and raises many interesting questions for future study. First of all, it is not clear to us whether the additional projection $\{P, P^{\perp}\}$ before the EPM, which is somewhat akin to a 'quantum expurgation' step, is truly necessary as a physical operation; i.e., whether a better error analysis might not need it and thus be able to prove that the EPM applied directly to the random code achieves the Holevo capacity. The reason we introduce that projection is for convenience of analysis, in particular, to facilitate the application of the Chebyshev's inequality to our modified Gram matrix, notwithstanding the fact that the Chebyshev's inequality is known to provide a weak bound in many of its applications. Since the Gram matrix of the original random code, a bound on whose minimum eigenvalue we seek for the USD error analysis, is a random matrix, it might benefit from an application of Random Matrix Theory [28], which could provide a stronger bound on the minimum eigenvalue. Another immediate question that arises is how to generalize our result to general mixed-state cq channel [20]. It is also interesting to ask whether an *intermediate measurement*, i.e., one 'between' the perfect minimum-error state discrimination with no inconclusive outcome, and optimal USD (i.e., one with maximum average success probability allowing for an inconclusive outcome) could also attain the Holevo bound. The reason this question is interesting is that such intermediate measurements have been studied both in the general case, as well as in the context of optical implementations [21], [22]. Last but not the least, our result alludes to a new possibility to find implementable structured optical receivers, via the USD approach, to achieve the Holevo capacity of a lossy optical channel [29]. This is a truly practical example of a pure-state cq channel where the results of our paper applies. Moreover, the Holevo capacity of the pure-loss optical channel is known to be achievable using a coherent-state (ideal laser light) modulation, and none of the known collective decoding methods translate readily into a fully realizable circuit of known optical components, at the time of writing this paper.

Appendix

In this appendix, we list the properties of the typical subspace and the matrix inequality used in the main text.

Lemma 2 (Properties of the typical projectors): The typical projector defined in Eq. (5) satisfies the following properties for arbitrary $\epsilon, \delta > 0$ and sufficiently large n:

$$\begin{aligned} \mathrm{Tr}\left[\Pi_\delta \rho^n\right] &\geq 1-\epsilon, & (22) \\ 2^{-n(S(\rho)+\delta)}\Pi_\delta &\leq \Pi_\delta \rho^n \Pi_\delta \leq 2^{-n(S(\rho)-\delta)}\Pi_\delta, & (23) \end{aligned}$$

For more detailed discussions, see [30] for example.

Lemma 3: Let A, B be $N \times N$ Hermitian matrices and le the eigenvalues $\lambda_i(A)$, $\lambda_i(B)$, $\lambda_i(AB)$ and $\lambda_i(A+B)$ be arranged in increasing order. For each $k = 1, \cdots, N$, we have

$$\begin{aligned} \lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B)\,, \\ \lambda_1(A)\lambda_N(B) \leq \lambda_1(AB)\,. \qquad (24) \end{aligned}$$

For the proof, see [26] for example.

References

[1] A. S. Holevo, "The capacity of the quantum channel with general signal states", *IEEE Trans. Inform. Theory,* vol. 44, pp. 269–273, 1998.

[2] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels", *Phys. Rev. A*, vol. 56, pp. 131–138, 1997.

[3] P. Hausladen, R. Jozsa, B. Schumacher, M. D. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel", *Phys. Rev. A*, vo. 54, pp. 1869–1876, 1996.

[4] B. Schumacher, "Quantum coding", *Phys. Rev. A*, vol. 51, pp. 2738–2747, 1995.

[5] P. Hausladen and W. K. Wootters, "A 'pretty good' measurement for distinguishing quantum states", *J. Mod. Opt.*, vol. 41, pp. 2385–2390, 1994.

[6] S. M. Barnett and S. Croke, "Quantum state discrimination", *Adv. Opt. Photon.*, vol. 1, pp. 238–278, 2009, and references therein.

[7] H. P. Yuen, R. S. Kennedy, and M. Lax, "Optimum testing of multiple quantum hypotheses in quantum detection theory", *IEEE Trans. Inform. Theory,* 21, 125, 1975.

[8] Y. C. Eldar and G. D. Forney, Jr., "On quantum detection and the square-root measurement," *IEEE Trans. Inform. Theory,* vol. 47, pp. 858–872, 2001.

[9] S. Lloyd, V. Giovannetti, L. Maccone, "Sequential projective measurements for channel coding", *Phys. Rev. Lett.*, vol. 106, p. 250501, 2011.

[10] V. Giovannetti, S. Lloyd,L. Maccone, "Achieving the Holevo bound via sequential measurements", *Phys. Rev. A*, vol. 85, p. 012302, 2012.

[11] M. M. Wilde, S. Guha, S. H. Tan, S. Lloyd, "Explicit capacity-achieving receivers for optical communication and quantum reading," *Proc. of Int. Symp. on Inf. Th. (ISIT) 2012*, arXiv:1202.0518v2 [quant-ph].

[12] M. M. Wilde and S. Guha, "Polar Codes for Classical-Quantum Channels," *IEEE Trans. on Inf. Theory*, vol. 59, pp. 1175–1187, 2013.

[13] I. D. Ivanovic, "How to differentiate between non-orthogonal states", *Phys. Lett. A*, vol. 123, pp. 257–259, 1987.

[14] D. Dieks, "Overlap and distinguishability of quantum states", *Phys. Lett. A*, vol. 126, pp. 303–307, 1988.

[15] A. Peres, "How to differentiate between non-orthogonal states", *Phys. Lett. A*, vol. 128, p. 19, 1988.

[16] G. Jaeger and A. Shimony, "Optimal distinction between two non-orthogonal quantum states", *Phys. Lett. A*, vol. 197, pp. 83–87, 1995.

[17] A. Chefles, "Unambiguous discrimination between linearly independent quantum states", *Phys. Lett. A*, vol. 239, pp. 339–347, 1998.

[18] A. Chefles and S. M. Barnett, "Optimum unambiguous discrimination between linearly independent symmetric states", *Phys. Lett. A*, vol. 250, pp. 223–229, 1998.

[19] Y. Eldar, "A Semidefinite Programming Approach to Optimal Unambiguous Discrimination of Quantum States", *IEEE Trans. Inform. Theory*, vol. 49, pp. 446–456, 2003.

[20] Y. Eldar, M. Stojnic, and B. Hassibi, "Optimal quantum detector for unambiguous detection of mixed states", *Phys. Rev. A*, vol. 69, p. 062318, 2004.

[21] A. Chefles and S. M. Barnett, "Strategies for discriminating between non-orthogonal quantum states", *J. Mod. Opt.*, vol. 45, pp.1295–1302, 1998.

[22] J. Fiurášek and M. Ježek, "Optimal discrimination of mixed quantum states involving inconclusive results", *Phys. Rev. A*, vol. 67, p. 012321, 2003.

[23] S. J. van Enk, "Unambiguous state discrimination of coherent states with linear optics: Application to quantum cryptography", *Phys. Rev. A*, vol. 66, p. 042313, 2002.

[24] M. Dušek, M. Jahma, and N. Lütkenhaus, "Unambiguous state discrimination in quantum cryptography with weak coherent states", *Phys. Rev. A* vol. 62, p. 022306, 2000.

[25] H. E. Brandt, "Unambiguous State Discrimination in Quantum Key Distribution", *Quantum Inf. Process.*, vol. 4, pp. 387–398, 2005.

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1985.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[28] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review", *Statist. Sinica*, vol. 9, p. 611, 1999.

[29] 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", *Phys. Rev. Lett.*, vol. 92, p. 027902, 2004.

[30] M. M. Wilde, *From Classical to Quantum Shannon Theory*, arXiv:1106.1445 [quant-ph].