Explicit capacity-achieving receivers for optical communication and quantum reading

Citation: Wilde, Mark M., Saikat Guha, Si-Hui Tan, and Seth Lloyd. Explicit Capacity-achieving Receivers for Optical Communication and Quantum Reading. In 2012 IEEE International Symposium on Information Theory Proceedings, 551-555. Institute of Electrical and Electronics Engineers, 2012.

As Published: http://dx.doi.org/10.1109/ISIT.2012.6284251

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/79118

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Explicit capacity-achieving receivers for optical communication and quantum reading

Mark M. Wilde*, Saikat Guha†, Si-Hui Tan‡, and Seth Lloyd§
*School of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada
†Disruptive Information Proc. Tech. Group, Raytheon BBN Technologies, Cambridge, Massachusetts 02138, USA
‡Data Storage Institute, Agency for Science, Tech., & Research, 117608 Singapore
§Research Laboratory for Electronics and Department of Mechanical Engineering,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

***Abstract*—An important practical open question has been to design explicit, structured optical receivers that achieve the Holevo limit in the contexts of optical communication and "quantum reading." The Holevo limit is an achievable rate that is higher than the Shannon limit of any known optical receiver. We demonstrate how a sequential decoding approach can achieve the Holevo limit for both of these settings. A crucial part of our scheme for both settings is a non-destructive "vacuum-or-not" measurement that projects an n-symbol modulated codeword onto the n-fold vacuum state or its orthogonal complement, such that the post-measurement state is either the n-fold vacuum or has the vacuum removed from the support of the n symbols' joint quantum state. The sequential decoder for optical communication requires the additional ability to perform multimode optical phase-space displacements—realizable using a beamsplitter and a laser, while the sequential decoder for quantum reading also requires the ability to perform phase-shifting (realizable using a phase plate) and online squeezing (a phase-sensitive amplifier).**

One of the first accomplishments in quantum information theory was the upper bound (now known as the *Holevo bound*) on how much classical information can be encoded into a quantum system, such that another party can reliably recover it using a quantum measurement [1]. Subsequently, Holevo, Schumacher, and Westmoreland (HSW) proved that the Holevo bound is also an achievable rate for classical communication over a quantum channel [2], [3], establishing a lower bound on a quantum channel's classical capacity. These initial results were the impetus for the field of quantum information theory [4], a generalization of Shannon's classical information theory that takes into account the quantum-physical nature of the carrier of information, channel, and the receiver measurement. The main accomplishment of HSW was to provide a mathematical specification of a decoding measurement that a receiver, bound only by the laws of quantum mechanics, could perform on the output codeword to recover the classical data transmitted by a sender at any rate below the Holevo limit. The HSW decoder prescription in general leads to a collective measurement on the codeword's joint quantum state, which may not be doable by detecting each individual symbol of the codeword separately.

For the single-mode lossy bosonic channel—which can be used to construct a wide class of practical free-space and fiber optical channels—it was shown that the single-letter Holevo bound is in fact the ultimate channel capacity [5], given by

$$g\left(\eta N_S\right) \equiv \left(\eta N_S + 1\right) \log\left(\eta N_S + 1\right) - \eta N_S \log\left(\eta N_S\right) \quad (1)$$

bits per channel use, where N_S is the mean transmitted photon number per channel use, and $\eta \in (0, 1]$ is the input-output power transmissivity. Furthermore, conventional laser-light (coherent-state) modulation with symbols chosen i.i.d. from an isotropic Gaussian prior distribution, can achieve this capacity (i.e., it is not necessary to use exotic non-classical states, such as squeezed or entangled states). The lossy bosonic channel preserves a coherent state ($|\alpha\rangle \rightarrow |\sqrt{\eta}\alpha\rangle$), thus preserving its purity. The average output state is a zero-mean circularly-symmetric Gaussian mixture of coherent states, which is a thermal state with mean photon number ηN_S, which saturates the entropy bound $g\left(\eta N_S\right)$. A converse proof shows that no other choice of modulation states and/or priors can exceed this capacity [5]. This result enabled comparing the ultimate channel capacity with the ideal Shannon limits of the classical channels induced by the quantum noise-characteristics of standard optical receivers, such as homodyne, heterodyne and direct detection receivers [5]. In spite of this accomplishment, it remains unclear how one could construct an implementation of the HSW decoding measurement for the bosonic channel using known optical components.

The theory of HSW also applies in the setting of "quantum reading" [6], where one can obtain a quantum advantage in the rate of read out of classical information stored in a digital memory. Classical bits are encoded into the reflectivity and phase of memory cells. A transmitter irradiates the memory with light that in turn is modulated by a passive linear reflection from the memory cells (each cell is a single-mode lossy bosonic channel, but this time information is encoded in the memory cell's transmissivity and phase). A mono-static receiver gathers the reflected light for measurement and processing. The above is a bare-bone model for optical disks such as CDs or DVDs. Pirandola originally considered this task in the context of quantum channel discrimination and demonstrated a quantum advantage. He and his collaborators later considered a coded strategy (in the information-theoretic sense) [7]. Later work [8], [9] improved upon Ref. [7], by demonstrating how to achieve the Holevo limit $g\left(N_S\right)$ bits/cell, where N_S is the mean number of photons available at

the transmitter to shine on each memory cell on an average. It turns out however, that the strategy for achieving $g(N_S)$ is different from that of the lossy bosonic channel, and surprisingly, a coherent-state probe fails to achieve the Holevo capacity [8], [9]. The classical information is encoded into the phase of the cells (with each having perfect reflectivity). The symbols of the phase code are chosen i.i.d. and uniformly at random from the interval $[0, 2\pi)$. The transmitter shines each cell with the single-mode quantum superposition state:

$$|\phi_{\text{II}}\rangle \equiv \sum_{n=0}^{\infty} \sqrt{N_S^n / (N_S+1)^{n+1}}\, |n\rangle\,, \tag{2}$$

and the receiver performs a collective measurement on the received codeword ($|n\rangle$ is a photon number state [10]). The average state of the received ensemble is a completely dephased version of $|\phi_{\text{II}}\rangle$, yet again, a thermal state with mean photon number N_S, which saturates the entropy bound $g(N_S)$. Again, the authors of Ref. [8] left open the question of a structured capacity-achieving receiver measurement.

In this paper, we address the open questions from Refs. [5], [8], [9], by detailing a structured quantum measurement that can achieve both of the above Holevo limits (for optical communication and quantum reading). The measurement is a sequential decoder, in the sense that it is a sequence of binary-outcome measurements that ask, "Was the received quantum state produced from the first codeword? the second codeword? the third?" etc., proceeding until the answer to one of the questions is "yes." Our work builds on recent insights of Giovannetti *et al*. [11] and Sen [12] in sequential decoding for quantum channels. Our primary contribution here is to show how to construct these measurements in an optical setting.

Our sequential decoding scheme for the lossy bosonic channel requires two capabilities at the receiver. First, the receiver should be able to apply a "displacement operator," which simply requires highly reflective beamsplitters and a strong laser local oscillator [13]. Second, the receiver should be able to perform a quantum non-demolition measurement to determine whether an n-mode state is in the vacuum state or not. That is, the measurement operators are of the form $\{|0\rangle\langle 0|^{\otimes n}, I^{\otimes n} - |0\rangle\langle 0|^{\otimes n}\}$, where $|0\rangle$ is the vacuum state and I is the identity operator. After performing such a measurement on an n-mode state $|\psi\rangle$, the post-measurement state should be either $|0\rangle^{\otimes n}$ or $(|\psi\rangle - c\,|0\rangle^{\otimes n})/\sqrt{1-|c|^2}$, with $c = \langle 0|^{\otimes n}|\psi\rangle$. The key aspect of this measurement is that its disturbance to an n-mode state becomes asymptotically negligible as n becomes large, as long as the number of codewords is no larger than $\sim 2^{ng(\eta N_S)}$. Our sequential decoding scheme for quantum reading requires the "vacuum-or-not" measurement described above, and the ability to perform phase shifting and online squeezing [10].

We structure this paper as follows. Section I reviews standard definitions and notation that are helpful for understanding the rest of the paper. Section II describes how a sequential decoder operates when decoding classical information transmitted over a pure-state classical-quantum channel, and for completeness, Appendix B provides a proof that this scheme achieves the Holevo capacity. Section III provides a summary of the operations needed for sequential decoding of the lossy bosonic channel. Section IV details an implementation of a sequential decoder for quantum reading. We conclude in Section V with a summary and a list of open questions.

I. Definitions and Notation

We denote quantum systems as A, B, and C and their corresponding Hilbert spaces as $\mathcal{H}^A$, $\mathcal{H}^B$, and $\mathcal{H}^C$ with respective dimensions d_A, d_B, and d_C. We denote pure states of the system A with a *ket* $|\phi\rangle^A$ and the corresponding density operator as $\phi^A = |\phi\rangle\langle\phi|^A$. All kets that are quantum states have unit norm, and all density operators are positive semi-definite with unit trace. We model our lack of access to a quantum system with the partial trace operation. That is, given a two-qubit state ρ^{AB} shared between Alice and Bob, we can describe Alice's state with the reduced density operator: $\rho^A = \text{Tr}_B\{\rho^{AB}\}$, where Tr_B denotes a partial trace over Bob's system. Let $H(A)_\rho \equiv -\text{Tr}\{\rho^A \log \rho^A\}$ be the von Neumann entropy of the state ρ^A.

II. Sequential Decoding

In this section, we describe the operation of a sequential decoder that can reliably recover classical information encoded into a pure state ensemble. Appendix B contains a full error analysis, demonstrating that the scheme achieves capacity.

Suppose that a classical-quantum channel of the form $x \to |\phi_x\rangle$ connects a sender Alice to a receiver Bob. For our purposes here, it does not matter whether the classical input x is discrete or continuous.

Theorem 1: Let $x \to |\phi_x\rangle$ be a classical-quantum channel and let $\rho \equiv \sum_x p_X(x)\,|\phi_x\rangle\langle\phi_x|$ for some distribution $p_X(x)$. Then the rate $H(\rho)$ bits per channel use is achievable for communication over this channel by having the receiver employ a sequential decoding strategy.

Proof: We break the proof into several steps.

Codebook Construction. Before communication begins, Alice and Bob agree upon a codebook. We allow them to select a codebook randomly according to the distribution $p_X(x)$. So, for every message $m \in \mathcal{M} \equiv \{1, \ldots, 2^{nR}\}$, generate a codeword $x^n(m) \equiv x_1(m) \cdots x_n(m)$ randomly and independently according to

$$p_{X^n}(x^n) \equiv \prod_{i=1}^{n} p_X(x_i)\,.$$

Sequential Decoding. Transmitting the codeword $x^n(m)$ through n uses of the channel $x \to |\phi_x\rangle$ leads to the following quantum state at Bob's output:

$$\left|\phi_{x^n(m)}\right\rangle \equiv \left|\phi_{x_1(m)}\right\rangle \otimes \cdots \otimes \left|\phi_{x_n(m)}\right\rangle.$$

Upon receiving the quantum codeword $\left|\phi_{x^n(m)}\right\rangle$, Bob performs a sequence of binary-outcome quantum measurements to determine the classical codeword $x^n(m)$ that Alice transmitted. He first "asks," "Is it the first codeword?" by performing the measurement $\{\left|\phi_{x^n(1)}\right\rangle\left\langle\phi_{x^n(1)}\right|, I^{\otimes n} - \left|\phi_{x^n(1)}\right\rangle\left\langle\phi_{x^n(1)}\right|\}$. If he receives the outcome "yes," then

he performs no further measurements and concludes that Alice transmitted the codeword $x^n(1)$. If he receives the outcome "no," then he performs the measurement $\{|\phi_{x^n(2)}\rangle\langle\phi_{x^n(2)}|, I^{\otimes n} - |\phi_{x^n(2)}\rangle\langle\phi_{x^n(2)}|\}$ to check if Alice sent the second codeword. Similarly, he stops if he receives "yes," and otherwise, he proceeds along similar lines.

The above concludes the description of the operation of the sequential decoder. We provide an error analysis demonstrating that this scheme works well in Appendix B, i.e., the word error goes to zero as $n \to \infty$, as long as $R < H(\rho)$. Note that Sen [12] and Giovannetti *et al*. [11] already gave a proof that a sequential decoder works, but our proof in Appendix B is a bit simpler because it is specialized to the case of pure-state ensembles (which is sufficient to consider for our settings of pure-loss optical communication and quantum reading).

III. Sequential Decoding for Optical Communication

We now provide a physical realization of the sequential decoding strategy in the context of optical communications. In this setting, we suppose that a lossy bosonic channel, specified by the following Heisenberg relations, connects Alice to Bob:

$$\hat{b} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{e}, \tag{3}$$

where $\hat{a}$, $\hat{b}$, and $\hat{e}$ are the respective field operators for Alice's input mode, Bob's output mode, and an environmental input mode (assumed to be in its vacuum state). The transmissivity $\eta \in [0,1]$ is the fraction of Alice's input photons that make it to Bob on average. We assume that Alice is constrained to using mean photon number N_S per channel use.

The strategy for achieving the classical capacity of this channel is for Alice to induce a classical-quantum channel, by selecting $\alpha \in \mathbb{C}$ and preparing a coherent state $|\alpha\rangle$ [10] at the input of the channel in (3). The resulting induced classical-quantum channel to Bob is of the following form:

$$\alpha \to |\sqrt{\eta}\alpha\rangle.$$

By choosing the distribution $p_X(x)$ in Theorem 1 to be an isotropic, complex Gaussian with variance N_S:

$$p_{N_S}(\alpha) \equiv (1/\pi N_S)\exp\left\{-|\alpha|^2/N_S\right\},$$

we have that $g(\eta N_S)$ is an achievable rate for classical communication. The quantity $g(\eta N_S)$ is the entropy of the average state of the ensemble $\{p_{N_S}(\alpha),\ |\sqrt{\eta}\alpha\rangle\}$:

$$\int d^2\alpha\ p_{N_S}(\alpha)\,|\sqrt{\eta}\alpha\rangle\langle\sqrt{\eta}\alpha|,$$

which is a thermal state with mean photon number ηN_S [10].

Each quantum codeword selected from the ensemble $\{p_{N_S}(\alpha),\ |\alpha\rangle\}$ has the following form:

$$|\alpha^n(m)\rangle \equiv |\alpha_1(m)\rangle \otimes \cdots \otimes |\alpha_n(m)\rangle.$$

We assume $\eta = 1$ above and for the rest of this section without loss of generality. Thus, the sequential decoder consists of measurements of the following form for all $m \in \mathcal{M}$:

$$\left\{|\alpha^n(m)\rangle\langle\alpha^n(m)|,\ I^{\otimes n} - |\alpha^n(m)\rangle\langle\alpha^n(m)|\right\}. \tag{4}$$

Observing that

$$|\alpha^n(m)\rangle = D(\alpha_1(m)) \otimes \cdots \otimes D(\alpha_n(m))\,|0\rangle^{\otimes n},$$

where $D(\alpha) \equiv \exp\{\alpha\hat{a}^\dagger - \alpha^*\hat{a}\}$ is the well-known unitary "displacement" operator from quantum optics [10] and $|0\rangle^{\otimes n}$ is the n-fold tensor product vacuum state, it is clear that the decoder can implement the measurement in (4) in three steps:

1) Displace the n-mode codeword state by
$$D(-\alpha_1(m)) \otimes \cdots \otimes D(-\alpha_n(m)),$$
by employing highly asymmetric beam-splitters with a strong local oscillator [13].
2) Perform a "vacuum-or-not" measurement of the form
$$\left\{|0\rangle\langle 0|^{\otimes n},\quad I^{\otimes n} - |0\rangle\langle 0|^{\otimes n}\right\}.$$
If the vacuum outcome occurs, decode as the m^{th} codeword. Otherwise, proceed.
3) Displace by $D(\alpha_1(m)) \otimes \cdots \otimes D(\alpha_n(m))$ with the same method as in Step 1.

The receiver just iterates this strategy for every codeword in the codebook, and Theorem 1 states this strategy is capacity-achieving.

Remark 2: The above strategy is reminiscent of the class of conditional pulse nulling receivers [14], which are useful in discriminating M-ary pulse-position-modulation coded states with $|\alpha\rangle$ in the i^{th} slot and vacuum states $|0\rangle$ in the other $M-1$ slots. In this strategy, the receiver hypothesizes at first that the transmitted codeword is the first codeword $|\alpha\rangle\,|0\rangle^{\otimes M-1}$, nulls the first mode by applying $D^\dagger(\alpha)$, and direct-detects the first mode. If the sender in fact transmitted the first codeword, then the resulting state is ideally $|0\rangle^{\otimes m}$, and direct detection of the first mode should ideally produce no "clicks." If there is no click, then the receiver direct detects the other modes to confirm the original hypothesis. If there are no further clicks, then the receiver declares that the sender transmitted the first codeword. If there is a further click, then the receiver guesses the codeword corresponding to the position of the click. If on the first mode there is a click, then the receiver hypothesizes that the transmitted codeword is the second one and repeats the above algorithm on the next $M-1$ modes.

The difference between the sequential decoding strategy and conditional pulse nulling is that the codewords are different, and the vacuum-or-not measurement in the sequential decoding strategy is much more difficult to perform in practice than direct detection, which annihilates the detected quantum state. Ideally, the vacuum-or-not should be a non-demolition measurement such that the post-measurement state is $|0\rangle^{\otimes n}$ or $(|\psi\rangle - c\,|0\rangle^{\otimes n})/\sqrt{1-|c|^2}$, with $c = \langle 0|^{\otimes n}|\psi\rangle$, if the pre-measurement state is $|\psi\rangle$, with probabilities $p_0 = |c|^2$ and $p_1 = 1 - p_0$, respectively, of the two possible outcomes.

Remark 3: The crucial (and most difficult) step in sequential decoding for the lossy bosonic channel is the vacuum-or-not measurement. Oi *et al*. have provided a method for performing this measurement, by interacting the light field with a three-level atom in a STIRAP process [15]. This

approach would likely be quite lossy in practice, so it would be ideal to determine an all-optical vacuum-or-not measurement.

Remark 4: If the mean input photon number $N_S \ll 1$, then one does not require a full Gaussian distributed codebook in order to achieve capacity. A simpler method, called binary phase-shift keying, suffices to approach capacity very closely. In this approach, the ensemble for generating a codebook randomly is just $\{1/2,\ |\pm\alpha\rangle\}$. This also simplifies the sequential decoder because the only displacements required for implementation are $D(\pm\alpha)$. An additional advantage is that a random linear encoder should achieve the capacity, by an argument similar to that on pages 3-14 and 3-15 of Ref. [16]. BPSK polar codes are capacity-achieving for low-photon number as well [17].

Remark 5: Tan proved a variation of Theorem 1 for the lossy bosonic channel in her thesis [18], but the analysis in Appendix B demonstrates that it is actually not necessary to perform a measurement onto the average typical subspace. We avoided having to do so by demonstrating that it is sufficient to code for a typical-projected version of the channel and applying Sen's non-commutative union bound from Ref. [12].

Remark 6: The above sequential decoding approach also works well in the context of private classical communication over a lossy bosonic channel [19], [20]. The private classical capacity of the channel in (3) is $g(\eta N_S) - g((1-\eta)N_S)$ (compare to its public classical capacity of $g(\eta N_S)$), and the strategy for encoding is again to choose coherent states randomly according to an isotropic Gaussian prior. The sequential decoder can just test for all codewords in a codebook of size $2^{ng(\eta N_S)}$ and recover the transmitted private message correctly. The privacy in the scheme comes about by choosing $2^{ng((1-\eta)N_S)}$ codewords corresponding to each message and selecting one of these uniformly at random in order to randomize Eve's knowledge of the transmitted message [19].

IV. Sequential Decoding for Quantum Reading

The sequential decoding strategy also finds application in "quantum reading" [6]. In this setting, we suppose that information is encoded into passive memory cells of an optically-readable memory, which a transceiver can read out by irradiating them with laser (or quantum) light and detecting the reflected light. More specifically, we can model the i^{th} optical memory cell as a beamsplitter of the following form:

$$\hat{b}_i = \exp\{\mathrm{i}\theta_i\}\sqrt{\eta_i}\hat{a}_i + \sqrt{1-\eta_i}\hat{e}_i,$$

where the parameters η_i and θ_i are the respective reflectivity and phase of the i^{th} cell, and $\hat{a}_i$, $\hat{b}_i$, and $\hat{e}_i$ are the respective field operators for the transmitter's i^{th} input mode, the i^{th} reflected mode, and an environmental mode (assumed to be in its vacuum state). We assume perfect channels from the transmitter to the optical memory cells and from the cells back to the receiver (which is co-located with the transmitter).

The objective is for the transmitter to interrogate each optical memory cell with some quantum state of light with mean photon number N_S. The receiver then collects all of the reflected light and performs some measurement to recover the classical information encoded in the memory cells. If we use a coherent-state transmitter to interrogate each cell, we call it the Type I setting [8]. If we do not allow the transmitter to retain any state entangled with the transmitted light, but allow it to send any quantum state (entangled spatially across modes or an unentangled non-classical product state), then this is termed the Type II setting [8]. Finally, if we do allow for entanglement assistance, in the sense that the transmitter can prepare two modes in an entangled state for each of the n memory cells, send one to a memory cell while retaining the other, then this is termed a Type III setting [8]. In each of the three settings, the receiver is always allowed to perform a general (collective) quantum measurement on the reflected n modes (and the retained n modes, in case of Type III). It is straightforward to prove that $g(N_S)$ is the Holevo (upper) bound on the capacity of quantum reading in the Type I and Type II settings, while it is unknown whether $g(N_S)$ could be exceeded in the Type III setting [8].

Recently, Guha *et al*. proved that the following strategy achieves the $g(N_S)$ bound for quantum reading using a Type II transmitter [8], [9]. The transmitter interrogates each memory cell with a quantum state of light of the form in (2). It is straightforward to compute that the mean number of photons in this state is N_S: $\langle\phi_{\text{II}}|\ \hat{n}\ |\phi_{\text{II}}\rangle = N_S$, where $\hat{n} = \hat{a}^\dagger\hat{a}$ is the photon number operator [10]. Each memory cell has classical information encoded into only the phase variable θ_i (with $\eta_i = 1$), so that a randomly chosen code in the sense of Theorem 1 is selected from the following ensemble:

$$\{1/2\pi,\ |\phi_{\text{II},\theta}\rangle\}, \tag{5}$$

where

$$|\phi_{\text{II},\theta}\rangle \equiv \sum_{n=0}^{\infty}\sqrt{N_S^n/(N_S+1)^{n+1}}\exp\{\mathrm{i}n\theta\}|n\rangle, \tag{6}$$

and each θ is selected uniformly at random from the interval $[0, 2\pi)$. The average state of this code ensemble is

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta\ |\phi_{\text{II},\theta}\rangle\langle\phi_{\text{II},\theta}|\ = \sum_{n=0}^{\infty} N_S^n/(N_S+1)^{n+1}|n\rangle\langle n|,$$

which is a thermal state with mean photon number N_S. (The effect of phase-randomizing the state $|\phi_{\text{II}}\rangle$ is simply to dephase it to a thermal state.) Thus, a random code constructed from the ensemble in (5) along with a sequential decoder saturates the entropy bound $g(N_S)$ because the average state is a thermal state.

It is not clear to us at the moment how to implement a sequential decoder for the above Type II strategy. Though, if we allow for a Type III transmitter, the strategy is straightforward to specify. First, the transmitter interrogates each optical memory cell with one mode of a two-mode squeezed vacuum state [10] of the following form:

$$|\phi_{\text{III}}\rangle \equiv \sum_{n=0}^{\infty}\sqrt{N_S^n/(N_S+1)^{n+1}}|n\rangle|n\rangle,$$

while retaining the other mode. The encoding in the optical memory cells is the same as above, such that the memory cells have classical information encoded only into a uniformly random phase. The code ensemble is then $\{1/2\pi,\ |\phi_{\text{III},\theta}\rangle\}$, with $|\phi_{\text{III},\theta}\rangle$ defined similarly as in (6). The authors of Ref. [8] showed that this ensemble also saturates the $g(N_S)$ bound. Consider the m^{th} quantum codeword to have the form:

$$|\phi_{\text{III},\theta^n(m)}\rangle \equiv |\phi_{\text{III},\theta_1(m)}\rangle \otimes \cdots \otimes |\phi_{\text{III},\theta_n(m)}\rangle.$$

Consider further that each of the states in the above tensor product can be written as

$$|\phi_{\text{III},\theta_i(m)}\rangle = \left(P\left(\theta_i\left(m\right)\right) \otimes I\right) S\left(r\right) |0\rangle^{\otimes 2},$$

where $P\left(\theta_i\left(m\right)\right) = \exp\{\mathrm{i}\hat{n}\theta_i\left(m\right)\}$ is a phase shifter, $S\left(r\right)$ is a two-mode squeezing operator [10] with the squeezing strength r, s.t. $N_S = \sinh^2 r$, and $|0\rangle^{\otimes 2}$ is a two-mode vacuum state. This then leads us to specify the m^{th} step of the sequential decoder, which proceeds as follows:

1) Apply the operator $\left(P^\dagger\left(\theta_i\left(m\right)\right) \otimes I\right)$ by phase-shifting the first mode of the i^{th} pair by $-\theta_i\left(m\right)$.
2) Apply the unsqueezing operator $S^\dagger\left(r\right)$. The receiver can accomplish this with a phase-sensitive amplifier.
3) Perform a "vacuum-or-not" measurement of the same form as in Step 2 in the previous section. If the vacuum outcome occurs, decode as the m^{th} codeword. Otherwise, proceed.
4) Apply the squeezing operator $S\left(r\right)$.
5) Apply the operator $\left(P\left(\theta_i\left(m\right)\right) \otimes I\right)$ by phase-shifting the first mode of the i^{th} pair by $\theta_i\left(m\right)$.

The receiver again iterates this strategy for all codewords in the codebook, and Theorem 1 states that this strategy is Holevo-capacity-achieving, i.e., it achieves $g(N_S)$ bits/cell.

Remark 7: At $N_S \ll 1$, a binary phase-shift keying code approximately achieves the Holevo limit of $g\left(N_S\right)$, i.e., $C_{\text{BPSK}}(N_S) = H_2\left((1 \pm e^{-2N_S})/2\right)$. The code ensemble for this case is just $\{(1/2, |\phi_{\text{III}}\rangle), (1/2, |\phi_{\text{III},\pi}\rangle)\}$, and the sequential decoder only needs to have phase shifts of 0 or π. Interestingly enough, with binary phase modulation, even a coherent-state (Type I) transmitter can achieve $C_{\text{BPSK}}(N_S)$.

V. Conclusion

We have demonstrated that a sequential decoding strategy achieves the Holevo capacity for optical communication and quantum reading, by building on information-theoretic works on sequential decoding in Refs. [11], [12]. Both schemes employ a "vacuum-or-not" measurement which distinguishes coherently and in a non-demolition way between the vacuum or "not vacuum," so that the disturbance on the encoded state is asymptotically negligible for long codewords (as long as the code rate is less than the Holevo limit). For optical communication, the only other operation needed is implementing a displacement operator, while the sequential quantum reading receiver requires phase shifting and online squeezing.

The most important open problems going forward concern making the scheme more practical. In this vein, it might be helpful to realize an all-optical implementation of the "vacuum-or-not" measurement—which could help both on the scalability front, and relative ease of implementation as compared to a system that uses atom-light interaction [15]. Also, the sequential decoding scheme given here is impractical from a computational perspective because it requires an exponential number of measurements (there are an exponential number of codewords). It would be better to have a sequential decoder that decodes one bit at a time and would thus require only a linear number of measurements. The polar decoder for classical-quantum channels is one such sequential decoder [17], but it remains unclear to us how to implement it with optical devices.

We thank J. P. Dowling, V. Giovannetti, P. Hayden, L. Maccone, and J. H. Shapiro for useful discussions.

References

[1] 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.
[2] ——, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, 1998.
[3] 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.
[4] M. M. Wilde, *From Classical to Quantum Shannon Theory*, June 2011, arXiv:1106.1445.
[5] 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, no. 2, p. 027902, January 2004.
[6] S. Pirandola, "Quantum reading of a classical digital memory," *Physical Review Letters*, vol. 106, p. 090504, March 2011.
[7] S. Pirandola, C. Lupo, V. Giovannetti, S. Mancini, and S. L. Braunstein, "Quantum reading capacity," *New Journal of Physics*, vol. 13, no. 11, p. 113012, November 2011, arXiv:1107.3500.
[8] S. Guha, Z. Dutton, R. Nair, J. H. Shapiro, and B. J. Yen, "Information capacity of quantum reading," in *Frontiers in Optics*, 2011.
[9] S. Guha *et al.*, "Achieving the Holevo limit in quantum reading," 2012, in preparation.
[10] C. Gerry and P. Knight, *Introductory Quantum Optics*. Cambridge University Press, November 2004.
[11] V. Giovannetti, S. Lloyd, and L. Maccone, "Achieving the Holevo bound via sequential measurements," December 2010, arXiv:1012.0386.
[12] P. Sen, "Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding," September 2011, arXiv:1109.0802.
[13] M. G. A. Paris, "Displacement operator by beam splitter," *Physics Letters A*, vol. 217, pp. 78–80, July 1996.
[14] S. Guha, J. L. Habif, and M. Takeoka, "PPM demodulation: On approaching fundamental limits of optical communications," in *Proceedings of the 2010 IEEE International Symposium on Information Theory*, Austin, Texas, USA, June 2010, pp. 2038–2042, arXiv:1001.2447.
[15] D. Oi, V. Potoček, and J. Jeffers, Private communication, 2012.
[16] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," January 2010, arXiv:1001.3404.
[17] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," September 2011, arXiv:1109.2591.
[18] S.-H. Tan, "Quantum state discrimination with bosonic channels and gaussian states," Ph.D. dissertation, Massachussetts Institute of Technology, September 2010.
[19] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, pp. 44–55, 2005.
[20] S. Guha, J. H. Shapiro, and B. I. Erkmen, "Capacity of the bosonic wiretap channel and the entropy photon-number inequality," in *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Ontario, Canada, July 2008, pp. 91–95, arXiv:0801.0841.

APPENDIX A
IMPORTANT LEMMAS

In order to describe the "distance" between two quantum states, we use the notion of *trace distance*. The trace distance between states σ and ρ is $\|\sigma-\rho\|_1 = \mathrm{Tr}\,|\sigma-\rho|$, where $|X| = \sqrt{X^\dagger X}$. Two states that are similar have trace distance close to zero, whereas states that are perfectly distinguishable have trace distance equal to two.

Two states can substitute for one another up to a penalty proportional to the trace distance between them:

Lemma 8: Let $0 \leq \rho, \sigma, \Lambda \leq I$. Then

$$\mathrm{Tr}\left[\Lambda\rho\right] \leq \mathrm{Tr}\left[\Lambda\sigma\right] + \left\|\rho-\sigma\right\|_1. \tag{7}$$

Proof: This follows from a variational characterization of trace distance as the distinguishability of the states under an optimal measurement M [4]: $\|\rho-\sigma\|_1 = 2\max_{0\leq M\leq I}\mathrm{Tr}\left[M(\rho-\sigma)\right]$. ∎

Consider a density operator ρ with the following spectral decomposition:

$$\rho = \sum_x p_X(x)\left|x\right\rangle\left\langle x\right|.$$

The weakly typical subspace is defined as the span of all vectors such that the sample entropy $\overline{H}(x^n)$ of their classical label is close to the true entropy $H(X)$ of the distribution $p_X(x)$ [4]:

$$T_\delta^{X^n} \equiv \mathrm{span}\left\{\left|x^n\right\rangle : \left|\overline{H}(x^n) - H(X)\right| \leq \delta\right\},$$

where

$$\overline{H}(x^n) \equiv -\frac{1}{n}\log\left(p_{X^n}(x^n)\right),$$
$$H(X) \equiv -\sum_x p_X(x)\log p_X(x).$$

The projector $\Pi_{\rho,\delta}^n$ onto the typical subspace of ρ is defined as

$$\Pi_{\rho,\delta}^n \equiv \sum_{x^n\in T_\delta^{X^n}} \left|x^n\right\rangle\left\langle x^n\right|,$$

where we have "overloaded" the symbol $T_\delta^{X^n}$ to refer also to the set of δ-typical sequences:

$$T_\delta^{X^n} \equiv \left\{x^n : \left|\overline{H}(x^n) - H(X)\right| \leq \delta\right\}.$$

The three important properties of the typical projector are as follows:

$$\mathrm{Tr}\left\{\Pi_{\rho,\delta}^n\rho^{\otimes n}\right\} \geq 1-\epsilon,$$
$$\mathrm{Tr}\left\{\Pi_{\rho,\delta}^n\right\} \leq 2^{n[H(X)+\delta]},$$
$$2^{-n[H(X)+\delta]}\Pi_{\rho,\delta}^n \leq \Pi_{\rho,\delta}^n\rho^{\otimes n}\Pi_{\rho,\delta}^n \leq 2^{-n[H(X)-\delta]}\Pi_{\rho,\delta}^n,$$

where the first property holds for arbitrary $\epsilon,\delta > 0$ and sufficiently large n.

Lemma 9 (Gentle Operator Lemma for Ensembles): Given an ensemble $\{p_X(x),\rho_x\}$ with expected density operator $\rho \equiv \sum_x p_X(x)\rho_x$, suppose that an operator Λ such that $I \geq \Lambda \geq 0$ succeeds with high probability on the state ρ:

$$\mathrm{Tr}\left\{\Lambda\rho\right\} \geq 1-\epsilon.$$

Then the subnormalized state $\sqrt{\Lambda}\rho_x\sqrt{\Lambda}$ is close in expected trace distance to the original state ρ_x:

$$\mathbb{E}_X\left\{\left\|\sqrt{\Lambda}\rho_X\sqrt{\Lambda}-\rho_X\right\|_1\right\} \leq 2\sqrt{\epsilon}.$$

A proof of the above lemma is available in Ref. [4].

APPENDIX B
ERROR ANALYSIS FOR SEQUENTIAL DECODING

In general, if Alice transmits the m^{th} codeword, then the probability for Bob to decode correctly with this sequential decoding strategy is as follows:

$$\mathrm{Tr}\left\{\phi_{x^n(m)}\hat{\Pi}_{m-1}\cdots\hat{\Pi}_1\phi_{x^n(m)}\hat{\Pi}_1\cdots\hat{\Pi}_{m-1}\phi_{x^n(m)}\right\},$$

where we make the abbreviations

$$\phi_{x^n(m)} \equiv \left|\phi_{x^n(m)}\right\rangle\left\langle\phi_{x^n(m)}\right|,$$
$$\hat{\Pi}_i \equiv I^{\otimes n} - \left|\phi_{x^n(i)}\right\rangle\left\langle\phi_{x^n(i)}\right|.$$

So the probability that Bob makes an error when decoding the m^{th} codeword is just

$$1-\mathrm{Tr}\left\{\phi_{x^n(m)}\hat{\Pi}_{m-1}\cdots\hat{\Pi}_1\phi_{x^n(m)}\hat{\Pi}_1\cdots\hat{\Pi}_{m-1}\phi_{x^n(m)}\right\}.$$

To further simplify the error analysis, we consider the expectation of the above error probability, under the assumption that Alice selects a message uniformly at random according to a random variable M and that the codeword x^n is selected at random according to the distribution $p_{X^n}(x^n)$ (as described above):

$$1-\mathop{\mathbb{E}}_{X^n,M}\mathrm{Tr}\left\{\phi_{X^n(M)}\hat{\Pi}_{M-1}\cdots\hat{\Pi}_1\phi_{X^n(M)}\hat{\Pi}_1\cdots\hat{\Pi}_{M-1}\right\}. \tag{8}$$

For the rest of the proof, it is implicit that the expectation $\mathbb{E}$ is with respect to random variables X^n and M.

Our first observation is that, for the purposes of our error analysis, we can "smooth" the channel $x^n \to \phi_{x^n}$, by imagining instead that we are coding for a projected version of the channel $\Pi\,\phi_{x^n}\,\Pi$, where Π is the typical projector for the average state $\rho \equiv \sum_x p_X(x)\phi_x$. Doing so simplifies the error analysis by cutting off large eigenvalues that reside outside of the high-probability typical subspace. Furthermore, we expect that doing so should not affect the error analysis very much because most of the probability tends to concentrate in this subspace anyway. That we can do so follows from the fact that

$$\begin{aligned}
1 &= \mathbb{E}\mathrm{Tr}\left\{\phi_{X^n(M)}\right\}\\
&= \mathbb{E}\mathrm{Tr}\left\{\Pi\phi_{X^n(M)}\right\} + \mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}\phi_{X^n(M)}\right\}\\
&= \mathbb{E}\mathrm{Tr}\left\{\Pi\phi_{X^n(M)}\Pi\right\} + \mathrm{Tr}\left\{\hat{\Pi}\mathbb{E}\phi_{X^n(M)}\right\}\\
&= \mathbb{E}\mathrm{Tr}\left\{\Pi\phi_{X^n(M)}\Pi\right\} + \mathrm{Tr}\left\{\hat{\Pi}\rho^{\otimes n}\right\},
\end{aligned}$$

where $\hat{\Pi} \equiv I - \Pi$. Furthermore, we know that

$$\begin{aligned}
&\mathbb{E}\mathrm{Tr}\left\{\phi_{X^n(M)}\hat{\Pi}_{M-1}\cdots\hat{\Pi}_1\phi_{X^n(M)}\hat{\Pi}_1\cdots\hat{\Pi}_{M-1}\right\} \\
&= \mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}_1\cdots\hat{\Pi}_{M-1}\phi_{X^n(M)}\hat{\Pi}_{M-1}\cdots\hat{\Pi}_1\phi_{X^n(M)}\right\} \\
&\geq \mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}_1\cdots\hat{\Pi}_{M-1}\phi_{X^n(M)}\hat{\Pi}_{M-1}\cdots\hat{\Pi}_1\Pi\phi_{X^n(M)}\Pi\right\} \\
&\qquad - \mathbb{E}\left\|\phi_{X^n(M)} - \Pi\phi_{X^n(M)}\Pi\right\|_1,
\end{aligned}$$

where the inequality follows from Lemma 8. Using the above observations and the facts that

$$\mathbb{E}\left\|\phi_{X^n(M)} - \Pi\phi_{X^n(M)}\Pi\right\|_1 \leq 2\sqrt{\epsilon}, \tag{9}$$

$$\mathrm{Tr}\left\{\hat{\Pi}\rho^{\otimes n}\right\} \leq \epsilon, \tag{10}$$

for all $\epsilon > 0$ whenever n is sufficiently large (these are from the properties of typicality and Lemma 9), we obtain the following upper bound on (8):

$$\begin{aligned}
&\mathbb{E}\mathrm{Tr}\left\{\Pi\phi_{X^n(M)}\Pi\right\} - \\
&\mathbb{E}\mathrm{Tr}\left\{\phi_{X^n(M)}\hat{\Pi}_{M-1}\cdots\hat{\Pi}_1\Pi\phi_{X^n(M)}\Pi\hat{\Pi}_1\cdots\hat{\Pi}_{M-1}\phi_{X^n(M)}\right\} \\
&\qquad + \epsilon + 2\sqrt{\epsilon}.
\end{aligned} \tag{11}$$

(In the next steps, we omit the terms $\epsilon + 2\sqrt{\epsilon}$ as they are negligible.) The most important step of this error analysis is to apply Sen's non-commutative union bound (Lemma 3 of Ref. [12]), which holds for any subnormalized state σ ($\sigma \geq 0$ and $\mathrm{Tr}\{\sigma\} \leq 1$) and sequence of projectors $\Pi_1, \ldots, \Pi_N$:

$$\mathrm{Tr}\left\{\sigma\right\} - \mathrm{Tr}\left\{\Pi_N\cdots\Pi_1\sigma\Pi_1\cdots\Pi_N\right\} \leq 2\sqrt{\sum_{i=1}^{N}\mathrm{Tr}\left\{\left(I-\Pi_i\right)\sigma\right\}}$$

For our case, we take $\Pi\phi_{X^n(M)}\Pi$ as σ and $\phi_{X^n(M)}$, $\hat{\Pi}_{M-1}$, $\ldots$, $\hat{\Pi}_1$ as the sequence of projectors. Applying Sen's bound and concavity of the square root function leads to the following upper bound on (11):

$$2\sqrt{\mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}_M\Pi\phi_{X^n(M)}\Pi\right\} + \mathbb{E}\sum_{i=1}^{M-1}\mathrm{Tr}\left\{\phi_{X^n(i)}\Pi\phi_{X^n(M)}\Pi\right\}}$$

where $\hat{\Pi}_M = I^{\otimes n} - \phi_{X^n(M)}$ and $\phi_{X^n(i)} = I^{\otimes n} - \hat{\Pi}_i$. We now bound each of the above two terms individually. For the first term, consider that

$$\begin{aligned}
&\mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}_M\Pi\phi_{X^n(M)}\Pi\right\} \\
&\leq \mathbb{E}\mathrm{Tr}\left\{\hat{\Pi}_M\phi_{X^n(M)}\right\} + \mathbb{E}\left\|\phi_{X^n(M)} - \Pi\phi_{X^n(M)}\Pi\right\|_1 \\
&\leq 2\sqrt{\epsilon}.
\end{aligned}$$

where the last inequality follows from applying (9) and because

$$\begin{aligned}
\mathrm{Tr}\left\{\hat{\Pi}_M\phi_{X^n(M)}\right\} &= \mathrm{Tr}\left\{\left(I^{\otimes n} - \phi_{X^n(M)}\right)\phi_{X^n(M)}\right\} \\
&= 0.
\end{aligned}$$

For the second term, consider that

$$\begin{aligned}
&\mathbb{E}\sum_{i=1}^{M-1}\mathrm{Tr}\left\{\phi_{X^n(i)}\Pi\phi_{X^n(M)}\Pi\right\} \\
&\leq \mathbb{E}_M\sum_{i\neq M}\mathbb{E}_{X^n}\mathrm{Tr}\left\{\phi_{X^n(i)}\Pi\phi_{X^n(M)}\Pi\right\} \\
&= \mathbb{E}_M\sum_{i\neq M}\mathrm{Tr}\left\{\mathbb{E}_{X^n}\left\{\phi_{X^n(i)}\right\}\Pi\mathbb{E}_{X^n}\left\{\phi_{X^n(M)}\right\}\Pi\right\} \\
&= \sum_{i\neq M}\mathrm{Tr}\left\{\rho^{\otimes n}\Pi\rho^{\otimes n}\Pi\right\} \\
&\leq 2^{-n[H(\rho)-\delta]}\sum_{i\neq M}\mathrm{Tr}\left\{\rho^{\otimes n}\Pi\right\} \\
&\leq 2^{-n[H(\rho)-\delta]}\ \ |\mathcal{M}|
\end{aligned}$$

The first inequality follows by just adding in all of the future terms $i > M$ to the sum. The first equality follows because the random variables $X^n(i)$ and $X^n(M)$ are independent, due to the way that we selected the code (each codeword is selected independently of a different one). The second equality follows from averaging the state ϕ_{X^n} with respect to the distribution p_{X^n}, and we drop the expectation $\mathbb{E}_M$ because the quantities inside the trace no longer have a dependence on the message M. The second inequality follows from the entropy bound for the eigenvalues of $\rho^{\otimes n}$ in the typical subspace. The final inequality follows because $\mathrm{Tr}\{\rho^{\otimes n}\Pi\} \leq 1$.

Thus, the overall upper bound on the error probability with this sequential decoding strategy is

$$\epsilon' \equiv \epsilon + 2\sqrt{\epsilon} + 2\sqrt{2\sqrt{\epsilon} + 2^{-n[H(\rho)-\delta]}\ \ |\mathcal{M}|},$$

which we can make arbitrarily small by choosing $|\mathcal{M}| = 2^{n[H(\rho)-2\delta]}$ and n sufficiently large. The next arguments are standard. We proved a bound on the expectation of the average probability, which implies there exists a particular code that has arbitrarily small average error probability under the same choice of $|\mathcal{M}|$ and n. For this code, we can then eliminate the worst half of the codewords, ensuring that the error probability of the resulting code is no larger than $2\epsilon'$. Furthermore, it should be clear that it is only necessary for the sequential decoder to process the remaining codewords when decoding messages. ■

Remark 10: Sen's proof applies to the more general case of classical-quantum channels $x \to \rho_x$, with ρ_x a mixed state, by employing conditionally typical projectors [12]. For pure-state classical-quantum channels, the conditionally typical projector is just the pure state itself, and the proof simplifies as seen above.