Upper and Lower Bounds on Quantum Codes

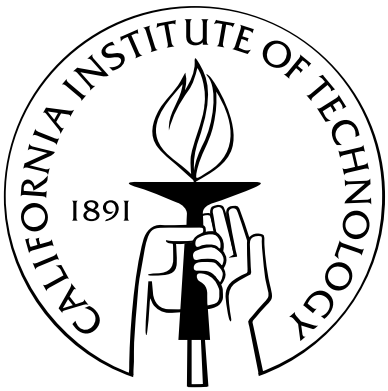
Thesis by

Graeme Stewart Baird Smith

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2006

(Defended May 8, 2006)

Dedicated to the Reader

Acknowledgements

It seems to me that the IQI has provided the best environment imaginable for learning a new subject—full of sharp, hard-working people, not to mention the constant flux of visitors. I'm grateful to my advisor, John Preskill, for fostering this environment and giving me the opportunity to be a part of it. He's also been the source of many thoughtful questions and suggestions.

Besides John, at least three others took on advisor-like roles for me: Debbie Leung, Patrick Hayden, and John Smolin have all had a lasting impact on me, both in terms of techniques and attitudes.

I'm also grateful to Igor Devetak, Robert McEliece, and John Smolin for agreeing to be on my thesis committee, to Alexei Kitaev and Barry Simon for being on my candidacy committee, and to Oskar Painter for being on both. Their comments and questions helped make this a better thesis.

My collaborators on the work contained in this thesis have taught me an awful lot. Chapter 3 is based on joint work with Debbie Leung, Chapter 4 on work with John Smolin, Chapter 5 on work with Joe Renes, Chapter 6 on work with John Smolin and Joe Renes, and Chapter 7 on work with John Smolin and Andreas Winter. I enjoyed working with them, and I'm proud of what we found.

While they're primarily my friends, Ben Toner, Carlos Mochon, Mike Zwolak, Paul Barclay, and Panos Aliferis also talked a lot of physics.

A lot of people who I've met professionally have also been quite friendly. Each one of Renato Renner, Mary-Beth Ruskai, Dave Bacon, Daniel Gottesman, Aram Harrow, Andrew Childs, Jon Yard, Frank Verstraete, Sergey Bravyi, Andrew Doherty, Matthias Christandl, Oscar Dahlsten, Guifre Vidal, Dominic Berry, Jonathan Oppenheim, Robin Blume-Kohout, Pawel Wocjan, Rob Spekkens, Richard Cleve, and Joseph Emerson took some time to explain something to me, or help figure something out. And during my summer at IBM, David DiVincenzo, Barbara Terhal, Roberto Oliveira, Charlie Bennett, and John Smolin all made me feel welcome, both scientifically and socially.

Even with all this help, I probably couldn't have done it without the support of my parents, my sister, Allyson, brother-in-law, Shawn, and my very good friends Sarah and Judy. I wouldn't have wanted to if it wasn't for my Uncle Baird, who's the one that introduced me to physics in the first place. Of course, things would have been entirely different without Marie, whose patience and encouragement were essential. Thank you.

Abstract

This thesis provides bounds on the performance of quantum error correcting codes when used for quantum communication and quantum key distribution. The first two chapters provide a bare-bones introduction to classical and quantum error correcting codes, respectively. The next four chapters present achievable rates for quantum codes in various scenarios. The final chapter is dedicated to an upper bound on the quantum channel capacity.

Chapter 3 studies coding for adversarial noise using quantum list codes, showing there exist quantum codes with high rates and short lists. These can be used, together with a very short secret key, to communicate with high fidelity at noise levels for which perfect fidelity is impossible.

Chapter 4 explores the performance of a family of degenerate codes when used to communicate over Pauli channels, showing they can be used to communicate over almost any Pauli channel at rates that are impossible for a nondegenerate code and that exceed those of previously known degenerate codes. By studying the scaling of the optimal block length as a function of the channel's parameters, we develop a heuristic for designing even better codes.

Chapter 5 describes an equivalence between a family of noisy preprocessing protocols for quantum key distribution and entanglement distillation protocols whose target state belongs to a class of private states called "twisted states."

In Chapter 6, the codes of Chapter 4 are combined with the protocols of Chapter 5 to provide higher key rates for one-way quantum key distribution than were previously thought possible.

Finally, Chapter 7 presents a new upper bound on the quantum channel capacity that is both additive and convex, and which can be interpreted as the capacity of the channel for communication given access to side channels from a class of zero capacity "cloning" channels. This "clone assisted capacity" is equal to the unassisted capacity for channels that are degradable, which we use to find new upper bounds on the capacity of a depolarizing channel.

Preface

It has been a little over a decade since the discovery of quantum error correcting codes. Before the work of [Sho95], it was not even clear that quantum error correction was possible *in principle*. The no-cloning theorem [WZ82], which states that quantum information cannot be copied, seemed to preclude any sort of redundancy, which is the essence of classical error correction. Nevertheless, with the appearance in October 1995 of [Sho95] there began a period of such remarkable discovery that by the end of 1996 it was not only clear that quantum error correcting codes exist, but that they could be constructed using fairly straightforward modifications of classical codes [Sho95, CS96, Got96, BDSW96, EM96, Ste96]. Work in the intervening years has taken full advantage of this observation, but there are some ways in which correcting quantum noise is fundamentally different from classical error correction, most notably when it comes to degenerate codes. This thesis represents my attempt to understand these differences.

Contents

Chapter 1

Classical Codes

1.1 The Repetition Code

Error correcting codes are designed to encode data in a form that is resilient to noise. The simplest such code is an n-bit repetition code, wherein a single (logical) bit is mapped into a string of n bits. If the logical bit is a 0, we map it to the string of n zeros, whereas if it is 1, we map it to the string of n ones. For example, a three bit repetition code would be given by the following mapping:

$$\begin{aligned} 0 &\rightarrow 000 \\ 1 &\rightarrow 111. \end{aligned}$$

If one of the three bits is accidentally flipped, we will still be able to determine the value of the logical bit by decoding every 3-bit string to whichever value it takes on most:

$$\begin{aligned} 0 &\rightarrow 000 \overset{\text{noise}}{\rightarrow} 001 \overset{\text{correct}}{\rightarrow} 0 \\ 1 &\rightarrow 111 \overset{\text{noise}}{\rightarrow} 101 \overset{\text{correct}}{\rightarrow} 1. \end{aligned}$$

It's possible to correct as many errors as you like with a code like this—by making the code longer you can correct more errors. Sometimes this is actually a good solution. The trouble with this code is that it only encodes a single logical bit, and it may have to be quite long. If storage is expensive, it's a good idea to encode as much data as possible into your bits. The rate of a code is the ratio of logical bits to the code's block length, so what we'd like is to find codes with the highest possible rate that can still handle the noise levels we'd like to correct. In the next section we'll study a family of codes that, in general, can do much better than a repetition code, and includes the repetition codes as a special case.

1.2 Linear Codes

An $[n, k]$ linear code maps k-bit messages into n-bit codewords. We specify such a code with an $(n-k) \times n$ binary matrix, H, called the parity check matrix. A code with parity check matrix H contains all n-bit strings $\mathbf{x} \in \{0,1\}^n$ satisfying

$$H\mathbf{x} = \mathbf{0}, \tag{1.1}$$

where arithmetic is done modulo 2. If $H\mathbf{x}_1 = 0$ and $H\mathbf{x}_2 = 0$ it is also the case that $H(\mathbf{x}_1 + \mathbf{x}_2) = 0$, so that the codewords span a linear space. As long as the rows of H are linearly independent, the dimension of this space is k, and the total number of vectors it contains is 2^k.

Now suppose we've encoded some data into an n-bit code with parity check matrix H. If our original codeword was $\mathbf{x}$ and some noise process has flipped some of the bits, we'll be holding the vector $\mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ has 1's in the positions where a bit was flipped. By computing

$$H(\mathbf{x} + \mathbf{e}) = H\mathbf{e}, \tag{1.2}$$

we end up with an $n-k$ "syndrome" with which we can hopefully "diagonose" the error. The point is that if all of the errors we're trying to correct have distinct syndromes, we can just calculate the syndrome of our string and flip back the bits of the associated error.
The repetition codes of the previous section are linear codes, with parity check matrices of the form

$$\begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ & & \vdots & & \vdots & \\ 1 & 0 & \dots & 0 & 1 & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}. \tag{1.3}$$

The parity checks just ensure that for every codeword, the first and the second bits agree, the first and the third agree, etc. Translating the 3-bit example of the previous section into this formalism, we have

$$0 \to \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \overset{\text{noise}}{\to} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \overset{\text{checkparities}}{\to} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.4}$$

which tells us that the first and the second bits of our string agree, whereas the third disagrees with the first, from which we conclude that the third bit should be flipped back.

1.3 Random Linear Codes

How many bit flip errors can an $[n, k]$ code tolerate before it becomes impossible to retrieve the encoded data? The answer to this question depends on the distance of the code. We say a code has distance d, and call it an $[n, k, d]$ code if the minimum Hamming distance between pairs of distinct codewords is d. In other words,

$$d(C) = \min_{\mathbf{x}_1 \neq \mathbf{x}_2 \in C} |\mathbf{x}_1 + \mathbf{x}_2|, \tag{1.5}$$

where $|\mathbf{x}|$ is the number of 1's in $\mathbf{x}$, which is known as the Hamming weight. An $[n, k, d]$ code can correct $\lfloor (d-1)/2 \rfloor$ errors, since any such error can be corrected by decoding every $\mathbf{x} \in \{0,1\}^n$ to the codeword it is closest to.

To see what values of n, k, and d are possible we will use a standard trick in information theory – choosing a random code. This will allow us to prove the following theorem.

Theorem 1. [Gilbert-Varshamov Bound] *For sufficiently large n there exist $[n, k, d]$ codes of rate R and relative distance $\delta_{\text{rel}} = \frac{d}{n}$ as long as*

$$R < R_{\text{GV}} = 1 - H\left(\delta_{\text{rel}}\right), \tag{1.6}$$

where $H(p) = -p \log_2 p - (1-p)\log_2(1-p)$ is the binary entropy.

Proof. An $[n, k]$ code with parity check matrix H has distance at least d as long as there is no string with weight less than or equal to d that maps one codeword to another. In other words, for all $\mathbf{e} \in \{0,1\}^n$ with weight no greater than d we must have $H\mathbf{e} \neq \mathbf{0}$.

Fixing $\mathbf{e} \neq 0$ with weight no greater than d, an $(n-k) \times n$ parity check matrix H with entries chosen independently and uniformly from $\{0, 1\}$, we have

$$\begin{aligned} \Pr\left(H\mathbf{e} = \mathbf{0}\right) &= \prod_{i=1}^{n-k} \Pr\left(\mathbf{h}_i \cdot \mathbf{e} = 0\right) & (1.7) \\ &= \frac{1}{2^{n-k}}, & (1.8) \end{aligned}$$

where $\mathbf{h}_i$ is the ith row of H. Furthermore, the total number of errors with weight less than or equal to d is given by

$$N_d = \sum_{l=1}^{d} \binom{n}{l} \leq d 2^{nH(d/n)}, \tag{1.9}$$

where we have used $\binom{n}{pn} \leq 2^{nH(p)}$, which can be found using Stirling's approximation. Now, by the

union bound[1] the probability that *any* of these N_d errors satisfies $H\mathbf{e} = \mathbf{0}$ is upper bounded by

$$\begin{aligned} \Pr\left(\exists \mathbf{e} \text{s.t.} H\mathbf{e} = \mathbf{0}\right) &\leq N_d \Pr\left(H\mathbf{e} = \mathbf{0}\right) && (1.10) \\ &= \frac{N_d}{2^{n-k}}. && (1.11) \end{aligned}$$

Since $N_d \leq d2^{nH(d/n)}$, the probability our code has distance less than d is no greater than

$$\Pr\left(\mathbf{dist}(C_H) < d\right) \leq d2^{k-n+n(H(d/n))}, \tag{1.12}$$

which is less than 1 as $n \to \infty$ as long as $k/n > 1 - H(d/n)$. Since the probability is less than 1, there must be at least one $[n, k]$ code with distance at least d. □

While its proof is straightforward, the previous theorem gives the best known lower bound on the rate of a code in terms of its blocklength and distance. Finding upper bounds on possible rates is much more difficult, but it is generally believed that there do not exist codes with asymptotic rates in excess of R_{GV}.

1.4 Noisy Channel Coding Theorem

Suppose now that the noise we are trying to correct consists of bit flips that occur independently on each bit with probability p (ref Figure). With high probability, the weight of the resulting error will be roughly pn, so that the Gilbert-Varshamov bound tells us we can achieve a rate of at least $1 - H(2p)$ —by using a code of distance a little more than $2pn$ we will be able to correct all of the "typical" errors of our channel, with the probability of an "atypical" error becoming very small as n gets large.

It turns out that if we are willing to tolerate an arbitrarily small probability of error in the decoding, using a distance $2pn$ code is overkill. Before we see why, we must define the capacity of a channel that, roughly speaking, is the highest rate at which it is possible to communicate with an arbitrarily small probability of error. More formally,

Definition 2. *Let $\mathcal{N} : X \to Y$ be a channel, described by transition probabilities $p(x|y)$. Then a rate R is said to be achievable if there is a family of codes $C_n \subset X^n$ of rate R_n and with encoding and decoding operations $\mathcal{E}_n$ and $\mathcal{D}_n$ such that $\lim_{n\to\infty} R_n \geq R$ and $\Pr\left(\mathcal{D}_n(\mathcal{N}(\mathcal{E}_n(c_i))) = i\right) > 1 - \delta_n$ for all $c_i \in C_n$ and where $\delta_n \to \infty$. The supremum over all achievable rates is the capacity of $\mathcal{N}$.*

We now show that the capacity of the binary symmetric channel is much larger than $1 - H(2p)$. In fact,

[1] The union bound tells us that for any events A_i, $\Pr(\cup_i A_i) \leq \sum_i \Pr(A_i)$.

Theorem 3. *The capacity of the binary symmetric channel with crossover probability p is $C = \sup_X I(X;Y) = 1 - H(p)$, where $I(X;Y) = H(X) + H(Y) - H(XY)$ and $Y = \mathcal{N}_{\text{bsc}}(X)$.*

We divide the proof into two parts – the direct part, in which we show that $1-H(p)$ is achievable, and the converse, where we show that it is the maximum achievable rate. The direct part will rely on the following key lemma:

Lemma 4. [Typical Sequences] *Let $\mathcal{T}^n_{p,\delta} = \{\mathbf{x} \in \{0,1\}^n \,|\, ||\mathbf{x}| - np| < \delta\sqrt{n}\sqrt{p(1-p)}\}$. Then for n sufficiently large*

$$\Pr(\mathcal{T}^n_{p,\delta}) > 1 - \frac{2}{\delta^2} \tag{1.13}$$

and

$$|\mathcal{T}^n_{p,\delta}| \leq 2^{nH(p)+2K\delta\sqrt{n}} \tag{1.14}$$

where $K = 2\log(e)/e$.

Proof. **[Direct]** As with the proof of the Gilbert-Varshamov bound, we will consider a random $[n,k]$ linear code with parity check matrix H. We will aim to correct all of the errors in the set $\mathcal{T}^n_{p,\delta}$. We will be able to correct an error $\mathbf{x}$ if its syndrome is unique among the elements of $\mathcal{T}^n_{p,\delta}$. In other words, we will be unable to correct $\mathbf{x}_1$ exactly when there is an $\mathbf{x}_2 \in \mathcal{T}^n_{p,\delta}$ such that $H(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0}$. For a fixed $\mathbf{x}_1$ the probability that this occurs is given by

$$\begin{aligned} \Pr\left(\exists \mathbf{x}_2 \in \mathcal{T}^n_{p,\delta} \text{s.t.} H(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0}\right) &\leq |\mathcal{T}^n_{p,\delta}| \Pr\left(H(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0}\right) &(1.15)\\ &\leq \frac{2^{n(H(p)+2K\delta\sqrt{n})}}{2^{n-k}}. &(1.16)\end{aligned}$$

The probability of decoding error is thus given by

$$\begin{aligned} \Pr\left(\mathbf{x}\text{uncorrectable}\right) &\leq \Pr\left(\mathbf{x} \notin \mathcal{T}^n_{p,\delta}\right) + \sum_{\mathbf{x} \in \mathcal{T}^n_{p,\delta}} \Pr(\mathbf{x}) \Pr\left(\exists \mathbf{x}' \in \mathcal{T}^n_{p,\delta} \text{s.t.} H(\mathbf{x} + \mathbf{x}') = \mathbf{0}\right) &(1.17)\\ &\leq \frac{2}{\delta^2} + \frac{2^{n(H(p)+2K\delta\sqrt{n})}}{2^{n-k}}, &(1.18)\end{aligned}$$

so that by choosing any $R < 1 - H(p) - \frac{2K\delta_n}{\sqrt{n}}$ with $\frac{\delta_n}{\sqrt{n}} \to 0$ and $\delta_n \to \infty$ we get a vanishing probability of error. □

We have shown that $1 - H(p)$ is an achievable rate for the binary symmetric channel, but we must also show that it is the maximal achievable rate. To do this, we will use the following three lemmas.

Lemma 5. [Fano's Inequality] *Let X and Y be random variables, each taking values in $\{1, \ldots r\}$*

with $\Pr(X \neq Y) < p_e$. *Then the conditional entropy,* $H(X|Y) = H(XY) - H(Y)$, *satisfies*

$$H(X|Y) \leq H(p_e) + p_e \log(r-1). \tag{1.19}$$

Lemma 6. [Data Processing Inequality] *Let* $U \to X \to Y \to V$ *be a Markov chain. Then*

$$I(U;V) \leq I(X;Y). \tag{1.20}$$

Lemma 7. [Subadditivity of Mutual Information] *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be a random variable,* $\mathcal{N}$ *be a channel, and* $\mathbf{Y} = (Y_1, \ldots, Y_n) = (\mathcal{N}(X_1), \ldots, \mathcal{N}(X_n))$. *Then,*

$$I(\mathbf{X};\mathbf{Y}) \leq \sum_{i=1}^{n} I(X_i, Y_i). \tag{1.21}$$

Proof. **[Converse]** Let $C_n = \{c_i^n\}_{i=1}^{2^{nR}}$ be an $[n, Rn]$ code with error probability p_{e}, and let U be uniformly distributed on $\{1 \ldots 2^{nR}\}$. Then, letting $\mathbf{X}(U) = c_U^n$, $\mathbf{Y} = \mathcal{N}^{\otimes n}(\mathbf{X})$, and $\hat{U} = \mathcal{E}_n(\mathbf{Y})$ we have

$$\begin{aligned}
nR &= H(U) = H(U|\hat{U}) + I(U;\hat{U}) && (1.22)\\
&leq\ H(U|\hat{U}) + I(\mathbf{X};\mathbf{Y}) && (1.23)\\
&\leq H(p_{\mathrm{e}}) + p_{\mathrm{e}} Rn + I(\mathbf{X};\mathbf{Y}) && (1.24)\\
&\leq H(p_{\mathrm{e}}) + p_{\mathrm{e}} Rn + \sum_{j=1}^{n} I(X_j;Y_j) && (1.25)\\
&\leq H(p_{\mathrm{e}}) + p_{\mathrm{e}} Rn + nC, && (1.26)
\end{aligned}$$

where the second line is by the data processing inequality, the third is by Fano's inequality, and the fourth is by subadditivity of mutual information. This leads us to the conclusion that

$$R \leq \frac{C}{1-p_{\mathrm{e}}} + \frac{H(p_{\mathrm{e}})}{n(1-p_{\mathrm{e}})}, \tag{1.27}$$

so that in order to have p_{e} arbitrarily small as $n \to \infty$, we must have $R \leq C$. □

In the next chapter we will study the quantum analogues of the Gilbert-Varshamov bound and the noisy channel coding theorem. In the case of the coding theorem, we will be able to find achievable rates that are quite similar to the above formula for the capacity of the binary symmetric channel. However, the quantum analogue of the subadditivity of mutual information will fail, and the rates found are not optimal. Understanding this failure will be a central theme of the rest of the thesis.

Chapter 2

Quantum Codes

2.1 Quantum States, Channels, and Measurements

A d-dimensional quantum system is described by a nonnegative trace 1 operator on the complex vector space $\mathbb{C}^d$ called the state of the system. We call such an operator, $\rho \in \mathcal{B}(\mathbb{C}^d)$ where $\mathcal{B}(\mathbb{C}^d)$ denotes the set of bounded operators on $\mathbb{C}^d$, pure if it is rank one, and mixed otherwise. Any state ρ can be written as

$$\rho = \sum_i \lambda_i |i\rangle\langle i|, \tag{2.1}$$

with the λ_i representing the probability that the system is in pure state $|i\rangle$.

The physical operations that can be implemented on such states are completely positive trace preserving (CPTP) maps, which we will also refer to as quantum channels. They must be trace preserving, since the output of the operation should be a quantum state (and therefore trace one). The completely positivity requirement means the tensor product of a quantum channel with the identity must always map nonnegative operators to nonnegative operators—it must map states to states. CPTP maps are exactly those operations consisting of an isometry followed by a partial trace. In other words, a quantum operation $\mathcal{N}$ can always be written in the form

$$\mathcal{N}(\rho) = \mathrm{Tr}_E \left(U \rho U^\dagger \right), \tag{2.2}$$

where $U : \mathbb{C}^{d_A} \to \mathbb{C}^{d_B} \otimes \mathbb{C}^{d_E}$ is an isometry (meaning it maps $\mathbb{C}^{d_A}$ unitarily to a subspace of $\mathbb{C}^{d_B} \otimes \mathbb{C}^{d_E}$). Alternatively, CPTP maps are exactly those that can be written as

$$\mathcal{N}(\rho) = \sum_i A_i \rho A_i^\dagger, \tag{2.3}$$

where the $A_i : \mathbb{C}^{d_A} \to \mathbb{C}^{d_B}$ are linear maps satisfying $\sum_i A_i^\dagger A_i = \mathbb{I}$. The A_i's in this formuation are called the Kraus operators of the map.

A special kind of quantum channel is a complete projective measurement, which is just a quantum channel with orthogonal rank one Kraus operators:

$$\mathcal{N}(\rho) = \sum_{i=1}^{d_A} |i\rangle\langle i|\rho|i\rangle\langle i| = \sum_{i=1}^{d_A} \langle i|\rho|i\rangle |i\rangle\langle i|, \tag{2.4}$$

where $|i\rangle \in \mathbb{C}^{d_A}$ with $\langle i|j\rangle = \delta_{ij}$, and $\langle i|\rho|i\rangle$ can be interpreted as the probability of finding measurement outcome i. When the $\{|i\rangle\}_i$ are eigenvectors of a hermitian operator O, it is sometimes said that this channel "measures O" and the measurement outcomes are labeled by the eigenvectors of O. A more general type of measurement is a rank one positive operator valued measure (POVM), which is a channel with Kraus operators of the form $A_i = |\varphi_i\rangle|i\rangle\langle\varphi_i|$, where $\{|\varphi_i\rangle\}$ is a set of (possibly nonorthogonal and subnomalized) pure states satisfying $\sum_i \langle\varphi_i|\varphi_i\rangle|\varphi_i\rangle\langle\varphi_i| = \mathbb{I}$. In general, a measurement is any channel whose Kraus operators satisfy $A_j^\dagger A_i = 0$ for $i \neq j$.

2.2 The Pauli Group

If we are interested in qubit channels, which map $\mathcal{B}(\mathbb{C}^2)$ to itself, we can expand each A_i as a linear combination of the matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.5}$$

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.6}$$

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{2.7}$$

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.8}$$

Using the fact that $Y = iXZ$, we can then express such an $\mathcal{N}$ as

$$\mathcal{N}(\rho) = \sum_{u_1,u_2=0}^{1} \sum_{v_1,v_2=0}^{1} \sum_i \alpha^i_{u_1 v_1} (\alpha^i_{u_2 v_2})^* X^{u_1} Z^{v_1} \rho Z^{v_2} X^{u_2}. \tag{2.9}$$

Similarly, n copies of the channel acting on $\mathcal{B}((\mathbb{C}^2)^{\otimes n})$ can be expressed as

$$\mathcal{N}(\rho) = \sum_{\mathbf{u}_1,\mathbf{u}_2 \in \{0,1\}^n} \sum_{\mathbf{v}_1,\mathbf{v}_2 \in \{0,1\}^n} \sum_{\mathbf{i}} \alpha^{\mathbf{i}}_{\mathbf{u}_1 \mathbf{v}_1} (\alpha^{\mathbf{i}}_{\mathbf{u}_2 \mathbf{v}_2})^* X^{\mathbf{u}_1} Z^{\mathbf{v}_1} \rho Z^{\mathbf{v}_2} X^{\mathbf{u}_2}, \tag{2.10}$$

where $X^{\mathbf{u}} = X^{u_1} \otimes X^{u_2} \otimes \ldots \otimes X^{u_n}$ represents the operator that applies an X to the lth qubit when $u_l = 1$ and similarly for $Z^{\mathbf{v}}$. Because our encoding and decoding operations will be linear (as any quantum operation must be) we can focus on correcting errors of the form $X^{\mathbf{u}}Z^{\mathbf{v}}$, with the $\alpha^{\mathbf{i}}_{\mathbf{u}_1\mathbf{v}_1}$ in Eq. (2.9) playing the role of the probability (or, more accurately, amplitude) of the error $X^{\mathbf{u}_1}Z^{\mathbf{v}_1}$.

The matrices X, Y, and Z are called Pauli matrices, and the set of n-fold tensor products of Pauli matrices, together with phase factors ± 1, $\pm i$, make up the Pauli group, $\mathcal{G}_n = \{\pm 1, \pm i\} \otimes \{I, X, Y, Z\}^{\otimes n}$. Notice that the inclusion of the phase factors, together with the relations $Y = iXZ$, $Z = iYX$, and $X = iZY$ and the fact that any one of X, Y, and Z anticommutes with the other two ensures that this set is, in fact, closed under multiplication. This group structure will be important for the construction of stabilizer codes, in section 2.4.

One usually works in the Z basis, $\{|0\rangle, |1\rangle\}$, where the actions of the Paulis are

$$X|0\rangle = |1\rangle \qquad Z|1\rangle = |0\rangle, \tag{2.11}$$

$$Y|0\rangle = i|1\rangle \qquad Z|1\rangle = -i|0\rangle, \tag{2.12}$$

$$Z|0\rangle = |0\rangle \qquad Z|1\rangle = -|1\rangle, \tag{2.13}$$

so that X is often refered to as a "bit-flip error" or "amplitude error," Z is called a "phase" error, and Y is "both an amplitude and phase error."

We now turn to the problem of correcting these errors.

2.3 A Quantum Code

In the first chapter, we saw that an easy way to protect a classical bit from noise is to encode it into a repetition code. For instance, a three bit repetition code is able to correct a single bitflip error. The most obvious quantum generalization of this code would be to encode any state $|\psi\rangle$ into $|\psi\rangle^{\otimes 3}$. However, this mapping is not linear, as any quantum operation must be, and so it is impossible to implement—this is the content of the no-cloning theorem.

What we *will* be able to do is copy states with respect to some fixed basis. For example, we can choose the isometry defined by

$$|0\rangle \quad \rightarrow \quad |0\rangle|0\rangle|0\rangle \tag{2.14}$$

$$|1\rangle \quad \rightarrow \quad |1\rangle|1\rangle|1\rangle \tag{2.15}$$

to encode into what we will call a *repetition code in the Z basis*, or *Z repetition code*. A qubit

state, $|\psi\rangle = a|0\rangle + b|1\rangle$, which we'll call the *logical state*, will be encoded into into the state $|\bar{\psi}\rangle = a|0\rangle^{\otimes 3} + b|1\rangle^{\otimes 3}$ by such a code.

Much like a classical repetition code, a Z repetition code can be used to correct a single X error on any qubit. If our original encoded state $|\psi\rangle = a|0\rangle|0\rangle|0\rangle + b|1\rangle|1\rangle|1\rangle$ is acted on by an X error on the second qubit, the corrupted state is given by

$$a|0\rangle|1\rangle|0\rangle + b|1\rangle|0\rangle|1\rangle. \tag{2.16}$$

Notice that this state is an eigenvector of ZZI with eigenvalue -1 and ZIZ with eigenvalue 1, so that if we measure these ZZI and ZIZ their respective outcomes will deterministically be -1 and 1. These outcomes lead us to conclude there has been a bit error on the second qubit: On the one hand, the -1 outcome for ZZI tells us that there has been an X error on one of the first and second qubits, since this outcome tells us that the value of the two qubits in the Z basis "disagrees." On the other hand, the 1 outcome for ZIZ tells us that the first and third qubits agree.

This code is a total failure when it comes to correcting Z errors, though. To see this, notice that we'd like to be able to protect *all* of the pure states in our code from errors, and in particular the following two states:

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes 3} + |1\rangle^{\otimes 3}\right) \tag{2.17}$$

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes 3} - |1\rangle^{\otimes 3}\right). \tag{2.18}$$

The trouble is that $ZII|\psi_1\rangle = |\psi_2\rangle$ (and similarly for IZI and IIZ), so we are unable to distinguish the situation where $|\psi_1\rangle$ is corrupted by a single Z error from when $|\psi_2\rangle$ is sent and left uncorrupted. In fact, ZII, IZI, and IIZ all have an effect on an encoded state that is equivalent to applying a Z operation on the logical state before encoding. That is, letting $U_{\mathrm{e,z}}$ denote the encoding map in Eq. (2.14), we have

$$U_{\mathrm{e,z}}Z|\psi\rangle = ZIIU_{\mathrm{e,z}}|\psi\rangle \tag{2.19}$$

$$= IZIU_{\mathrm{e,z}}|\psi\rangle \tag{2.20}$$

$$= IIZU_{\mathrm{e,z}}|\psi\rangle. \tag{2.21}$$

Of course, we didn't have to use a repetition code in the Z basis. We could have just as easily used this X repetition code:

$$U_{\mathrm{e,x}}|\pm\rangle = |\pm\rangle|\pm\rangle|\pm\rangle, \tag{2.22}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ are the ± 1 eigenvectors of X. This would have allowed us to correct any

single qubit Z error, but we'd have the same problem as before—now the logical X operators are just the single qubit X errors, so we've got no way to correct those.

Since the X repetition code succeeds where the Z code fails, and vice versa, we can concatenate the two codes into a nine qubit code that corrects any single error: First we encode our logical qubit into a three qubit X repetition code, after which we encode each of the three qubits into a three qubit Z repetition code:

$$\begin{aligned} a|0\rangle + b|1\rangle &\rightarrow a|+\rangle^{\otimes 3} + b|-\rangle^{\otimes 3} && (2.23)\\ &\rightarrow a\left(|0\rangle^{\otimes 3} + |1\rangle^{\otimes 3}\right)^{\otimes 3} + b\left(|0\rangle^{\otimes 3} - |1\rangle^{\otimes 3}\right)^{\otimes 3}. && (2.24)\end{aligned}$$

Since the logical X of the outer X-repetition code is XXX, we can find the syndrome of the code by measuring its logical syndromes:

$$XXXXXXIII \text{ and } XXXIIIXXX. \tag{2.25}$$

By doing so, we will be able to correct a logical Z error on one of the three blocks making up our X code. Furthermore, by measuring the syndromes of the three inner Z-repetition codes,

$$ZZIIIIIII, \quad ZIZIIIIII, \tag{2.26}$$

$$IIIZZIIII, \quad IIIZIZIII, \tag{2.27}$$

$$IIIIIIZZI, \quad IIIIIIZIZ, \tag{2.28}$$

we will be able to correct any single qubit X error. If there is an error on one of the nine qubits, the first level of codes can correct the amplitude part of the error, but any phase error will be propagated to the next level as a logical Z. When this happens, the other two blocks will be error free (since we're only concerned with correcting single qubit errors) and the outer repetition code will be able to correct it.

2.4 Stabilizer Codes

We now look at a family of codes that is the quantum analogue of the classical codes of Chapter 1.

Let S be an abelian subgroup of the Pauli group, $\mathcal{G}_n$. Any such S has size 2^{n-k} for some integer $0 \leq k \leq n$, and is generated by a set of size $n-k$, which we'll call $\{S_1, \ldots S_{n-k}\}$. We say that a state $|\psi\rangle$ is stabilized by S if $s|\psi\rangle = \psi$ for all $s \in S$, and call the 2^k dimensional subspace of $(\mathbb{C}^2)^{\otimes n}$ that is stabilized by such an S an $[n, k]$ stabilizer code. The decoding operation for this code will just be to measure some generating set of S, say $\{S_1, \ldots S_{n-k}\}$, and perform some recovery operation based

on the outcomes. The measurement outcomes for a given error are referred to as the "syndrome" of that error.

The logical operations on the codespace are unitaries that map codewords to codewords. For a stabilizer code with stabilizer S, which we call C_S, the collection of all such unitaries is the *normalizer* of S, which is given by

$$N(S) = \{U \in \mathbb{U}(2^n) | USU^\dagger = S\}. \tag{2.29}$$

The point is that for any $U \in N(S)$ and $|\psi\rangle \in C_S$, we also have $U^\dagger \in N(S)$ so that for all $s \in S$

$$sU|\psi\rangle = UU^\dagger sU|\psi\rangle = Us_U|\psi\rangle = U|\psi\rangle, \tag{2.30}$$

where we have let s_U be the element of S that s is mapped to by conjugation by $U^\dagger$. Since this means $U|\psi\rangle$ is stabilized by every element of S, it must also belongs to C_S. Notice that $S \subset N(S)$, since S being abelian implies that

$$sts^\dagger = ss^\dagger t = t \tag{2.31}$$

for all $s, t \in S$.

Now, if we want to correct a set of errors, $\mathcal{E} \subset \mathcal{G}_n$, we could get into trouble if there are E_1 and E_2 in $\mathcal{E}$ such that

$$E_1^\dagger E_2 \in N(S). \tag{2.32}$$

The problem is that for any state $|\psi\rangle \in C_S$, we would then have

$$S_j E_1^\dagger E_2 = (-1)^{\omega(S_j,E_1)+\omega(S_j,E_2)} E_1^\dagger E_2 S_j, \tag{2.33}$$

where by definition we let $PQ = (-1)^{\omega(P,Q)}QP$ for $P, Q \in \mathcal{G}_n$, which implies that

$$\left(E_1^\dagger E_2\right)^\dagger S_j E_1^\dagger E_2 = (-1)^{\omega(S_j,E_1)+\omega(S_j,E_2)} S_j. \tag{2.34}$$

Since $E_1^\dagger E_2 \in N(S)$, the only way this could happen is if $\omega(S_j, E_1) = \omega(S_j, E_2)$ for $j = 1 \ldots n-k$. When we measure S_j on $E|\psi\rangle$, our syndrome outcomes are $\omega(S_j, E)$, so this means that E_1 and E_2 would be assigned the same syndrome. Because they have the same syndrome, we'll need to be able to reverse them with the same recovery operation, which will only be possible if $E_1|\psi\rangle = E_2|\psi\rangle$ for all $|\psi\rangle \in C_S$. This will be true only if $E_1^\dagger E_2 \in S$, which leads us to the error correction conditions for stabilizer codes:

Theorem 8. *A set of errors $\mathcal{E} \subset \mathcal{G}_n$ can be corrected by a stabilizer code exactly when every pair of errors, $E_1, E_2 \in \mathcal{E}$, satisfies*

$$E_1^\dagger E_2 \notin N(S) - S. \tag{2.35}$$

One way to achieve this condition is to require that any $E_1, E_2 \in \mathcal{E}$ satisfies $E_1^\dagger E_2 \notin N(S)$. Then every error in $\mathcal{E}$ would be assigned a different syndrome, and when we found the syndrome corresponding to a particular error, we could just apply its inverse and recover the original state. Codes like this are called *non-degenerate*, and are fairly well understood. *Degenerate* codes, which also allow $E_1^\dagger E_2 \in S$ (and must have *many* such pairs to behave significantly differently from a non-degenerate code) are more poorly understood, and will be the topic of Chapters 4 and 6. There we will see degenerate codes that can tolerate noise levels for which all non-degenerate codes will fail.

2.5 Quantum Gilbert-Varshamov Bound

In the first chapter, we saw that it's a pretty good idea to study random linear codes—we got the Gilbert-Varshamov bound and Shannon's theorem for a binary symmetric channel. Stabilizer codes are sort of the quantum version of linear codes, so looking at random stabilizer codes is an obvious thing to do.

We first need to look a little more closely at $N(S)$, the normalizer. In particular, we'd like to know how many elements there are in $N(S) \cap \mathcal{G}_n$—how many Pauli elements are there in the normalizer of a stabilizer with $n-k$ generators? First off, $N(S) \cap \mathcal{G}_n$ will contain the logical Pauli operations on the k qubits in C_S, of which there are 2^{2k}— the size of $\mathcal{G}_k$ (omitting phases ± 1, $\pm i$). Furthermore, for each logical Pauli operation on the codespace, there will be 2^{n-k} different elements of $N(S)$ with that action. To see this, notice that if, say $P \in N(S)$ acts as a logical X on the first qubit of C_S, so will sP for all $s \in S$. Which means there must be $2^{2k}2^{n-k} = 2^{n+k}$ Paulis in $N(S)$.

Now, if we'd like to be able to correct all errors of weight up to t, one way to do this is to show there is a stabilizer S such that for every pair $E_1, E_2 \in \mathcal{E} := \{E \in \mathcal{G}_n | \mathrm{wt}(E) \leq t\}$ we have $E_1^\dagger E_2 \notin N(S)$. Choosing a random S of size 2^{n-k} and fixing $E_1, E_2 \in \mathcal{E}$, the probability that this fails to be the case is

$$\Pr\left(E_1^\dagger E_2 \in N(S)\right) = \frac{2^{n+k}-1}{2^{2n}-1} \leq \frac{1}{2^{n-k}}, \tag{2.36}$$

since the total number of nonidentity elements in $N(S)$ and $\mathcal{G}_n$ are $2^{n+k}-1$ and $2^{2n}-1$, respectively. Letting $\mathcal{E}^{(2)} = \{E_1^\dagger E_2 | E_1, E_2 \in \mathcal{E}\}$ be the set of all products of pairs of errors in $\mathcal{E}$, we then see that

$$\begin{aligned}
\Pr\left(\exists E \in \mathcal{E}^{(2)} | \in N(S)\right) &= \Pr\left(\cup_{i=1}^{|\mathcal{E}^{(2)}|}\{E \in N(S)\}\right) && (2.37)\\
&\leq \sum_{i=1}^{|\mathcal{E}^{(2)}|} \Pr\left(E_i \in N(S)\right) && (2.38)\\
&\leq \frac{|\mathcal{E}^{(2)}|}{2^{n-k}}, && (2.39)
\end{aligned}$$

so that we will have a nonzero probability of having all pairs of errors be outside of the stabilizer as long as

$$k < n - \log_2 \left| \mathcal{E}^{(2)} \right|. \tag{2.40}$$

Since $\mathcal{E}^{(2)}$ is the set of all products of pairs of errors with weight up to t, we have

$$\left| \mathcal{E}^{(2)} \right| = \sum_{w=1}^{2t} \binom{n}{w} 3^w \leq 2t 2^{nH(2t/n)} 3^{2t}, \tag{2.41}$$

so that there will be distance $2t+1$ codes of rate $\frac{k}{n}$ as long as

$$k < n - nH(2t/n) - 2t \log_2 3 - \log_2(2t), \tag{2.42}$$

which gives us

Theorem 9. [Quantum Gilbert-Varshamov Bound] *For sufficiently large n there are stabilizer codes of relative distance $\delta_{\text{rel}} = \frac{d}{n}$ and rate R for all*

$$R < 1 - H\left(\delta_{\text{rel}}\right) - \delta_{\text{rel}} \log_2 3. \tag{2.43}$$

Such a code can be used to correct arbitrary errors affecting up to a fraction $\approx \delta_{\text{rel}}/2$ of the qubits in a block.

2.6 Probabilistic Quantum Errors

Just like in the classical case, it will turn out that correcting probabilistic quantum noise, we'll be able to do much better than using a bounded distance code. The noise model we'll use is the depolarizing channel, which acts on qubits as follows:

$$\mathcal{N}(\rho) = (1-p)\rho + \frac{p}{3} X\rho X + \frac{p}{3} Y\rho Y + \frac{p}{3} Z\rho Z. \tag{2.44}$$

If we use a large number of such channels, we will be able to restrict our attention to a set of typical errors:

$$\mathcal{T}^n_{\mathbf{p},\delta} = \left\{ \mathbf{x} \in \{0,1,2,3\}^n \,\middle|\, |\#\{l|x_l = i\} - p_i| \leq \delta\sqrt{n}\sqrt{p_i(1-p_i)} \right\}, \tag{2.45}$$

which has the property that

$$\Pr\left(\mathcal{T}^n_{\mathbf{p},\delta}\right) \geq 1 - \frac{4}{\delta^2}, \tag{2.46}$$

and has size $2^{nH(\mathbf{p})}$, up to corrections of order $2^{\delta\sqrt{n}}$.

For a randomly chosen code with stabilizer S, the probability of a particular typical error,

$$E_i \in \mathcal{E}_{\rm typ} := \left\{ X^{\mathbf{u}} Z^{\mathbf{v}} \Big| ((u_1, v_1), \ldots (u_n, v_n)) \in \mathcal{T}^n_{\mathbf{p},\delta} \right\}, \tag{2.47}$$

being uncorrectable is exactly

$$\Pr\left(E_i \text{uncorrectable}\right) = \Pr\left(\exists E_j \neq E_i, E_j \in \mathcal{E}_{\rm typ} | E_j^\dagger E_i \in N(S) - S\right). \tag{2.48}$$

Since the elements of $\mathcal{E}_{\rm typ}$ will all occur with roughly equal probabilities, the average (over S) of the probability of error of the code defined by S is no larger than

$$\begin{aligned}
\langle P_{\rm e} \rangle_S &\leq \frac{1}{|\mathcal{E}_{\rm typ}|} \sum_{i=1}^{|\mathcal{E}_{\rm typ}|} \Pr\left(E_i \text{uncorrectable}\right) && (2.49)\\
&\leq \frac{1}{|\mathcal{E}_{\rm typ}|} \sum_{i=1}^{|\mathcal{E}_{\rm typ}|} \Pr\left(\exists E_j \in \mathcal{E}_{\rm typ} | E_j^\dagger E_i \in N(S)\right) && (2.50)\\
&\leq |\mathcal{E}_{\rm typ}| \Pr\left(E_j^\dagger E_i \in N(S)\right) && (2.51)\\
&\leq |\mathcal{E}_{\rm typ}| \frac{1}{2^{n-k}}. && (2.52)
\end{aligned}$$

Since $\log_2 |\mathcal{E}_{\rm typ}| \approx nH(\mathbf{p})$, we see that $\langle P_{\rm e} \rangle_S$ falls off exponentially with n as long as

$$R = \frac{k}{n} < 1 - H(\mathbf{p}) = 1 - H(p) - p \log_2 3. \tag{2.53}$$

2.7 Quantum Noisy Channel Coding Theorem

In contrast to linear codes for the binary symmetric channel, it turns out that the rate achieved by random stabilizer codes over the depolarizing channel is not optimal. Actually, this was known well before it was clear what the quantum channel capacity formula was. The rate that we showed is achievable in the previous section, which is sometimes called the hashing rate, is exactly

$$Q_1(\mathcal{N}) = \max_{\phi_B} I^{\rm c}(\mathcal{N}, \phi_B), \tag{2.54}$$

where $I^{\rm c}(\mathcal{N}, \phi_B) = I^{\rm c}\left(I \otimes \mathcal{N}(|\phi^{\mathcal{AB}}\rangle\langle\phi^{\mathcal{AB}}|)\right)$, $|\phi_{AB}\rangle$ is a purification of ϕ_B, and $I^{\rm c}(\rho_{AB}) = S(\rho_B) - S(\rho_{AB})$ with $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$, known as the *coherent information*, is a measure of quantum correlations playing a similar role to the mutual information of the classical capacity theorem. In the case of the depolarizing channel, the optimal ϕ_B is the maximally mixed state, $\mathbb{I}/2$. In fact, for the depolarizing channel $Q_1(\mathcal{N})$ is exactly the maximum rate achievable with a nondegenerate code. It was shown in [DSS98, SSa] that it is possible to do better than Q_1 by using degenerate codes, which is reflected

in the quantum capacity formula (proved in [Sho, Dev05])

$$Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\phi_{B_n}} I^c(\mathcal{N}^{\otimes n}, \phi_{B_n}) = \lim_{n\to\infty} \frac{1}{n} Q_1(\mathcal{N}^{\otimes n}), \tag{2.55}$$

by the necessity of optimizing the coherent information over an arbitrarily large number of channel uses. The findings of [DSS98, SSa] can be interpreted as showing that, in contrast to the subadditivity of mutual information, we can sometimes have

$$Q_1(\mathcal{N}^{\otimes n}) > nQ_1(\mathcal{N}). \tag{2.56}$$

This leaves us in kind of a bad situation since, unless we are dealing with a very special channel, there's no way to evaluate the formula in Eq. (2.55).

Chapter 3

The Adversarial Channel and Quantum List Codes

In this chapter we study quantum communication in the presence of adversarial noise. In this setting, communicating with perfect fidelity requires using a quantum code of bounded minimum distance, for which the best known rates are given by the quantum Gilbert-Varshamov (QGV) bound. By asking only for arbitrarily high fidelity and allowing the sender and receiver to use a secret key with length logarithmic in the number of qubits sent, we achieve a dramatic improvement over the QGV rates. In fact, we find protocols that achieve arbitrarily high fidelity at noise levels for which perfect fidelity is impossible. To achieve such communication rates, we introduce fully quantum list codes, which may be of independent interest.

3.1 Introduction

Effectively dealing with noise is a major challenge faced by all proposals for the coherent manipulation of quantum information. In addition to quantum communication, sending a quantum state over a noisy channel models noisy storage and, as such, characterizing communication rates over quantum channels is a central question in the study of both quantum information and computation.

Various asymptotic capacities of quantum channels have been studied [Dev05, DHW04, DS05, Sho, Win99, Win04, BKN00, SW97, Hol98, Llo97, BSSA02]. However, this work has been almost exclusively concerned with discrete memoryless channels (DMCs), wherein a sender and receiver use many independent and identical copies of a channel. In this scenario, one studies the asymptotic communication rate possible using an operation of the form $\mathcal{N}^{\otimes n}$, where $\mathcal{N}$ is the channel under consideration and its rate is given by $R = k/n$ where k is the number of high fidelity logical qubits sent. Relatively little is known outside of the DMC scenario, with notable exceptions found in [BDM05, KW05, HN03, Ren05, BD].

In this chapter, we study communication over an adversarial quantum channel (AC), which is

perhaps as different from a DMC as one can imagine. When sending n qubits over an AC, rather than errors on different qubits occuring independently, an adversary who knows what protocol is being used tries to foil the communication by maliciously choosing a superposition of errors, subject only to a restriction on the number of qubits each error affects. We will call this channel $\mathcal{N}_{p,n}^{\mathrm{adv}}$, where p is the fraction of the n qubits that the adversary is permitted to corrupt. $\mathcal{N}_{p,n}^{\mathrm{adv}}$ is the natural quantum generalization of the classical adversarial channel that was considered in [Lan04, Gur03] and whose roots go back to [Ham50].

If the receiver is required to reconstruct the logical state exactly, communicating over $\mathcal{N}_{p,n}^{\mathrm{adv}}$ requires using a quantum error correcting code (QECC) of distance $2\lceil np\rceil+1$. The quantum Gilbert-Varshamov bound guarantees the existence of such a code with a rate of at least [Got]

$$1-H(2p)-2p\log 3, \tag{3.1}$$

where logarithms are taken base 2 here and throughout. Communication beyond this rate is possible only if QECCs beating the Gilbert-Varshamov bound exist, which is a question that has been quite difficult to resolve. Furthermore, Rains has shown that there are no quantum codes with distance greater than $n/3$ [Rai99], so that it is impossible to send even a single qubit for $p \geq 1/6$.

Surprisingly, if we ask only for a high fidelity reconstruction, and allow the sender and receiver to share a secret key of size $O(\log n)$ it is possible to communicate at rates much higher than the Gilbert-Varshamov and Rains bounds suggest. Below, we present a coding strategy for this scenario that achieves a rate of

$$1 - H(p) - p\log 3, \tag{3.2}$$

which is significantly larger than the Gilbert-Varshamov rate for all values of p and remains nonzero up to $p \approx 0.189$.

There are three ingredients in achieving such rates with negligible secret keys. We employ two coding techniques. The first ingredient is a predetermined quantum list code that is known to the adversary. This alone allows high-rate but low-fidelity transmission. To improve on the fidelity, a random subcode is further chosen according to a secret key unknown to the adversary. Finally, the subcode is derandomized by using ϵ-biased sets.

Informally, a quantum list code is an error correcting code with the relaxed reconstruction requirement that the decoded state be equal to the original state acted on by a superposition of a small number of errors. The number of errors is called the "list length." This relaxation allows a considerable increase in rate compared to QECCs, and by using a standard probabilistic argument we show there are list codes with constant-length lists and rate approaching $1 - H(p) - p\log 3$, which tolerate pn errors. Then, by using $O(\log n)$ bits of secret key to choose a pseudorandom, large subcode of the list code, the receiver is able to distinguish between the various errors in the list and

communicate with high fidelity at the rate of the list code being used.

Note that a single level of random code could also be used, but the secret key required would be $O(n^2)$ bits. One could also achieve a rate of $1 - H(p) - p\log 3$ by using secret key to determine a permutation of the n channel uses (see, e.g., [SP00] or [Ren05]), at a cost of $O(n \log n)$ bits that, unfortunately, also leads to a divergent secret key rate. We further note that key recycling as a technique to lower the amortized cost cannot be used in a straighforward manner in our adversarial scenario. In a sense, our list-code construction can be viewed as a derandomization these key-inefficient protocols, achieving the same result with a much shorter secret key.

After the initial presentation of this result [Smi06], we learned of two independent studies of list codes, both in settings quite different from our own. Reference [KY] studied decoding of *classical list codes* using quantum algorithms, and Ref. [Hay] studied list codes for sending *classical* messages via iid quantum channels.

In the next section we review some basic background material, then present the details of our construction after which we discuss an application to entanglement distillation from states with adversarial errors, as well as a few open problems.

3.2 Background and Definitions

Throughout, our sender, receiver, and adversary will be named Alice, Bob, and Eve, respectively. The encoding of a k-qubit state $|\psi\rangle$ into an error correcting code will be written as $|\bar{\psi}\rangle$. We call the Pauli group acting on n qubits $\mathcal{G}_n$ and write its elements in the form

$$P = i^t X^{\mathbf{u}} Z^{\mathbf{v}}, \tag{3.3}$$

where $t \in \{0, 1, 2, 3\}$, $\mathbf{u}, \mathbf{v}$ are binary vectors of length n,

$$X^{\mathbf{u}}(Z^{\mathbf{v}}) \tag{3.4}$$

denotes

$$X^{u_1} \otimes \cdots \otimes X^{u_n}(Z^{v_1} \otimes \cdots \otimes Z^{v_n}), \tag{3.5}$$

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.6}$$

and

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.7}$$

The (anti)commutation relation between $P_1, P_2 \in \mathcal{G}_n$ is determined by

$$P_1 P_2 = (-1)^{\omega(P_1,P_2)} P_2 P_1 \tag{3.8}$$

with

$$\omega(P_1, P_2) = \mathbf{u}_1 \cdot \mathbf{v}_2 + \mathbf{u}_2 \cdot \mathbf{v}_1, \tag{3.9}$$

where the dot products and sum are computed in arithmetic modulo 2. We let $\langle P_l \rangle$ denote the subgroup of $\mathcal{G}_n$ generated by a set of Pauli elements $\{P_l\}$.

A state $|\psi\rangle$ is said to be stabilized by a Pauli matrix P when $P|\psi\rangle = |\psi\rangle$. An $[n, k]$ stabilizer code is a 2^k-dimensional space of n-qubit states simultaneously stabilized by all elements of a size 2^{n-k} abelian subgroup of $\mathcal{G}_n$. The abelian subgroup is typically called S and is referred to as the code's stabilizer, and has $n-k$ generators denoted by $\{S_i\}_{i=1}^{n-k}$. For any $E \in \mathcal{G}_n$ we refer to the $(n-k)$-bit string $\omega(E, S_i)$ as the syndrome of E (see, e.g., [Got, NC04]). The weight of a Pauli matrix P, which we denote by $\mathrm{wt}(P)$, is the number of qubits on which P acts nontrivially, and we call a stabilizer code an $[n, k, d]$ code if it can detect all errors of weight less than the distance d, which is equivalent to being able to correct all errors of weight less than $\lfloor (d-1)/2 \rfloor$.

For any real number r, let $\mathcal{E}^r$ be the set of all Pauli matrices of weight no larger than $\lfloor r \rfloor$. Let $N(S)$ be the set of all unitaries that leave S invariant under conjugation. Then, S defines an $[n, k, d]$ code if and only if for every pair of errors $E_i, E_j \in \mathcal{E}^{(d-1)/2}$ we have

$$E_i^\dagger E_j \notin N(S) - S. \tag{3.10}$$

We now define the channel we wish to study.

Definition 10. *The n-qubit adversarial quantum channel with error rate p, which we call $\mathcal{N}_{p,n}^{\mathrm{adv}}$, acts on a state of n qubits, ρ, and is of the form*

$$\mathcal{N}_{p,n}^{\mathrm{adv}}(\rho) = \sum_i A_i \rho A_i^\dagger \text{ with } A_i = \sum_{E \in \mathcal{E}^{pn}} \alpha_E^i E \tag{3.11}$$

subject to the requirement that $\sum_i A_i^\dagger A_i = I$ and where $\mathcal{E}^{np} = \{E \in \mathcal{G}_n \,|\, wt(E) \leq pn\}$ is as defined before. The particular choice of the $\{A_i\}$'s is made by Eve only after Alice and Bob have decided on a communication strategy.

In particular, notice that to communicate effectively over $\mathcal{N}_{p,n}^{\mathrm{adv}}$ one must find a strategy that works with high fidelity for *all* channels described by Eq. (3.11). To do this, we will use quantum list codes, which are defined below.

3.3 Quantum List Codes

Definition 11. *We say that an $[n,k]$ stabilizer code, $\mathcal{C}$, is an $[n,k,t,L]$-list code if there is a decoding operation, $\mathcal{D}$, such that for every $E_i \in \mathcal{E}^t$ and $|\bar{\psi}\rangle \in \mathcal{C}$, the decoded k-qubit state, along with the syndrome s, is given by $\mathcal{D}(E_i|\bar{\psi}\rangle\langle\bar{\psi}|E_i^\dagger) = \sum_s \sum_j A_j^s|\psi\rangle\langle\psi|A_j^{s\dagger} \otimes |s\rangle\langle s|$ where $\sum_{sj} A_j^{s\dagger}A_j^s = I$, and each A_j^s is a linear combination of the 2^L elements of $\langle P_l^s\rangle_{l=1}^L$, where $\{P_l^s\}_{l=1}^L$ is a list of logical errors on the codespace and $\langle P_l^s\rangle_{l=1}^L$ is the group they generate.*

We now show that, asymptotically, there exist $[n,k,t,L]$-list codes with favorable parameters. We proceed by considering random stabilizer codes, arguing along the lines of [BDSW96, Got]. In particular, we'll show that if we choose a random stabilizer code with rate as below, in the limit of large n the probability of it failing to be L-list decodable is less than 1.

Theorem 12. *$[n, \lfloor Rn\rfloor, \lfloor pn\rfloor, L]$-list codes exist for sufficiently large n and for*

$$R < 1 - \left(1 + \frac{1}{L}\right)\left(H(p) + p\log 3\right) . \tag{3.12}$$

Proof. Let $N_E = |\mathcal{E}^{pn}|$ and $\mathcal{E}^{pn} = \{E_i\}_{i=1}^{N_E}$. Two errors E_i and E_j have the same syndrome iff $E_i^\dagger E_j \in N(S)$. A code fails to be L-list decodable only if there are $L+1$ *independent* errors $E_0, \cdots, E_L$ having the same syndrome. Mathematically, $E_i^\dagger E_j \in N(S)$ for $0 \le i,j \le L$ (or equivalently, $E_0^\dagger E_j \in N(S)$ for $1 \le j \le L$).

We procede with a simple result that there are

$$\Pi_{a=0}^{n-k-1}\ \left(2^{2n-a} - 2^a\right) \tag{3.13}$$

unique generating sets for stabilizers with $n-k$ generators (we omit the overall factors $\pm 1, i$ of the Pauli matices). This is because S_1 can be chosen from the $2^{2n}-1$ nontrivial Pauli matrices, and S_2 then has to be chosen from the set of 2^{2n-1} Pauli matrices commuting with S_1 but outside of the multiplicative group generated by S_1. Similarly, each S_i is chosen from the $2^{2n-(i-1)}$ Pauli matrices commuting with $S_1, \cdots, S_{i-1}$ but not from the multiplicative group generated by them, and thus there are

$$2^{2n-(i-1)} - 2^{i-1} \tag{3.14}$$

choices. Furthermore, any stabilizer of size 2^{n-k} has

$$\prod_{b=0}^{n-k-1} \left(2^{n-k-b} - 1\right) \tag{3.15}$$

different generating sets, so we also have found the total number of stabilizers of this size.

Now, back to the arbitrary and fixed list $E_0, \cdots, E_L$ of independent errors. It follows that

$\{E_0^\dagger E_j\}_{j=1,\cdots,L}$ are also independent. Thus there are 2^{2n-L} Pauli operators commuting with them. Adapting the above counting argument to the present case, there are

$$\Pi_{a=0}^{n-k-1} \left(2^{2n-L-a}-2^a\right) \tag{3.16}$$

sets of generators for stabilizers with all $n-k$ generators commuting with $E_0^\dagger E_j$ for all j. Together with the unconstrained stabilizer count, a randomly chosen S will give the same syndrome for $\{E_j\}_{j=0,\cdots,L}$ with probability

$$\frac{\Pi_{a=0}^{n-k-1} \left(2^{2n-L-a}-2^a\right)}{\Pi_{a=0}^{n-k-1} \left(2^{2n-a}-2^a\right)} \leq 2^{-L(n-k)}. \tag{3.17}$$

Applying the union bound for the choice of the $L+1$ E_j's, the probability that a random $[n,k]$ code is not L-list decodable is upper bounded by

$$\binom{N_E}{L+1} 2^{-L(n-k)}, \tag{3.18}$$

which is less than

$$N_E^{L+1}\, 2^{-L(n-k)}. \tag{3.19}$$

This is in turns less than 1 if $k \leq n - (1+1/L)\log N_E$. For every $\epsilon > 0$, $\exists n_\epsilon$ s.t. whenever

$$n \geq n_\epsilon, \tag{3.20}$$

$$\log N_E \leq n(H(p) + p\log 3 + \epsilon) \tag{3.21}$$

so choosing

$$k = n\left[1 - \left(1+\frac{1}{L}\right)(H(p) + p\log 3) - 2\epsilon\right] \tag{3.22}$$

completes the proof. □

3.4 Coding Strategy

Theorem 12 tells us that for any rate $R < 1 - H(p) - p\log 3$, there exist $[n, Rn, pn, L]$-list codes for large enough n and L. (For example, let $\eta = 1 - H(p) - p\log 3 - R$, and choose $\epsilon = \eta/3$, $L \geq (H(p) + \log 3)/\epsilon$, and $n \geq n_\epsilon$ in Thm. 12.) Using such a code, $\mathcal{C}^{n,L}$, we can correct any error in $\mathcal{E}^{pn}$ to get a state of the form

$$\sum_i B_i^s |\psi\rangle\langle\psi| B_i^{s\dagger}, \tag{3.23}$$

where $|\psi\rangle$ is the state the sender wished to communicate and

$$B_i^s = \sum_{f=1}^{2^L} \beta_{if} G_f^s \tag{3.24}$$

subject to

$$\sum_i B_i^{s\dagger} B_i^s = I, \tag{3.25}$$

and where $G_f^s \in \langle P_l^s \rangle$.

We would now like to add a few more stabilizers to $\mathcal{C}^{n,L}$ so that the receiver can reconstruct $|\psi\rangle$ unambiguously. These additional stabilizers will be determined by a secret key shared by the sender and receiver, and are thus unknown to the adversary.

It will follow from proof of Thm. 13 below that adding

$$(1/\log(4/3))(2L + \log(1/\epsilon)) \tag{3.26}$$

random stabilizers to the code $\mathcal{C}^{n,L}$ would allow us to distinguish among the $\{G_f^s\}_{j=1}^{2^L}$ possible errors in the list-decoded state with a probability of $1-\epsilon$. This would require

$$2n(2L + \log(1/\epsilon))/\log(4/3) \tag{3.27}$$

bits of shared key.

A much smaller key can be used if ϵ-biased sets are used to choose stabilizers pseudo-randomly. A subset of $\{0,1\}^m$, denoted A, is said to be an ϵ-biased set of *length* m if for each $e \in \{0,1\}^m$, roughly half of the elements of A has odd/even parity with e:

$$\left| \Pr_{a \in A} (e \cdot a = 0) - \Pr_{a \in A} (e \cdot a = 1) \right| \leq \epsilon. \tag{3.28}$$

Let

$$\{S_i\}_{i=1}^{n-k} \tag{3.29}$$

be the stabilizers of $\mathcal{C}^{n,L}$. We add K stabilizers

$$T_1, \cdots, T_K. \tag{3.30}$$

After each addition, we get a subcode of the previous code, and the number of encoded qubits decreases. In particular, suppose $j-1$ stabilizers have been added, resulting in a subcode $\mathcal{C}_{j-1}^{n,L}$ with $k-j+1$ encoded qubits. The next stabilizer T_j is chosen according to a random element $(\mathbf{u_j}, \mathbf{v_j})$ of

an ϵ-biased set A_j of length $2(k-j+1)$, with

$$T_j = \overline{X}_{j-1}^{\mathbf{u_j}} \overline{Z}_{j-1}^{\mathbf{v_j}} \tag{3.31}$$

where $\overline{X}_{j-1}$ and $\overline{Z}_{j-1}$ are logical operations of $\mathcal{C}_{j-1}^{n,L}$. The following theorem shows that using this procedure to add

$$K = O(L \log 1/\epsilon) \tag{3.32}$$

stabilizers allows the receiver to reconstruct the encoded state with high probability (and thus, decode with high fidelity). There are efficient constructions of ϵ-biased sets of length m with only $O(\frac{m^2}{\epsilon})$ elements [NN90, AGHP90], so that the amount of secret key used in this construction is

$$O\left((2L + \log(1/\epsilon)) \log(n^2/\epsilon)\right) \tag{3.33}$$

bits.

Theorem 13. *Let $\mathcal{C}^{n,L}$ be an $[n, Rn, pn, L]$-list code of rate R and let $\mathcal{C}_K^{n,L}$ be the code obtained from $\mathcal{C}^{n,L}$ by progressively adding $K = (1/\log(4/3))(2L + \log(1/\epsilon))$ stabilizers determined by ϵ-biased sets $A_1, \cdots, A_K$ (of decreasing length) as described above. By using a secret key of fewer than $O(K(\log(\frac{n^2}{\epsilon})))$ bits to select $\mathcal{C}_K^{n,L}$, $nR - K = n(R - o(n))$ qubits can be sent over $\mathcal{N}_{p,n}^{\text{adv}}$ with fidelity at least $1 - \epsilon$ for all $\epsilon < 1/2$.*

Proof. Since we use an $[n, Rn, pn, L]$-list code, the adversary's power is reduced to choosing the probability distribution for s, and the corresponding superposition of the list

$$\{G_f^s\}_{f=1}^{2^L} = \langle P_l^s \rangle \tag{3.34}$$

of error operations. So, if for all syndromes of the list code, the probability (over the choice of $T_1, \cdots, T_K$) that there is a pair of list elements that have the same commutation relations with the T_j stabilizers is less than ϵ, the fidelity of the decoded state with the original will be at least $1 - \epsilon$.

Let $f_{1,2}$ be fixed and define the events M_j as

$$M_j = \{\omega(G_{f_1}^s, T_j) = \omega(G_{f_2}^s, T_j)\}. \tag{3.35}$$

Then, the probability of $G_{f_1}^s$ and $G_{f_2}^s$ being assigned the same syndrome by all T_j is

$$\Pr\left(\cap_{j=1}^K M_j\right) = \prod_{j=1}^K \Pr\left(M_j | M_{j-1} \ldots M_1\right). \tag{3.36}$$

Since each T_j is chosen using an ϵ-biased string of encoded operations $\bar{X}_{j-1}$ and $\bar{Z}_{j-1}$ of the code

$\mathcal{C}_{j-1}^{n,L}$, we have

$$\Pr\left(M_j | M_{j-1} \dots M_1\right) \leq \tfrac{1+\epsilon}{2}, \tag{3.37}$$

which immediately implies that

$$\Pr\left(\cap_{j=1}^{K} M_j\right) \leq \left(\frac{1+\epsilon}{2}\right)^K. \tag{3.38}$$

By a union bound over the choice of $f_{1,2}$, the probability of *any* pair f, f' having the same commutation relations for all j is less than

$$2^{2L}\left(\frac{1+\epsilon}{2}\right)^K. \tag{3.39}$$

By choosing

$$K = (1/\log(4/3))(2L + \log(1/\epsilon)) \tag{3.40}$$

we make this failure probability less than ϵ for any $\epsilon < 1/2$ so that with probability at least $1-\epsilon$, G_f^s can be unambiguously identified and the state reconstructed. The output state is thus of the form $(1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\varphi$ for some state φ. □

3.5 Discussion

We have introduced the adversarial quantum channel and shown that by using a logarithmic length secret key it is possible to communicate over this channel with a rate of $1-H(p)-p\log 3$. This is much higher than would be naively expected based on existing error correcting codes, and is quite close to what is known to be possible using n independent depolarizing channels with error probability p.

Our construction involves using quantum list codes, which we defined and showed to exist with favorable parameters. We expect quantum list codes to be useful in other contexts.

The scenario considered in this chapter and the spirit of our protocols are closely related to those of [CGS05]. Comparing their result with ours points to interesting open questions to consider. Reference [CGS05] constructed *approximate* quantum error correcting codes of length n capable of correcting up to $(n-1)/2$ errors with high probability (compared to at most $n/4$ correctable errors for an exact code). Thus, the fraction of errors that can be tolerated in [CGS05] approaches $1/2$ as n gets large, which is much higher than in our current scheme. Furthermore, unlike our scheme, no secret key is required. Instead, randomizing parameters are sent as part of the message via carefully constructed secret sharing schemes. However, the alphabet size of the codes in [CGS05] grows as a function of both the blocklength and the code's accuracy, which severely limits the transmission rate. Also, when their large dimensional channel is viewed as a block of qubit channels, the adver-

sary considered in [CGS05] is much more restricted than ours, being limited to the corruption of *continuous blocks* of qubits.

It is an interesting question whether there are *qubit* approximate QECCs that achieve the rates of our codes but don't require the use of a secret key, or, less ambitiously, require only a key of constant size. More generally, the tradeoff between distance, rate, and key required remains to be studied.

As a side remark, the secret key is used in our scheme as a randomizing parameter that is unaccessible to the adversary. Since the adversary must corrupt the transmitted state before it is received by Bob, if Bob is allowed to send a "receipt" of the quantum states to Alice, Alice can simply disclose the random code afterwards and no key is required. In other words, one bit of back communication together with logarithmic forward classical communication (all authenticated) can be used to replace the key requirement.

Our result also finds application to a different problem—entanglement distillation with bounded weight errors. In this problem, a state is already distributed between Alice and Bob, so that the adversary has already acted and randomizing parameters can be sent in public without a "receipt." In [AG], it was shown that n noisy EPR pairs with errors of weight up to pn could be purified to $n(1 - H(p) - p\log 3)$ perfect EPR pairs using a two-way distillation procedure. Our construction allows us to distill high fidelity EPR pairs at the same rate using only forward classical communication. In fact, the authors of [AG] speculated that it would be possible to reduce the computational complexity of their protocols by using quantum list codes—almost exactly the approach taken here, though in our case with an eye towards reducing the communication required. The question of efficent encoding and decoding via list codes has not yet been resolved.

There remain several potentially fruitful avenues of inquiry about adversarial quantum channels. The most obvious question we have left unanswered regards the capacity of $\mathcal{N}_{p,n}^{\mathrm{adv}}$ assisted by a negligible length secret key. It seems quite likely that this is equal to the capacity of the depolarizing channel with error rate p, which would be in analogy with the classical result of [Lan04]. It may also be interesting to consider how restricting the computational power of our adversary affects the channel's capacity, which is another topic we will leave to future work.

Chapter 4

Degenerate Coding I—Repetition Codes

One of the most striking features of quantum error correcting codes is that they can sometimes be used to correct more errors than they can uniquely identify. That is, quantum codes may be degenerate. In this chapter we study a family of such codes and show they can be used to communicate over almost any Pauli channel at rates that are impossible for a nondegenerate code and that exceed those of previously known degenerate codes. We also identify a channel for which none of our codes outperforms the best nondegenerate code and show that it is nevertheless quite unlike any channel for which nondegenerate codes are known to be optimal.

4.1 Introduction

It was Shannon [Sha48] who discovered, by a random coding argument, the beautiful fact that the capacity of a noisy channel $\mathcal{N}$ is equal to the maximal mutual information between an input variable, X, and its image under the action of the channel:

$$C = \max_X I(X;\mathcal{N}(X)). \tag{4.1}$$

It is remarkable that this maximization is over a single input to the channel; it does not require consideration of inputs correlated over many channel uses. This reflects the fact that a random code whose codewords have letters chosen independently according to the distribution maximizing Eq. (4.1) will with high probability have vanishingly small error probability for any rate less than C.

The natural quantum generalization of Eq. (4.1) is

$$Q_1 = \max_{\phi_B} I^c(\mathcal{N}, \phi_B), \tag{4.2}$$

where

$$I^c(\mathcal{N}, \phi_B) = I^c\left(I \otimes \mathcal{N}(|\phi^{AB}\rangle\langle\phi^{AB}|)\right), \tag{4.3}$$

$|\phi_{AB}\rangle$ is a purification of ϕ_B, and

$$I^c(\rho_{AB}) = S(\rho_B) - S(\rho_{AB}), \tag{4.4}$$

known as the *coherent information*, is a measure of quantum correlations playing a similar role to the mutual information of Eq. (4.1). One would hope that a random code with any rate less than the capacity chosen on the typical subspace of $\phi_B^{\otimes n}$, where ϕ_B maximizes the coherent information, would have transmission fidelity close to 1 with high probability. While we can achieve Q_1 in this way, it has been known for almost a decade that this rate is suboptimal in some settings [SSa, DSS98]. In fact, there are examples of codes that, concatenated with a random code, achieve rates beyond Q_1 for the very noisy depolarizing channel. The correct quantum capacity formula is not Q_1, and is given by [Dev05, Sho, Llo97]

$$Q = \lim_{n\to\infty} \frac{1}{n} \max_{\phi_{B_n}} I^c\left(\mathcal{N}^{\otimes n}, \phi_{B_n}\right), \tag{4.5}$$

where the need for regularization reflects the fact that we must consider the behaviour of the channel on inputs entangled across many uses.

While it is in general intractable to perform the optimization in Eq. (4.5), it is usually possible to evaluate Q_1. For instance, a Pauli channel $\mathcal{N}_{\mathbf{p}}$, which applies X, Y, and Z errors with probabilities $\mathbf{p} = (p_x, p_y, p_z)$, has a hashing rate (as Q_1 is often called) of $1 - H(\mathbf{p})$, where $H(\mathbf{p})$ is the entropy of $(1 - p_x - p_y - p_z, p_x, p_y, p_z)$. While we will show one can often achieve higher rates, there *is* a sense in which this is optimal—it's the maximum rate for a code that completely determines the identity of the errors it corrects; that is, Q_1 is the best you can do with a *nondegenerate* code. For a Pauli channel, a random stabilizer code achieves Q_1, so it is both easy to evaluate and achievable by a straightforward code.

It was found in [SSa, DSS98] that there can exist *degenerate* codes (codes wherein multiple typical errors with the same action on the codespace are assigned the same syndrome) that achieve rates beyond Q_1. They showed that an m-qubit repetition code, sometimes called a "cat code" because the code space is spanned by $|0\rangle^{\otimes m}$ and $|1\rangle^{\otimes m}$, concatenated with a random stabilizer code achieves a higher rate than hashing alone. For a depolarizing channel ($p_x = p_y = p_z = \frac{p}{3}$), Q_1 is nonzero only up to $p \approx 0.18929$ while a 5-qubit cat code followed by hashing has positive rate up to $p \approx 0.19036$. In [DSS98] a 5-qubit cat code in the Z basis concatenated with 5-qubit cat code in the X basis and finally with a random stabilizer code was shown to have positive rate up to $p \approx 0.19056$. While these codes beat hashing, they do so by a small amount over a tiny range, and very little has been

understood about why, besides the notion that degeneracy is involved.

In this chapter we study the performance of degenerate codes for general Pauli channels. We first provide an explicit formula for the rate achieved over $\mathcal{N}_{\mathbf{p}}$ by an m-qubit repetition code concatenated with a random stabilizer code, finding a channel for which the benefit of degenerate codes over the hashing rate is dramatic—its hashing rate goes to zero at $p_x{+}p_y{+}p_z{=}p \approx 0.274$ whereas repetition codes allow nonzero rates up to $p \approx 0.295$. While the optimal repetition length and basis vary, as does the magnitude of the benefit, it is a generic fact that using such a code is beneficial in the regime where the hashing rate is near zero. By studying the scaling of the optimal repetition length as a function of $\mathbf{p}$ we arrive at an intuitive understanding of the role of degeneracy in our codes. We also find a channel for which all of our codes fail to outperform hashing, and show it is nevertheless quite unlike any channel for which hashing is known to be optimal. Finally, we use our improved understanding of degenerate coding to find codes for the depolarizing channel that outperform those of [SSa, DSS98] and mention some ideas for codes that may be even better.

4.2 Cat Codes for Pauli Channels

The code we will consider has stabilizers

$$Z_1Z_2, Z_1Z_3, \ldots Z_1Z_m \tag{4.6}$$

and logical operators

$$\bar{X} \quad = \quad X^{\otimes m} \text{ and} \tag{4.7}$$

$$\bar{Z} = Z_1, \tag{4.8}$$

so that an error of the form $X^{\mathbf{u}}Z^{\mathbf{v}}$ leads to syndrome $\{u_1 \oplus u_2, \ldots, u_1 \oplus u_m\}$ and in the absence of a recovery operation gives a logical error of $\bar{X}^{u_1}\bar{Z}^{\oplus_l v_l}$. By encoding half of $|\phi_{00}^{AB}\rangle$ $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ in our repetition code, we get the state $|\phi_m^{AB}\rangle = \frac{1}{\sqrt{2}}(|0_A\rangle|0_B\rangle^{\otimes m} + |1_A\rangle|1_B\rangle^{\otimes m})$, which is exactly the state for which the coherent information in Eq. (4.5) will be more than m times Q_1. Sending the B system of $|\phi_m^{AB}\rangle$ through $\mathcal{N}_{\mathbf{p}}{}^{\otimes m}$ and subsequently measuring the stabilizers $\{Z_1Z_l\}_{l=2}^m$ leads to the state

$$\rho_{AB_m} = \sum_{\mathbf{r}\in\{0,1\}^{m-1}} \Pr(\mathbf{r}) I \otimes \mathcal{N}^{\mathbf{r}}(|\phi_{00}\rangle\langle\phi_{00}|) \otimes |\mathbf{r}\rangle\langle\mathbf{r}|, \tag{4.9}$$

where $\mathbf{r}$ is the syndrome measured, $\mathcal{N}^{\mathbf{r}}$ is the induced channel given $\mathbf{r}$ (which is also a Pauli channel), and $\Pr(\mathbf{r})$ is the probability of measuring $\mathbf{r}$. Concatenating our repetition code with a random

stabilizer code allows communication with high fidelity at a rate of

$$\frac{1}{m}I^{\rm c}(\rho_{AB_m}) = \frac{1}{m}\sum_{\mathbf{r}} \Pr(\mathbf{r}) I^{\rm c}(I\otimes\mathcal{N}^{\mathbf{r}}(|\phi_{00}\rangle\langle\phi_{00}|)). \tag{4.10}$$

Because the repetition code is highly symmetric we can find explicit formulas for both $\Pr(\mathbf{r})$ and $\mathcal{N}^{\mathbf{r}}$, and thus a fairly compact expression for $I^{\rm c}(\rho_{AB_m})$. The joint probabilities of logical errors and syndrome outcomes are, as proved in the appendix, given by

$$\begin{aligned}\Pr(\bar{X}^u\bar{Z}^v,\mathbf{r}) &= \frac{1}{2}\Big((p_x{+}p_y)^{u(m-2r)+r}(1{-}p_x{-}p_y)^{(1-u)(m-2r)+r} &(4.11)\\ &\quad +(-1)^v(p_x{-}p_y)^{u(m-2r)+r}(1{-}p_x{-}p_y{-}2p_z)^{(1-u)(m-2r)+r}\Big), &(4.12)\end{aligned}$$

where $r=|\mathbf{r}|$, and which allows us to find both $\Pr(r)$ and the error probabilities of $\mathcal{N}^{\mathbf{r}}$, $\Pr(\bar{X}^u\bar{Z}^v|r) = \Pr(\bar{X}^u\bar{Z}^v,r)/\Pr(r)$. A salient feature of this formula is that it depends on r but has no other dependence on $\mathbf{r}$.

By evaluating Eq. (4.10) for the probabilities of Eq. (4.12), we find that for almost all Pauli channels there is some repetition code that offers a nonzero rate at the hashing point. When $p_x \gg p_z$ the best code is in the Z basis with length scaling like $1/p_z$, which we'll study in detail in the next section. In general, for $p_x \geq p_z \geq p_y$ it is a good rule of thumb to use a Z repetition code of length $m \approx 1/p_z$, with the largest increase in rate for fairly asymmetrical channels (e.g., Fig. 4.1).

4.3 The Almost Bitflip Channel

In order to develop an understanding of how to choose a repetition code length, we will study their performance for channels with independent phase and amplitude error probabilities. An error X^uZ^v is said to be a phase error if $v=1$ and an amplitude error if $u=1$ (note that when $u=1$ and $v=1$ it is both). Throughout, we define $q_x{=}p_x{+}p_y$ and $q_z{=}p_y{+}p_z$ to be the amplitude and phase error probabilities, respectively, and in a slight abuse of terminology refer to amplitude and phase errors as X and Z errors, with a Y error being "both X and Z." Independence of phase and amplitude errors requires $p_x = q_x(1{-}q_z)$, $p_y = q_xq_z$, and $p_z = q_z(1{-}q_x)$. When $q_x \gg q_z$ we find that the repetition code with the best zero-rate noise threshold has $m \approx 1/q_z$, which can be understood by considering the effective channels induced by the code.

The independence of phase and amplitude, together with our generators involving only Z's tells us that the probability of a logical phase error is independent of $\mathbf{r}$, and given by

$$q_{\bar{Z}} = \Pr\left(\oplus_{l=1}^m v_l{=}1\right) = [1{-}(1{-}2q_z)^m]/2, \tag{4.13}$$

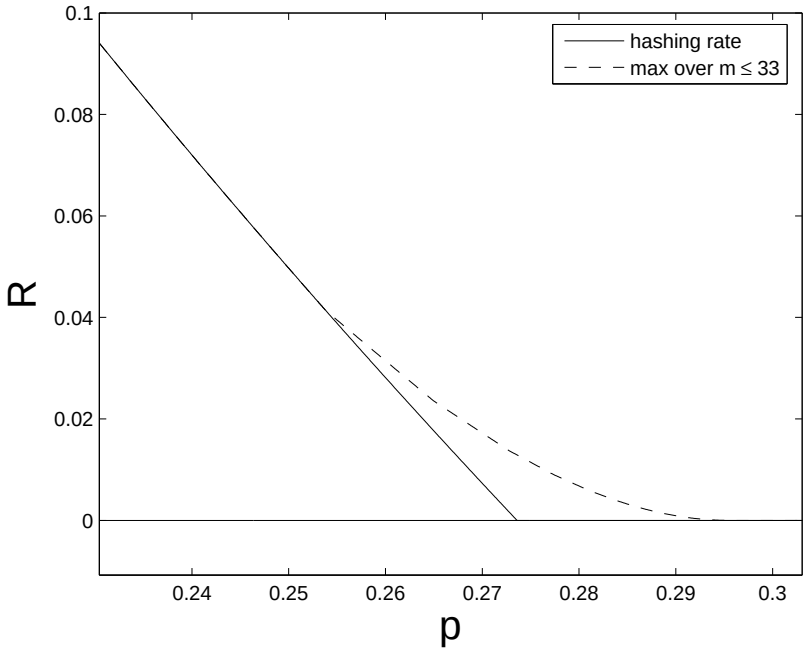


Figure 4.1: Best Z-cat code rates for independent phase and amplitude errors with $q_z = q_x/9$ (and where $p = p_x{+}p_y{+}p_z$). The optimal m increases with p. $m = 33$ gives the best threshold of $\approx .295$, compared to a hashing threshold less than .274.

which also follows from Eq. (4.12).

As we have already seen, the probability of a logical amplitude error depends only on $r = |\mathbf{r}|$, not on $\mathbf{r}$ itself. If m is large, the probability distribution of r becomes concentrated near $r_o \equiv (m{-}1)q_x$ and $r_1 \equiv (m{-}1)(1{-}q_x)$. This is because there are typically $(m{-}1)q_x$ X errors on qubits 2 through m and these qubits all get flipped if qubit 1 has an X error. So, the measured value of r tells us whether or not a logical X error has occurred, at least with high probability. The probability that this is incorrect, leading to a misdiagnosed amplitude error, is

$$q_{\bar{X}|r_0} = \frac{1}{1+\left((1-q_x)/q_x\right)^{m-2r_0}} \tag{4.14}$$

for the $r{=}r_0$ syndromes and

$$q_{\bar{X}|r_1} = \frac{1}{1+\left((1-q_x)/q_x\right)^{2r_1-m}} \tag{4.15}$$

for $r{=}r_1$, both of which scale like

$$\tilde{q}_{\bar{X}} \equiv \left(\frac{q_x}{1-q_x}\right)^{m(1-2q_x)} \tag{4.16}$$

for large m. One can see from this, together with the $q_{\bar{Z}}$ above, that as m increases we learn more about the logical X error at the expense of knowing less about the logical Z.

The optimal repetition length will minimize the entropy in the logical qubits conditioned on r, given by

$$H_m \approx H(\tilde{q}_{\bar{X}}) + H(q_{\bar{Z}}). \tag{4.17}$$

We will have beaten the hashing rate when

$$(1-H_m)/m > 1-H(q_x)-H(q_z), \tag{4.18}$$

so that at the hashing point, where $H(q_x)+H(q_z) = 1$, our goal is to have $H_m < 1$.

If $q_x \gg q_z$, at the hashing point we will have

$$q_x \approx 1/2-\epsilon \tag{4.19}$$

with

$$\epsilon = (1/2)\sqrt{q_z \ln(1/q_z)}, \tag{4.20}$$

so that the entropy left in the logical qubits will be roughly

$$H(e^{-8\epsilon^2 m}) + H(q_{\bar{Z}}). \tag{4.21}$$

For repetition length m, the expected number of Z errors on the block is mq_z, which seems to suggest that if m is allowed to be large it will be completely unclear whether a Z error occurred, leaving us with a bit of entropy in the Z errors alone. However, because the number of Z errors will be roughly Poisson distributed with mean 1 for $m \approx 1/q_z \gg 1$, although the expected number of Z errors is 1, the probability of 0, 1, and 2 errors are all quite comparable, leaving well under a bit of entropy in the logical Z error. The probability of an X error for $m \approx 1/q_z$ is roughly $e^{-2\ln(1/q_z)} = q_z^2$, so that there will be very little entropy left in the X's, leading to an overall entropy of under a bit. The point is that until the repetition length gets to be around $1/q_z$, increasing m allows us to gain information about the logical X error faster than we destroy our knowledge of the logical Z's.

A more detailed analysis of this case shows the entropy is minimized for

$$m \approx \ln\ln(1/q_z)/(2q_z \ln(1/q_z)), \tag{4.22}$$

which leads to

$$H(\bar{X}) \approx H(\bar{Z}) = O\left(\ln\ln(1/q_z)/\ln(1/q_z)\right), \tag{4.23}$$

giving an overall rate of

$$(1/m)\left(1-H(\bar{X})-H(\bar{Z})\right) \approx 2q_z \ln(1/q_z)/\ln\ln(1/q_z). \tag{4.24}$$

Note that essentially all of the entropy in the X errors is removed by the best code, with the optimal length determined by a tradeoff between the reduction of entropy in the X errors and the increase of entropy in the Z errors. This sort of tradeoff also determines the optimal repetition code length for a general Pauli channel.

4.4 Concatenated Repetition Codes

We can immediately apply this analysis to design even better codes by using concatenation. By adapting a second level of repetition code to the error probabilities of the channels induced by the first level we can exceed the performance of any single level cat code. We have used this approach for the depolarizing channel with the results shown in Fig. 4.2, where we plot the probabilities at which the rate of a concatenated 3 in m and 5 in m code goes to zero as a function of m, the size of the outer cat code. If we first use a 3-cat code in the Z basis, followed by an m-cat code in the X basis, we find the highest threshold for a 3 in 19 code, with a nonzero rate up to $p \approx 0.19086$, surpassing the codes of [DSS98]. Starting with a 5-cat code the threshold increases up to $p \approx 0.19088$ for $m = 16$, the best known code for this channel, but for higher values of m the computation of this probability is quite slow. Based on the character of the channels induced by the inner repetition code, together with the behaviour for $m \leq 16$ we expect that the threshold increases until something like 5 in 25, at which point a larger m begins to reduce the effectiveness of the code.

4.5 A Special Channel

Besides the one-Pauli channels, the only channels for which we can find no code offering an advantage near the hashing point are tightly concentrated near

$$\mathcal{N}_p^{\mathrm{tp}}(\rho) \equiv (1-p)\rho + \frac{p}{2} X\rho X + \frac{p}{2} Z\rho Z. \tag{4.25}$$

While Rains has shown [Rai99] that hashing is optimal for one-Pauli channels, $\mathcal{N}_p^{\mathrm{tp}}$ is not known to have additive coherent information, which is equivalent to the optimality of hashing. Furthermore, we will show that unlike all channels known to be additive this channel is not degradable [DS05].

Every channel, $\mathcal{N}$, can be expressed as an isometry followed by a partial trace, which is to say there is an isometry $U_{\mathcal{N}} : A \to BE$ such that

$$\mathcal{N}(\rho) = \mathrm{Tr}_E U_{\mathcal{N}} \rho U_{\mathcal{N}}^\dagger. \tag{4.26}$$

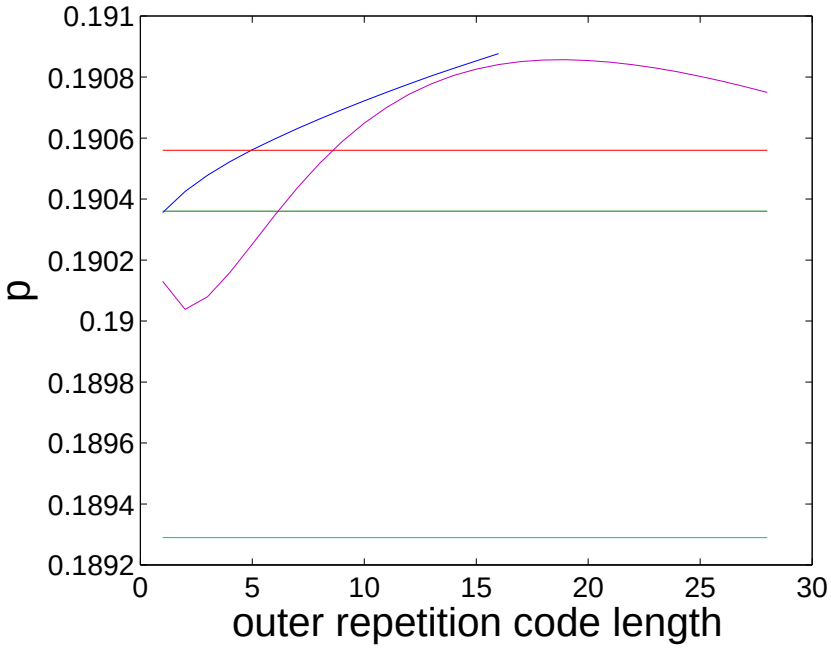


Figure 4.2: Error probability where the rate goes to zero, as a function of length of second level cat code. The bottom line shows the threshold for hashing, the middle line for the 5-cat code, and the upper line for the concatenated 5 in 5 cat code. The lower curve is the threshold of a 3 in m concatenated cat code as a function of m, while the upper curve shows the threshold for a 5 in m concatenated code.

The complementary channel of $\mathcal{N}$, called $\mathcal{N}^C$, results by tracing out system B rather than E:

$$\mathcal{N}^C(\rho)=\mathrm{Tr}_B U_{\mathcal{N}} \rho U_{\mathcal{N}}^\dagger. \tag{4.27}$$

A channel is called degradable if there is a completely positive map, $\mathcal{D} : B \to E$, which "degrades" $\mathcal{N}$ to $\mathcal{N}^C$, so that

$$\mathcal{D} \circ \mathcal{N}=\mathcal{N}^C. \tag{4.28}$$

The existence of such a map immediately implies the additivity of I^c[DS05], which can be seen by noting that

$$I^c(\mathcal{N}^{\otimes(n_1+n_2)}, \phi_{B_{n_1}B_{n_2}}) \leq I^c(\mathcal{N}^{\otimes n_1}, \phi_{B_{n_1}}) + I^c(\mathcal{N}^{\otimes n_2}, \phi_{B_{n_2}}) \tag{4.29}$$

exactly when

$$I(E_{n_1}; E_{n_2}) \leq I(B_{n_1}; B_{n_2}) \tag{4.30}$$

and recalling that $I(B_{n_1}; B_{n_2})$ cannot increase under local operations. We now show there is no such $\mathcal{D}$ for $\mathcal{N}_p^{\mathrm{tp}}$ when $0 < p < 1$.

Letting

$$\mathcal{N}_p^{\mathrm{tp}}(|i\rangle\langle j|) = \sum_{kl} N_{ij;kl}|k\rangle\langle l| \tag{4.31}$$

define N and

$$\mathcal{N}_p^{\mathrm{tp}\,C}(|i\rangle\langle j|) = \sum_{kl} N^C_{ij;kl}|k\rangle\langle l| \tag{4.32}$$

define N^C, we find

$$N = \begin{pmatrix} 1{-}p/2 & 0 & 0 & p/2 \\ 0 & 1{-}3p/2 & p/2 & 0 \\ 0 & p/2 & 1{-}3p/2 & 0 \\ p/2 & 0 & 0 & 1{-}p/2 \end{pmatrix} \tag{4.33}$$

and

$$N^C = \begin{pmatrix} p/2 & 0 & 0 & 0 & p/2 & \alpha & 0 & \alpha & 1{-}p \\ 0 & -p/2 & \alpha & p/2 & 0 & 0 & \alpha & 0 & 0 \\ 0 & p/2 & \alpha & -p/2 & 0 & 0 & \alpha & 0 & 0 \\ p/2 & 0 & 0 & 0 & p/2 & -\alpha & 0 & -\alpha & 1{-}p \end{pmatrix}, \tag{4.34}$$

where $\alpha = \sqrt{p(1-p)/2}$. If $\mathcal{N}_p^{\mathrm{tp}}$ is degradable, there must be a CPTP map $\mathcal{D}$ such that $\mathcal{D}\circ\mathcal{N} = \mathcal{N}^C$, which is equivalent to $ND = N^C$, with D defined by $\mathcal{D}(|s\rangle\langle t|) = \sum_{uv} D_{st;uv}|u\rangle\langle v|$. For N and N^C as above, this gives

$$D = \begin{pmatrix} p/2 & 0 & 0 & 0 & p/2 & \beta & 0 & \beta & 1{-}p \\ 0 & -\gamma & \beta & \gamma & 0 & 0 & \beta & 0 & 0 \\ 0 & \gamma & -\beta & \gamma & 0 & 0 & \beta & 0 & 0 \\ p/2 & 0 & 0 & 0 & p/2 & -\beta & 0 & -\beta & 1{-}p \end{pmatrix} \tag{4.35}$$

with $\beta = \sqrt{p/(2-2p)}$ and $\gamma = p/(2-4p)$. The Choi matrix [Cho75] of $\mathcal{D}$, $C^{\mathcal{D}}_{ik;jl} = D_{ij;kl}$, is thus

$$C^{\mathcal{D}} = \begin{pmatrix} p/2 & 0 & 0 & 0 & -\gamma & \beta \\ 0 & p/2 & \beta & \gamma & 0 & 0 \\ 0 & \beta & 1{-}p & \beta & 0 & 0 \\ 0 & \gamma & \beta & p/2 & 0 & 0 \\ -\gamma & 0 & 0 & 0 & p/2 & -\beta \\ \beta & 0 & 0 & 0 & -\beta & 1{-}p \end{pmatrix}, \tag{4.36}$$

which contains the subblock

$$\begin{pmatrix} p/2 & -\gamma \\ -\gamma & p/2 \end{pmatrix}. \tag{4.37}$$

This has a negative eigenvalue for all $0 < p < 1$, so that $C^{\mathcal{D}}$ cannot be nonnegative and thus $\mathcal{D}$ is not CP.

Besides repetition codes, we have explored concatenated repetition codes for $\mathcal{N}_p^{\mathrm{tp}}$, all of which performed worse than the hashing rate of $1-H(p)-p$. This suggests the capacity of $\mathcal{N}_p^{\mathrm{tp}}$ is exactly $1-H(p)-p$, and in light of its nondegradability we hope a proof of this conjecture will point towards a new sufficient criterion for the additivity of coherent information.

4.6 Discussion

We have left many questions unanswered, but there are several lines of inquiry we believe are ripe for further progress.

A most tantalizing possibility is that there is a simpler characterization of the quantum channel capacity than is provided by Eq. (4.5). In particular, contrary to what is sometimes claimed, the results of [DSS98, SSa] and this work *do not rule out a single letter formula for the capacity*—what is ruled out is the possibility that the single letter optimized coherent information is the correct formula. It could be that there *is* a single letter formula for the capacity, or less ambitiously simply an efficiently calculable expression, which takes degeneracy into account. The characterization of capacity in terms of coherent information is fundamentally nondegenerate, and it may be this which leads to the necessity of regularization, rather than an inherent superadditivity of quantum information.

On a similar note, it would be nice to find families of quantum codes that are capacity approaching with high probability. This is not the case for random stabilizer codes, nor for random codes on the typical subspace of a state maximizing the single letter coherent information, but perhaps by explicitly considering the codes' degeneracy progress could be made.

More concretely, the two-Pauli channel with equal probabilities seems to be somehow different from other Pauli channels. Given their success with almost all other Pauli channels, the failure of cat codes to beat Q_1 in this case suggests that hashing is optimal. Resolving this conjecture seems to be a manageable problem whose solution may lead to a better understanding of additivity questions for quantum channels in general.

The ideas explored here are also useful for quantum key distribution. In particular, using repetition codes one can improve the noise threshold for BB84 with one-way classical post-processing from 12.4% to 12.9%, which is discussed further in Chapter 6.

Finally, we hope the coding approach suggested by the almost bitflip channel will lead to codes with rates beyond what we have presented here. Focusing on reducing the amplitude error rate with an inner code while trying to avoid scrambling the phase errors more than necessary and following this up with a random stabilizer code (or perhaps a second similarly chosen code reversing the roles

of amplitude and phase) offers an appealing heuristic for code design. Viewed in this way, the inner codes we have considered are quite primitive—a repetition code is the simplest code there is—and it seems likely more sophisticated codes will perform better.

Appendix

Proof. (of Eq. (4.12)) All four expressions above follow by similar reasoning, but for concreteness we focus on the first. The probability of a physical error, $X^{\mathbf{u}}Z^{\mathbf{v}}$ leading to a logical operation I is exactly

$$\begin{aligned}
\Pr(I,\mathbf{r}) &= \sum_{\mathbf{u},\mathbf{v}|u_1=0,\oplus_l v_l=0 \text{and}(u_2,\ldots u_m)=\mathbf{r}} \Pr(X^{\mathbf{u}}Z^{\mathbf{v}}) && (4.38)\\
&= \sum_{\mathbf{v}|\oplus_l v_l=0} \Pr(X^{(0,\mathbf{r})}Z^{\mathbf{v}}) && (4.39)\\
&= \sum_{l,t|l+t\equiv 0 \mod 2} \binom{r}{l}\binom{m-r}{t} p_x^{r-l}p_y^l p_z^t (1-p_x-p_y-p_z)^{m-(r+t)} && (4.40)\\
&= \sum_{l=0}^{r}\sum_{t=0}^{m-r} \frac{1+(-1)^{t+l}}{2}\binom{r}{l}\binom{m-r}{t} p_x^{r-l}p_y^l p_z^t (1-p_x-p_y-p_z)^{m-(r+t)} && (4.41)\\
&= \frac{(p_x-p_y)^r}{2}\sum_{t=0}^{m-r}(-1)^t\binom{m-r}{t} p_z^t (1-p_x-p_y-p_z)^{m-(r+t)} && (4.42)\\
&\quad + \frac{(p_x+p_y)^r}{2}\sum_{t=0}^{m-r}\binom{m-r}{t} p_z^t (1-p_x-p_y-p_z)^{m-(r+t)} && (4.43)\\
&= \frac{1}{2}\left((p_x+p_y)^r(1-p_x-p_y)^{m-r} + (p_x-p_y)^r(1-p_x-p_y-2p_z)^{m-r}\right), && (4.44)
\end{aligned}$$

where we have used the identity $\sum_{n=0}^{N}\binom{N}{n}x^n y^{N-n} = (x+y)^N$ repeatedly. □

Chapter 5

Noisy Preprocessing and Twisted State Distillation

We provide a Shor-Preskill-type security proof for prepare and measure quantum key distribution protocols employing noisy preprocessing and one-way postprocessing of the key. This is achieved by showing that the security of such a protocol is equivalent to that of an associated key distribution protocol in which, instead of the usual maximally entangled states, a more general type of private state called a twisted state is distilled. Except for the more general target state, normal means of entanglement distillation are employed, with the crucial difference that noisy preprocessing allows some phase errors to be left uncorrected without compromising the privacy of the key.

5.1 Introduction

Quantum key distribution (QKD) holds the promise of communication with a level of security that is impossible in a classical world. As such, an enormous amount of experimental and theoretical work on QKD has been carried out in recent years, with such rapid progress in both that widespread use of QKD may not be far off [GRTZ02].

Entanglement has been the cornerstone of many security proofs to date: A prepare & measure protocol by which Alice and Bob generate a secret key is shown to be secure exactly when an associated entanglement distillation protocol succeeds in producing a high fidelity maximally entangled state. Secrecy of the key then follows because maximal entanglement can only be shared between two parties [LC99, SP00, Lo01, GP01, TKI03, GL03, BTB$^+$05, RG]. This paradigm links the information-theoretic concept of security to more concrete physical concepts, such as the monogamy of entanglement [CKW00, KW04]. The resulting proofs are intuitive, somewhat constructive, and allow the designers of QKD schemes to incorporate current methods of quantum error correction and entanglement distillation.

Renner, Gisin, and Kraus adopt a more information-theoretic approach to QKD security with

the surprising result that secure key can be established at noise levels beyond what seems possible in the entanglement-based picture [KGR05, RGK05]. By including a step in which Alice adds noise to her sifted key before proceeding to error correction and privacy amplification, the overall key rate can actually increase. Of course, adding noise damages the correlations held by Alice and Bob but the key observation of [KGR05, RGK05] is that this noise may damage Eve's correlations even more. It is puzzling that this preprocessing can generate key in the presence of noise so strong as to preclude the distillation of even a single Einstein-Podolsky-Rosen (EPR) pair via any one-way protocol. In fact, it was suggested in [KGR05, RGK05] that entanglement-based security proofs would be bound to fail for their protocols.

We find a possible resolution in the observation of [HHHO05] that maximally entangled states are not strictly necessary for generating secret keys. Instead, states that lead to secret keys belong to the class of so-called twisted states. Such states are composed of completely correlated systems A and B containing the uniformly distributed key, along with "shield" systems A' and B'. More precisely, $\gamma_{ABA'B'}$ is said to be a twisted state if there is a set of unitaries $U^{(j)}$ and a "twisting operator" of the form

$$U_{\text{twist}} = \sum_j |jj\rangle_{AB}\langle jj| \otimes U^{(j)}_{A'B'}, \tag{5.1}$$

such that

$$\gamma_{ABA'B'} = U_{\text{twist}} \left(|\Phi\rangle_{AB}\langle\Phi| \otimes \rho_{A'B'} \right) U^{\dagger}_{\text{twist}} \tag{5.2}$$

for some $\rho_{A'B'}$, where $|\Phi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. The twisting operator ensures that, while Alice and Bob may not share a maximally entangled state, Eve's reduced state is independent of the key value. This construction recalls an earlier result [AB02] that the secrecy of key pairs created from entangled systems is not diminished by phase noise in the devices performing the entanglement distillation. Ultimately, Eve's state will factor out, ensuring her ignorance of the key.

A key distribution scheme whose security is based on the distillation of twisted states was presented in [HLLO06], proving that even when an untrusted party provides Alice and Bob with a state whose distillable entanglement is arbitrarily small, it may still be possible for them to establish a secret key. However, the protocol considered there has the crucial drawback that it is not a prepare and measure scheme—to succeed, both Alice and Bob need quantum computers. It is not clear what class of QKD protocols can be cast in the form considered by [HLLO06], and in particular the security proof presented there does not apply the protocols discovered in [KGR05, RGK05].

In the following, we will show that a prepare and measure QKD scheme with noisy preprocessing and one-way postprocessing is secure exactly when an associated twisted state distillation protocol succeeds with high fidelity. This requires only minor modifications to the standard entanglement distillation procedure, and in particular Alice and Bob will make use of Calderbank-Shor-Steane (CSS)-like quantum error-correcting codes. We can establish key at bit error rates as high as 12.4%

for the Bennett-Brassard-84 (BB84) protocol [BB84], and 14.1% for the six-state protocol, matching the rates of [KGR05, RGK05], and surpassing all previous thresholds from entanglement-based proofs. The auxiliary system purifying the noise introduced by Alice will function as a shield, and after bit error correction and a suitable amount of phase error correction, they will be left with a twisted state. The crucial difference from previous entanglement based security proofs is that *Alice and Bob need not correct every phase error in order to guarantee security*, and this savings will often more than compensate for the associated increase in the number of bit errors they must correct.

We first provide a description of a general QKD protocol with noisy preprocessing and a proof sketch of its security, followed by a more detailed analysis of the steps differing from a standard entanglement-based proof.

5.2 Twisted State Distillation

We begin with the quantum reformulation of the BB84 and six-state protocols [LC99, SP00], noting that other protocols can be handled in a similar manner [RG]. In the quantum version of both protocols, Alice begins by preparing the state $|\Phi\rangle_{AB}$ and sending the B system to Bob. In BB84, each party then randomly measures in the X or Z basis (for the six-state protocol, they randomly choose X, Y, or Z), and by public discussion they sift out only those outcomes that correspond to the same basis choice. This is equivalent to Alice (Bob) sending a random bit in (measuring in) one of the bases at random, since the statistics of measurements as well as an eavesdropper Eve's dependence on their outcomes are identical in both cases. Alice and Bob then publicly compare a small fraction of the sifted key to estimate the noise parameters of the channel.

If the noise level is found to be zero, the resulting length-n sifted key can be described by the state $|\Phi\rangle^{\otimes n}$. Otherwise, the most general noisy channels we need to consider are Pauli channels, since all the subsequent operations performed will commute with a (hypothetical) measurement in the Bell-basis, which digitizes the actual noise into this form [LC99, GL03]. Attributing the noise to Eve, we can write the state of the key as

$$\sum_{\mathbf{u},\mathbf{v}} \sqrt{p_{\mathbf{u},\mathbf{v}}}\,(\mathbb{I}_A \otimes X_B^{\mathbf{u}} Z_B^{\mathbf{v}})|\Phi\rangle_{AB}^{\otimes n}|\mathbf{u}\rangle_{E_1}|\mathbf{v}\rangle_{E_2}, \tag{5.3}$$

where $p_{\mathbf{u},\mathbf{v}}$ is the probability of error pattern $X^{\mathbf{u}}Z^{\mathbf{v}}$ described by length-n bit strings $\mathbf{u}$ and $\mathbf{v}$. Furthermore, if Alice and Bob randomly permute their n systems, it is sufficient to consider noise that is independent and identical for each transmitted qubit, given by rate $p_{u,v}$ (see, e.g., Lemma 3 of [GL03], or [SP00, GP01]).

By performing bit error correction and privacy amplification (phase error correction), Alice and Bob proceed to distill the key. To guarantee privacy, the protocol must be secure against all Pauli

channels consistent with the outcome of the parameter estimation phase.

First, however, Alice adds independent and identically distributed noise to her system, randomly applying X at rate q. This procedure can be described in a coherent way by first adding an auxiliary system A' prepared in the state $|\varphi\rangle_{A'} = \sqrt{1-q}|0\rangle_{A'} + \sqrt{q}|1\rangle_{A'}$ and then using this as the control system in a controlled-NOT gate, all of which results in the state

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{f}} \sqrt{p_{\mathbf{u},\mathbf{v}} q_{\mathbf{f}}}\, |\mathbf{f}\rangle_{A'} (X_A^{\mathbf{f}} \otimes X_B^{\mathbf{u}} Z_B^{\mathbf{v}}) |\Phi\rangle_{AB}^{\otimes n} |\mathbf{u}\rangle_{E_1} |\mathbf{v}\rangle_{E_2}, \tag{5.4}$$

where $q_{\mathbf{f}} = q^{|\mathbf{f}|}(1-q)^{n-|\mathbf{f}|}$ for length-n bit string $\mathbf{f}$ and $|\mathbf{f}|$ its Hamming weight. We can also think of Alice's error operator acting on Bob's system, since $X \otimes XZ$ and $\mathbb{I} \otimes XZX$ have the same effect on $|\Phi\rangle$.

Now Alice and Bob perform bit error correction using a standard linear error correcting code. This step is the same as the usual analysis, since all bit errors must be corrected in the final key, no matter their source. The bit error rate is

$$\tilde{p} = p_x(1-q) + q(1-p_x) \tag{5.5}$$

for $p_x = \sum_v p_{1,v}$, meaning Alice and Bob need to measure $nH_2(\tilde{p})$ parity syndromes, where H_2 is the binary Shannon entropy, in order to uniquely identify the error pattern with high probability. To simplify the resulting expressions, we use the method of decoupling error correction and privacy amplification [Lo03], itself based on the breeding entanglement distillation protocol [BBP$^+$96], whereby the syndromes are collected in auxiliary shared entangled pairs.

In the classical description of the protocol, this amounts to encrypting the syndromes with a one-time pad before transmission, preventing information leakage to Eve. Since this encryption requires a key, which in the quantum description is a twisted state, Alice and Bob generally collect the parity syndromes in the key subsystems of twisted states, not maximally entangled states. Fortunately, this raises no additional complications, as pointed out in [HLLO06], and there will be no loss of generality in taking the preshared ancilla state to be maximally entangled in what follows[1].

Alice collects the bit parities in her halves of the ancilla states, measures them, and sends the result to Bob. Taking this information, Bob coherently corrects system B and records the error in an ancilla system B', producing

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{f}} \sqrt{p_{\mathbf{u},\mathbf{v}} q_{\mathbf{f}}}\, Z_{A'}^{\mathbf{v}} |\mathbf{f}\rangle_{A'} |\mathbf{u}+\mathbf{f}\rangle_{B'} Z_B^{\mathbf{v}} |\Phi\rangle_{AB}^{\otimes n} |\mathbf{u}\rangle_{E_1} |\mathbf{v}\rangle_{E_2}, \tag{5.6}$$

[1] Note that if we use the key part of a twisted state to measure a bit parity of a noisy EPR pair, the control from the key system to the shield is transfered to the noisy EPR pair, which implies that the noisy EPR pair can be corrected to a twisted state exactly when using a perfect EPR to measure the parity would have made this possible.

where the $Z^{\mathbf{v}}_{A'}$ comes from the need to commute $X^{\mathbf{f}}_B$ through $Z^{\mathbf{v}}_B$ before correcting.

At this stage, the normal entanglement-based proof would proceed to correct the phase errors, corresponding to privacy amplification in the prepare and measure scheme. This would not give the error thresholds found in [KGR05, RGK05], since the only effect of the extra noise would be to reduce the key rate and lower the threshold. Instead, we come to the central observation of this chapter: not all phase errors need to be corrected. After correcting enough of them, the resulting state will be very close to a twisted state, resulting in approximately private key.

Examining the state shared by Alice and Bob will make clear how this comes about. Tracing out Eve's systems, they hold the state

$$\rho = C_{A'B'}\left(\sum_{\mathbf{u},\mathbf{v}} p_{\mathbf{u},\mathbf{v}}[\mathbf{u}]_{B'}[\varphi^{\mathbf{v}}]_{A'}Z^{\mathbf{v}}_B[\Phi]_{AB}Z^{\mathbf{v}}_B\right)C^{\dagger}_{A'B'}, \tag{5.7}$$

where $[\theta] = |\theta\rangle\langle\theta|$,

$$|\varphi^{\mathbf{v}}\rangle = Z^{\mathbf{v}}|\varphi\rangle^{\otimes n}, \tag{5.8}$$

and we have used a controlled-NOT $C_{A'B'}$ to write

$$|\mathbf{f}\rangle_{A'}|\mathbf{u}{+}\mathbf{f}\rangle_{B'} \tag{5.9}$$

as

$$C_{A'B'}|\mathbf{f}\rangle_{A'}|\mathbf{u}\rangle_{B'}. \tag{5.10}$$

By performing phase error correction at a reduced rate, the pattern of phase errors will not be uniquely identified, but rather narrowed to a set $\mathcal{V}_{\mathbf{s}}$ indexed by the syndrome $\mathbf{s}$:

$$\mathcal{V}_{\mathbf{s}} = \{\mathbf{v} \,|\, \text{syndrome}(\mathbf{v}) = \mathbf{s}\}. \tag{5.11}$$

The key point here is that if the vectors $|\varphi^{\mathbf{v}}\rangle$ for $\mathbf{v} \in \mathcal{V}_{\mathbf{s}}$ were mutually orthogonal, we could define the unitary operator

$$D_{A'B} = \sum_{\mathbf{v}\in\mathcal{V}_{\mathbf{s}}} [\varphi^{\mathbf{v}}]_{A'} \otimes Z^{\mathbf{v}}_B \tag{5.12}$$

and use $U_{BA'B'} = D_{A'B}C_{A'B'}$ to untwist the state:

$$\begin{aligned}\rho' &= U_{BA'B'}\rho U^{\dagger}_{BA'B'} \\ &= [\Phi]^{\otimes n}_{AB} \otimes \Big(\sum_{\mathbf{u}} p_{\mathbf{u}}[\mathbf{u}]_{B'} \sum_{\mathbf{v}\in\mathcal{V}_s} p_{\mathbf{v}|\mathbf{u}}\,[\varphi^{\mathbf{v}}]_{A'}\Big).\end{aligned} \tag{5.13}$$

Since D is a controlled-Z gate, either system can be thought of as the control, so

$$D_{A'B} = \sum_{\mathbf{j}} U^{(\mathbf{j})}_{A'} \otimes [\mathbf{j}]_B \tag{5.14}$$

for some unitaries $U^{(j)}$. $U_{BA'B'}$ is a twisting operation, so that Alice and Bob would share a twisted state. Keys derived from this state would be secret.

5.3 Detailed Analysis

To rigorously establish the secrecy of keys generated from ρ, recall the univerally-composable definition of security formulated in [KR05]. A key K is called ϵ-secure if the state ρ_{KE} of the key and eavesdropper satisfies

$$||\rho_{KE} - \kappa \otimes \rho_E||_1 \leq 2\epsilon, \tag{5.15}$$

where κ is a uniform mixture of all possible key values shared by Alice and Bob. The latter state is a perfect key and this formulation ensures that ρ_{KE} can safely be used for any further cryptographic purpose.

In the present context, the key is created by measuring systems A and B of ρ in the Z basis. As the untwisting operation is unitary and commutes with the measurement by definition, whether it is performed before the measurement or after does not affect the security of the key. When performing the untwisting operation on the unmeasured state results in a maximally entangled state on AB, the generated key will be perfectly secure. By the same token, if there exists an untwisting operation that maps the AB subsystem to within 2ϵ of a maximally entangled state, then the key is ϵ-secure.

For simplicity we first consider the case of independent amplitude and phase errors, with the case of correlated $\mathbf{u}$ and $\mathbf{v}$ following along similar lines. To construct an untwisting operation, it suffices to find a rank-one POVM having elements $E_{\mathbf{v}}$ that can distinguish the $|\varphi^{\mathbf{v}}\rangle$ with average error P_{e} no larger than $\epsilon^2/2$:

$$P_{\mathrm{e}} = \langle P^{\mathbf{v}}_{\mathrm{e}} \rangle = \sum_{\mathbf{v},\mathbf{v}' \neq \mathbf{v}} p_{\mathbf{v}} \langle \varphi^{\mathbf{v}} | E_{\mathbf{v}'} | \varphi^{\mathbf{v}} \rangle \leq \epsilon^2/2, \tag{5.16}$$

where $P^{\mathbf{v}}_{\mathrm{e}}$ is probability of decoding input state $|\varphi^{\mathbf{v}}\rangle$ incorrectly. This problem was considered by [HJS$^+$96] in the context of transmitting classical information over a quantum channel. Letting

$$\sigma = (1 - p_z)|\varphi\rangle\langle\varphi| + p_z Z|\varphi\rangle\langle\varphi|Z, \tag{5.17}$$

$p_z = \sum_u p_{u1}$, and $S(\sigma)$ be the entropy of σ, their results imply that with probability $1 - \epsilon^2/2$, the

elements of a randomly chosen subset

$$\mathcal{V}_{\mathbf{s}} \subset \mathcal{V} \tag{5.18}$$

of size

$$2^{n(S(\sigma)-\delta)} \tag{5.19}$$

can be distinguished by the pretty-good measurement (PGM) with average error probability $\epsilon^2/2$ where ϵ decreases exponentially with n for arbitrarily small positive δ.

The PGM has rank-one elements by construction [HW94], so we can write

$$E_{\mathbf{v}} = |\widetilde{\theta}^{\mathbf{v}}\rangle\langle\widetilde{\theta}^{\mathbf{v}}| \tag{5.20}$$

for unnormalized $|\widetilde{\theta}^{\mathbf{v}}\rangle$. Then we can append another auxiliary system A'' and consider the Neumark extension consisting of orthonormal states

$$|\theta^{\mathbf{v}}\rangle_{A'A''} \tag{5.21}$$

in the joint Hilbert space $A'A''$ such that

$${}_{A'A''}\langle\theta^{\mathbf{v}}|\varphi^{\mathbf{v}'}\rangle_{A'}|0\rangle_{A''} = {}_{A'}\langle\widetilde{\theta}^{\mathbf{v}}|\varphi^{\mathbf{v}}\rangle_{A'} \tag{5.22}$$

(see, e.g., [NC04]). With this, we can finally construct the untwisting operator

$$U = (\sum_{\mathbf{v}} [\theta^{\mathbf{v}}]_{A'A''} \otimes Z_B^{\mathbf{v}}) C^{\dagger}_{A'B'}. \tag{5.23}$$

Letting $\widetilde{\rho} = |0\rangle\langle 0| \otimes \rho$, the fidelity of $U\widetilde{\rho}U^{\dagger}$ with

$$\rho' = [\Phi]_{AB}^{\otimes n} \otimes \sum_{\mathbf{u},\mathbf{v}} p_{\mathbf{u},\mathbf{v}} [\theta^{\mathbf{v}}]_{A'A''} \otimes [\mathbf{u}]_{B'} \tag{5.24}$$

is given by

$$F(U\widetilde{\rho}U^{\dagger}, \rho') = \sum_{\mathbf{u},\mathbf{v}} p_{\mathbf{u},\mathbf{v}} \, |\langle\varphi^{\mathbf{v}}|\widetilde{\theta}^{\mathbf{v}}\rangle| = \langle\sqrt{P_{\mathrm{s}}^{\mathbf{v}}}\rangle, \tag{5.25}$$

where $P_{\mathrm{s}}^{\mathbf{v}}$ is the conditional probability of successful transmission of $\mathbf{v}$. Since

$$\begin{aligned} \langle\sqrt{P_{\mathrm{s}}^{\mathbf{v}}}\rangle &\geq \langle P_{\mathrm{s}}^{\mathbf{v}}\rangle & (5.26)\\ &= 1 - P_{\mathrm{e}} & (5.27)\\ &\geq 1 - \epsilon^2/2, & (5.28) \end{aligned}$$

using the relation between trace norm and fidelity [FvdG99], we find

$$\begin{aligned} ||U\widetilde{\rho}U^\dagger - \rho'||_1 &\leq 2\sqrt{1-F^2} && (5.29)\\ &\leq 2\sqrt{\epsilon^2 - \epsilon^4/4} && (5.30)\\ &\leq 2\epsilon, && (5.31) \end{aligned}$$

proving ϵ-security.

A subtlety arises in the use of the Neumark extension in that we have moved beyond the usual definition of a twisted state. Rather than showing that ρ is close to a state of the form

$$U_{\text{twist}}\left(|\Phi\rangle_{AB}\langle\Phi| \otimes \rho_{A'B'}\right) U_{\text{twist}}^\dagger \tag{5.32}$$

with

$$U_{\text{twist}} = \sum_j |jj\rangle_{AB}\langle jj| \otimes U^{(j)}_{A'B'}, \tag{5.33}$$

we have shown that there is an auxiliary space A'' such that

$$\widetilde{\rho} = |0\rangle\langle 0|_{A''} \otimes \rho \tag{5.34}$$

is close to such a state. However, the privacy of the key is uncompromised: While Eve may have knowledge of the shield system, as long as Alice and Bob hold the key and shield, the fact that they could be untwisted implies that Eve is ignorant of the key.

In the above, we took $\mathbf{u}$ and $\mathbf{v}$ to be independent. When they are not, randomly choosing sets $\mathcal{V}_{\mathbf{s}}$ of size $2^{n(S(\sigma|u)-\delta)}$, where $S(\sigma|u)$ is the conditional entropy of σ given u, will lead to an exponentially small average probability of decoding error for the PGM, and the rest of the argument remains unchanged (see, e.g., [Lo01]).

Putting this all together, by using a random code Alice and Bob can select a subset $\mathcal{V}_s$ of size $\approx 2^{nS(\sigma|u)}$. Then, with probability exponentially close to unity, the untwisting operation can be constructed from the pretty-good measurement, ensuring the final key is ϵ-secure.

5.4 Achievable Key Rates

What key generation rates can be achieved by the protocols considered above? The bit error correction step consumes $nH_2(\tilde{p})$ previously established secret key bits, but in so doing produces n error-free bits. The phase error correction must reduce the number of errors to $2^{nS(\sigma|u)}$ in order to

ensure that Alice and Bob could untwist the state, so we find an overall rate of

$$1 - H_2(\tilde{p}) - (H(v|u) - S(\sigma|u)). \tag{5.35}$$

This can be written as

$$R = 1 - H_2(\tilde{p}) - \sum_u p_u \left(H_2(p_{1|u}) - H_2(\lambda_u^+) \right), \tag{5.36}$$

where

$$\lambda_u^+ = \frac{1}{2}(1 + \sqrt{1 - 16q(1-q)p_{1|u}(1-p_{1|u})}) \tag{5.37}$$

is the larger eigenvalue of

$$\sigma_u = (1 - p_{1|u})|\varphi\rangle\langle\varphi| + p_{1|u} Z|\varphi\rangle\langle\varphi|Z. \tag{5.38}$$

In the BB84 protocol, bit and phase errors are equal but uncorrelated, meaning $p_{1|u} = p_z = p_x = p_{1|v}$. From this one immediately finds a maximum sustainable error rate of 12.4% by letting $q \to 1/2$. In the six-state protocol all Pauli errors occur at the same rate, from which we find a threshold error rate of 14.1%. Both of these figures agree with those found in [KGR05, RGK05].

5.5 Discussion

We have shown that one-way key distribution protocols employing noisy preprocessing can be understood as distillation protocols for twisted states, extending the entanglement distillation paradigm initiated in [LC99, SP00]. This resolves two outstanding issues surrounding entanglement-based proofs of security—the role of twisted states, and the apparent impossiblity of achieving the key rates of [KGR05, RGK05] with an entanglement-based proof. By formulating the protocol in this manner, we gain insight into the mechanism by which addition of noise improves the key rate, namely by "deflecting" Eve's correlations with Alice and Bob to the shield and away from the key.

In the security proof of the six-state protocol [Lo01], building on the work of [DSS98] in the context of entanglement distillation, Lo showed that a degenerate error-correcting code could be used to improve the threshold error rate from 12.6% to 12.7%. Further progress in this direction can be found in the next chapter, where we report on the combination of that method with the noisy preprocessing studied here, showing that the threshold error rate of BB84 can be increased from 12.4% to 12.9%. We believe our findings will point towards new methods of key distillation and analagous methods of twisted state distillation, furthering the fruitful exchange between privacy amplification and entanglement distillation.

Chapter 6

Degenerate Coding II—Better Codes for BB84

In this chapter we study achievable secret key rates for the Bennett-Brassard-84 (BB84) quantum key distribution protocol with one-way classical postprocessing. Specifically, we characterize the performance of a family of error correcting codes when used in the information reconciliation phase of BB84. When combined with noisy preprocessing, these codes allow secure key to be established for quantum bit error rates up to 0.129. Taken together, our information reconciliation and privacy amplification stages can be described by a massively degenerate CSS code whose improvement over the previous best noise threshold of 0.124 is analogous to the benefit of degenerate codes over random stabilizer codes when communicating over a very noisy quantum channel.

6.1 Introduction

Quantum key distribution (QKD) allows two parties using public channels to remotely establish a secret key whose security is not predicated on the difficulty of some computational task. Rather, the security of the key generated by a QKD protocol depends only on fundamental laws of physics. For this reason there has recently been an enormous amount of work on practical and theoretical aspects of QKD, and corresponding rapid progress in both (e.g., [GRTZ02]).

The first QKD protocol was proposed by Bennett and Brassard in 1984 [BB84], and like all QKD schemes, it is based on the tradeoff between information gain and disturbance in quantum mechanics. To establish a bit of raw key, the sender (Alice) encodes a random bit into one of two conjugate bases (X or Z), chosen at random, and transmits it to a receiver (Bob). Bob measures in either the X or Z basis, also chosen at random. After generating a large number (say, $2n$) of bits in this fashion, Alice and Bob can sift out the bits for which they both chose the same basis by public discussion, leaving them with roughly n bits.

Alice then randomly permutes her remaining bits and announces the permutation to Bob, after

which they perform parameter estimation by comparing a small fraction of their remaining bits to determine the error rate of the sifted key. If the fraction p of bits on which Alice and Bob's strings disagree is sufficiently small, they proceed with information reconciliation and privacy amplification to finally arrive at a secret key. Otherwise they abort the protocol. The essence of the protocol is that if an eavesdropper Eve, who is assumed to have control of the quantum channel, examines the signals in order to determine the key, she will necessarily cause some disturbance that manifests itself as errors in the sifted key. Thus p also characterizes how much information an eavesdropper could have gained about the key.

An important property of any QKD protocol is the amount of noise that can be tolerated without compromising the security of the resulting key. The entanglement-based security proof of Shor and Preskill [SP00] showed that BB84 can be used to generate a secure key for detected error rates as high as $p \approx 0.11$, basically by showing there exist Calderbank-Shor-Steane (CSS) [Ste96, CS96] codes correcting noise up to this level. Remarkably, it was recently found [KGR05, RGK05] that this can be improved to $p \approx 0.124$ if Alice performs a preprocessing step in which she adds noise to her sifted key before performing the distillation steps. In the following, we push this threshold to $p \approx 0.129$ by finding improved error correcting codes for the information reconciliation phase.

Taken together our information reconciliation and privacy amplification steps can be described by a highly degenerate CSS code. A quantum code is called degenerate if its syndrome does not uniquely identify the errors that it corrects. This is a uniquely quantum effect – there is no such thing as a degenerate classical code – and there are many unanswered questions about such codes. However, it is known that in many cases degenerate codes are strictly necessary to achieve the capacity of very noisy channels [SSa, DSS98, SSb]. Degenerate codes have been used in security proofs before; in particular, the noise threshold of the six-state protocol was improved from 12.6 to 12.7 percent [Lo01]. Our result combines the use of degenerate codes with the noisy preprocessing of [KGR05, RGK05], leading to an improvement over [KGR05, RGK05] that is analogous to the improvement of degenerate codes over random stabilizer codes found in [SSa, DSS98, SSb] for quantum communication over noisy channels.

6.2 Analytic Key Rate Expression

To determine the secret key generation rate of the modified protocol, we follow the proof method outlined in [KGR05, RGK05, Ren05]. First, the prepare and measure protocol can be converted to an equivalent scheme in which Alice prepares the maximally entangled state $|\Phi^+\rangle_{AB}^{\otimes mn}$ and sends the latter half to Bob. Each party then randomly and independently measures either X or Z on each signal, saving the outcomes for use in parameter estimation and key generation. Denoting these outcomes K_A and K_B, respectively, it then follows from [Ren05] that for any m-bit preprocessing step

$K_A^m \to U^m$ and $U^m \to V^m$ it is possible to use standard error correction and privacy amplification methods to distill secret key from the sifted key at rate

$$r = \frac{1}{m} \inf_{\sigma_{AB} \in \Gamma_p} \left[S(U^m|V^m E^m) - S(U^m|V^m K_B^m) \right], \tag{6.1}$$

where Γ_p is the set of Bell-diagonal states σ_{AB} passing the parameter estimation phase of the protocol and E^m is Eve's system. The rate expression in [KGR05, RGK05] is similar, except that by using the de Finetti Theorem as in [Ren05], we avoid any difficulties arising from the use of blockwise processing. Since the X and Z bases are randomly used to create the sifted key, the error estimation also provides an estimate of the bit- and phase-flip noise rates in the physical channel. Hence, the allowable σ_{AB} are of the form

$$\sigma_{AB} = (1 + t - 2p)|\Phi^+\rangle\langle\Phi^+| + (p - t)(|\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+|) + t|\Psi^-\rangle\langle\Psi^-| \tag{6.2}$$

for some $t \in [0, p]$.

In the following, we will choose a particular $K_A^m \to U^m \to V^m$ for which the rate of Eq. (6.1) outperforms all previously known protocols for large p. The measurements leading to K_A and K_B will be the same as for the usual BB84 protocol, with the preprocessing step chosen as follows. For each m bit block of K_A, $(x_1, x_2, \ldots, x_m)$, Alice independently flips each bit with probability q, resulting in $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_m)$. She then computes

$$U^m = (\tilde{x}_1, \tilde{x}_1 \oplus \tilde{x}_2, \ldots, \tilde{x}_1 \oplus \tilde{x}_m) \tag{6.3}$$

and sends

$$V^m = (\tilde{x}_1 \oplus \tilde{x}_2, \ldots, \tilde{x}_1 \oplus \tilde{x}_m) \tag{6.4}$$

to Bob, after which they proceed with error correction and privacy amplification as usual. The key rate they achieve is given by the following theorem.

Theorem 14. *The key rate achieved using the preprocessing $X^m \to U^m \to V^m$ with $U^m = (\tilde{x}_1, \tilde{x}_1 \oplus \tilde{x}_2, \ldots, \tilde{x}_1 \oplus \tilde{x}_m)$, $V^m = (\tilde{x}_1 \oplus \tilde{x}_2, \ldots, \tilde{x}_1 \oplus \tilde{x}_m)$, where $\tilde{\mathbf{x}} = \mathbf{x} \oplus \mathbf{f}$ and $\mathbf{f}$ is a string of independent 0–1 random variables, each with probability q of being 1, is given by*

$$r = \frac{1}{m}\left(1 - \sum_{\mathbf{s}} P_m^{\tilde{p}}(\mathbf{s}) H(P_m^{\tilde{p}}(u|\mathbf{s})) + mS(\rho_{p,q}) - S\left(\frac{1}{2}\rho_{p,q}^{\otimes m} + \frac{1}{2}Z^{\otimes m}\rho_{p,q}^{\otimes m}Z^{\otimes m}\right)\right). \tag{6.5}$$

Here $\rho_{p,q} = (1-q)|\varphi_+\rangle\langle\varphi_+| + q|\varphi_-\rangle\langle\varphi_-|$ with $|\varphi_\pm\rangle = \sqrt{1-p}|0\rangle \pm \sqrt{p}|1\rangle$, $\tilde{p} = p(1-q) + q(1-p)$, while $P_m^{\tilde{p}}(u, \mathbf{s})$ is defined in Lemma 15.

The proof will proceed by noting that in the entanglement picture, our preprocessing step is equivalent to Alice first adding independent amplitude errors to her halves of the noisy EPR pairs, measuring the stabilizers of an m qubit repetition code, and then sending her syndrome outcomes to Bob. Hence we can then apply the following lemma, which follows immediately from Chapter 4.

Lemma 15. *The m qubit repetition code with stabilizers $Z_1Z_2, \ldots, Z_1Z_m$ maps the error $X^{\mathbf{u}}Z^{\mathbf{v}}$ to the logical error $X^{u_1}Z^{\oplus_{l=1}^m v_l}$ and syndrome $\mathbf{s} = (u_1 \oplus u_2, \ldots u_1 \oplus u_m)$. When used to correct independent amplitude errors of probability p, the probability of a logical amplitude error u and syndrome $\mathbf{s}$ is given by*

$$P_m^p(u, \mathbf{s}) = \left(p^{m-s}(1-p)^s\right)^u \left(p^s(1-p)^{m-s}\right)^{1-u}, \tag{6.6}$$

for $s = |\mathbf{s}|$.

Proof. (of Theorem 14) To evaluate Eq. (6.1), first let

$$\sigma_{AB}^{\otimes m} = \sum_{\mathbf{u},\mathbf{v}} p_{\mathbf{uv}} X_B^{\mathbf{u}} Z_B^{\mathbf{v}} \left[|\Phi^+\rangle\langle\Phi^+|\right]_{AB}^{\otimes m} Z_B^{\mathbf{v}} X_B^{\mathbf{u}}, \tag{6.7}$$

with $p_{\mathbf{u},\mathbf{v}}$ such that

$$p_{\mathbf{u}} = \sum_{\mathbf{v}} p_{\mathbf{u},\mathbf{v}} = p^{|\mathbf{u}|}(1-p)^{m-|\mathbf{u}|}, \tag{6.8}$$

for measured bit error rate p, and similarly for $p_{\mathbf{v}}$.

Alice adds independent noise at error rate q to the A register, so the state of the Alice-Bob-Eve system can be described as

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{f}} \sqrt{p_{\mathbf{uv}} q_{\mathbf{f}}}\, |\mathbf{f}\rangle_{A'} X_B^{\mathbf{u}} Z_B^{\mathbf{v}} X_B^{\mathbf{f}} |\Phi^+\rangle_{AB}^{\otimes m} |\mathbf{u}\rangle_{E_1} |\mathbf{v}\rangle_{E_2}. \tag{6.9}$$

Note that Eve's system is determined by the fact that, in the worst case, she holds the purification of the state after it emerges from the channel. However, she does not hold the purification of the noise Alice adds.

Alice and Bob then measure the stabilizers of the m-qubit repetition code ($Z_1Z_2, \ldots Z_1Z_m$) and Alice sends her measurement outcomes to Bob. This is equivalent to Alice first measuring and sending the syndromes and then Bob performing coherent operations between the message and his syndrome registers such that the two syndromes agree. Renaming Bob's $m-1$ syndrome qubits system B', the state they'll share is thus

$$\sum_{\mathbf{u},\mathbf{v},\mathbf{f}} \sqrt{p_{\mathbf{uv}} q_{\mathbf{f}}}\, |\mathbf{f}\rangle_{A'} X_B^{u_1 \oplus f_1} Z_B^{\oplus_{l=1}^m v_l} |\Phi^+\rangle_{AB} |\left((u_1 \oplus f_1)\mathbf{1}\right) \oplus \mathbf{u}' \oplus \mathbf{f}'\rangle_{B'} |\mathbf{u}\rangle_{E_1} Z_{E_2}^{\mathbf{f}} |\mathbf{v}\rangle_{E_2}, \tag{6.10}$$

where $\mathbf{u}' = (u_2, u_3, \ldots, u_m)$, $\mathbf{f}' = (f_2, f_3, \ldots, f_m)$, $\mathbf{1}$ is the length-$(m-1)$ vector of ones and the $Z^{\mathbf{f}}$ acting on Eve's second system comes from the commutation of $Z_B^{\mathbf{v}}$ and $X_B^{\mathbf{f}}$.

Getting rid of the A' system (but keeping it from Eve), we now let Alice and Bob measure systems A and BB' in the computational basis, respectively. According to Eq. (6.1), the difference of conditional entropies for the resulting state will give us the key rate. This will be simpler to analyze by first rewriting the lower bound as

$$r \geq \frac{1}{m} \inf_{\sigma_{AB} \in \Gamma_p} I(A;BB') - I(A;E). \tag{6.11}$$

The first term, $I(A;BB')$, is the mutual information of the state

$$\rho_{ABB'} = \frac{1}{2} \sum_{x=0}^{1} |x\rangle\langle x|_A \otimes \rho^x_{B'B}, \tag{6.12}$$

where

$$\begin{aligned} \rho^x_{B'B} &= \sum_{\mathbf{f}} \sum_{\mathbf{u}} q_{\mathbf{f}} p_{\mathbf{u}} \, |x{+}f_1{+}u_1\rangle\langle x{+}f_1{+}u_1|_B \otimes |\, ((u_1 \oplus f_1)\mathbf{1}) \oplus \mathbf{u}' \oplus \mathbf{f}'\rangle\langle ((u_1 \oplus f_1)\mathbf{1}) \oplus \mathbf{u}' \oplus \mathbf{f}'|_{B'} \\ &= \sum_{\mathbf{s}} P_m^{\tilde{p}}(\mathbf{s}) \sum_{u=0}^{1} P_m^{\tilde{p}}(u|\mathbf{s}) |x{+}u\rangle\langle x{+}u|_B \otimes |\mathbf{s}\rangle\langle \mathbf{s}|_{B'}, \end{aligned} \tag{6.13}$$

and the $P_m^{\tilde{p}}(u, \mathbf{s})$ are given by Lemma 15. From this, we see that the mutual information, $I(A;BB')$, is exactly

$$1 - \sum_{\mathbf{s}} P_m^{\tilde{p}}(\mathbf{s}) H(P_m^{\tilde{p}}(u|\mathbf{s})). \tag{6.14}$$

Notice that this term only depends on $p_{\mathbf{u}}$, which is determined by the parameter estimation phase, so it will be the same for all $\sigma_{AB} \in \Gamma_p$.

Turning to the second term in Eq. (6.11), we want to find the mutual information of the state obtained by tracing out Bob's systems,

$$\rho_{AE_1E_2} = \frac{1}{2} \sum_{x=0}^{1} |x\rangle\langle x|_A \otimes \rho^x_{E_1E_2}, \tag{6.15}$$

where

$$\rho^x_{E_1E_2} = \left(Z_{E_2}^{\otimes m}\right)^x \left(\sum_{\mathbf{u},\mathbf{v}_1,\mathbf{v}_2,\mathbf{f}} q_{\mathbf{f}} \sqrt{p_{\mathbf{u}|\mathbf{v}_1} p_{\mathbf{u}|\mathbf{v}_2}} \, |\mathbf{u}\rangle\langle \mathbf{u}|_{E_1} \otimes \sqrt{p_{\mathbf{v}_1} p_{\mathbf{v}_2}} \, Z^{\mathbf{f}} |\mathbf{v}_1\rangle\langle \mathbf{v}_2| Z^{\mathbf{f}} \right) \left(Z_{E_2}^{\otimes m}\right)^x \tag{6.16}$$

Note that the $(Z_{E_2}^{\otimes m})^x$ comes from the action of $Z^{\oplus_{l=1}^m v_l}$ on system B. When Eve's amplitude and

phase errors are independent, this expression can be further simplified. Defining

$$\mu = \sum_{\mathbf{u}} p_{\mathbf{u}} |\mathbf{u}\rangle\langle\mathbf{u}| \tag{6.17}$$

and

$$\rho_{p,q} = (1-q)|\varphi_+\rangle\langle\varphi_+| + q|\varphi_-\rangle\langle\varphi_-| \tag{6.18}$$

with

$$|\varphi_\pm\rangle = \sqrt{1-p}|0\rangle \pm \sqrt{p}|1\rangle, \tag{6.19}$$

we can write

$$\rho^x_{E_1,E_2} = \mu_{E_1} \otimes \left(Z^{\otimes m}_{E_2}\right)^x \left[\rho^{\otimes m}_{p,q}\right]_{E_2} \left(Z^{\otimes m}_{E_2}\right)^x. \tag{6.20}$$

Actually, while we have to maximize $I(A;E_1E_2)$ over all possible $p_{\mathbf{uv}}$ corresponding to states in $\sigma_{AB} \in \Gamma_p$, we can see from the above that the largest value is attained for independent phase and amplitude errors. In particular, if Eve starts with the independent $\mathbf{u},\mathbf{v}$ state, by tracing out the E_1 system and using the isometry

$$U = \sum_{\mathbf{v},\mathbf{u}} \sqrt{p_{\mathbf{u}|\mathbf{v}}} |\mathbf{u}\rangle_{E_3} |\mathbf{v}\rangle_{E_2} \langle\mathbf{v}|_{E_2}, \tag{6.21}$$

then completely dephasing the E_3 system, she can construct a $\rho_{AE_2E_3}$ with the same mutual information as if the errors were distributed according to $p_{\mathbf{u}|\mathbf{v}}p_{\mathbf{v}}$. Since mutual information cannot be increased by local operations, the independent noise state must have the largest value. Moreover, as the E_1 system is uncorrelated with the rest, the mutual information between Alice and Eve can be easily computed, yielding

$$I(A;E) = S\left(\frac{1}{2}\rho^{\otimes m}_{p,q} + \frac{1}{2}Z^{\otimes m}\rho^{\otimes m}_{p,q}Z^{\otimes m}\right) - mS(\rho_{p,q}). \tag{6.22}$$

Taking the difference between $I(A;BB')$ and $I(A;E)$, keeping in mind we must send m qubits for each m-block, leads to the overall key rate of Eq. (6.5). □

6.3 Numerical Evaluation of Key Rates

We would now like to evaluate the key rate in Eq. (6.5) for particular values of p, q, and m. The expression $S(\rho_{p,q})$ can be easily calculated and the second term can be evaluated efficiently using Lemma 15. The most difficult term is

$$S\left(\frac{1}{2}\rho^{\otimes m}_{p,q} + \frac{1}{2}Z^{\otimes m}\rho^{\otimes m}_{p,q}Z^{\otimes m}\right), \tag{6.23}$$

but it can be handled as follows. Due to the permutation-invariance of the state $\rho_{p,q}^{\otimes m}$, it can be compactly expressed as a direct sum of states on the $SU(2)$ irreducible representations (irreps). Each irrep occurs with some degeneracy, giving rise to a permutation factor, which by Schur's lemma is maximally mixed. Using the expression for multiple copies of a general qubit mixed state from [BMG$^+$], which describes the irreducible states of $\rho_{p,q}^{\otimes m}$ as a function of its Bloch vector, and doing the same for $Z^{\otimes m}\rho_{p,q}^{\otimes m}Z^{\otimes m}$, we can compute $S\left(\frac{1}{2}\rho_{p,q}^{\otimes m} + \frac{1}{2}Z^{\otimes m}\rho_{p,q}^{\otimes m}Z^{\otimes m}\right)$ for values of m up to several hundred.

In general, larger m allow us to get higher thresholds with the optimal value of $q \approx 0.3$ increasing slowly with m (e.g., Figure 6.1). Choosing $m = 400$ and $q = 0.32$ allows a nonzero key generation rate up to $p = .1292$, but for larger m the computation becomes quite slow.

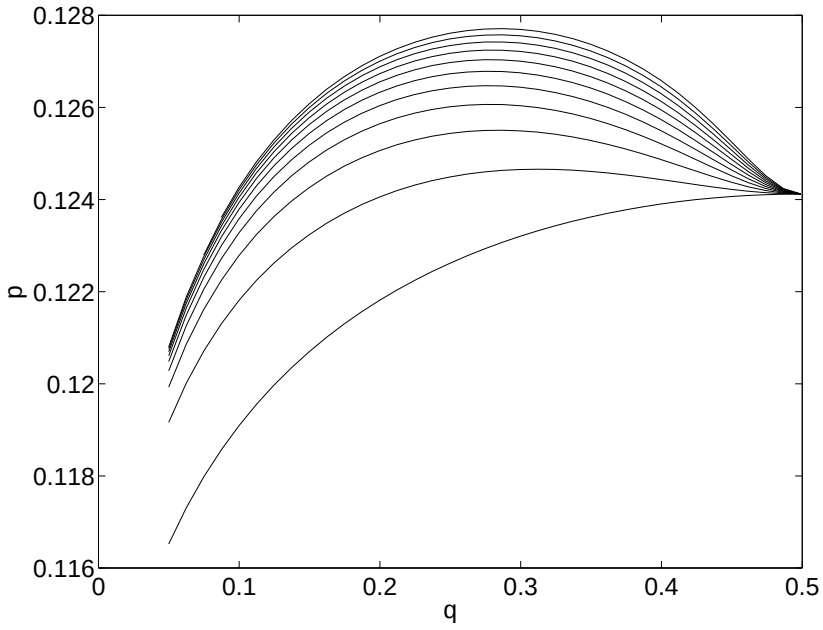


Figure 6.1: Bit error rate p at which the key rate goes to zero as a function of preprocessing noise q when using various-sized repetition codes in the BB84 protocol. The curves are, from bottom to top, $m = 1$,$m = 10, 20, \ldots 100$, illustrating the fact that a longer repetition code allows a higher threshold. As m is increased, the optimal q also grows. Taking $m = 400$ and $q = 0.32$ gives our best threshold of 0.1292.

6.4 Discussion

Given the pattern of improving thresholds with larger repetition lengths, it is tempting to suggest that the best threshold within the family of codes we have described will be achieved when $m \to \infty$ while $q \to 0.5$. While we have not yet been able to perform such an analysis, it seems likely that an

asymptotic analysis of our key rates in the limit of large m would be tractable.

It is important to mention that the codes we have considered are highly restricted, and it is not at all clear that the key rates they allow should be optimal. One promising idea for higher key rates would be to adapt the concatenation of repetition codes in conjugate bases used in [DSS98] and Chapter 4 to the problem of secret key generation, with the repetition code in the X basis providing improved privacy amplification. A more ambitious approach would be to develop new degenerate codes for this problem, perhaps designed using the heuristic suggested in Chapter 4.

The one-way protocols we have presented bear a striking resemblance to two-way protocols using advantage distillation [GL03]. In particular, an advantage distillation protocol can be described as using a repetition code, with Bob sending the syndromes of the repetition code back to Alice. Error correction and privacy amplification are then performed on blocks for which no error is detected, and the blocks for which an error is detected are thrown away. Notice that without the back communication from Bob, Alice would not know the syndromes, and thus be unable to discard the blocks in which Bob had detected an error. Our findings show that even in this case, when Alice is ignorant of the syndromes, and thus unable to discard bad blocks, there is still a benefit in using a repetition code. In a sense, the repetition code works better than expected, because it collapses many phase errors to a single logical phase error, while still providing information about bit errors. This benefit should also appear when the repetition code is used for advantage distillation in a two-way protocol with noisy preprocessing.

As was pointed out in Chapter 5, one-way protocols with noisy preprocessing are closely related to distillation protocols for the class of twisted states[HHHO05]. In that work it was shown that noisy preprocessing can be interpreted as the deflection of Eve's correlations away from the sifted key into a "shield" system, which purifies the noise added by Alice. This is analagous to the ancient martial art, *jujitsu*, wherein one uses leverage to deflect an attacker's force away from oneself rather than opposing it directly. Viewed in this way, the benefit of a repetition code is that it allows us to combine the "soft" approach of deflecting phase errors and the "hard" approach of correcting amplitude errors—while learning about bit errors that we must correct, we are simultaneously decreasing Eve's correlation with the key, reducing the need for privacy amplification later.

Chapter 7

Clone Assisted Capacity

We present an upper bound for the quantum channel capacity that is both additive and convex. Our bound can be interpreted as the capacity of a channel for high fidelity communication when assisted by a family of zero capacity channels made up of all channels that map symmetrically to their output and environment. The bound seems to be quite tight, and for degradable quantum channels it coincides with the unassisted channel capacity. Furthermore, we will use the clone assisted capacity to find new upper bounds on the capacity of the depolarizing channel.

7.1 Introduction

The archetypical problem in information theory is finding the capacity of a noisy channel to transmit high fidelity messages. Already in [Sha48], Shannon provided a simple formula for the capacity in the case of a discrete memoryless channel. Results for more general channels have also been found (e.g., [Ver98]).

The status of the quantum channel capacity question is not nearly so nice. While there has recently been significant progress in finding expressions for the various capacities of a quantum channel [Dev05, Llo97, Sho, BSSA02], with the exeption of the entanglement assisted capacity formula of [BSSA02] and results for some very special channels (e.g., amplitude damping, dephasing, and erasure channels) these studies have arrived capacity expressions that cannot be evaluated in any tractable way. For instance, it was shown in [Dev05, Llo97, Sho] that the capacity of a quantum channel $\mathcal{N}$ is given by

$$Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \sup_{\phi_{B_n}} I^{\mathrm{coh}}(\mathcal{N}^{\otimes n}, \phi_{B_n}), \tag{7.1}$$

where $I^{\mathrm{coh}}(\mathcal{N}, \phi_B) = S(\mathcal{N}(\phi_B)) - S(I \otimes \mathcal{N}(|\phi\rangle\langle\phi|)))$ is known as the *coherent information*. In order to evaluate this regularized formula one must perform an optimization over an infinite number of variables, a notoriously difficult computational problem. Furthermore, it is known that the limit on the right is in general strictly larger than the corresponding single letter expression [DSS98, SSa,

SSb](i.e., $Q^{(1)} \sup_{\phi_B} I^{\mathrm{coh}}(\mathcal{N}, \phi_B)) < Q(\mathcal{N})$).

In the absence of an explicit formula for the quantum capacity, it is desirable to find upper and lower bounds for Eq. (7.1). Unfortunately, most known bounds seem to be as difficult to evaluate in general as Eq.(7.1). Examples of upper bounds that *can* be easily evaluated, at least in some special cases, are given by the no-cloning-based arguments of [BDE$^+$98, Cer00], the semi-definite programming bounds of Rains[Rai99] and the closely related relative entropy of entanglement ([VP98], others). None of these is expected to be particularly tight – the last two are also upper bounds for the capacity assisted by two-way classical communication (which can be much larger than one way), whereas the first is based solely on reasoning about where the channel's capacity must be zero. As such, it would be useful to find new upper bounds for the quantum capacity that are both free of regularization and fundamentally one way. In the following we present just such a bound.

Inspired by the fact that allowing free forward classical communication does not increase the quantum channel capacity [BDSW96, BST98], we will consider the capacity of a quantum channel assisted by the use of a quantum channel that maps symmetrically to the reciever (Bob) and the environment (Eve). Such assistance maps, which we call *cloning channels*, can be used for forward classical communication but are apparently somewhat stronger. They can, however, immediately be seen to have zero quantum capacity, so that while the assisted capacity we find may in general be larger than the usual quantum capacity, one expects that it will provide a fairly tight upper bound. In particular, the *clone assisted capacity* we find will not be an upper bound for the capacity assisted by two-way classical communication.

The expression we find for the assisted capacity, which we'll call Q_{ca}, has several nice properties and turns out to be much easier to deal with than Eq. (7.1). Most importantly, our expression is free of the regularization present in so many capacity formulas. We will also see that Q_{ca} is convex, additive, and that it is equal to Q for the family of degradable channels [DS05]. We will be able to use these properties to find upper bounds on Q_{ca} of the depolarizing channel, which, in turn, will give a significant improvement over other known bounds for its usual capacity.

It should be emphasized that we have not found an upper bound on the dimension of the cloning channel needed to attain the assisted capacity, which in general prevents us from evaluating Q_{ca} explicitly. While we cannot rule out such a bound, the arguments we use to establish several of Q_{ca}'s nice properties rely explicitly on the availablility of an unbounded dimension. This suggests dealing with an assistance channel of unbounded dimension may be the price we pay to get such desirable properties as additivity and convexity, which is reminiscent of the findings of [BHLS03, CW04].

7.2 Properties of Q_{ca}

Before studying the ca-capacity, we must first make explicit some definitions. Letting

$$S_d \subset U \otimes V \tag{7.2}$$

be the $(d(d+1))/2$-dimensional symmetric subspace between d-dimensional spaces U and V, we call the inclusion map

$$\mathcal{A}_d : \mathbb{C}^{(d^2+d)/2} \hookrightarrow S_d \tag{7.3}$$

the d-dimensional cloning channel.

We say that a rate R is ca-achievable if for all $\epsilon > 0$ there is a $n_0(\epsilon)$ such that for any $n > n_(\epsilon)$ there is a dimension d_n, code $C_n \subset \tilde{A}^{\otimes n} \otimes S_{d_n}$ with $\dim C_n > Rn$, and a decoding operation $\mathcal{D}_n$ such that for all states $|\psi\rangle \in C_n$, the reconstructed state $\mathcal{D}_n \circ \mathcal{N}^{\otimes n} \otimes \mathcal{A}_{d_n}(|\psi\rangle\langle\psi|)$ has a fidelity with the original state $|\psi\rangle$ of at least $1-\epsilon$. The ca-capacity, which we will denote by $Q_{ca}(\mathcal{N})$, is defined as the supremum of all ca-achievable rates.

We are now in a position to introduce a quantity that will play a central role in our study of the clone assisted capacity. Letting $\mathcal{N} : \tilde{A} \to B$ be a completely positive map, we define $Q_{ca}^{(1)}(\mathcal{N})$ to be the supremum over all states $|\phi_{A\tilde{A}UV}\rangle$ that are invariant under the permutation of U and V of the coherent information of A given BV, evaluated after the $\tilde{A}$ register of ϕ is acted on by $\mathcal{N}$. That is, we let

$$Q_{ca}^{(1)}(\mathcal{N}) = \sup_{|\phi_{ABUV}\rangle, U \leftrightarrow V} I^c(A\rangle BV)_{\mathcal{N}_B(\phi)}. \tag{7.4}$$

It will turn out that $Q_{ca}^{(1)}(\mathcal{N})$ is exactly the clone assisted capacity of $\mathcal{N}$, which we show using the following two lemmas.

Lemma 16. *$Q_{ca}^{(1)}$ is additive. That is, $Q_{ca}^{(1)}(\mathcal{N}_1 \otimes \mathcal{N}_2) = Q_{ca}^{(1)}(\mathcal{N}_1) + Q_{ca}^{(1)}(\mathcal{N}_2)$.*

Proof. First notice that

$$\begin{aligned} I^c(A\rangle B_1B_2V) &= \frac{1}{2}\left(S(A|E_1E_2U) - S(A|B_1B_2V)\right) && (7.5)\\ &= \frac{1}{2}\left(S(A|E_1E_2V) - S(A|B_1E_2V) + S(A|B_1E_2V) - S(A|B_1B_2V)\right) && (7.6)\\ &\leq Q_{ca}^{(1)}(\mathcal{N}_1) + Q_{ca}^{(1)}(\mathcal{N}_2), && (7.7) \end{aligned}$$

where the final inequality follows from the fact that for any $|\phi_{A\tilde{A}_1\tilde{A}_2UV}\rangle$ that is U-V permutation invariant, we can define

$$|\phi_{A\tilde{A}_1\tilde{U}\tilde{V}}\rangle = \frac{1}{\sqrt{2}}U_{\mathcal{N}}^{\tilde{A}_2}|\phi_{A\tilde{A}_1\tilde{A}_2UV}\rangle|0\rangle_{C_1}|1\rangle_{C_2} + \frac{1}{\sqrt{2}}\Pi_{\tilde{A}_2E_2}U_{\mathcal{N}}^{\tilde{A}_2}|\phi_{A\tilde{A}_1\tilde{A}_2UV}\rangle|1\rangle_{C_1}|0\rangle_{C_2} \tag{7.8}$$

which is symmetric in $\tilde{U}$-$\tilde{V}$, where $\Pi_{\tilde{A}_2 E_2}$ permutes $\tilde{A}_2$ and E_2, we let $\tilde{U} = U\tilde{A}_2 C_2$, $\tilde{V} = V\tilde{E}_2 C_1$ that has the property that

$$\frac{1}{2}\left(S(A|E_1E_2V) - S(A|B_1E_2V)\right) = \frac{1}{2}\left(S(A|E_1\tilde{V}) - S(A|B_1\tilde{V})\right), \tag{7.9}$$

and similarly for the last two terms. This shows

$$Q^{(1)}_{ca}(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq Q^{(1)}_{ca}(\mathcal{N}_1) + Q^{(1)}_{ca}(\mathcal{N}_2). \tag{7.10}$$

Furthermore, by restricting the optimization in Eq. (7.4) to states of the form $|\phi_{A_1\tilde{A}_1U_1V_1}\rangle|\phi_{A_2\tilde{A}_2U_2V_2}\rangle$ we see that

$$Q^{(1)}_{ca}(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq Q^{(1)}_{ca}(\mathcal{N}_1) + Q^{(1)}_{ca}(\mathcal{N}_2). \tag{7.11}$$

□

The other ingredient we need is the following expression for the clone assisted capacity, which follows by standard arguments (e.g., [Dev05]).

Lemma 17.

$$Q_{ca}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} Q^{(1)}_{ca}(\mathcal{N}^{\otimes n}). \tag{7.12}$$

Proof. To see that the ca-capacity is no less than the right hand side, note that for any $|\phi_{AB^nUV}\rangle$ that is symmetric under the interchange of U and V, the rate

$$\frac{1}{n} I^c(A\rangle B^n V)_{\mathcal{N}^{\otimes n}(\phi)} \tag{7.13}$$

is achievable by the quantum noisy channel coding theorem applied to the channel $\mathcal{N}^{\otimes n} \otimes \mathcal{A}_{d_U}$.
To prove the converse, fix ϵ, let $C \subset \tilde{A}^n S$ be a (n, ϵ) code of rate R making use of a symmetric broadcast channel with output dimension d_V^2 and let $|\phi^C\rangle$ be a state that is maximally entangled with C. Then

$$\begin{aligned} I^c(A\rangle B^n V)_{\mathcal{N}^{\otimes n}(\phi^C)} &\geq I^c(A\rangle B^n V)_{\mathcal{D}_{BV}\circ\mathcal{N}^{\otimes n}(\phi^C)} && (7.14)\\ &\geq Rn - \frac{2}{e} - 8\log(d_C)\sqrt{\epsilon} && (7.15)\\ &= Rn - \frac{2}{e} - 8Rn\sqrt{\epsilon}, && (7.16)\end{aligned}$$

so that

$$\frac{1}{n} Q^{(1)}_{ca}(\mathcal{N}^{\otimes n}) + \frac{2}{ne} \geq R(1 - 8\sqrt{\epsilon}). \tag{7.17}$$

□

Lemmas 16 and 17 immediately imply the following expression for $Q_{ca}(\mathcal{N})$:

Theorem 18. $Q_{ca}(\mathcal{N}) = Q_{ca}^{(1)}(\mathcal{N})$.

From this we can easily show

Proposition 19. *Q_{ca} is convex.*

Proof. Because $I^c(A\rangle BV)_{\rho_{ABV}}$ is convex in ρ_{ABV}, we have

$$I^c(A\rangle BV)_{(p\mathcal{N}_1+(1-p)\mathcal{N}_2)(\phi)} \leq pI^c(A\rangle BV)_{\mathcal{N}_1(\phi)} + (1-p)I^c(A\rangle BV)_{\mathcal{N}_2(\phi)}, \tag{7.18}$$

so that

$$\sup_\phi I^c(A\rangle BV)_{(p\mathcal{N}_1+(1-p)\mathcal{N}_2)(\phi)} \quad \leq \quad p\sup_\phi I^c(A\rangle BV)_{\mathcal{N}_1(\phi)} + (1-p)\sup_\phi I^c(A\rangle BV)_{\mathcal{N}_2(\phi)}, \tag{7.19}$$

which tells us exactly that

$$Q_{ca}(p\mathcal{N}_1 + (1-p)\mathcal{N}_2) \leq pQ_{ca}(\mathcal{N}_1) + (1-p)Q_{ca}(\mathcal{N}_2). \tag{7.20}$$

□

7.3 Implications for Unassisted Quantum Capacities

In this section we will explore some of the limitations the clone assisted capacity, $Q_{ca}(\mathcal{N})$, places on the standard capacity of a quantum channel, $Q(\mathcal{N})$. As noted in the introduction, by simply not using the cloning channel provided, it is possible to communicate over a channel at the unassisted rate. In other words,

$$Q(\mathcal{N}) \leq Q_{ca}(\mathcal{N}). \tag{7.21}$$

Furthermore, as we will now see, this upper bound is actually an equality for the class of channels known as *degradable*[DS05]. Every channel, $\mathcal{N}$, can be expressed as an isometry followed by a partial trace, which is to say there is always an isometry

$$U_{\mathcal{N}} : A \to BE \tag{7.22}$$

such that

$$\mathcal{N}(\rho) = \mathrm{Tr}_E\, U_{\mathcal{N}}\rho U_{\mathcal{N}}^\dagger. \tag{7.23}$$

The complementary channel of $\mathcal{N}$, which we call $\mathcal{N}^C$, is the channel that results by tracing out system B rather than the environment:

$$\mathcal{N}^C(\rho) = \mathrm{Tr}_B\, U_{\mathcal{N}} \rho U_{\mathcal{N}}^\dagger. \tag{7.24}$$

A channel is degradable if there exists a completely positive map, $\mathcal{D} : B \to E$, that "degrades" the channel $\mathcal{N}$ to $\mathcal{N}^C$. In other words, $\mathcal{D} \circ \mathcal{N} = \mathcal{N}^C$. The capacity of a degradable channel is given by the single letter maximization of the coherent information, as shown in [DS05]. Furthermore, we will now show that the ca-capacity of a degradable channel is given by the same formula. That is, the assistance channels we have been considering are of no use at all for a degradable channel.

Theorem 20. *If $\mathcal{N}$ is degradable,*

$$Q_{ca}(\mathcal{N}) = Q(\mathcal{N}). \tag{7.25}$$

Proof. Fix $|\phi_{A\tilde{A}S}\rangle$. Then

$$I^c(A\rangle BV)_{I_A \otimes \mathcal{N} \otimes \mathcal{A}(\phi)} \leq I^c(AUV\rangle B)_{I_A \otimes \mathcal{N} \otimes \mathcal{A}(\phi)} + I^c(ABE\rangle V)_{I_A \otimes \mathcal{N} \otimes \mathcal{A}(\phi)} \tag{7.26}$$

exactly when

$$I(E;U) \leq I(B;V), \tag{7.27}$$

which is true if $\mathcal{N}$ is degradable by the monotonicity of mutual information under local operations. This implies that the maximum value of the left hand side of Eq. (7.26) is no larger than the maximum of the right hand side. The maximum of the first term on the right is exactly

$$\sup_{\phi_{A\tilde{A}}} I^c(A\rangle B)_{I_A \otimes \mathcal{N} \otimes \mathcal{A}(\phi)} = Q(\mathcal{N}), \tag{7.28}$$

whereas the maximum of the second is zero, so that

$$I^c(A\rangle BV)_{I_A \otimes \mathcal{N} \otimes \mathcal{A}(\phi)} \leq Q(\mathcal{N}). \tag{7.29}$$

Furthermore, by choosing $|\phi_{A\tilde{A}S}\rangle = |\phi_{A\tilde{A}}\rangle|\phi_S\rangle$ the left hand side can achieve $Q(\mathcal{N})$. □

Theorem 20 allows us to calculate the ca-capacity of any degradable channel. If a channel $\mathcal{N}$ can be written as a convex combination of degradable channels this theorem, together with the convexity of Q_{ca}, provides an upper bound for $Q_{ca}(\mathcal{N})$ and therefore also $Q(\mathcal{N})$. For instance, the

depolarizing channel can be written as a convex combination of dephasing-type channels,

$$\begin{aligned} \mathcal{N}_p^{\text{dep}}(\rho) &= (1-p)\rho + \frac{p}{3}X\rho X + \frac{p}{3}Y\rho Y + \frac{p}{3}Z\rho Z & (7.30)\\ &= \frac{1}{3}\mathcal{N}_p^X(\rho) + \frac{1}{3}\mathcal{N}_p^Y(\rho) + \frac{1}{3}\mathcal{N}_p^Z(\rho), & (7.31)\end{aligned}$$

where $\mathcal{N}_p^X(\rho) = (1-p)\rho + pX\rho X$ and similarly for $\mathcal{N}_p^Y$ and $\mathcal{N}_p^Z$. From this we can conclude that

$$\begin{aligned} Q_{ca}(\mathcal{N}_p^{\text{dep}}) &\leq \frac{1}{3}Q_{ca}(\mathcal{N}_p^X) + \frac{1}{3}Q_{ca}(\mathcal{N}_p^Y) + \frac{1}{3}Q_{ca}(\mathcal{N}_p^Z) & (7.32)\\ &= 1 - H(p), & (7.33)\end{aligned}$$

where we have used the fact that $\mathcal{N}_p^X$ is degradable and has ca-capacity $1 - H(p)$. This reproduces the upper bounds of [VP98, Rai99], which have been the best known for small p.

We can also evaluate $Q_{ca}(\mathcal{N}_p^{\text{dep}})$ for $p = \frac{1}{4}$ as follows. For this value of p, there is a CP map that can be composed with the complementary channel, $\mathcal{N}_p^C$, to generate $\mathcal{N}_p$ [BDE$^+$98]. This immediately implies $Q_{ca}(\mathcal{N}_{1/4}) = 0$, since otherwise Bob and Eve could both reconstruct the encoded state with high fidelity, giving a violation of the no-cloning theorem. More explicitly, for any state $|\phi_{A\tilde{A}UV}\rangle$ we have

$$\begin{aligned} I^c(A\rangle BV)_{I\otimes\mathcal{N}(\phi)} &= -I^c(A\rangle EV) & (7.34)\\ &\leq -I^c(A\rangle BV), & (7.35)\end{aligned}$$

from which we conclude

$$Q_{ca}(\mathcal{N}_{1/4}) = 0, \tag{7.36}$$

and where the second line is due to the quantum data processing inequality[SN96]. This reproduces the bound of [BDE$^+$98], and furthermore because we know that the clone assisted capacity is convex we find that

$$Q(\mathcal{N}_p^{\text{dep}}) \leq Q_{ca}(\mathcal{N}_p^{\text{dep}}) \leq \text{conv}(1 - H(p), 1 - 4p). \tag{7.37}$$

It is important to note that the quantum capacity Q is not known to be convex and, indeed, based on numerical evidence for NPT bound entangled states [SST01] it is not expected to be convex. Thus, while the two bounds above were already known, it was not clear that the convex hull of these was also an upper bound.

We will now provide a tighter bound for $Q_{ca}(\mathcal{N}_p^{\text{dep}})$, by expressing the depolarizing channel as a convex combination of amplitude damping channels, which were shown to be degradable in [GF05].

The amplitude damping channel can be expressed as

$$\mathcal{N}_{\gamma}^{\text{amp}}(\rho) = A_0 \rho A_0 + A_1 \rho A_1, \tag{7.38}$$

where

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \tag{7.39}$$

and

$$A_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \tag{7.40}$$

From this we find that

$$\frac{1}{2}\mathcal{N}_{\gamma}^{\text{amp}}(\rho) + \frac{1}{2} Y \mathcal{N}_{\gamma}^{\text{amp}}(Y\rho Y) Y = \mathcal{N}_{(q,q,p_z)}(\rho), \tag{7.41}$$

where

$$\mathcal{N}_{(q,q,p_z)}(\rho) = (1 - 2q - p_z)\,\rho + qX\rho X + qY\rho Y + p_z Z\rho Z \tag{7.42}$$

with

$$q = \frac{\gamma}{4} \tag{7.43}$$

and

$$p_z = \frac{1}{2}\left(1 - \frac{\gamma}{2} - \sqrt{1-\gamma}\right). \tag{7.44}$$

The depolarizing channel can now be expressed as

$$\mathcal{N}_{2q+p_z}^{\text{dep}}(\rho) = \frac{1}{3}\mathcal{N}_{(q,q,p_z)}(\rho) + \frac{1}{3}\mathcal{N}_{(q,p_z,q)}(\rho) + \frac{1}{3}\mathcal{N}_{(p_z,q,q)}(\rho), \tag{7.45}$$

so that $\mathcal{N}_{p}^{\text{dep}}(\rho)$ is a convex combination of amplitude damping channels with

$$\gamma_p = 2\sqrt{4 - 2p - 3p^2} - 2(2 - p). \tag{7.46}$$

This gives us an upper bound (shown in Figure) of

$$Q(\mathcal{N}_{p}^{\text{dep}}) \leq Q_{ca}(\mathcal{N}_{p}^{\text{dep}}) \leq \text{conv}(Q(\mathcal{N}_{\gamma_p}^{\text{amp}}), 1 - 4p), \tag{7.47}$$

where $Q(\mathcal{N}_{\gamma_p}^{\text{amp}})$ is equal to[GF05]

$$\max_{x\in[0,1]}\left(H_2(\gamma_p x) - H_2((1-\gamma_p)x)\right). \tag{7.48}$$

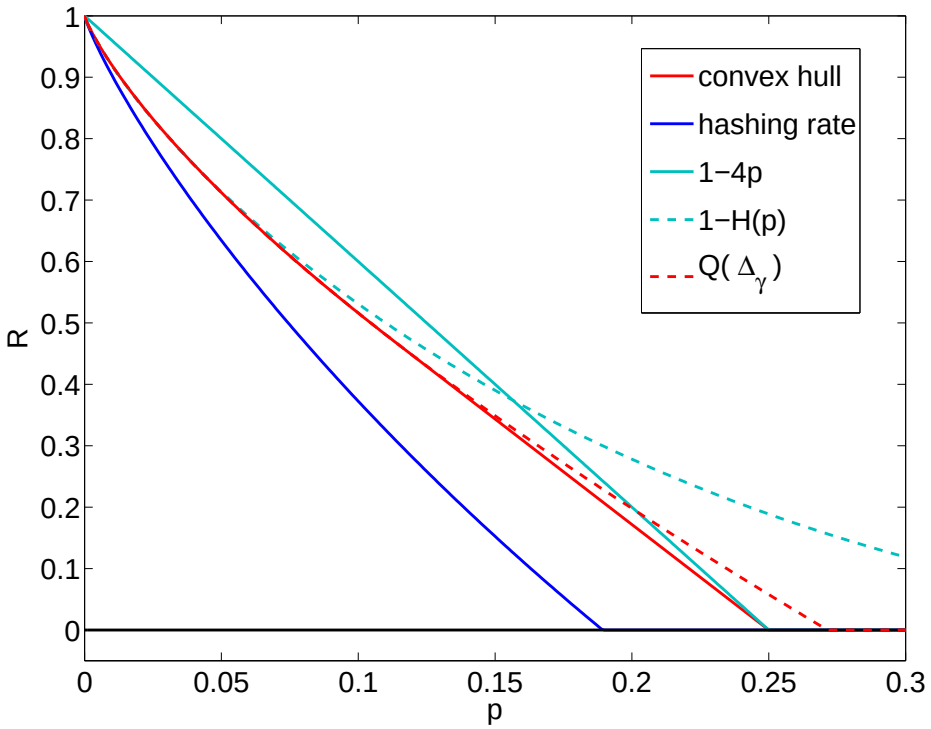


Figure 7.1: Our upper bound evaluated for the depolarizing channel: The straight solid light-grey line comes from no-cloing, the broken light-grey line is the capacity of a dephasing channel, and the broken dark-grey line is the capacity of the amplitude damping channel; finally, the solid dark-grey line is the convex hull of the first three, our best upper bound on $Q_{ca}(\mathcal{N}_p^{\mathrm{dep}})$ and $Q(\mathcal{N}_p^{\mathrm{dep}})$ so far; The solid black line is the hashing (lower) bound, $1 - H(p) - p \log 3$.

7.4 Discussion

We have studied the capacity of a quantum channel given the assistance of an arbitrary cloning channel. The capacity formula we find is in many ways more manageable than the known expression for the (unassisted) quantum capacity, and we are able to establish that the clone assisted capacity is both convex and additive. By taking advantage of the convexity of Q_{ca} and the fact that Q_{ca} and Q coincide for degradable channels, we presented a general method for finding upper bounds to Q and in particular provided a bound for the capacity of the depolarizing channel that is stronger than any previously known result. We have left many questions unanswered.

The most pressing question is whether it is possible to find bounds on the dimension of the cloning channel necessary to achieve the ca-capacity. Finding such a bound would allow us to evaluate $Q_{ca}(\mathcal{N})$ efficiently, which we expect would provide very tight bounds on Q in many cases.

So far, we have not been able to find a channel for which the ca-capacity and capacity differ. We expect that such channels exist, and a better understanding of when the two capacities differ may point towards simplifications of the quantum capacity formula in Eq.(7.1).

It is worth mentioning that we first discovered that an unsymmetrized version of the quantity $Q_{ca}^{(1)}$ is an upper bound for Q while attempting to find the entanglement analogue of the upper bound on distillable key presented in [KGR05].

It was only later that it became clear the formula could be made symmetric and interpreted as the quantum capacity of a channel given the family of assistance channels we have considered. The upper bound of [KGR05] can be understood similarly as the one-way distillable key (starting from a ccq-state) assisted by cq-channels mapping symmetrically from Alice's (classical) data to states of Bob/Eve.

Finally, it should be noted that the approach we have taken here is qualitatively similar to the work of [VP98, Rai99] in the two-way scenario. In that work, it was found that *enlarging* the set of operations allowed for entanglement distillation from LOCC to the easier-to-deal-with set of separable or PPT-preserving operations made it possible to establish tighter bounds on two-way distillable entanglement than was possible by considering LOCC protocols directly. Similarly, we have shown that by augmenting a channel with a zero capacity cloning channel, a dramatically simplified capacity formula can be found that allows us to establish tighter bounds on the unassisted capacity than were possible by direct considerations. To what extent this approach can be used in general, the reason such an approach works at all, and the tightness of the bounds achieved in this way are all questions that we leave wide open.

Bibliography

[AB02] H. Aschauer and H. J. Briegel, *Security proof of quantum cryptography based entirely on entanglement purification*, Phys. Rev. A **66**, 032302 (2002).

[AG] A. Ambainis and D. Gottesman, Two-Way Entanglement Purification for Finite Block Size, arXiv quant-ph/0310097.

[AGHP90] N. Alon, O. Goldreich, J. Hastad and R. Peralta, Simple Constructions of Almost k-Wise Independent Random Variables, in *IEEE Symposium on Foundations of Computer Science*, pages 544–553, 1990.

[BB84] C. Bennett and G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing , 175 (1984).

[BBP+96] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin and W. K. Wootters, *Purification of noisy entanglement and faithful teleportation via noisy channels*, Phys. Rev. Lett. **76** (1996).

[BD] G. Bowen and N. Datta, Beyond i.i.d. in Quantum Information Theory, arXiv quant-ph/0604013.

[BDE+98] D. Bruss, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello and J. A. Smolin, *Optimal Universal and State-Dependent Quantum Cloning*, Phys. Rev. A **57**, 2368 (1998), arXiv:quant-ph/9705038.

[BDM05] G. Bowen, I. Devetak and S. Mancini, *Bounds on classical information capacities for a class of quantum memory channels*, Phys. Rev. A. **71**, 034310 (2005), arXiv quant-ph/0312216.

[BDSW96] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin and W. K. Wootters, *Mixed state entanglement and quantum error correction*, Phys. Rev. A. **54**, 3824–3851 (1996), arXiv quant-ph/9604024.

[BHLS03] C. Bennett, A. Harrow, D. Leung and J. Smolin, *On the capacities of bipartite Hamiltonians and unitary gates*, IEEE Trans. Inf. Theory **49**, 1895–1911 (2003), arXiv quant-ph/0205057.

[BKN00] H. Barnum, E. Knill and M. A. Nielsen, *On quantum fidelities and channel capacities*, IEEE Trans. Inf. Theory **46**, 1317–1329 (2000).

[BMG$^+$] E. Bagan, M.A.Ballester, R. Gill, A.Monras and R. Munoz-Tapia, Optimal full estimation of qubit mixed states, quant-ph/0510158.

[BSSA02] C. Bennett, P. Shor, J. Smolin and A.V.Thapliyal, *Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem*, IEEE Trans. Inf. Theory **48**, 2637–2655 (2002).

[BST98] H. Barnum, J. Smolin and B. Terhal, *Quantum capacity is properly defined without encodings*, Phys. Rev. A **58**, 3496–3501 (1998).

[BTB$^+$05] J. C. Boileau, K. Tamaki, J. Batuwantudawe, R. Laflamme and J. M. Renes, *Unconditional security of a three state quantum key distribution protocol*, Phys. Rev. Lett. **94**, 040503 (2005).

[Cer00] N. Cerf, *Quantum cloning and the capacity of the Pauli channel*, Phys. Rev. Lett. **84**, 4497 (2000).

[CGS05] C. Crepeau, D. Gottesman and A. Smith, *Approximate Quantum Error-Correcting Codes and Secret Sharing Schemes*, Advances in Cryptology – EUROCRYPT (2005), arXiv quant-ph/0503139.

[Cho75] M. Choi, *Completely positive linear maps on complex matrices*, Linear Algebra and its Applications **10**, 285–290 (1975).

[CKW00] V. Coffman, J. Kundu and W. K. Wootters, *Distributed entanglement*, Phys. Rev. A **61**, 052306 (2000).

[CS96] A. Calderbank and P. Shor, *Good Quantum Error-Correcting Codes Exist*, Phys. Rev. A **54**, 1098–1106 (1996).

[CW04] M. Christandl and A. Winter, *Squashed entanglement: An additive entanglement measure*, J. Math. Phys. **45**(3), 829–840 (2004), arXiv:quant-ph/0308088.

[Dev05] I. Devetak, *The private classical capacity and quantum capacity of a quantum channel*, IEEE Trans. Inf. Theory **51**(1), 44–55 (2005), arXiv quant-ph/0304127.

[DHW04] I. Devetak, A. W. Harrow and A. Winter, *A family of quantum protocols*, Phys. Rev. Lett. **93**(23), 230504 (2004), arXiv quant-ph/0308044.

[DS05] I. Devetak and P. Shor, *The capacity of a quantum channel for simultaneous transmission of classical and quantum information*, Communications in mathematical physics **256**(2), 287–303 (2005), arXiv quant-ph/0311131.

[DSS98] D. DiVincenzo, P. Shor and J. Smolin, *Quantum-channel capacity of very noisy channels*, Phys. Rev. A **57**, 830–839 (1998).

[EM96] A. Ekert and C. Macchiavello, *Quantum error correction for communication*, Phys. Rev. Lett. **77**(12), 2585–2588 (1996).

[FvdG99] C. Fuchs and J. van de Graaf, *Cryptographic distinguishability measures for quantum mechanical states*, IEEE Trans. Inf. Theory **45**, 1216–1227 (1999).

[GF05] V. Giovannetti and R. Fazio, *Information-capacity description of spin-chain correlations*, Phys. Rev. A. **71**, 032314 (2005), arXiv:quant-ph/0405110.

[GL03] D. Gottesman and H.-K. Lo, *Proof of security of quantum key distribution with two-way classical communications*, IEEE Trans. Inf. Theory **49**, 457 (2003).

[Got] D. Gottesman, Stabilizer Codes and Quantum Error Correction, Caltech Ph.D. Thesis.

[Got96] D. Gottesman, *Class of quantum error-correcting codes saturating the quantum Hamming bound*, Phys. Rev. A **54**, 1862–1868 (1996).

[GP01] D. Gottesman and J. Preskill, *Secure quantum key distribution using squeezed states*, Phys. Rev. A **63**, 022309 (2001).

[GRTZ02] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden, *Quantum Cryptography*, Reviews of Modern Physics **74**, 145–195 (2002).

[Gur03] V. Guruswami, *List Decoding with Side Information*, IEEE Conference on Computational Complexity , 300–309 (2003).

[Ham50] R. Hamming, *Error detecting and error correcting codes*, Bell System Technical Journal **29**, 147–160 (1950).

[Hay] M. Hayashi, Channel capacities of classical and quantum list decoding, arXiv:quant-ph/0603031.

[HHHO05] K. Horodecki, M. Horodecki, P. Horodecki and J. Oppenheim, *Secure key from bound entanglement*, Phys. Rev. Lett. **94**, 160502 (2005), quant-ph/0309110.

[HJS+96] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland and W. K. Wootters, *Classical information capacity of a quantum channel*, Phys. Rev. A **54**, 1869–1876 (1996).

[HLLO06] K. Horodecki, D. Leung, H.-K. Lo and J. Oppenheim, *Quantum key distribution based on arbitrarily-weak distillable entangled states*, Phys. Rev. Lett. **96**, 070501 (2006), quant-ph/0510067.

[HN03] M. Hayashi and H. Nagaoka, *General formulas for capacity of classical-quantum channels*, IEEE Trans. Inf. Theory **49**, 1753–1768 (2003), arXiv quant-ph/0206186.

[Hol98] A. Holevo, *The Capacity of Quantum Channel with General Signal States*, IEEE Trans. Inf. Theory **44**, 269–273 (1998), arXiv:quant-ph/9611023.

[HW94] P. Hausladen and W. K. Wootters, *A pretty good measurement for distinguishing quantum states*, J. Mod. Opt. **41**, 2385–2390 (1994).

[KGR05] B. Kraus, N. Gisin and R. Renner, *Lower and upper bounds on the secret key rate for quantum key distribution protocols using one-way classical communication*, Phys. Rev. Lett. **95**, 080501 (2005), quant-ph/0410215.

[KR05] R. König and R. Renner, Universally composable privacy amplification against quantum adversaries, in *Proc. of TCC 2005*, 2005, Proc. of TCC 2005, LNCS, Springer, vol. 3378 (2005).

[KW04] M. Koashi and A. Winter, *Monogamy of entanglement and other correlations*, Phys. Rev. A **69**, 022309 (2004).

[KW05] D. Kretschmann and R. F. Werner, *Quantum Channels with Memory*, Phys. Rev. A. **72**, 062323 (2005), arXiv:quant-ph/0502106.

[KY] A. Kawachi and T. Yamakami, Quantum Hardcore functions by Complexity-Theoretical Quantum List Decoding, arXiv:quant-ph/0602088.

[Lan04] M. Langberg, *Private codes or Succinct random codes that are (almost) perfect*, Proceedings of FOCS , 325–334 (2004).

[LC99] H.-K. Lo and H. F. Chau, *Unconditional Security Of Quantum Key Distribution Over Arbitrarily Long Distances*, Science **283**, 2050–2056 (1999), quant-ph/9803006.

[Llo97] S. Lloyd, *Capacity of the noisy quantum channel*, Phys. Rev. A **55**, 1613–1622 (1997).

[Lo01] H.-K. Lo, *Proof of unconditional security of six-state quantum key distribution scheme*, Quantum Information and Computation **1**, 81 (2001).

[Lo03] H.-K. Lo, *Method for decoupling error-correction from privacy amplification*, New Journal of Physics **5**, 36.1 (2003), quant-ph/0201030.

[NC04] M. Nielsen and I. Chuang, *Quantum computation and quantum information*, Cambridge, 2004.

[NN90] J. Naor and M. Naor, Small-bias Probability Spaces: Efficient Constructions and Applications, in *ACM Symposium on Theory of Computing*, pages 213–223, 1990.

[Rai99] E. Rains, *Quantum shadow enumerators*, IEEE Trans. Inf. Theory **45**, 2361–2366 (1999), arXiv: quant-ph/9611001.

[Ren05] R. Renner, Security of Quantum Key Distribution, Ph.D. Thesis, Swiss Federal Institute of Technology, 2005.

[RG] J. M. Renes and M. Grassl, Generalized decoding, effective channels, and simplified security proofs in quantum key distribution, quant-ph/0505061.

[RGK05] R. Renner, N. Gisin and B. Kraus, *An information theoretic security proof for QKD protocols*, Phys. Rev. A **72**, 012332 (2005), quant-ph/0502064.

[Sha48] C. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. **27**, 379–423 and 623–656 (1948).

[Sho] P. Shor, The quantum channel capacity and coherent information., lecture notes, MSRI Workshop on Quantum Computation, 2002.
http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.

[Sho95] P. Shor, *Scheme for reducing decoherence in quantum computer memory*, Phys. Rev. A **52**, R2493–R2496 (1995).

[Smi06] G. Smith, "Communicating Over Adversarial Quantum Channels," in QIP 2006, Paris. http://www.lri.fr/qip06/slides/smith.pdf, 2006.

[SN96] B. Schumacher and M. Nielsen, *Quantum data processing and error correction*, Physical Review A **54**, 2629 (1996).

[SP00] P. Shor and J. Preskill, *Simple Proof of Security of the BB84 Quantum Key Distribution Protocol*, Phys. Rev. Lett. **85**, 441–444 (2000), quant-ph/0003004.

[SSa] P. Shor and J. Smolin, Quantum Error-Correcting Codes Need Not Completely Reveal the Error Syndrome, arXiv quant-ph/9604006.

[SSb] G. Smith and J. Smolin, Degenerate coding for Pauli channels, quant-ph/0604107.

[SST01] P. Shor, J. Smolin and B. Terhal, *Nonadditivity of Bipartite Distillable Entanglement Follows from a Conjecture on Bound Entangled Werner States*, Phys. Rev. Lett. **86**, 2681–2684 (2001).

[Ste96] A. Steane, *Multiple Particle Interference and Quantum Error Correction*, Proc. Roy. Soc. Lond. A **452**, 2551 (1996), quant-ph/9601029.

[SW97] B. Schumacher and M. Westmoreland, *Sending classical information via noisy quantum channels*, Phys. Rev. A **56**, 131–138 (1997).

[TKI03] K. Tamaki, M. Koashi and N. Imoto, Phys. Rev. Lett. **90**, 167904 (2003).

[Ver98] S. Verdu, *Fifty years of Shannon theory*, IEEE Trans. Inf. Theory **44**, 2057–2077 (1998).

[VP98] V. Vedral and M. Plenio, *Entanglement measures and purification procedures*, Phys. Rev. A **57**, 1619 (1998), arXiv:quant-ph/9707035.

[Win99] A. Winter, *Coding theorem and strong converse for quantum channels*, IEEE Trans. Inf. Theory **45**, 2481–2485 (1999).

[Win04] A. Winter, *"Extrinsic" and "intrinsic" data in quantum measurements: asymptotic convex decomposition of positive operator valued measures*, Comm. Math. Phys. **244**, 157–185 (2004).

[WZ82] W. Wooters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299 (5886)**, 802–803 (1982).