A Review of the Quantum Reverse Shannon Theorem

Theodore Rogozinski
Department of Physics, Institute for Quantum Computing
Waterloo, ON

(Dated: March 19, 2018)

The Quantum reverse Shannon Theorem is an interesting result in quantum information which allows one to simulate a quantum channel using another channel of greater entanglement assisted capacity. We present a simplified explanation of M. Berta's the One-Shot Quantum Reverse Shannon Theorem. We begin with a brief explanation of the Classical Reverse Shannon Theorem followed by an overview of the Quantum Reverse Shannon Theorem (QRST). We collect the major concepts required to prove the theorem in one place and give a proof outline of the QRST. The purpose of this paper will be to bridge the gap between [3] which may be too technical for beginners and [5] which may be too sparse.

I. INTRODUCTION

First published by Bennett et al. in [1], the quantum reverse Shannon Theorem (QRST) is an exciting new development in Quantum Information Theory. Where Shannon's Noisy Channel Coding Theorem gives a prescription for how much information can be sent reliably via a noisy channel, the classical reverse theorem uses a noiseless channel and a reservoir of shared randomness in order to simulate noisy communication between 2 parties. The QRST accomplishes a similar task to this, but we must substitute shared randomness with quantum entanglement in the form of embezzlement states (definition III.3 on page 4).

The QRST was conjectured in 2001 in [2] when Bennett et al. proved that the entanglement assisted capacity of a quantum channel takes a similar form to the classical capacity of a channel. Because of the similarities between the classical and quantum cases, we will review the classical case as well.

We begin with a proof outline of the Classical Reverse Shannon Theorem as given in [1], we then follow up with some definitions and intuitive explanations about the concepts which contribute to the QRST. This will lead up to giving an outline of the proof of the QRST as given in [5]. M. Berta et al. published an outline of the proof in [4], however that may be too sparse for the reader to gain meaningful information. It is the goal of this paper to bridge the gap between these two publications as a primer for reading [5].

II. CLASSICAL REVERSE SHANNON THEOREM

Throughout this section on classical communication, all messages will be assumed to be in bits.

A. Classical Shannon Theorem

In his 1948 paper Shannon defined the mutual information of 2 messages X and Y of length N and M respectively are

$$I(X:Y) = S(X) + S(Y) - S(XY) \tag{1}$$

where S is the Shannon entropy and the mutual entropy where applicable and X and Y are sequences of bits. We define a noisy channel $\Lambda_r : X \to \Lambda_r(X)$ with transmission rate r. In the noisy coding theorem, Shannon proved that a sender Alice (A) may transmit a message X of length N to a receiver Bob (B) through Λ_r with an error $\epsilon(N)$ such that $\epsilon \to 0$ as $N \to \infty$ if

$$r < C = \max_X I(X : \Lambda_r(X)). \tag{2}$$

In order to do this, Shannon employs error correction codes given in [9]. It is worth noting that if both A and B have access to the same string of random bits R then this does not improve either the size of the message they are able to send reliably or the rate at which they are able to send it. Also, if B is able to send a message to A, then A is still not able to improve the transmission rate. Both of these were proved by Shannon in [9], but will be omitted here for brevity.

In [2], along with the QRST, Bennett et al. gave a corollary to Shannon's result which is now known as the Classical Reverse Shannon Theorem (CRST). Let us assume that A and B both have a noiseless communication channel with $r = 0$. Then using a shared randomness R, A and B are able to effectively simulate a noisy channel using Λ_0.

It is easy to see that systems like the CRST might be useful for examples like a simplistic private key cryptography system; Shannon himself noted the possible usefulness of these systems for encryption [3] but never produced the CRST.

B. Protocol and Proof Outline

To gain an intuition as to how the QRST will work, it is useful for us to consider the CRST over a channel for bits. Because of this simplification we can use concepts like the Hamming distance (The number of deletions an exchanges required to turn one bit string into another) without losing generality of the type of messages that can be sent. The following definitions and protocol come from combining the protocols given in [2, pg. 36-39]. [1]

Definition II.1. *For a Channel N let the stochastic transition matrix be the matrix whose elements N_{x_0,y_0} give the probability that any bit $x_0 \in \{0,1\}$ is changed to $y_0 \in \{0,1\}$. Applying this matrix to every bit in our message $x \in \{0,1\}^n$ gives our output $y \in \{0,1\}^n$. We denote the probability of N mapping x to y in this way by $N_{x,y}$.*

If N is a discrete memoryless channel (DMT), this matrix completely characterizes N. We assume that N is known by Alice so she can simulate the channel locally.

Theorem II.2 (CRST on Binary Channel)**.** *There exists a simulation protocol S_n that can simulate a channel N_n with capacity C for some message x of length n. Let R be a sequence of randomly chosen bits which are shared between the sender and the receiver (Alice and Bob respectively). let $\epsilon > 0$ be a constant. S_n simulates the channel N_n in the precise sense that*

$$\forall_{n,x,y}(S_n)_{x,y} = (N_n)_{x,y} \tag{3}$$

Denote the number of bits sent from Alice to Bob $m_n(x)$. There is a tight bound on m_n given by

$$\lim_{n\to\infty} \max_{x\in\{0,1\}^n} P(m_n(x) > n(C+\epsilon)) = 0 \tag{4}$$

Proof Outline. For each string s, let $Z(s)$ be the count the number of 0's in s. Let C_k be the channel capacity for an input with k zeros. We define the protocol via the following steps.

1. Before Alice receives the message, Alice uses her knowledge of N and the reservoir of shared randomness to form a set of $2^{n(C_k-\epsilon/2)}$ tuples $\mathcal{S}_k = \{(s, N(s))_{i,k} : Z(s) = k\}$ with s a random string of length n which she shares with Bob. She repeats this process until there is 1 set of tuples for each possible $Z(x)$. So then Alice and Bob have a list of outputs of N, for each value the entropy of the message can take.

2. Alice and Bob separate and Alice obtains the message x.

3. Alice calculates $Z(x)$ and sends $Z(x)$ to Bob using $o(n)$ bits of communication.

4. Alice runs the message through the Channel simulation, obtaining $N(x)$. [2]

5. Alice calculates the Hamming distance, $d = |x - N(x)|$.

6. Alice attempts to find an element in the prearranged set $\mathcal{S}_{Z(x)}$ such that $|x-y| = |x-N(x)| = d$.
   (a) If one exists, she sends Bob the index of the string $N(x')$ with 0 appended to the front. The message $0i$ is of order $n(C_{Z(x')} + \epsilon/2)$.
   (b) If multiple strings are found then she picks a random one and performs (a).
   (c) If there are no suitable strings she sends Bob the string $1N(x)$.

7. Bob either uses the index and $Z(x)$ to obtain $N(x')$ or uses $N(x)$.

Firstly, we must make plausible is that S_n is, in fact, a simulation. If Alice sends $1N(x)$ then it is trivial. In the other case Bob receives a string $N(x')$ which has a hamming distance equal to that of $N(x)$. The hamming distance d measures the number of errors so strings, so strings with the same hamming distance as x are equally likely. Thus, we always produce a string which has equal likelihood of being produced as $N(x)$ satisfying equation (3).

Next, we look at the claim equation (4). We know that so long as Alice does not have to send $N(x')$, equation (4) is satisfied. So we must make plausible that the set $\mathcal{S}_{Z(x)}$ contains a string with $|x - N(x')| = d$ as $n \to \infty$. This can proved with a straightforward but calculationally strenuous counting argument by counting the number of errors and the probability that one of the guesses in step 4 will have the correct hamming distance. Then we can see that as $n \to \infty$, $2^{n(C_k-\epsilon/2)}$ guesses will be sufficient for the probability of failure to go to 0. □

C. Related Results

Corollary II.2.1 (Simulate Noise with Noisy Channel)**.** *Let N_1 and N_2 be channels with capacities C_1 and C_2 respectively. With an infinite reservoir of shared information Alice and Bob may simulate N_2 with N_1 if and only if $C_2 \leq C_1$ with unit asymptotic efficiency.*

This can be justified by noting that equation (2) on the preceding page gives the capacity as a maximum over possible inputs. By simulating noise locally as in theorem II.2 we are only limiting the amount of messages

[1] If the reader finds these proof outlines unsatisfactory they are referred to the mathematically precise proofs given in [1, pg. 15-19] although they use the theory of types which has been explicitly avoided here because of its complexity.

[2] One might think that Alice would simply send this message to Bob but that would require more communication between them than promised.

that can be sent; thus the ability of N_1 to send simulated randomness $C_{1,R_{N_2}}$ is less than C_1. However, if $C_2 > C_1$ then information must necessarily be lost when transmitting the message over N_1, so we must have $C_1 \leq C_2$.

From this we can define the capacity of N_1 to simulate N_2 in the presence of infinite shared randomness via the simple formula

$$C_R = \frac{C_1}{C_2}. \tag{5}$$

A similar corollary exists for the quantum case and follow directly from the use of entangled capacity and the QRST.

III. INGREDIENTS FOR THE QUANTUM REVERSE SHANNON THEOREM

A. General Definitions

Let $P(\mathcal{H})$ denote the set of positive semi-definite operators on $\mathcal{H}$ and take $S_{=}(\mathcal{H}) = \{\rho \in P(\mathcal{H}) : \mathrm{tr}(\rho) = 1\}$. We define $S_{\leq}(\mathcal{H})$ analogously so that we may express certain tensor products more simply and avoid normalization in some cases until it is needed. For $\rho \in S_{=}(\mathcal{H})$ define the Von Neumann entropy S as

$$S(\rho) = -\,\mathrm{tr}(\rho \log(\rho)) \tag{6}$$

and the mutual information of a density matrix $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ as

$$I(\rho_{AB}) = I(A:B)_\rho = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \tag{7}$$

For notational brevity we will also use $|A| = \dim(\mathcal{H}_A)$, $\rho_A = \mathrm{tr}_A(\rho_{AB})$, and $|\phi\rangle\langle\phi|_A = \phi_A$ when applicable. The fidelity of 2 density matrices ρ and σ

$$F(\rho, \sigma) = \mathrm{tr}(\sqrt{\rho, \sigma}) \tag{8}$$

can be used in the definition of fidelity for a quantum channel

$$F(\mathcal{E}, \rho) = \inf_{|\phi\rangle_{\rho,R}} \{F(|\phi\rangle_{\rho,R}, (\mathcal{E} \otimes \mathcal{I}_R)\,|\phi\rangle_{\rho,R})\}. \tag{9}$$

Here $|\phi\rangle_{\rho,R}$ is a purification of ρ with respect to some reference system R and $\mathcal{I}_R$ is the identity operator over R.

In contrast to the classical theory, the capacity of quantum channels cannot be characterized by a single parameter. It was only in 2001 when Bennet et al. [2] found that if unlimited entanglement[3] between Alice and Bob is allowed the capacity or the so called takes a familiar form

$$\begin{aligned} C_E(N) &= \max_{\rho \in \mathcal{H}_{in}} S(\rho) + S(N(\rho)) - S((N \otimes \mathcal{I})(\phi_\rho)) \\ &= I(A:B)_\rho. \end{aligned} \tag{10}$$

Here $\mathcal{H}_{in}$ is the Hilbert space of all possible density matrices that Alice can send and the quantity $(N \otimes I)(\phi_\rho)$ is a purification of Bob's half of the initial entangled state. In other words, for some reference Hilbert Space $\mathcal{H}_{ref}$ we have $\phi_\rho = \mathcal{H}_{in} \otimes \mathcal{H}_{\mathrm{ref}}$ and we run Alice's portion of the pure state through N and do nothing to Bob's portion. Also note that we have shifted over from the Shannon Entropy to the Von Neumann entropy. Although equation (10) is a statement about quantum channels it is a capacity with respect to classical bits, so it has the same meaning as the capacities discussed in section II on page 1.

The quantum relative entropy for $\rho \in S_{\leq}(\mathcal{H})$ and $\sigma \in P(\mathcal{H})$

$$D(\rho||\sigma) = H(\rho) - \mathrm{tr}(\rho \log(\sigma)), \tag{11}$$

the mutual information can be extended for $\rho \in S_{\leq}(\mathcal{H})$ without change and extend the definition

$$I(A:B)_\rho = D(\rho_{AB}||\rho_A \otimes \rho_B). \tag{12}$$

Here we see our first use of $S_{\leq}(\mathcal{H})$ in this rather simple formula.

Although many of these definitions will not be used in this proof they are abundant in [3] and thus will be very useful if a more rigorous proof is desired.

B. The Tools of One-Shot Proofs

In classical and quantum information theory, quantities are often defined and developed around notions of arbitrarily large communication. However, when dealing with one-shot theorems, we wish to find equivalent notions for a single communication like the relative max entropy

$$D_{max}(\rho||\sigma) = \inf\{\lambda \in \mathbb{R} : (2^\lambda \sigma - \rho) \in P(\mathcal{H})\}. \tag{13}$$

where $\rho \in S_{\leq}(\mathcal{H})$ and $\sigma \in P(\mathcal{H})$. We also define the conditional min-entropy where $\rho_{AB} \in S_{\leq}(\mathcal{H})$ as

$$H_{min}(A|B)_\rho = -\inf_{\sigma_B} D_{max}(\rho_{AB}||\mathcal{I}_A \otimes \sigma_B) \tag{14}$$

These are a subset of entropy measures for one-shot systems known as As one can see, these measures of entropy depend on extremizing a quantity over a set of states. If we are going to make statements about the space where we take extremes, it is highly useful to have a notion of distance. This can be obtained via the generalized fidelity

$$\tilde{F}(\rho, \sigma) = F(\rho, \sigma) + \sqrt{(1 - \mathrm{tr}(\rho))(1 - \mathrm{tr}(\sigma))}. \tag{15}$$

[3] In this case Bennett et al. meant a pair of shared Bell states, but we will use this result to prove things about embezzlement states later.

where $\rho, \sigma \in S_{\leq}(\mathcal{H})$. From this we obtain the notion of distance given by

$$P(\rho, \sigma) = \sqrt{1 - \tilde{F}^2}. \tag{16}$$

In this paper, when we make statements about equality it is generally done in terms of this distance operator. It is clear to see that in the case of $\mathrm{tr}[\rho] = 1$ these definitions simplify to the normalized equivalent. If $\rho \approx_\delta \sigma$ then $\tilde{F}(\rho, \sigma) < \delta$ and visa versa.

Lemma III.1 (Bound on Purified Trace distance)**.**

$$\frac{1}{2}||\rho - \sigma||_1 \leq P(\rho, \sigma) \tag{17}$$

Because we will be making statements about channels, it will be useful to have a notion of distance on channels as well.

Definition III.2. *(Diamond Norm) The Diamond norm is a norm of CPTP maps defined by*

$$||\mathcal{E}||_\diamond = \sup_\rho ||(\mathcal{E} \otimes \mathcal{I}_K)(\rho)||_1 \tag{18}$$

and this norm induces a distance measure on the space of CPTP maps called the Diamond Distance given by

$$D_\diamond(\mathcal{E}, \mathcal{F}) = ||\mathcal{E} - \mathcal{F}||_\diamond \tag{19}$$

Where $\mathcal{E}$ and $\mathcal{F}$ are quantum channels (CPTP maps) and $\rho \in S_{\leq}(\mathcal{H})$.

Proof of the norm axioms will be omitted for brevity however one may find a great explanation in [12]. [4] As with many of the measures in one-shot quantum information, the presence of a maximum in this formula can be quite cumbersome. Thus we will endeavor to find a bound for this norm which we can use to bound the Diamond Distance. This is the purpose of the Post-Selection technique found in subsection III D on page 6. [5]

C. Quantum State Splitting

One of the biggest hurdles of proving the QRST is the proper use of entanglement. Normally, entanglement is shared between Alice and Bob in the form of ebits (i.e. shared Bell states). However, [1] proved that these states are not sufficient to prove the QRST because entanglement cannot be discarded without excessive communication or introducing error.[6] The Quantum State Merging (QSM) algorithm given by [5] solves this problem by using embezzlement states and local quantum operations to reduce the amount of communication needed between Alice and Bob.

Definition III.3. *We define the embezzlement state of index k as*

$$|\mu(k)\rangle_{AB} = \frac{1}{\sqrt{G(k)}} \sum_{j=1}^{k} \frac{1}{\sqrt{j}} |jj\rangle_{AB} \tag{20}$$

where

$$G(k) = \sum_{j=1}^{k} \frac{1}{j}. \tag{21}$$

We have defined these embezzlement states because of their ability to produce entanglement in other states via the following theorem proved in [11].

Proposition III.4 (Entanglement Without Communication)**.** *Let $|\phi\rangle_{AB}$ be an ebit with Schmidt rank m. For $\epsilon > 0$ and the local transformations $X_{A\to AA'}$ and $X_{B\to BB'}$,*

$$(X_{A\to AA'} \otimes X_{B\to BB'})\mu(k)_{AB}(X_{A\to AA'} \otimes X_{B\to BB'})^\dagger \approx_\delta \mu(k)_{AB} \otimes |\phi\rangle\langle\phi|_{AB} \tag{22}$$

as $k \to \infty$ and with $\delta > 0$.

Proof Outline. The proof of this relies on showing that there exists a state $|w(k)\rangle$ which both approximates $|\mu\rangle_k$ in the sense that as $k \to \infty$, $|\langle\mu(k)_{AB}|w(k)\rangle| \to 1$ and $|w(k)\rangle$ can be re-arranged into $|\mu(k)\rangle_{AB} \otimes |\phi\rangle_{AB}$ via local linear operations[7]. This will leave the embezzlement state unchanged in the presence of infinite entanglement and give us the Bell state we required by local operations. The full proof of this is delegated to [11]. This process can be repeated to produce N ebits with error $N\delta$. □

Since we assume that we have "unlimited entanglement", we generally take this to mean that this error is 0.[8]

In the following subsection we describe Berta's proof of Quantum State Splitting with embezzling states. This technique uses entanglement and quantum communication in order to change two bipartite states $\rho_{\mathcal{H}_A} \otimes \rho_{\mathcal{H}_B}$ shared by Alice and Bob into a general state $\rho_{\mathcal{H}_A\mathcal{H}_B}$ for relatively small communication costs.[9] We will use this later as a method of transferring a density matrix $\mathcal{E}(\rho_A)$ which Alice has simulated to Bob.

Theorem III.5 (Quantum State Splitting with Embezzling States [5])**.** *Let $\epsilon > 0, \epsilon' \geq 0, \delta > 0$ and $\rho_{AA'R} \in S_{\leq}(\mathcal{H}_{AA'R})$ be a pure state. A Completely Positive Trace Preserving (CPTP) map $\mathcal{E}$ is called a One-Shot State-Splitting Protocol of $\rho_{AA'R}$ if it allows only*

[4] Watrous calls the diamond norm the 'bounded trace norm' in these course notes.

[5] A complete, yet sparse proof can be found in [6]

[6] A proof of this has been omitted here in favor of brevity.

[7] specifically by re-arranging basis vectors in the Schmidt composition of $|\mu(k)\rangle_{AB} \otimes |\phi\rangle_{AB}$

[8] [3] has a much more rigorous treatment of this.

[9] State Splitting is often described in [5] as the dual to quantum state merging. A process which essentially accomplishes the inversion of state merging.

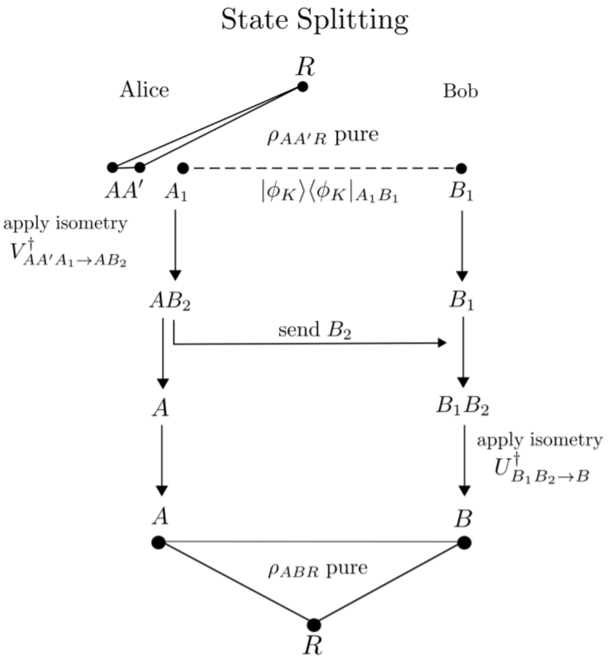


Figure 1: A Graphical representation of the state splitting process. Alice starts with systems A A' and A_1 and uses local isometries to create the entangled part of Bob's state. She sends this to Bob who uses local isometries to merge these two subsystems to create B. Credit for this figure goes to [5].

local operations for Alice and Bob using a δ-embezzling state μ_{AB} and communication of

$$q \le \frac{1}{2} I^{\epsilon'}_{max}(A:R)_\rho + 2 * \log\frac{1}{\epsilon} + 4 + \log\log|A'| \qquad (23)$$

qubits to produce

$$(\mathcal{E}\otimes I_R)(\rho_{AA'R}\otimes\mu_{AB}) \approx_\epsilon \rho_{ABR}. \qquad (24)$$

Where A' is the part of Alice's system which is decoupled from the reference system R and sent to Bob via some identity channel $I_{A'\to B}\otimes I_{AR}$. ρ_{ABR} is a pure state.

Because of the complicated nature of this proof, at times the reader will be referred to [5, pg. 5-13] for concrete proofs of lemmas which will be made plausible here.

Proof. This protocol can be broken up into several steps

1. Decompose $\rho_{AA'R}$ into its eigenvectors while ignoring the eigenvectors whose eigenvalues are small. The reader may recognize this as a sort of approximate spectral decomposition.

2. Categorize these eigenvectors into classes based on the size of their eigenvalues and label these classes by i. Let $|\rho^i\rangle$ be the linear combination of all the eigenvectors in class i weighted by their eigenvalues.

3. Alice uses her embezzlement states to create the appropriate amount of Bell states for each of the new vectors created.

4. Alice uses a local isometry to rearrange the states ρ^i and the Bell states into the system that she wants to send to Bob called $\rho^i_{B_2}$.

5. Alice sends the resulting systems $\rho^i_{B_2}$ to Bob along with the information as to how to use his embezzlement states to create the other half of the Bell states the Alice created.

6. Now that Bob has these systems he can use his embezzlement states in order to create the proper Bell states.

7. Bob applies a local isometry on his Bell states and $\rho^i_{B_2}$ to construct $\tilde\rho_B$. $\tilde\rho_B$ is a state which approximate the spectral decomposition of ρ_B while leaving off small eigenvalues as in step 1. Thus, our whole system when purified will be ρ_{ABR}.

Because of the simplicity of the proof, we will show that steps 1 and 2 create an acceptable approximation to ρ.

Let $P_{A',i}$ be the projectors onto the eigenstates of A' with eigenvalues in $\chi_i = [2^{-(i+1)}, 2^{-i}]$ where $0 \le i \le i_{max} = \lceil 2\log|A'| - 1\rceil$ and let P be the projector onto all the other eigenvectors.[10] Let E_i be the set of eigenvalues in χ_i. It can be easily seen via spectral decomposition that

$$\mathrm{tr}(P_{A',i}\rho_{A'}) = \sum_{x\in E_i} x = p_i \qquad (25)$$

Now we define $|\rho^i\rangle = p_i^{-1/2} P^i_{A'} |\rho\rangle_{AA'R}$ and

$$|\bar\rho\rangle_{AA'R} = \frac{1}{\sqrt{\gamma}} \sum_{i=0}^{i_{max}} \sqrt{p_i}\, |\rho^i\rangle_{AA'R} \qquad (26)$$

where γ is simply a normalizing factor. Next we calculate

$$P(\bar\rho_{AA'R}\rho_{AA'R}) = \sqrt{1 - \sum_{i=0}^{i_{max}} p_i} = \sqrt{p} \qquad (27)$$

where $p = \mathrm{tr}(P\rho_{A'})$. It is easy to see from equation (25) that

$$p = \sum_{x\in E_i} x \le \sum_{x\in E_i} 2^{-2log|A'|} = |A'|2^{-2log|A'|} = \frac{1}{|A'|} \qquad (28)$$

[10] The ones whose eigenvalues are in $[2^{-2\log|A'|}, 0]$ because the sum of the eigenvalues of $\rho_{A'}$ must be less than 1.

So now we have that

$$\rho_{AA'R} \approx_{|A'|^{-\frac{1}{2}}} \bar{\rho}_{AA'R} \tag{29}$$

exactly as in [5, pg. 9].

The next step in the proof is to characterize the cost of sending these states from Alice to Bob. This process is rather lengthy and is best left to the original proof. In [3] Berta notes that this process is the inverse of quantum state merging, a more well known process. By using this process in reverse for each of the parts of the system ρ^i he obtains the quantum communication cost

$$\begin{aligned} q = \max_i \left\lceil \frac{1}{2}(H_0(A')_{\rho^i} - H_{min}(A'|R)_{\rho^i}) + 2\log\frac{1}{\epsilon} \right\rceil \\ + \log\lceil 2\log|A'|\rceil. \end{aligned} \tag{30}$$

This can then be simplified to get the final bound on the communication

$$\begin{aligned} q \leq \max_i \left\lceil \frac{1}{2} I^{\epsilon'}_{max}(B:RR')_{(\mathcal{E}^{\otimes n}\otimes\mathcal{I})\zeta^n)} \right\rceil \\ + 2\log\frac{1}{\epsilon'} + 4 + \log n + \log\log|B|. \end{aligned} \tag{31}$$

□

D. The Post-Selection Technique

The Post-Selection Technique (PST) was first given in 2008 by Christandl et al. in [6]. The PST allows one to, in some sense, determine the probability that two channels $\mathcal{E}$ and $\mathcal{F}$ will be distinguishable from one another by testing their action on a purification of DiFinetti states.

Definition III.6 (Difinetti States [6]). *Define the DiFinetti State ζ^n as follows where $\sigma \in S_=(\mathcal{H})$*

$$\zeta^n = \int \sigma^{\otimes n}\mu(\sigma) \in S_=(\mathcal{H}). \tag{32}$$

and μ is the probability measure over states.

It is not important for the reader to understand this formula completely, however it is important to see that this state has symmetry in exchanging σ meaning it is symmetric under permutations.

The following proof is a simplification of the proof of the Post Selection technique that can be found in [6]. For our purposes it is sufficient to define the DiFinetti state to be the state which satisfies this property.

Lemma III.7. *Let $\Pi_{\mathcal{H}}$ be the set of all permutation maps on a Hilbert Space $\mathcal{H}$. Define $Sym(\mathcal{H}) = \{\rho \in \mathcal{H} : \pi(\rho) = \rho \ \forall\pi \in \Pi_{\mathcal{H}}\}$. Now take $\mathcal{K} \cong \mathcal{H}$ and let $\rho_{\mathcal{H}^n\mathcal{K}^n} \in Sym(\mathcal{H}\otimes\mathcal{K})$ and we will have that the for a purification of the DiFinetti State $\zeta^n_{\mathcal{H}^n\mathcal{K}^n\mathcal{R}}$*

$$\rho_{\mathcal{H}^n\mathcal{K}^n} = g_{n,d}(\mathcal{I}_{\mathcal{H}^n\mathcal{K}^n}\otimes\mathcal{T})(\zeta^n_{\mathcal{H}^n\mathcal{K}^n\mathcal{R}}). \tag{33}$$

Here R is simply a reference system which purifies $\zeta^n_{\mathcal{H}^n\mathcal{K}^n\mathcal{R}}$ and $\mathcal{T}$ is a CPTP which does not increase the trace norm. $g_{n,d}$ is a constant which depends only on n and $d = \dim(\mathcal{H})$.

Next, we give a statement and proof outline of the Post Selection Technique for bounding the Diamond Norm

Theorem III.8 (Post Selection [6]). *Let $\mathcal{E} \in \mathcal{H}$ have there property that $\forall\pi \in \Pi_{\mathcal{H}} \ \exists K_\pi \in \mathcal{K} \cong \mathcal{H}$ such that $\mathcal{E}(\pi(\rho)) = K_\pi(\mathcal{E}(\rho))$ for all density matricies ρ. Then we have that*

$$||\mathcal{E}||_\diamond \leq g_{n,d}||(\mathcal{E}\otimes\mathcal{I}''_R)(\zeta_{\mathcal{H}R''})||_1. \tag{34}$$

Where $d = \dim\mathcal{H}$ and R'' is a purifying reference system.

Proof. Define

$$\bar{\rho}_{\mathcal{H}^nRR'} = \frac{1}{n!}\sum_\pi(\pi\otimes\mathcal{I}_R)(\rho_{\mathcal{H}^nR})\otimes|\pi\rangle\langle\pi|_{R'}. \tag{35}$$

Where the summation is over all the possible permutations and $|\pi\rangle$ are simply orthogonal basis vectors in the reference space chosen to have the correct dimension. Note that $\bar{\rho}_{\mathcal{H}^nRR'}$ can be purified using R and R'. It is easy to see that this new density matrix will be independent of permutations and thus, so are its partial traces. By inspection we note that $\bar{\rho}_{\mathcal{H}^n}$ will be permutation invariant. Thus we may purify it using elements of $\mathcal{K}$ so that it can be written in terms of matrices in $Sym((\mathcal{H}\otimes\mathcal{K})^{\otimes n})$. Because $\bar{\rho}_{\mathcal{H}^n}$ corresponds to 2 separate purifications we know there is a local isometry between the two which we will label $(\mathcal{I}_{\mathcal{H}}\otimes\mathcal{G})$. All that is left is to make an inequality chain and use lemma III.7.

$$\begin{aligned} &||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_R)\rho_{\mathcal{H}R}||_1 \\ &= ||((\mathcal{E}_{\mathcal{H}^n}\circ\pi)\otimes\mathcal{I}_R)\rho_{\mathcal{H}R}||_1 \\ &= \frac{1}{n!}\sum_\pi ||((\mathcal{E}_{\mathcal{H}^n}\circ\pi)\otimes\mathcal{I}_R)\rho_{\mathcal{H}R}||_1 \end{aligned} \tag{36}$$

By the definition of $\bar{\rho}_{\mathcal{H}^nRR'}$ we have

$$\begin{aligned} &||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_R)\rho_{\mathcal{H}R}||_1 \\ &= ||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_R)\bar{\rho}_{\mathcal{H}^nRR'}||_1 \\ &= ||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{G})\bar{\rho}_{\mathcal{H}^n\mathcal{K}^n}||_1 \\ &\leq ||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_{\mathcal{K}^n})\bar{\rho}_{\mathcal{H}^n\mathcal{K}^n}||_1 \end{aligned} \tag{37}$$

where the last one in the chain is because CPTP maps cannot increase the trace norm [12]. Finally we apply lemma III.7 to obtain

$$\begin{aligned} &||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_{\mathcal{K}^n})\bar{\rho}_{\mathcal{H}^n\mathcal{K}^n}||_1 \\ &= g_{n,d}||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{T}_{\mathcal{K}^nR''})\zeta_{\mathcal{H}^n\mathcal{K}^nR''}||_1 \\ &\leq g_{n,d}||(\mathcal{E}_{\mathcal{H}^n}\otimes\mathcal{I}_{\mathcal{K}^nR''})\zeta_{\mathcal{H}^n\mathcal{K}^nR''}||_1 \end{aligned} \tag{38}$$

Where the last step here is the same as the last step in equation (37). □

E. Symmetrization of Quantum Channels

In our proof of theorem III.8 on the previous page we used the fact that our channel $\mathcal{E}$ is permutation invariant. Of course, this does not apply to all quantum channels so we must employ a technique for making a channel permutation invariant in the sense given in theorem III.8 on the preceding page. This technique is called Symmetrization and was first given in [6].

This is done by applying a random permutation to an input of the channel $\mathcal{E}$ which is known by both Alice and Bob. Alice and Bob may generate shared randomness by creating ebits using embezzlement states and measuring each state. We want $\mathcal{E}$ to satisfy the following condition as given in theorem III.8 on the previous page

$$\mathcal{E}(\pi(\rho)) = K_\pi(\mathcal{E}(\rho)) \tag{39}$$

with the same notation. Let us assume that $\mathcal{E}$ is not permutation invariant and split it up into $\mathcal{E} = \pi_{\mathcal{E}} \circ \tilde{\mathcal{E}}$ where $\tilde{\mathcal{E}}$ is permutation invariant and $\pi_{\mathcal{E}}$ is the permutation done by $\mathcal{E}$. Next, introduce the random permutation $\bar{\pi}$ generated from shared randomness between Alice and Bob and we obtain

$$\pi_{\mathcal{E}}(\tilde{\mathcal{E}}(\bar{\pi}(\rho))) = \pi_{\mathcal{E}}(K_{\bar{\pi}}(\tilde{\mathcal{E}}(\rho))) = G_{\bar{\pi},\pi_{\mathcal{E}}}(\pi_{\mathcal{E}}(\tilde{\mathcal{E}}(\rho)) \tag{40}$$

by the symmetry of $\tilde{\mathcal{E}}$ where $G_{\pi,\pi_{\mathcal{E}}}(\rho) = \pi_{\mathcal{E}}(K_{\bar{\pi}}(\pi_{\mathcal{E}}^{-1}(\rho)))$. Simplifying, we get

$$\mathcal{E}(\bar{\pi}(\rho)) = G_{\bar{\pi},\pi_{\mathcal{E}}}(\mathcal{E}(\rho)) \tag{41}$$

This equation is of the form equation (39) and thus satisfies symmetrization. Note that Bob may reconstruct $G_{\bar{\pi},\pi_{\mathcal{E}}}$ so long as he uses the shared random bits and has knowledge of $\pi_{\mathcal{E}}$.

F. Carathéodory's theorem and DiFinetti States

The following is taken from lemma D.5 and D.6 of [5, pg. 29].

The Post Selection Technique is an extremely powerful tool for proving bounds on quantum channels. However, in the original statement of the QRST Bennet et al. put thier bound in terms of a maximum over a set of states. In order to convert back from purification of DiFinetti states to a maximum over a set of states Berta uses Carathéodory's theorem.

Theorem III.9 (Carathéodory's Theorem [7]). *let $Conv(P)$ be the convex hull of P a set of points in $\mathbb{R}^n$ space. For $x \in Conv(P)$ we may find $Conv(P') \subset Conv(P)$ such that $x \in Conv(P')$*

This is an elementary result in convexity theory so the proof will be delegated to [7].

Lemma III.10. *For the a purification of the DiFinetti state, we may write*

$$\zeta_{AR}^n = \sum_{i=0}^{(n+1)^{2|A||R|-2}} p_i * (\omega^i)_{AR}^{\otimes n} \tag{42}$$

where $(\omega^i)_A R^{\otimes n} \in S_=(\mathcal{H})$ and the $0 < p_i$, $\sum_i p_i = 1$ are a probability distribution.

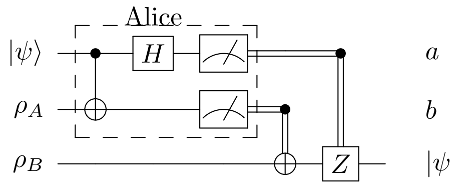
Proof. By inspection one may see that $\zeta_{AR}^n \in Sym((H_A \otimes H_R)^{\otimes n})$. From [6, pg. 3] we see that the dimension of this set is bounded by $|Sym((H_A \otimes H_R)^{\otimes n})| \leq (n+1)^{|A||R|-1}$. Because this space is complex it is isomorphic to a real dimensional space of $(n+1)^{2|A||R|-2}$ dimensions. Let ξ_{AR}^n be the equivalent of normalized density matrices in the real space we have just constructed. By theorem III.9 we may write real space analog of ζ_{AR}^n as a sum of ξ's. Thus, we have that we may write equation (42). □

G. Quantum Teleportation

It is fairly easy to see that quantum state splitting will take care of the communication costs for the QRST. However Bennett et al. [2] found that C_E contains all the interesting properties. Thus, we must convert our quantum communication costs into classical communication costs using a well known technique: Quantum Teleportation. A short review is given below [11]

Theorem III.11 (Quantum Teleportation). *We may send one qubit $|\psi\rangle$ from Alice to Bob in exchange for 2 qubits of classical communication and 1 Bell state.*

Proof. Let ρ_A be the half of the Bell state held by Alice and similarly for Bob. The algorithm is given by the following simple quantum circuit.



Where the dashed box represents all the operations on Alice's side and outside the box are operations on Bob's side. It can be shown by simple matrix analysis that this circuit will yield $|\psi\rangle$ on Bob's side of communication. □

By proposition III.4 on page 4, embezzlement states allow Alice and Bob to create as many Bell states as they need. So we may use theorem III.11 to trade quantum communication for double the classical communication ad infinitum.

[11] For full explanation see [8, pg. 26] for details.

H. Stinespring Dilation

This section will serve as a brief reminder of Stinespring's Dilation Theorem for the expression of a quantum channel. We may represent a quantum channel in its Kraus operator form

$$\mathcal{E}(\rho) = \sum_{i=0}^{m} A_i \rho A_i^\dagger. \tag{43}$$

Then it is easy to see that we may re-arrange these operators into block vectors

$$U^\dagger = (A_1 A_2 ... A_m) \tag{44}$$

So that we have

$$\mathcal{E}(\rho) = \mathrm{tr}_1(U \rho U^\dagger). \tag{45}$$

This is called the Stinespring form and it will be used in our proof of the QRST.

IV. THE QUANTUM REVERSE SHANON THEOREM

A proof of the Quantum Reverse Shannon theorem given in [5] will be reviewed in this section. The goal of this will be the review the proof without weighing down the concepts with technical lemmas.

Definition IV.1 (One-Shot Reverse Shannon Simulation)**.** *Let $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$ be a channel between the Hilbert spaces of Alice and Bob respectively. A CPTP map $\mathcal{P}$ with*

$$||\mathcal{E} - \mathcal{P}||_\diamond \leq \epsilon \tag{46}$$

for $\epsilon > 0$ is called a One Shot Reverse Shannon Simulation of $\mathcal{E}$ with error ϵ. This channel will make use of a δ-ebit embezzling state and classical communication c between Alice and Bob.

Theorem IV.2. *For $\mathcal{P}$ a One Shot Reverse Shannon Simulation of $\mathcal{E}$ error ϵ after an asymptotic number of uses*[12] *denoted $\mathcal{P}^{\otimes n}$ with*

$$\lim_{n\to\infty} \epsilon_n = 0 \tag{47}$$

where n is the number of uses of $\mathcal{P}$ and $\mathcal{E}$ and ϵ_n are the errors as defined in equation (46) with $\mathcal{E}$ and $\mathcal{P}$ replaced with $\mathcal{E}^{\otimes n}$ and $\mathcal{P}^{\otimes n}$ respectively. We will also have that

$$\lim_{n\to\infty} \frac{c_n}{n} \leq C_E \tag{48}$$

with C_E defined in equation (10) on page 3 as the capacity of $\mathcal{E}$ and c_n is the number of bits sent after n single uses of $\mathcal{P}$.

Proof Outline. First, we will construct a CPTP map $\mathcal{P}$ which satisfies equation (47) in the limit of repeated uses. In order to simplify the condition imposed by equation (46) we use the Post Selection Technique

$$||(\mathcal{E}^{\otimes n} - \mathcal{P}^{\otimes n} \otimes \mathcal{I}_{RR'})\zeta^n_{ARR'}||_1 < \epsilon g^{-1}_{n,k} \tag{49}$$

Where $\zeta^n_{ARR'}$ is a purification of the DiFinetti State tensored to itself n times and where R and R' are the reference spaces used to purify ζ^n. To show 47 we will look at the Stinespring form of $\mathcal{E}^{\otimes n}$ on ζ^n

$$\zeta^n_{BCRR'} = (U^n_{A\to BC} \otimes \mathcal{I}'_R R')\zeta^n_{ARR'}(U^n_{A\to BC} \otimes \mathcal{I}'_R R')^\dagger \tag{50}$$

Where we have simplified the action of the unitary on the reference system to the identity due to its lack of effect on the overall calculation. This step is exactly the process of Alice locally simulating the channel. Unfortunately, M. Berta does not give much explanation beyond this as to how the channel would be simulated locally. However, I would speculate that it would be akin to the classical case where some sort of matrix operation which is known by Alice is performed on her system.

From the proof of theorem III.5 on page 4 we get that it is possible to create a protocol to create $\mathcal{P}$ with the following trace distance

$$\begin{aligned} P((\mathcal{E}^{\otimes n} \otimes \mathcal{I}_{RR'})\zeta^n_{ARR'}, (\mathcal{P}^{\otimes n} \otimes \mathcal{I}_{RR'})\zeta^n_{ARR'}) \\ \leq \epsilon + \epsilon' + \delta n \log|B| + |B|^{n/2} \end{aligned} \tag{51}$$

and from lemma III.1 on page 4 we see that

$$\begin{aligned} ||((\mathcal{E}^{\otimes n} - \mathcal{P}^{\otimes n}) \otimes \mathcal{I}_{RR'})\zeta^n_{ARR'}||_1 \\ \leq 2 * (\epsilon + \epsilon' + \delta n \log|B| + |B|^{n/2}) \end{aligned} \tag{52}$$

Because these constants were chosen arbitrarily we may re-define them in terms of each other to get the following expression

$$\begin{aligned} ||((\mathcal{E}^{\otimes n} - \mathcal{P}^{\otimes n}) \otimes \mathcal{I}_{RR'})\zeta^n_{ARR'}||_1 \\ \leq \epsilon g_{n,d}. \end{aligned} \tag{53}$$

For a more rigorous treatment of this step see [5, pg. 15-16]. Thus, by the post selection technique (theorem III.8 on page 6) we have equation (49).

Next we worry about the communication costs of our protocol. The proof of these bounds requires many lemmas and actually constitutes a decent chunk of the lemmas presented in [5]. Rather than bog ourselves down in the details of how this is done a review will be given here so that the reader may more easily understand the proof given in [5].

We obtain from the state splitting process that

$$\begin{aligned} q_n \leq \frac{1}{2} I^{\epsilon'}_{max}(B : RR')_{(\mathcal{E}^{\otimes n} \otimes \mathcal{I})\zeta^n)} \\ + 2\log\frac{1}{\epsilon'} + 4 + \log n + \log\log|B| \end{aligned} \tag{54}$$

[12] A "use" of a channel appears to be loosely defined as sending a single density matrix across the channel. But the overall definition seems to be ambiguous and M. Berta's paper could benefit from some clarification on this matter.

is a bound on the quantum communication costs. We use quantum teleportation to put this in terms of the classical communication costs

$$c_n \leq I_{max}^{\epsilon'}(B:RR')_{(\mathcal{E}^{\otimes n}\otimes\mathcal{I})\zeta^n)} + 4\log\frac{1}{\epsilon'} + 8 + 2\log n + 2\log\log|B|. \quad (55)$$

Berta then uses several lemmas (See [3, pg. 16-17]) to prove that

$$c_n \leq N * \max_{\phi} I(B:R)_{(\mathcal{E}\otimes\mathcal{I})\phi} + O(\sqrt{N} + \sqrt{-\log\epsilon'}) \quad (56)$$

Where we have used lemma III.10 on page 7 to replace ζ with a maximum over $\phi \in S_=(\mathcal{H}_{AR})$. Thus when we take the limit as $n \to \infty$ we get

$$\limsup_{\epsilon'\to 0}\limsup_{n\to\infty}\frac{c_n}{n} \leq \max_{\phi} I(B:R)_{(\mathcal{E}\otimes\mathcal{I})\phi} = C_E \quad (57)$$

By the definition of I we have equation (48) on the preceding page and the theorem is proved. □

Now we justify the analogous result to corollary II.2.1 on page 2.

Corrolary IV.2.1. *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be channels with entanglement assisted capacities C_1 and C_2 respectively. With an infinite reservoir of shared information Alice and Bob may simulate $\mathcal{E}_2$ with $\mathcal{E}_1$ if and only if $C_2 \leq C_1$ with unit asymptotic efficiency.*

proof outline. Because the QRST uses classical capacities we may justify this by the same logic as before. We note that the capacity is a maximum of all the possible density matrices that can be sent over a channel. Then it follows from $C_2 \leq C_1$ that $\mathcal{E}_2$ can only transfer less information than $\mathcal{E}_1$ for any input. The QRST allows us to simulate any noise for low cost and thus, we may add the randomness into the channel required to simulate the channel $\mathcal{E}_2$. □

V. CONCLUSION

It is a testament to the conceptual simplicity of Berta's proof of the QRST that, once it's lemmas are proved, the overall argument can be reduced to a single page. The ability to simulate a quantum channel is a fairly surprising result given the diversity of errors that may arise in quantum communication. Because of this diversity of errors, it perhaps is not so surprising that there are many parts to the proof of the Quantum Reverse Shannon Theorem. Berta's proof in [5] makes the proof accessible, however it suffers from the use of a few lesser known procedures. Without explanations of each of these procedures readily available, his proof appears to lack simplicity. Hopefully, having a justification for all of these techniques in one paper has made this proof more comprehensible, perhaps at the cost of precision

Acknowledgments

I would like to thank Peirs Lillystone for his help in understanding the concepts involved in theorem III.8 on page 6. I would also like acknowledge Joshua Ruebeck for his review of this paper. Finally, I would like to thank the Institute for Quantum Computing and the University of Waterloo for their support which allowed for the writing of this paper and my continued research.

[1] Charles H Bennett, Igor Devetak, Aram W Harrow, Peter W Shor, and Andreas Winter. The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014.

[2] Charles H Bennett, Peter W Shor, John A Smolin, and Ashish V Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655, 2002.

[3] Mario Berta. Classical and quantum channel simulations. http://www.statslab.cam.ac.uk/biid2013/slides/Berta.pdf, 2013.

[4] Mario Berta, Matthias Christandl, and Renato Renner. A conceptually simple proof of the quantum reverse shannon theorem. In *Conference on Quantum Computation, Communication, and Cryptography*, pages 131–140. Springer, 2010.

[5] Mario Berta, Matthias Christandl, and Renato Renner. *A Conceptually Simple Proof of the Quantum Reverse Shannon Theorem*, pages 131–140. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[6] Matthias Christandl, Robert König, and Renato Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical review letters*, 102(2):020504, 2009.

[7] PM Gruber, JM Wills, and GM Ziegler. Handbook of convex geometry. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 98(4), 1996.

[8] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[9] Claude E Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[10] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min-and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, 2010.

[11] Wim van Dam and Patrick Hayden. Universal entanglement transformations without communication. *Phys.*

Rev. A, 67:060302, Jun 2003.

[12] John Watrous. Lecture 20: Channel distinguishability and the completely bounded trace norm, fall 2011.