On the Capacity of the Erasure Channel and the Construction of an ϵ-randomizing Map

by

Joungkeun Lim

B.S., Seoul National University, 2003

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
SEP 29 2008
LIBRARIES

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

©Joungkeun Lim, 2008. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author .....
Department of Mathematics
August 18, 2008

Certified by. .....
Peter Shor
Morss Professor of Applied Mathematics
Thesis Supervisor

Accepted by .....
Alar Toomre
Chairman, Applied Mathematics Committee

Accepted by .....
David Jerison
Chairman, Department Committee on Graduate Students

ARCHIVES

ARCHIVES

On the Capacity of the Erasure Channel and the Construction of an ϵ-randomizing Map

by

Joungkeun Lim

Submitted to the Department of Mathematics
on August 18, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The quantum information theory is the counterpart of the classical information theory in quantum computation, and it has raised many questions regarding the transmission and security of the information in quantum computers. This thesis studies the efficiency of such processes and contributes to two separate area of quantum information theory.

The first half of this thesis presents a communication protocol for the erasure channel assisted by backward classical communication, which achieves a significantly better rate than the best prior result. In addition, we reduce the proof of a new upper bound for the capacity of the channel to a conjecture. The proposed upper bound is smaller than the capacity of the erasure channel when it is assisted by two-way classical communication. Hence, the proof of the separation between quantum capacities assisted by backward classical communication and two-way classical communication is also reduced to the conjecture.

The second half of this thesis studies the construction of an ϵ-randomizing map that uses Pauli operators. An ϵ-randomizing map transforms any n-qubit state to an almost random state - a state that is within ϵ-distance of the completely random state, in the trace norm. We show that at least $\Theta(\frac{2^n}{\epsilon^{log_2 3}})$ Pauli operators are required for the construction of an ϵ-randomizing map. This proves the lower bound on the length of a private key required for a private communication as $min\{2n, n+log_2 3\, log(1/\epsilon)\}+O(1)$. Our result matches the previous upper bound of $n + 2log(1/\epsilon) + O(1)$ for the optimal key length, in the order of n.

Thesis Supervisor: Peter Shor
Title: Morss Professor of Applied Mathematics

0.1 Acknowledgments

Professor Peter Shor has been instrumental in guiding me throughout my years in graduate school. This thesis would never be completed without his help. He enlightened and supported me, helped me refine my idea, and shared several of his ideas.

I would like to thank Professors Daniel J. Kleitman and Scott Aaronson for being my thesis committee.

Also I am grateful to Professor Debbie Leung for being a collaborator, a mentor, and a friend. In fact, she initially acquainted me with the ϵ-randomizing map problem that composes the second half of this thesis.

Professors Andrzej Grudka and Michal Horodecki pointed out an important mistake and suggested a solution that substantially simplifies a proof in Chapter 2.

I have been fortunate to have many friends in the math department at MIT. I wish to thank Victor Chen, Pasha Pylyavskyy, Alan Leung, Jaehyuk Choi, King Yick and Michael Baym for their support during the last five years.

My work in the first four years of graduate school was mostly funded by the Samsung Lee Kun Hee Scholarship Foundation. I am grateful for their generous support. I would like to thank the National Science Foundation for the support through grant CCF-0431787.

The Chisholm Fund and Akamai generously supported several of my travel opportunities.

I would like to thank my parents and my brother. Their support throughout my life made it possible for me to pursue study in the US.

Finally, I dedicate the thesis to my wife, Youjin, and my son, Joonsuh.

Contents

0.1 Acknowledgments 4

1 Motivation and contents **7**

2 Capacity of quantum erasure channel assisted by backward classical communication **11**

2.1 Preliminaries 12

2.1.1 Quantities and inequalities 12

2.1.2 Quantum capacities 13

2.1.3 Coherent teleportation 14

2.2 Introduction and previous results 15

2.3 Lower bound on $Q_{\mathrm{B}}(\mathcal{N}_p)$ 16

2.3.1 Communication protocol using coherent teleportation 16

2.3.2 Communication protocol using coherent superdense coding 17

2.3.3 Lower bound 19

2.4 Upper bound on $Q_{\mathrm{B}}(\mathcal{N}_p)$ 19

2.4.1 $Q_{\mathrm{RB}}(\mathcal{N}_p)$ and a conjecture 19

2.4.2 Mutual information bound 20

2.4.3 Proof of an upper bound on $Q_{\mathrm{RB}}(\mathcal{N}_p)$ 24

2.5 Discussion 25

3 Lower bound on the number of Pauli operators constructing an ϵ-randomizing map **27**

3.1 Introduction and previous results 28

3.2 Distribution of keys over $\{0,1\}^n \times \{0,1\}^n$ 30

3.2.1 The base case 31

3.2.2 The general case 32

3.3 Visual representation 36

3.3.1 Key set 36

3.3.2 Permutation P_t 36

3.3.3 Distribution of the key set 37

3.4 The lower bound 38

3.4.1 Row-wise distance on subarrays 39

3.4.2 Partitioning arrays 42

3.4.3 Proof of lower bound 45

3.5 Discussion 47

Chapter 1

Motivation and contents

Quantum information theory studies how to process the information stored in quantum states. An important issue in this field is to find *communication protocols* that encode and decode quantum states so that the quantum information is not damaged by the noise during the transmission from the sender to the receiver. A communication protocol should also focus on the *rate* of the coding - the asymptotic ratio of the size of the original information to the size of the encoded information. Hence, a protocol with a higher rate is preferred.

A *quantum channel* is a communication medium through which quantum information is transmitted. The *capacity* of a channel is the theoretical maximum of the rate of the channel over all possible communication protocols. Study of the capacity is important in that the communication protocol that matches its rate with the capacity is indeed the most efficient protocol, asymptotically. A lower bound of the capacity of a channel is given by the rate of an efficient communication protocol. The upper bound is proved by a mathematical argument regarding the nature of the noise of the channel. If the lower and upper bound match, the capacity is determined.

The *quantum erasure channel* is a channel which erases random qubits with a certain probability. The capacity of the quantum erasure channel can differ when the channel is assisted by various classical communications. The classical communi-

cations can be void, forward, backward, or two-ways. The capacities of the erasure channel with void, forward, and two-way communications are completely determined. However, the capacity is not determined when the channel is assisted by backward classical communication.

Chapter 2 discusses the capacity of the quantum erasure channel when it is assisted by backward classical communication. Our improvement over the previous results goes both ways - improving the lower bound of the capacity and presenting an idea to improve its upper bound. We present an improved communication protocol that has a significantly better rate than previous ones. Also we reduce the proof of a new upper bound to the validity of a conjecture. We believe the conjecture is true, and the intuition behind the conjecture is given in this chapter.

Another important issue in quantum information theory is a *secure encryption of quantum states.* One wishes to encrypt a quantum state so that another person can decrypt the state, but a third party can gain almost no information from eavesdropping on the encrypted state.

The *randomization of quantum states* is a scheme to encrypt quantum states so that, without access to the shared key, the encrypted state appears very close to the completely random state - $\frac{\mathcal{I}}{2^n}$ for an n-qubit quantum state. To be precise, by the encryption, all the quantum states are mapped to states less than ϵ-distance from the completely random state. We call this encryption the *ϵ-randomizing map.*

The best known method to construct an ϵ-randomizing map is to use Pauli operators. Assume that two parties share a private binary key (a, b) chosen out of a key set S. Given a quantum state ρ and a private key (a, b), the sender encrypts the state to $X^a Z^b \rho Z^b X^a$. The receiver applies the inverse operation to recover the initial state ρ as $Z^b X^a (X^a Z^b \rho Z^b X^a) X^a Z^b = \rho$. However, for a third party who does not have access to the key (a, b), the state appears as

$$R(\rho) = \frac{1}{|S|} \sum_{(a,b) \in S} X^a Z^b \rho Z^b X^a,$$

which is an almost random state with a well-chosen key set S.

Since the private key is an exhaustive resource, the key length needs to be minimized. Since the key length can be reduced to $log_2|S|$, a scheme with a set S of minimal size, i.e., a scheme with a minimal number of Pauli operators is the most efficient.

Chapter 3 explores the lower bound on the key length for the private communication. For n-qubit quantum states, our lower bound improves over the best prior lower bound and matches the best upper bound for the optimal construction, in the order of n.

Chapter 2

Capacity of quantum erasure channel assisted by backward classical communication

In this chapter, we study the capacity of quantum erasure channel, when unlimited amount of backward classical communication channel is allowed to use. Our result approaches in two ways: First, we present an efficient communication protocol giving a lower bound on the capacity. Second, we show an idea to improve the upper bound on the capacity. We reduce a proof of a new upper bound to a conjecture. If the conjecture holds, the new upper bound is smaller than the capacity assisted by two-way classical communication. Hence, the separation between two capacities - the capacity when the channel is assisted by backward classical communication and the capacity when it is assisted by two-way classical communication - is also reduced to the conjecture.

Section 1 introduces notions and facts needed to understand the result of this chapter. Section 2 describes the problem and the previous results. Section 3 shows a new lower bound on the capacity by giving an efficient communication protocol. Section 4 reduces a proof of a new upper bound on the capacity to a conjecture.

Section 5 discusses our result.

2.1 Preliminaries

This section introduces notions and facts in quantum information theory that are relevant to our result in this chapter. We also state and prove a lemma that is used for the proof of a theorem in Section 2.4.

2.1.1 Quantities and inequalities

Recall the definition of *von Neumann entropy* [14] $H(A) = H(\psi^A) = -tr(\psi^A \log \psi^A)$, where ψ^A is the density operator for system A. The log is of base 2. Suppose disjoint quantum systems A and B have a joint state ρ^{AB}. Then the following inequality is known as *subadditivity inequality*[14]:

$$H(A) + H(B) \geq H(AB),$$

where $H(AB) = H(\rho^{AB})$, $H(A) = H(\rho^A) = H(Tr_B \rho^{AB})$, and $H(B) = H(\rho^B) = H(Tr_A \rho^{AB})$. Similarly, for disjoint quantum systems A, B, and C, *strongsubadditivity inequality*[14] is defined as

$$H(AB) + H(BC) \geq H(B) + H(ABC).$$

We can further define *quantum mutual information* [9] and *coherent information* [15, 16] as

$$I(A;B) = H(A) + H(B) - H(AB) \text{ and}$$
$$I(A\rangle B) = H(B) - H(AB).$$

Nonnegativity of quantum mutual information is equivalent to the subadditivity in-

equality.

The following lemma shows some properties of quantum mutual information and coherent information, and will be used in the proof of a theorem in Section 2.4.

Lemma 1. *For disjoint systems A, B, and C,*
(i) $I(AB;C) - I(B;C) \leq I(A;BC)$.
(ii) $I(A\rangle B) \leq I(A\rangle BC)$.
(iii) $I(A\rangle C) + I(B\rangle C) \leq I(AB\rangle C)$.
(iv) $I(A\rangle BC) - I(A\rangle B) \leq 2H(CE)$, *where E is any subset of B.*

Proof. Subadditivity and strong subadditivity inequalities easily give $(i), (ii), (iii)$,

$$H(CDE) \leq H(D) + H(CE),$$
$$H(AD) \leq H(CE) + H(ADCE), \text{ and}$$
$$H(D) + H(ADE) \leq H(AD) + H(DE),$$

for $E \subset B$ and $D = B/E$. Adding these three inequalities yields (iv). □

2.1.2 Quantum capacities

The *capacity* $Q(\chi)$ of a channel χ is the theoretical maximum of the rate m/n that is achievable by a communication protocol that sends m-qubit information with n uses of the channel, where n tends to infinity.

The above definition of Q is functional for the case without auxiliary resources, and additional free classical communication may increase the capacity. We use Q, Q_1, Q_{B}, and Q_2 to denote the quantum capacities of a quantum channel when unassisted, assisted by unlimited forward, backward, and two-way classical communication, respectively. It was proved that classical forward communication alone does not increase the quantum capacity of any channel; in other words, $Q(\chi) = Q_1(\chi)$ for all channels χ [7]. In contrast, Q_2 is greater than Q for some channels [7]. Q_{B} is also

known to be greater than Q for some channels [6], but it has been an open question whether $Q_{\mathrm{B}}(\chi) = Q_2(\chi)$ for all χ.

The reliability of a quantum communication algorithm is measured by *fidelity*, a measure of similarity between input states and output states. The fidelity of states ρ_{in} and ρ_{out} is defined to be

$$F(\rho_{in}, \rho_{out}) \equiv \mathrm{tr}\sqrt{\rho_{in}^{1/2}\rho_{out}\rho_{in}^{1/2}}.$$

Hence the fidelity is 1 for two identical states, and 0 for two orthogonal states. The fidelity between the input and the output states is also equal to the probability that the latter would pass a test of being the former. In this paper, we consider near-perfect communication protocols that produce, with high probability, the output states of high fidelity with the input states.

2.1.3 Coherent teleportation

From now on, we call the sender Alice, the receiver Bob, and the environment Eve.

Given an unknown qubit state $|\psi\rangle = a|0\rangle + b|1\rangle$ in system M and an *ebit* (sometimes called an EPR pair or Bell state) $|\Phi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ between Alice and Bob, Alice can transmit $|\psi\rangle$ to Bob by teleportation [5]. In the original teleportation protocol, the change of basis takes the initial state $|\psi\rangle_M|\Phi\rangle_{AB}$ to

$$\tfrac{1}{2}\sum_{ij} |ij\rangle_{MA}\, X^i Z^j |\psi\rangle_B\,. \tag{2.1}$$

Reference [12] proposes a *coherent* variant of teleportation in which Alice does not measure $|ij\rangle_{MA}$ but instead, coherently copies $|ij\rangle_{MA}$ to two ancillary systems C_1C_2 and transmits them coherently to Bob. Mathematically, Alice and Bob share the joint state

$$\tfrac{1}{2}\sum_{ij} |ij\rangle_{MA}\, |ij\rangle_{C_1C_2}\, X^i Z^j |\psi\rangle_B\,.$$

After receiving C_1C_2, Bob can apply a control-X from C_1 to B and then a control-Z from C_2 to B. Alice and Bob then share the state

$$\tfrac{1}{2}\sum_{ij} |ij\rangle_{MA}\, |ij\rangle_{C_1C_2}\, |\psi\rangle_B\,,$$

with $|\psi\rangle$ transmitted *and* two ebits shared between Alice and Bob. The ebits saved here can be used as a resource for the future communication.

2.2 Introduction and previous results

We study the *quantum erasure channel*, which was first introduced in [11]. The quantum erasure channel of erasure probability p, denoted by $\mathcal{N}_p$, replaces the incoming qubit, with probability p, with an "erasure state" $|2\rangle$ orthogonal to both $|0\rangle$ and $|1\rangle$, thereby both erasing the qubit and informing the receiver that it has been erased. In an equivalent formulation, called the isometric extension, the channel exchanges the incoming qubit with the environmental system in state $|2\rangle$ with probability p. It was shown in [6] that the quantum capacities Q, Q_1, and Q_2 for $\mathcal{N}_p$ are given by

$$\begin{aligned} &Q(\mathcal{N}_p) = Q_1(\mathcal{N}_p) = \max\{0, 1-2p\} \text{ and} \\ &Q_2(\mathcal{N}_p) = 1-p. \end{aligned}$$

However, until the current investigation, little has been known about $Q_{\mathrm{B}}(\mathcal{N}_p)$ except for two lower bounds that follow straightforwardly from 1-way hashing [7] and teleportation [5] and an upper bound given by $Q_2(\mathcal{N}_p)$ as

$$\begin{aligned} &Q_{\mathrm{B}}(\mathcal{N}_p) \geq 1-2p, \text{ if } p \leq 2/5, \\ &Q_{\mathrm{B}}(\mathcal{N}_p) \geq (1-p)/3, \text{ if } p \geq 2/5, \text{ and} \\ &Q_{\mathrm{B}}(\mathcal{N}_p) \leq Q_2(\mathcal{N}_p) = 1-p. \end{aligned} \tag{2.2}$$

In this chapter, we present an efficient communication protocol that achieves a better lower bound of $Q_{\mathrm{B}}(\mathcal{N}_p)$, and we reduce a new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p)$ to a conjecture. If the conjecture is true, $Q_{\mathrm{B}}(\mathcal{N}_p) < Q_2(\mathcal{N}_p)$ for all p and the separation between Q_{B} and Q_2, the previously open question, is resolved.

2.3 Lower bound on $Q_{\mathrm{B}}(\mathcal{N}_p)$

we derive an improved lower bound for $Q_{\mathrm{B}}(\mathcal{N}_p)$ by providing a communication protocol. The protocol combines two subprotocols that utilize *coherent teleportation* introduced in [12].

2.3.1 Communication protocol using coherent teleportation

Suppose Alice and Bob already share an ebit, and Alice teleports $|\psi\rangle$ to Bob by attempting to use the erasure channel for coherent classical communication of each of $|i\rangle_{C_1}$ and $|j\rangle_{C_2}$ (see Section 2.1.3 on coherent teleportation). Bob tells Alice whether the communication is erased or not. If so, Alice copies and sends it again until Bob receives it. Note that the transmission is coherent if it is not erased in the first trial. If i and j are erased k and l times before they are sent successfully, the state becomes (after Bob's controlled-X and Z)

$$\begin{aligned}&\tfrac{1}{2}\sum_{ij}|ij\rangle_A|i\rangle_E^{\otimes k}|j\rangle_E^{\otimes l}|ij\rangle_B|\psi\rangle_B\\ &\sim|\Gamma\rangle_{ABE}^{\otimes(1_k+1_l)}|\Phi\rangle_{AB}^{\otimes(2-1_k-1_l)}|\psi\rangle_B,\end{aligned}$$

where $1_k = \begin{cases} 0 & \text{if } k=0 \\ 1 & \text{if } k>0 \end{cases}$ and similarly for 1_l, $|\Gamma\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, and $\sim$ denotes equivalence up to a unitary transformation on E.

Since the success probability of each transmission is $1-p$, Alice tries $\frac{1}{1-p}$ times on average to send each register i and j. Hence she transmits $\frac{2}{1-p}$ qubits through the

channel. Both 1_k and 1_l have expectation p. In *asymptotic resource inequality*[12]

$$\tfrac{2}{1-p}\,\mathcal{N}_p + \Phi_{AB} \;\geq\; 1 \text{ Qbit} + 2(1-p)\;\Phi_{AB} + 2p\;\Gamma_{ABE}, \tag{2.3}$$

where resources on the left-hand side simulate those on the right, $\mathcal{N}_p$ denotes one use of the erasure channel, and Qbit denotes one use of the noiseless qubit channel. We have used Φ and Γ as shorthand for $|\Phi\rangle\langle\Phi|$ and $|\Gamma\rangle\langle\Gamma|$. With free backward classical communication, one use of $\mathcal{N}_p$ can prepare one ebit with probability $1-p$. Hence,

$$1\,\mathcal{N}_p \geq (1-p)\;\Phi_{AB}. \tag{2.4}$$

We combine equations (2.3) and (2.4) to get

$$1\,\mathcal{N}_p \geq \tfrac{1-p}{2} \;\text{ Qbit, if } p \leq 1/2 \text{ and}$$
$$1\,\mathcal{N}_p \geq \tfrac{1-p}{1+2p} \;\text{ Qbit, if } p \geq 1/2.$$

Hence, the rate of the first subprotocol is

$$\tfrac{1-p}{2}\;, \text{ if } p \leq 1/2 \text{ and}$$
$$\tfrac{1-p}{1+2p}\;, \text{ if } p \geq 1/2.$$

2.3.2 Communication protocol using coherent superdense coding

This method only differs from the previous subprotocol in that $|ij\rangle$ will be sent using a coherent version of *superdense coding* [14]. More specifically, in this case, Alice and Bob first share an ebit $|\Phi\rangle_{C_1C_2}$ where C_1 belongs to Alice and C_2 belongs to Bob. After the change of basis (see equation (2.1)), Alice applies control-X from M to C_1 and control-Z from A to C_1, resulting in the joint state

$$\tfrac{1}{2}\sum_{ij} |ij\rangle_{MA}\, |\Phi_{ij}\rangle_{C_1C_2}\, X^iZ^j|\psi\rangle_B\,,$$

and sends C_1 to Bob using the erasure channel. The states $|\Phi_{ij}\rangle = X^iZ^j|\Phi\rangle$ are orthogonal (they form the *Bell basis*) [14]. In case of erasure, Bob and Eve share $|\Phi_{ij}\rangle_{C_1C_2}$ and Alice and Bob will take another ebit and repeat the superdense coding procedure, until Bob receives the transmission (call the two-qubit system in his possession D_1D_2). Then, Bob applies the transformation $|\Phi_{ij}\rangle_{D_1D_2} \to |ij\rangle_{D_1D_2}$ and coherently reverts the X^iZ^j not only in $X^iZ^j|\psi\rangle_B$ but also in all the $|\Phi_{ij}\rangle$ he shares with Eve (by acting only on his halves), so that the final state becomes

$$\tfrac{1}{2}\sum_{ij} |ij\rangle_{MA}|ij\rangle_{D_1D_2}\, |\Phi\rangle_{EB}^{\otimes k}\, |\psi\rangle_B,$$

where k again denotes the number of erasures before the successful transmission. In this method, Alice and Bob *always* share 2 ebits at the end.

Once again, Alice needs to apply superdense coding $\frac{1}{1-p}$ times on average. This gives the asymptotic resource inequality,

$$\begin{aligned}&\Phi_{AB} + \tfrac{1}{1-p}\left[\,\mathcal{N}_p + \Phi_{AB}\,\right]\\ &\quad\geq\ 1\ \text{Qbit}\ + 2\ \Phi_{AB} + (\tfrac{1}{1-p} - 1)\ \Phi_{BE}\,.\end{aligned}$$

Note that the above consumes more ebits than it produces for all p; thus, we use equation (2.4) to supply the needed ebits, and obtain

$$1\ \mathcal{N}_p \geq (1-p)^2\ \text{Qbit}.$$

Hence the rate of the second subprotocol is $(1-p)^2$.

2.3.3 Lower bound

Applying the two protocols selectively, the rate of the protocol is

$$\begin{aligned} &(1-p)^2 \text{ , if } p \leq 1/2 \text{ and} \\ &\tfrac{1-p}{1+2p} \text{ , if } p \geq 1/2. \end{aligned} \tag{2.5}$$

2.4 Upper bound on $Q_{\mathrm{B}}(\mathcal{N}_p)$

The purpose of this section is to propose a new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p) \leq \frac{1-p}{1+p}$. We reduce the proof of the proposed upper bound to a conjecture that $Q_{\mathrm{RB}}(\mathcal{N}_p) = Q_{\mathrm{B}}(\mathcal{N}_p)$

2.4.1 $Q_{\mathrm{RB}}(\mathcal{N}_p)$ and a conjecture

Let $Q_{\mathrm{RB}}(\mathcal{N}_p)$ be the capacity of the erasure channel $\mathcal{N}_p$ when the channel is assisted by backward classical communication, with a further restriction that Bob is not allowed to perform any measurement until the last transmission of n qubits from Alice to Bob. Since Bob's measurement in the middle of the transmission is not allowed, a choice of communication protocol is limited. Hence, the capacity may decrease with this restriction. Therefore,

$$Q_{\mathrm{RB}}(\mathcal{N}_p) \leq Q_{\mathrm{B}}(\mathcal{N}_p).$$

A measurement by Alice or Bob is performed to find out undiscovered information. For the erasure channel, the transmission from Alice is either intact to Bob or lost to Eve. Hence, as long as Alice is informed about what happened to the previous transmissions, all the information of the communication is open to Alice including Bob's current quantum state. With Backward classical communication, Bob can notify Alice if the transmission is successful or not. Hence, Alice is informed of all the necessary information about the communication, and there is no need for Bob to measure his state to find out undiscovered information.

The above intuition tells us that prohibiting Bob's measurement in the middle

of the communication does not exclude the most efficient communication protocol. Hence, we propose the following conjecture.

conjecture 2. *For the quantum erasure channel $\mathcal{N}_p$, $Q_{\mathrm{RB}}(\mathcal{N}_p) = Q_{\mathrm{B}}(\mathcal{N}_p)$.*

In the following sections, we prove $Q_{\mathrm{RB}}(\mathcal{N}_p) \leq \frac{1-p}{1+p}$. If the conjecture is true, the new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p) \leq \frac{1-p}{1+p}$ is proved.

2.4.2 Mutual information bound

In this section, we prove a theorem regarding the bound of mutual information between communication parties, which has to be satisfied for a successful communication protocol. We assume that Bob is allowed to perform a quantum measurement only at the end of the communication.

By the definition of the capacity, for each n, there is a protocol $\mathcal{P}_n$ that uses backward classical communication and $\mathcal{N}_p$ at most n times and transmits $n(Q_{\mathrm{RB}}(\mathcal{N}_p) - \delta_n)$ qubits from Alice to Bob with fidelity at least $1 - \epsilon_n$ and probability at least $1 - \epsilon_n$, where $\epsilon_n, \delta_n \longrightarrow 0$ as $n \longrightarrow \infty$.

It was shown that capacity for transmission of entanglement equals to the capacity for transmission of subspace [4]. In other words, if Alice can send m halves of ebits shared between Alice and a reference system R through the channel, she can also send arbitrary m-qubit state through the channel, and vice versa. Hence we assume that Alice starts with half of $|\Pi\rangle_{AR}^{\otimes m}$ that are maximally entangled with reference system R, and wants to send her half to Bob (recall that $|\Pi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$). And the final fidelity between the input state $\rho_{in} = |\Pi\rangle_{AR}^{\otimes m}$ and output state $\rho_{out} = (\mathcal{A} \otimes \mathcal{I})|\Pi\rangle_{AR}^{\otimes m}$ is almost 1. Note that $\mathcal{A}$ is the quantum operation by the communication on the half of the entanglements on Alice's side.

Our strategy to show the upper bound is as follows. We consider a protocol that transmits m qubits with n uses of the channel. If Alice transmits her halves of the ebits shared with R directly through the channel, any loss to Eve can never be recovered.

Thus, Alice has to transmit quantum states whose potential entanglement with R can be materialized or nullified depending on Bob's backward communication and Alice's future transmissions. The materializing or nullifying process requires further uses of the channel, giving an upper bound to the capacity.

To quantify the above idea, denote by $S_1, S_2, \cdots S_n$ the qubits transmitted by Alice through the channel. Each S_i is delivered to Bob with probability $1-p$ or lost to Eve with probability p. Let $\mathcal{B} = \{i | S_i \text{ sent to Bob}\}$ and $\mathcal{E} = \{i | S_i \text{ sent to Eve}\}$ be the index sets of qubits delivered to Bob and Eve. Furthermore, let $B_i = \bigcup_{1 \le j \le i,\, j \in \mathcal{B}} S_j$ be Bob's system after the ith channel use. Thus $B_i = B_{i-1} \cup S_i$ if S_i is delivered to Bob, and $B_i = B_{i-1}$ if S_i is lost to Eve. Similarly we define $E_i = \bigcup_{1 \le j \le i,\, j \in \mathcal{E}} S_j$ to be Eve's system after the ith transmission. After the final decoding operation, Bob produces an m-qubit system $B^{(1)}$ that is almost maximally entangled with the system R. We denote the rest of Bob's system by $B^{(2)}$. Bob's decoding operation can be assumed to be isometric by making his measurement operations coherent as shown in [12].

In the following theorem, $I(S_i; B_{i-1}R)$ is the amount of mutual information carried by each transmission S_i. Part (i) of the theorem states that a sufficient amount of mutual information ($2m$ for m ebits) has to be delivered to Bob. Part (ii) states that the more mutual information is lost to Eve, the more transmissions are needed to nullify the lost information.

Theorem 3. *If the fidelity between the input and output states is at least* $1-\epsilon_n$, *then*
(i) $\sum_{i \in \mathcal{B}} I(S_i; B_{i-1}R) \ge 2m - 2(2\sqrt{2}m\sqrt{\epsilon_n} + 1)$.
(ii) $\sum_{i \in \mathcal{E}} I(S_i; B_{i-1}R) \le n - m + 4(2\sqrt{2}m\sqrt{\epsilon_n} + 1)$.

Proof. (i) For each $i \in \mathcal{B}$, apply part (i) of lemma 1 on the systems S_i, B_{i-1}, and R to obtain

$$\begin{aligned} I(B_i; R) - I(B_{i-1}; R) &= I(B_{i-1}S_i; R) - I(B_{i-1}; R) \\ &\leq I(S_i; B_{i-1}R). \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{i \in \mathcal{B}} I(S_i; B_{i-1}R)) &\geq \sum_{i \in \mathcal{B}} (I(B_i; R) - I(B_{i-1}; R)) \\ &= I(B_n; R) = I(B^{(1)}B^{(2)}; R) \\ &\geq I(B^{(1)}; R) \\ &= H(B^{(1)}) + H(R) - H(B^{(1)}R) \\ &\geq 2(H(R) - H(B^{(1)}R)). \end{aligned}$$

Note that the fidelity between the state μ in $B^{(1)}R$ and $\Phi^{\otimes m}$ is at least $1-\epsilon_n$. Let $D = \frac{1}{2}\mathrm{tr}|\mu - \Phi^{\otimes m}|$ be the *trace distance* [14] between μ and $\Phi^{\otimes m}$. By page 415 of [14],

$$D \leq \sqrt{1 - F(\mu, \Phi^{\otimes m})^2} \leq \sqrt{2\epsilon_n}.$$

By Fannes' inequality [14],

$$\begin{aligned} H(B_1R) = |H(\mu) - H(\Phi^{\otimes m})| &\leq 2Dm - 2D\log(2D) \\ &\leq 2\sqrt{2}m\sqrt{\epsilon_n} + 1. \end{aligned}$$

(ii) Using 2, 3, and 4 to denote the use of parts $(ii), (iii)$, and (iv) of lemma 1 respectively, we have

$$\begin{aligned}
\sum_{i\in\mathcal{E}} I(S_i\rangle B_{i-1}R) &\overset{2}{\leq} \sum_{i\in\mathcal{E}} I(S_i\rangle B_n R) \\
&\overset{3}{\leq} I(\bigcup_{i\in\mathcal{E}} S_i\rangle B_n R) \\
&= I(E_n\rangle B^{(1)}B^{(2)}R) \\
&\overset{4}{\leq} I(E_n\rangle B^{(1)}B^{(2)}) + 2H(B^{(1)}R) \\
&\overset{3}{\leq} I(E_nR\rangle B^{(1)}B^{(2)}) - I(R\rangle B^{(1)}B^{(2)}) + 2H(B^{(1)}R) \\
&= I(E_nR\rangle B_n) - I(R\rangle B^{(1)}B^{(2)}) + 2H(B^{(1)}R),
\end{aligned}$$

where the equalities use the fact that Bob's decoding is isometric. $I(E_nR\rangle B_n)$ is upper bounded by $|B_n| = |\mathcal{B}| = n - |\mathcal{E}|$. $I(R\rangle B^{(1)}B^{(2)})$ is lower bounded as

$$\begin{aligned}
I(R\rangle B^{(1)}B^{(2)}) &\overset{2}{\geq} I(R\rangle B^{(1)}) \\
&\overset{4}{\geq} I(R\rangle B^{(1)}T) - 2H(T) \\
&= m - 2H(B^{(1)}R)
\end{aligned}$$

where T purifies $B^{(1)}R$. Putting together the two previous sets of inequalities,

$$\sum_{i\in\mathcal{E}} I(S_i\rangle B_{i-1}R) \leq n - |\mathcal{E}| - m + 4(2\sqrt{2}m\sqrt{\epsilon_n} + 1).$$

Hence,

$$\begin{aligned}\sum_{i\in\mathcal{E}} I(S_i; B_{i-1}R) &= \sum_{i\in\mathcal{E}} (H(S_i) + I(S_i\rangle B_{i-1}R)) \\ &\leq \sum_{i\in\mathcal{E}} (1 + I(S_i\rangle B_{i-1}R)) \\ &\leq n - m + 4(2\sqrt{2}m\sqrt{\epsilon_n} + 1).\end{aligned}$$

□

2.4.3 Proof of an upper bound on $Q_{\mathrm{RB}}(\mathcal{N}_p)$

Since Alice cannot predict whether Bob or Eve will receive the next transmission and a certain fraction of the transmission are lost to Eve, the same fraction of mutual information has to be lost to Eve. Combined with the theorem, the argument gives an upper bound of $Q_{\mathrm{RB}}(\mathcal{N}_p)$. To prove this rigorously, consider the following random variable.

$$X_i = \begin{cases} \frac{p}{2}\, I(S_i; B_{i-1}R) & \text{if } S_i \text{ is delivered to Bob} \\ \frac{-(1-p)}{2}\, I(S_i; B_{i-1}R) & \text{if } S_i \text{ is lost to Eve} \end{cases}$$

Then $|X_i| \leq 1$ and $E(X_i) = 0$. Note that the X_i's may not be independent variables. Let $Y_i = \sum_{j=1}^{i} X_j$ and $Y_0 = 0$. Then $Y_0, Y_1, \cdots, Y_n$ is a *martingale* [1] with $|Y_{i+1} - Y_i| \leq 1$. If the fidelity between the input and output states is at least $1 - \epsilon_n$, then from theorem 3

$$\begin{aligned}Y_n &= \tfrac{p}{2}\sum_{i\in\mathcal{B}} I(S_i; B_{i-1}R) - \tfrac{(1-p)}{2}\sum_{i\in\mathcal{E}} I(S_i; B_{i-1}R) \\ &\geq \tfrac{(1+p)}{2}m - \tfrac{(1-p)}{2}n - (2-p)(2\sqrt{2}m\sqrt{\epsilon_n} + 1).\end{aligned}$$

Assume by contradiction that $Q_{\mathrm{RB}}(\mathcal{N}_p) > \frac{1-p}{1+p}$. Then, for sufficiently large n, $\frac{m}{n} \geq \frac{1-p}{1+p} + 4k$ for some $k > 0$. The above expression for Y_n, which holds with probability

at least $1-\epsilon_n$, will exceed kn. Therefore

$$\lim_{n\to\infty} \Pr[\ |Y_n| \geq kn] = 1. \tag{2.6}$$

However, Azuma's inequality [1] applied to martingale Y_i gives

$$\Pr[|Y_n| \geq kn] \leq e^{-\frac{k^2}{2}n}.$$

Therefore,

$$\lim_{n\to\infty} \Pr[\ |Y_n| \geq kn] = 0,$$

which is a contradiction with equation (2.6). Hence,

$$Q_{\mathrm{RB}}(\mathcal{N}_p) \leq \tfrac{1-p}{1+p}. \tag{2.7}$$

2.5 Discussion

The previous lower and upper bounds of $Q_{\mathrm{B}}(\mathcal{N}_p)$ given in equation (2.2) are

$$\begin{aligned}
&Q_{\mathrm{B}}(\mathcal{N}_p) \geq 1-2p, \text{ if } p \leq 2/5,\\
&Q_{\mathrm{B}}(\mathcal{N}_p) \geq (1-p)/3, \text{ if } p \geq 2/5, \text{ and}\\
&Q_{\mathrm{B}}(\mathcal{N}_p) \leq Q_2(\mathcal{N}_p) = 1-p.
\end{aligned}$$

Our new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p)$ given in equation (2.5) is

$$\begin{aligned}
&Q_{\mathrm{B}}(\mathcal{N}_p) \geq (1-p)^2 \text{ , if } p \leq 1/2,\\
&Q_{\mathrm{B}}(\mathcal{N}_p) \geq \tfrac{1-p}{1+2p} \text{ , if } p \geq 1/2.
\end{aligned}$$

We proved an upper bound on $Q_{\mathrm{RB}}(\mathcal{N}_p)$ in equation (2.7) as

$$Q_{\mathrm{RB}}(\mathcal{N}_p) \leq \tfrac{1-p}{1+p}.$$

If the conjecture 2 holds, the new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p)$ is given as

$$Q_{\mathrm{B}}(\mathcal{N}_p) \leq \tfrac{1-p}{1+p}.$$

The new upper bound of $Q_{\mathrm{B}}(\mathcal{N}_p)$ is strictly less than $Q_2(\mathcal{N}_p)$. Hence the same conjecture is the reduced problem of the separation between Q_{B} and Q_2, the long-standing question raised in [6].

In the following figure, dashed lines (1) and (2) are previous lower and upper bounds. The lower solid line (3) is our new lower bound and the upper solid line (4) is our conjectured upper bound.

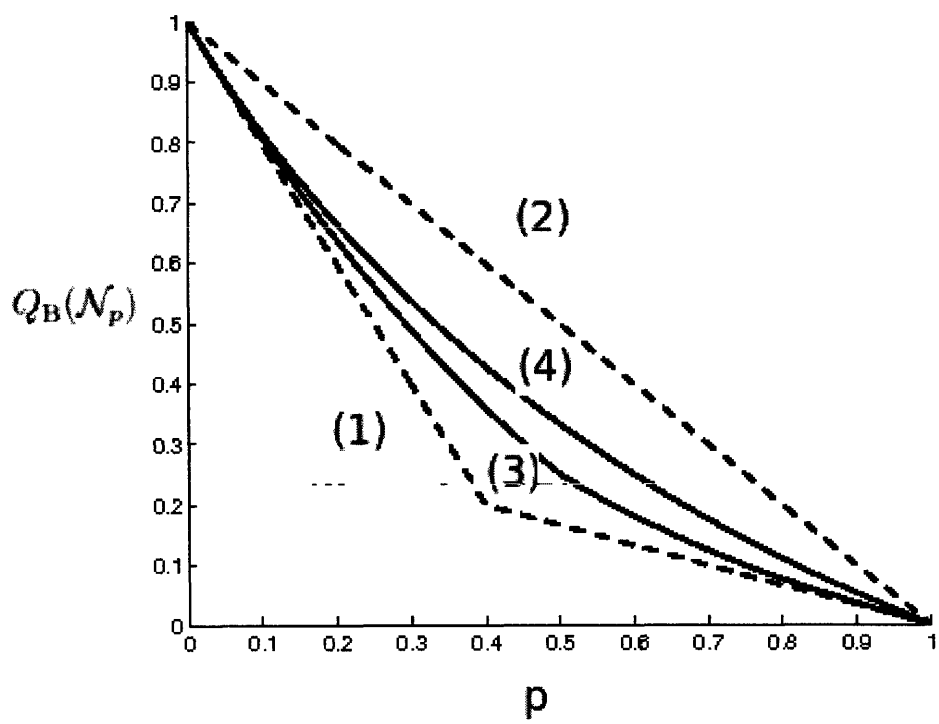

Figure 2-1: Lower and upper bounds on $Q_{\mathrm{B}}(\mathcal{N}_p)$

Chapter 3

Lower bound on the number of Pauli operators constructing an ϵ-randomizing map

In this chapter, we study the construction of ϵ-randomizing map with Pauli operators. We prove the lower bound on the number of Pauli operators needed for a construction of the map. This bound also implies the lower bound on the key length for a secure encryption of quantum states. Our lower bound is asymptotically better than the previous best result.

Section 1 introduces notions and terms, describe the problem and study the previous results. Section 2 shows that to construct an ϵ-randomizing map, the key set should be well-distributed. Section 3 introduce a visual tool to help us understand the property of key set. Section 4 proves the lower bound of key length. Section 5 compares our result to the previous results.

3.1 Introduction and previous results

When two parties exchange a secure information, they wish to encrypt the information such a way that a third-party obtains almost no information from an eavesdropping. The randomization of quantum states is such an encryption that, without access to the key, an eavesdropper is unable to distinguish the encrypted state from the completely random state.

It has been known that applying random Pauli operators to each qubit of a quantum state maps the state to the completely random state. More precisely,

$$\frac{1}{2^{2n}} \sum_{(a,b)\in\{0,1\}^n\times\{0,1\}^n} X^a Z^b \rho Z^b X^a = \frac{I}{2^n},$$

for n-qubit quantum states ρ, where the first entry a and the second entry b are n-digit binary numbers. Also $X^a = X^{a_n} \otimes \cdots \otimes X^{a_1}$ and $Z^a = Z^{b_n} \otimes \cdots \otimes Z^{b_1}$ when $a = a_n a_{n-1} \cdots a_1$, $b = b_n b_{n-1} \cdots b_1$, and $a_i, b_i \in \{0, 1\}$.

An encryption scheme utilizes the above fact. The sender and the receiver choose a random $2n$-bit key, and share the key before the encryption. For the quantum information in a quantum state ρ, the sender encrypts the state ρ to $X^a Z^b \rho Z^b X^a$, and the receiver decrypts the encrypted state to ρ by applying $Z^b X^a$ and $X^a Z^b$ on the left and right hand side of it. For a third party with no access to the shared key, the encrypted state appears as $\frac{I}{2^n}$ - an average state over the random choice of keys.

We call the above scheme the *perfect encryption*, since an eavesdropper gains no information about the initial state ρ. For the perfect encryption, the sender and the receiver should share a $2n$-bit key [8, 2].

A map R is an ϵ-*randomizing map*, when the trace distance between $R(\rho)$ and $\frac{I}{2^n}$ is at most ϵ, for any n-qubit quantum states ρ as

$$\left\| R(\rho) - \frac{I}{2^n} \right\|_{tr} \leq \epsilon.$$

$||M||_{tr}$, the trace norm of matrix M is defined as $Tr\sqrt{M^+M}$. Equivalently, it is the sum of the singular values of M. We are interested in the ϵ-randomizing map constructed with Pauli operators as

$$R(\rho) = \frac{1}{|S|} \sum_{(a,b)\in S} X^a Z^b \rho Z^b X^a, \tag{3.1}$$

where $S \subset \{0,1\}^n \times \{0,1\}^n$. Note that if $S = \{0,1\}^n \times \{0,1\}^n$, then $R(\rho) = \frac{I}{2^n}$ for all ρ.

An ϵ-randomizing map performs a *near-perfect encryption.* The near-perfect encryption is the randomization of quantum states to almost random states - states that are very close to the completely random state. A key (a, b) is chosen randomly from the key set S with equal probability, and encryption and decryption processes are the same as the perfect encryption. A quantum state of distance ϵ from the completely random state can be distinguished from it with probability at most ϵ. Hence, the near-perfect encryption sacrifices ϵ amount of security. However, it is known that the near-perfect encryption has a significantly smaller key length compared to the perfect encryption. [13, 3, 10]

If the ϵ-randomizing map in Equation (3.1) can be constructed with a key set S that is a strict subset of $\{0,1\}^n \times \{0,1\}^n$, then the $2n$-bit key length can be reduced to $log_2|S|$-bit. For n-qubit quantum states, it was proved that ϵ-randomizing map can be constructed with $n+log\, n+2log\, (1/\epsilon)+O(1)$ key length [13]. Subsequently, an efficient (quadratic time) scheme with $n+min\{2log\, n+2log\, (1/\epsilon), log\, n+3log\, (1/\epsilon)\}+O(1)$ key length was given [3]. Also in [10], they reduced the key length for the construction to $n + 2log\ (1/\epsilon) + O(1)$, and proved that at least $min\{2n, log\ n + 2log\ (1/\epsilon) - log\ log\ (1/\epsilon)\} + O(1)$ key length is required for the construction of an ϵ-randomizing map. Hence, there has been a gap between the upper bound and the lower bound on the key length for the optimal construction.

In this chapter, we prove a lower bound on the number of the key length needed

to construct the ϵ-randomizing map as

$$min\{2n, n + log_3 2 \ log \ (1/\epsilon)\} + O(1) \approx min\{2n, n + 0.631 log \ (1/\epsilon)\} + O(1),$$

which is an improvement over the previous results and matches the upper bound of the optimal construction in the order of n.

3.2 Distribution of keys over $\{0,1\}^n \times \{0,1\}^n$

From now on, we consider an ϵ-randomizing map R in the format of

$$R(\rho) = \frac{1}{|S|} \sum_{(a,b) \in S} X^a Z^b \rho Z^b X^a.$$

where $S \subset \{0,1\}^n \times \{0,1\}^n$.

Since the key set S determines the map, there are properties that the set S has to satisfy for the map R to be an ϵ-randomizing map. We investigate the properties of key set S in this section. We show that the keys $(a, b) \in S$ are well-distributed over the first entry a, and this property is preserved over many permutations on the set S. For this purpose, we input various n-qubit quantum states into the map R. Since any quantum state should be mapped to an almost random state, each of input gives a condition on the map R, equivalently a condition on the key set S. The number of conditions is as many as the number of input states.

3.2.1 The base case

Let's consider the following input state,

$$\psi = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

For this state, $R(\psi)$ is diagonal, and the trace distance with the completely random state is easily computed.

Since applying Z^b on ψ doesn't change the state, $X^a Z^b \psi Z^b X^a = X^a \psi X^a$. $X^a \psi X^a$ is a matrix with 1 in (a, a) position, and 0 otherwise. Hence, $R(\psi)$ is a diagonal matrix with $\frac{l}{|S|}$ in the position of (k, k), if set S has l elements with the first entry equal to k. Therefore,

$$\begin{aligned} \left|\left| R(\psi) - \frac{I}{2^n} \right|\right|_{tr} &= \left|\left| \frac{1}{|S|} \sum_{(a,b)\in S} X^a Z^b \psi Z^b X^a - \frac{I}{2^n} \right|\right|_{tr} \\ &= \left|\left| \frac{1}{|S|} \sum_{(a,b)\in S} X^a \psi X^a - \frac{I}{2^n} \right|\right|_{tr} \\ &= \sum_{a\in\{0,1\}^n} \left| \sum_{(a,b)\in S} \frac{1}{|S|} - \frac{1}{2^n} \right|. \end{aligned}$$

Since R is ϵ-randomizing map,

$$\sum_{a\in\{0,1\}^n} \left| \sum_{(a,b)\in S} \frac{1}{|S|} - \frac{1}{2^n} \right| \leq \epsilon \tag{3.2}$$

The implication of the above inequality is that the number of elements (a, b) for each first entry a is close to $\frac{|S|}{2^n}$ for all $a \in \{0, 1\}^n$. Hence, the elements of S are evenly distributed over the first key entry a.

If there are less than $2^n(1-\frac{\epsilon}{2})$ elements in S, then the best-distributed set S still makes the distance larger than ϵ. Hence, for Equation (3.2) to be satisfied,

$$|S| \geq 2^n \left(1 - \frac{\epsilon}{2}\right). \tag{3.3}$$

As ϵ converges to 0, a good lower bound should converge to 2^{2n}, since the near-perfect encryption converges to the perfect encryption. In this sense, the above lower bound in Equation (3.2) is not tight for small ϵ.

Note that the above lower bound is weak because it is derived from only one condition on the key set S. Later in this chapter, we input 3^n quantum states to the map, and 3^n conditions appear. Hence, the tighter lower bound is derived.

3.2.2 The general case

We input 3^n different states - that are variants of ψ - to the map R. Then we have 3^n different conditions. Combining these conditions, we get a stronger lower bound than Equation (3.3).

For this purpose, consider n-qubit quantum states ψ_t, variants of ψ. For each n-digit ternary(base-3) number $t = t_n \cdots t_1, t_i \in \{0, 1, 2\}$

, we define ψ_t as follows:

$$\psi_t = (M_1 \otimes \cdots \otimes M_n)\psi(M_1 \otimes \cdots \otimes M_n)^+,$$

where $M_i = \begin{cases} I & \text{if } t_i = 0 \\ H & \text{if } t_i = 1 \\ G & \text{if } t_i = 2 \end{cases}$, $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and $G = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$. Hence, for 3^n possible number t, there are 3^n different quantum states ψ_t.

The following theorem justifies the above selection of input states ψ_t.

Theorem 4. *For each ψ_t, there exists a corresponding permutation $P_t : \{0,1\}^n \times$*

$\{0,1\}^n \to \{0,1\}^n \times \{0,1\}^n$ *such that*

$$\left|\left|R(\psi_t) - \frac{I}{2^n}\right|\right|_{tr} = \sum_{a\in\{0,1\}^n} \left| \sum_{(a,b)\in P_t(S)} \frac{1}{|S|} - \frac{1}{2^n} \right|,$$

where $P_t(S)$ is the resulting set when P_t is applied on every elements in S.

Proof. We introduce a trick to simplify the computation of $\left|\left|R(\psi_t) - \frac{I}{2^n}\right|\right|_{tr}$. We multiply $(M_1 \otimes \cdots \otimes M_n)$ and its conjugate on the right and left side of $R(\psi_t) - \frac{I}{2^n}$. Since the multiplication only changes the eigenvectors of the matrix, not the eigenvalues, the norm of the matrix will be preserved. Hence,

$$\begin{aligned}
&\left|\left|R(\psi_t) - \frac{I}{2^n}\right|\right|_{tr} \\
&= \left|\left|\frac{1}{|S|}\sum_{(a,b)\in S} X^a Z^b \psi_t Z^b X^a - \frac{I}{2^n}\right|\right|_{tr} \\
&= \left|\left|(M_1 \otimes \cdots \otimes M_n)^+ \left(\frac{1}{|S|}\sum_{(a,b)\in S} X^a Z^b \psi_t Z^b X^a - \frac{I}{2^n}\right)(M_1 \otimes \cdots \otimes M_n)\right|\right|_{tr} \\
&= \left|\left|\frac{1}{|S|}\sum_{(a,b)\in S} (M_1 \otimes \cdots \otimes M_n)^+ X^a Z^b \psi_t Z^b X^a (M_1 \otimes \cdots \otimes M_n) - \frac{I}{2^n}\right|\right|_{tr} \\
&= \left|\left|\frac{1}{|S|}\sum_{(a,b)\in S} (M_1 \otimes \cdots \otimes M_n)^+ (X^{a_1} \otimes \cdots \otimes X^{a_n})(Z^{b_1} \otimes \cdots \otimes Z^{b_n})(M_1 \otimes \cdots \otimes M_n)\right.\right. \\
&\left.\left.\psi (M_1 \otimes \cdots \otimes M_n)^+ (Z^{b_1} \otimes \cdots \otimes Z^{b_n})(X^{a_1} \otimes \cdots \otimes X^{a_n})(M_1 \otimes \cdots \otimes M_n) - \frac{I}{2^n}\right|\right|_{tr} \\
&= \left|\left|\frac{1}{|S|}\sum_{(a,b)\in S} (M_1^+ X^{a_1} Z^{b_1} M_1) \otimes \cdots \otimes (M_n^+ X^{a_n} Z^{b_n} M_n)\psi\right.\right. \\
&\left.\left.(M_1^+ Z^{b_1} X^{a_1} M_1) \otimes \cdots \otimes (M_n^+ Z^{b_n} X^{a_n} M_n) - \frac{I}{2^n}\right|\right|_{tr}.
\end{aligned}$$

Note that

$$M_i^+ X^{a_i} Z^{b_i} M_i = \alpha X^{\tilde{a}_i} Z^{\tilde{b}_i},$$

for some $\tilde{a}_i, \tilde{b}_i \in \{0,1\}$ and a constant α. Table 3.1 gives the complete determination of $M_i^+ X^{a_i} Z^{b_i} M_i$, for different M_i, a_i, and b_i. For a fixed M_i (or fixed t_i), the map from (a_i, b_i) to $(\tilde{a}_i, \tilde{b}_i)$ is a permutation within $\{0,1\} \times \{0,1\}$.

t_i	M_i	(a_i, b_i)	$M_i^+ X^{a_i} Z^{b_i} M_i$	$(\tilde{a}_i, \tilde{b}_i)$
0	I	(0,0)	I	(0,0)
		(0,1)	Z	(0,1)
		(1,0)	X	(1,0)
		(1,1)	XZ	(1,1)
1	H	(0,0)	I	(0,0)
		(0,1)	X	(1,0)
		(1,0)	Z	(0,1)
		(1,1)	$-XZ$	(1,1)
2	G	(0,0)	I	(0,0)
		(0,1)	$-iXZ$	(1,1)
		(1,0)	X	(1,0)
		(1,1)	$-iZ$	(0,1)

Table 3.1: $M_i^+ X_{a_i} Z^{b_i} M_i = \alpha X^{\tilde{a}_i} Z^{\tilde{b}_i}$

Let $\tilde{a}$ and $\tilde{b}$ n-digit binary numbers as $\tilde{a} = \tilde{a}_n \cdots \tilde{a}_1$ and $\tilde{b} = \tilde{b}_n \cdots \tilde{b}_1$. Then

$$\begin{aligned}
&\left\| R(\psi_t) - \frac{I}{2^n} \right\|_{tr} \\
&= \left\| \frac{1}{|S|} \sum_{(a,b)\in S} X^{\tilde{a}_1} Z^{\tilde{b}_1} \otimes \cdots \otimes X^{\tilde{a}_n} Z^{\tilde{b}_n} \psi_t X^{\tilde{a}_1} Z^{\tilde{b}_1} \otimes \cdots \otimes X^{\tilde{a}_n} Z^{\tilde{b}_n} - \frac{I}{2^n} \right\|_{tr} \\
&= \left\| \frac{1}{|S|} \sum_{(a,b)\in S} X^{\tilde{a}} Z^{\tilde{b}} \psi_t Z^{\tilde{b}} X^{\tilde{a}} - \frac{I}{2^n} \right\|_{tr} .
\end{aligned}$$

Since, for a fixed t_i, the map from (a_i, b_i) to $(\tilde{a}_i, \tilde{b}_i)$ is a permutation, for a fixed $t = t_n \cdots t_1$, the map from (a, b) to $(\tilde{a}, \tilde{b})$ is a permutation. Name this permutation as $P_t : (a, b) \to (\tilde{a}, \tilde{b})$. Then,

$$\begin{aligned}
\left\| R(\psi_t) - \frac{I}{2^n} \right\|_{tr} &= \left\| \frac{1}{|S|} \sum_{(\tilde{a},\tilde{b})\in P_t(S)} X^{\tilde{a}} Z^{\tilde{b}} \psi_t Z^{\tilde{b}} X^{\tilde{a}} - \frac{I}{2^n} \right\|_{tr} \\
&= \sum_{\tilde{a}\in\{0,1\}^n} \left| \sum_{(\tilde{a},\tilde{b})\in P_t(S)} \frac{1}{|S|} - \frac{1}{2^n} \right| .
\end{aligned}$$

□

P_t is a permutation, and the size of set is invariant over P_t, as $|P_t(S)| = |S|$. Since R is an ϵ-randomizing map, for each of n-digit ternary number t,

$$\sum_{a\in\{0,1\}^n} \left| \sum_{(a,b)\in P_t(S)} \frac{1}{|S|} - \frac{1}{2^n} \right| \leq \epsilon. \tag{3.4}$$

Hence, each condition states that the elements in $P_t(S)$ are evenly distributed over the first key entry. The above 3^n inequalities relates to our final lower bound on $|S|$.

3.3 Visual representation

3.3.1 Key set

Let's think of an array of 2^n-by-2^n empty boxes. For a key set $S \in \{0,1\}^n \times \{0,1\}^n$, we mark the box in row a and column b if (a,b) is an element of S. Then the array is a visual representation of the key set S.

		×	
	×		×
×	×		
		×	

Figure 3-1: An example of visual representation when $n = 2$ and $S = \{(00,10),(01,01),(01,11),(10,01),(10,01),(11,10)\}$. The corresponding positions for elements in S are marked with $\times$.

3.3.2 Permutation P_t

When $t = 000\cdots0$, P_t is an identity permutation.

When $t = 100\cdots0$, P_t corresponds to the swap of the upper right quadrant and the lower left quadrant of the array. In other words, we mark the locations of elements in S in the array, then cut the upper right and lower left quadrants and switch them. The markings in these quadrants correspondingly moves their position. The new positions of markings corresponds to the elements in $P_t(S)$. From Table 3.1, $HX^{a_i}Z^{b_i}H = \alpha X^{\tilde{a}_i}Z^{\tilde{b}_i}$ makes the map from (a_i,b_i) to $(\tilde{a}_i,\tilde{b}_i)$ be a swap between $(0,1)$ and $(1,0)$. The permutation P_t is such a swap on the first digit of keys.

When $t = 0100\cdots0$, P_t is the swap between the upper right and the lower left. But this time, the swap is performed within each quadrant. We divide each quadrant again to four equal smaller divisions and swap the upper right and lower left divisions within each of quadrant. Similarly when t has 0 as its digits except 1 in mth digit from the left, then P_t corresponds to the swap between the upper right and lower left

in each of 2^{m-1}-by-2^{m-1} divisions.

When $t = 200\cdots0$, P_t is a swap between the upper right quadrant and the lower right quadrant. Similarly, when t has 0 as its digits except 2 in mth digit from the left, then P_t corresponds to the swap between the upper right and lower right in each of 2^{m-1}-by-2^{m-1} divisions. From Table 3.1, we can see that $G^{+}X^{a_i}Z^{b_i}G = \alpha X^{\tilde{a}_i}Z^{\tilde{b}_i}$ makes the map from (a_i, b_i) to $(\tilde{a}_i, \tilde{b}_i)$ be a swap between $(0,1)$ and $(1,1)$, the permutation is such swap on the digits of keys.

$P_{t_n\cdots t_1}$ is obtained by a sequence of permutations $P_{t_n0\cdots0}, P_{0t_{n-1}0\cdots0}, \cdots, P_{00\cdots t_1}$. Figure 3.2 shows an example of $P_{2100\cdots0}$.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

$\xrightarrow{P_{200\cdots0}}$

1	2	11	12
5	6	15	16
9	10	3	4
13	14	7	8

$\xrightarrow{P_{0100\cdots0}}$

1	5	11	15
2	6	12	16
9	13	3	7
10	14	4	8

Figure 3-2: Visual representation of $P_{2100\cdots0}$. The map is swapping 16 divisions of the array. The divisions are named from 1 to 16. $P_{200\cdots0}$ is applied first as it swaps the upper right quadrant and the lower right quadrant. Then, with $P_{0100\cdots0}$ each quadrant is divided to four smaller divisions and the upper right division is swapped with the lower left division within each of quadrants.

3.3.3 Distribution of the key set

We give a little bit of twist to the array defined above. Instead of simply marking the element's position, we write the assigned probability of the element to the position. Since we draw each key with the equal probability and there are $|S|$ elements in the key set, the probabilities are equal as $\frac{1}{|S|}$. Name this array $\mathcal{A}$. We also name the array for $P_t(S)$ as $\mathcal{A}^t$. Note that $\mathcal{A}^t$ can be obtained from dividing and swapping from $\mathcal{A}$ as introduced in the previous subsection. We also introduce an array of 2^n-by-2^n size, with $\frac{1}{2^{2n}}$ in all of the positions. We name this array as $\frac{\mathcal{I}_{2^n}}{2^{2n}}$ as $k \cdot \mathcal{I}_l$ is an l-by-l array with k in every position.

The terms between the pair of vertical lines in Equation (3.2) is equivalent to the

difference between a row sum of array $\mathcal{A}$ and a row sum of array $\frac{\mathcal{I}_{2^n}}{2^{2n}}$. Hence the left-hand side of the equation is absolute sum of this value over all the rows.

For the same size array $\mathcal{B}$ and $\mathcal{C}$, define $D_r(\mathcal{B}, \mathcal{C})$ as row-wise distance between two arrays as

$$D_r(\mathcal{B}, \mathcal{C}) = \sum_i \left| \left(\sum_j \mathcal{B}_{ij} \right) - \left(\sum_j \mathcal{C}_{ij} \right) \right|,$$

where B_{ij} and C_{ij} are the number in the row i and column j in array B and C. Hence Equation (3.2) is equivalent to

$$D_r\left(\mathcal{A}, \frac{\mathcal{I}_{2^n}}{2^{2n}}\right) \leq \epsilon.$$

Also Equation (3.4) is equivalent to

$$D_r\left(\mathcal{A}^t, \frac{\mathcal{I}_{2^n}}{2^{2n}}\right) \leq \epsilon. \tag{3.5}$$

3.4 The lower bound

In this section, we prove a lower bound on the size of the key set S using Equation (3.4). The equations state that the elements in key set S is well-distributed even after the key set is permuted by many different permutations. In other words, the elements are well-distributed over $\{0,1\}^n \times \{0,1\}^n$ in various directions. Hence, the distribution of elements in S shows no obvious pattern, like a random distribution. Exhibiting random-like distribution is difficult with a small density. Therefore a lower bound is derived.

We show that at least $\Theta\left(min\{2^{2n}, \frac{2^n}{\epsilon^{log_2 3}}\}\right)$ number of elements are required. With this, we prove the lower bound on the key length of $min\{2n, n+log_2 3\ log(1/\epsilon)\}+O(1)$.

3.4.1 Row-wise distance on subarrays

The shapes of arrays $\mathcal{A}_{0\cdots0}$, $\mathcal{A}_{10\cdots0}$, and $\mathcal{A}_{20\cdots0}$ are as $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}$, $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{C} \\ \hline \mathcal{B} & \mathcal{D} \\ \hline \end{array}$, and $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{D} \\ \hline \mathcal{C} & \mathcal{B} \\ \hline \end{array}$. Since above three arrays are close to the array $\frac{\mathcal{I}_{2^{2n}}}{2^{2n}}$ in terms of row-wise distance, it suggests that all four subarrays $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$ are close to the array $\frac{\mathcal{I}_{2^{2n-1}}}{2^{2n}}$ in terms of row-wise distance. Similarly we can divide each subarray $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$ to four subarrays and apply a similar argument on them. The following lemma links the distance between larger arrays to the distance between smaller arrays.

Lemma 5. *For $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$, subarrays of S, of size 2^m-by-2^m,*

$$(i)\ D_r\left(\mathcal{A}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{B}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{C}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{D}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right)$$
$$\leq D_r\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}, \frac{\mathcal{I}_{2^{m+1}}}{2^{2n}}\right) + D_r\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{C} \\ \hline \mathcal{B} & \mathcal{D} \\ \hline \end{array}, \frac{\mathcal{I}_{2^{m+1}}}{2^{2n}}\right) + D_r\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{D} \\ \hline \mathcal{C} & \mathcal{B} \\ \hline \end{array}, \frac{\mathcal{I}_{2^{m+1}}}{2^{2n}}\right).$$

$$(ii)\ D_r\left(\mathcal{A}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{B}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{C}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right) + D_r\left(\mathcal{D}, \frac{\mathcal{I}_{2^m}}{2^{2n}}\right)$$
$$\leq \frac{1}{2}D_r\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}, \frac{\mathcal{I}_{2^{m+1}}}{2^{2n}}\right) + \frac{1}{2}Sum\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}\right) + 2 \cdot \frac{2^{2m}}{2^{2n}},$$

where $Sum\left(\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}\right)$ *is the sum of all the numbers in the array* $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}$.

Proof. (i) Let $\mathcal{A}_i$,$\mathcal{B}_i$,$\mathcal{C}_i$, and $\mathcal{D}_i$ be the sum of the numbers at ith row of $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and

$\mathcal{D}$, then the equation above is equivalent to

$$\begin{aligned}
&\sum_{1\le i\le 2^m}\left|\mathcal{A}_i-\frac{2^m}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{B}_i-\frac{2^m}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{C}_i-\frac{2^m}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{D}_i-\frac{2^m}{2^{2n}}\right|\\
&\le\sum_{1\le i\le 2^m}\left|\mathcal{A}_i+\mathcal{B}_i-\frac{2^{m+1}}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{C}_i+\mathcal{D}_i-\frac{2^{m+1}}{2^{2n}}\right|\\
&+\sum_{1\le i\le 2^m}\left|\mathcal{A}_i+\mathcal{C}_i-\frac{2^{m+1}}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{B}_i+\mathcal{D}_i-\frac{2^{m+1}}{2^{2n}}\right|\\
&+\sum_{1\le i\le 2^m}\left|\mathcal{A}_i+\mathcal{D}_i-\frac{2^{m+1}}{2^{2n}}\right|+\sum_{1\le i\le 2^m}\left|\mathcal{B}_i+\mathcal{C}_i-\frac{2^{m+1}}{2^{2n}}\right|.
\end{aligned}$$

Hence, it is enough to show that for each i,

$$\begin{aligned}
&\left|\mathcal{A}_i-\frac{2^m}{2^{2n}}\right|+\left|\mathcal{B}_i-\frac{2^m}{2^{2n}}\right|+\left|\mathcal{C}_i-\frac{2^m}{2^{2n}}\right|+\left|\mathcal{D}_i-\frac{2^m}{2^{2n}}\right|\\
&\le\left|\left(\mathcal{A}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{B}_i-\frac{2^m}{2^{2n}}\right)\right|+\left|\left(\mathcal{C}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{D}_i-\frac{2^m}{2^{2n}}\right)\right|\\
&+\left|\left(\mathcal{A}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{C}_i-\frac{2^m}{2^{2n}}\right)\right|+\left|\left(\mathcal{B}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{D}_i-\frac{2^m}{2^{2n}}\right)\right|\\
&+\left|\left(\mathcal{A}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{D}_i-\frac{2^m}{2^{2n}}\right)\right|+\left|\left(\mathcal{C}_i-\frac{2^m}{2^{2n}}\right)+\left(\mathcal{B}_i-\frac{2^m}{2^{2n}}\right)\right|.
\end{aligned}$$

Therefore we only need to prove that for real numbers, a, b, c, and d,

$$|a|+|b|+|c|+|d|\le|a+b|+|a+c|+|a+d|+|b+c|+|b+d|+|c+d|.$$

Without loss of generality, assume that a is nonnegative and has the largest absolute value among a, b, c, and d. Also assume that $a\ge b\ge c\ge d$. We prove the validity of equation for the following four cases:

When $a\ge 0>b$, the equation is equivalent to $0\le a.$, which is straightforward.

When $b\ge 0>c$, the equation is equivalent to $-c-d\le 2a+|b+c|+|b+d|$. Since a has the largest absolute value, the equation holds.

When $c \geq 0 > d$, the equation is equivalent to $-2d \leq 2a+b+c+|b+d|+|c+d|$. Since a has the largest absolute value, the equation holds.

When $d > 0$, the equation is equivalent to $0 \leq 2a+2b+2c+2d$. Since a, b, c, and d are nonnegative, the equation holds.

(ii) Let $\mathcal{A}_i$, $\mathcal{B}_i$, $\mathcal{C}_i$, and $\mathcal{D}_i$ be the sum of numbers at ith rows of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$. The equation above is equivalent to

$$\begin{aligned}
&\sum_i \left|\mathcal{A}_i - \frac{2^m}{2^{2n}}\right| + \sum_i \left|\mathcal{B}_i - \frac{2^m}{2^{2n}}\right| + \sum_i \left|\mathcal{C}_i - \frac{2^m}{2^{2n}}\right| + \sum_i \left|\mathcal{D}_i - \frac{2^m}{2^{2n}}\right| \\
&\leq \frac{1}{2}\sum_i \left|\mathcal{A}_i + \mathcal{B}_i - \frac{2^{m+1}}{2^{2n}}\right| + \frac{1}{2}\sum_i \left|\mathcal{C}_i + \mathcal{D}_i - \frac{2^{m+1}}{2^{2n}}\right| \\
&+ \frac{1}{2}\sum_i (\mathcal{A}_i + \mathcal{B}_i + \mathcal{C}_i + \mathcal{D}_i) + 2 \cdot \frac{2^m}{2^{2n}}.
\end{aligned}$$

Hence, it is enough to prove that for each i,

$$\begin{aligned}
&\left|\mathcal{A}_i - \frac{2^m}{2^{2n}}\right| + \left|\mathcal{B}_i - \frac{2^m}{2^{2n}}\right| \leq \frac{1}{2}\left|\mathcal{A}_i + \mathcal{B}_i - \frac{2^{m+1}}{2^{2n}}\right| + \frac{1}{2}(\mathcal{A}_i + \mathcal{B}_i) + \frac{2^m}{2^{2n}} \text{ and} \\
&\left|\mathcal{C}_i - \frac{2^m}{2^{2n}}\right| + \left|\mathcal{D}_i - \frac{2^m}{2^{2n}}\right| \leq \frac{1}{2}\left|\mathcal{C}_i + \mathcal{D}_i - \frac{2^{m+1}}{2^{2n}}\right| + \frac{1}{2}(\mathcal{C}_i + \mathcal{D}_i) + \frac{2^m}{2^{2n}}.
\end{aligned}$$

Therefore, we only need to prove for real numbers $a, b, c \geq 0$,

$$|a-c| + |b-c| \leq \frac{1}{2}|a+b-2c| + \frac{1}{2}(a+b) + c.$$

Without loss of generality, assume that $a \geq b$,

When $a, b \geq c$, the equation is equivalent to $-2c \leq 0$, which is straightforward.

When $a \geq c > b$, the equation becomes $\frac{1}{2}(a+b-2c) \leq \frac{1}{2}|a+b-2c| + 2b$. Since b is nonnegative, the equation holds.

When $a, b < c$,the equation becomes $-a-b \leq 0$. Since a and b are nonnegative, the equation holds.

$\square$

3.4.2 Partitioning arrays

In this subsection we utilize the lemma in the previous subsection in more direct and organized format. For this purpose, we give each divisions of $\mathcal{A}^t$ a name.

Since the array $\mathcal{A}^t$ is 2^n-by-2^n, we can partition it to four smaller arrays of size 2^{n-1}-by-2^{n-1}. Each one of the four partitions can be partitioned again to four smaller arrays of size 2^{n-2}-by-2^{n-2}. Continue partitioning m times, then array $\mathcal{A}^t$ is partitioned to 4^m arrays of size 2^{n-m}-by-2^{n-m}. Then for these 4^m arrays, name the array located at ith from the top and jth from the left as $\mathcal{A}^t(i,j,m), 1 \leq i,j \leq 2^m$. Note that the array $\mathcal{A}^t(i,j,m)$ can be partitioned again to four arrays $\mathcal{A}^t(\bar{i},\bar{j},m+1)$, where $\bar{i} \in \{2i-1, 2i\}$ and $\bar{j} \in \{2j-1, 2j\}$. Figure 3-3 is an example of such partitioning.

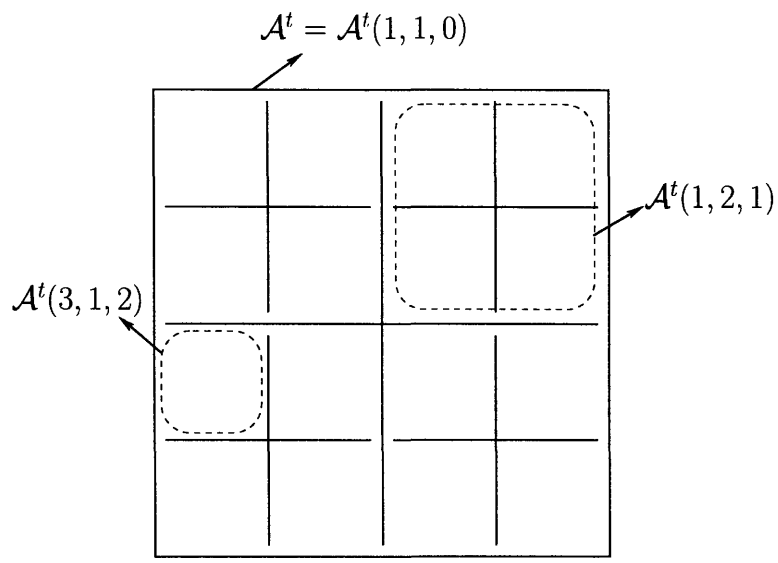

Figure 3-3: Partitioning A^t. When no partition is performed to A^t, $A^t(1,1,0)$ is the only subarray. Hence, $A^t(1,1,0) = A^t$. Also $A^t(1,2,1)$ is the right upper quadrant when A^t is partitioned to four subarrays. Similarly $A^t(3,1,2)$ is a subarray of A^t located third from the top, first from the left, when A^t is partitioned to 16 subarrays of the same size.

For each of arrays $\mathcal{A}^t(i,j,m)$, measure the row-wise distance from $\frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}$. Then

sum up the distance for all $1 \leq i, j \leq 2^m$. We call this value as $V^t(m)$ as follows:

$$V^t(m) = \sum_{1 \leq i,j \leq 2^m} \left\| D_r \left(\mathcal{A}^t(i,j,m), \frac{\mathcal{I}_{2^{n-m}}}{2^{2n}} \right) \right\|.$$

Hence, $V^t(m)$ is the sum of row-wise distances from $\frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}$ for arrays resulting from $\mathcal{A}^t$ when partitioned to subarrays m times. Then from Equation (3.5),

$$V^t(0) = D_r \left(\mathcal{A}^t(1,1,0), \frac{\mathcal{I}_{2^n}}{2^{2n}} \right) = D_r \left(\mathcal{A}^t, \frac{\mathcal{I}_{2^n}}{2^{2n}} \right) \leq \epsilon. \tag{3.6}$$

and

$$\begin{aligned} V^t(n) &= \sum_{1 \leq i,j, \leq 2^n} D_r \left(\mathcal{A}^t(i,j,n), \frac{\mathcal{I}_{2^0}}{2^{2n}} \right) \\ &= \sum_{1 \leq i,j, \leq 2^n} \left| \mathcal{A}^t_{ij} - \frac{1}{2^{2n}} \right| \\ &= \left| \frac{1}{|S|} - \frac{1}{2^{2n}} \right| \cdot |S| + \left| -\frac{1}{2^{2n}} \right| \cdot (2^{2n} - |S|) \\ &= 2 \left(1 - \frac{|S|}{2^{2n}} \right). \end{aligned} \tag{3.7}$$

Hence, an upper bound on $V^t(n)$ gives a lower bound on $|S|$. Note that $V^t(n)$ has the same value regardless of the value of t as $\mathcal{A}^t$'s are rearrangements of $\mathcal{A}$. The following lemma is derived from Lemma 5, and it states the upper bound of $V^t(m+1)$ in terms of $V^t(m)$. By cascading the lemmas, we prove the upper bound on $V^t(n)$, equivalently the lower bound on $|S|$.

Lemma 6. *(i)*

$$\sum_{t \in \{0,1,2\}^n} V^t(m+1) \leq 3 \cdot \sum_{t \in \{0,1,2\}^n} V^t(m).$$

(ii) For each $t \in \{0,1,2\}^n$,

$$V^t(m+1) \leq \frac{1}{2} \cdot V^t(m) + 1.$$

Proof. (i) For a given n-digit ternary number $t = t_n \cdots t_1$ and $1 \le m \le n$, consider three numbers, t^0, t^1, t^2 such that $t^0 = t_n \cdots t_{n-m+2}\ 0\ t_{n-m} \cdots t_1$, $t^1 = t_n \cdots t_{n-m+2}\ 1\ t_{n-m} \cdots t_1$, and $t^2 = t_n \cdots t_{n-m+2}\ 2\ t_{n-m} \cdots t_1$. Hence, t^0, t^1, and t^2 differ from t only by their mth highest digit.

Note that P_{t^0},P_{t^1} and P_{t^2} are permutations within $\{0,1\}^n \times \{0,1\}^n$, and their operation only differ by the action on mth bits of two entries of (a, b). Hence, the corresponding arrays, $\mathcal{A}^{t^0}$,$\mathcal{A}^{t^1}$, and $\mathcal{A}^{t^2}$ differ only by the mth highest digit of the positions. Therefore, $\mathcal{A}^{t^0}(i,j,m)$,$\mathcal{A}^{t^1}(i,j,m)$, and $\mathcal{A}^{t^2}(i,j,m)$ are different only by their highest digit of the positions. Hence, the shape of these three arrays, $\mathcal{A}^{t^0}(i,j,m)$,$\mathcal{A}^{t^1}(i,j,m)$, and $\mathcal{A}^{t^2}(i,j,m)$ are $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \\ \hline \end{array}$, $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{C} \\ \hline \mathcal{B} & \mathcal{D} \\ \hline \end{array}$, and $\begin{array}{|c|c|}\hline \mathcal{A} & \mathcal{D} \\ \hline \mathcal{C} & \mathcal{B} \\ \hline \end{array}$, where, $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$ in the above expression corresponds to $\mathcal{A}^t(2i-1, 2j-1, m+1)$, $\mathcal{A}^t(2i-1, 2j, m+1)$, $\mathcal{A}^t(2i, 2j-1, m+1)$, and $\mathcal{A}^t(2i, 2j, m+1)$. Using the Lemma 5 (i) on these arrays,

$$\begin{aligned}
&D_r\left(\mathcal{A}^t(2i-1,2j-1,m+1), \frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right) + D_r\left(\mathcal{A}^t(2i-1,2j,m+1), \frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right) \\
&+ D_r\left(\mathcal{A}^t(2i,2j-1,m+1), \frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right) + D_r\left(\mathcal{A}^t(2i,2j,m+1), \frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right) \\
&\le D_r\left(\mathcal{A}^{t^0}(i,j,m), \frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}\right) + D_r\left(\mathcal{A}^{t^1}(i,j,m), \frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}\right) + D_r\left(\mathcal{A}^{t^2}(i,j,m), \frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}\right).
\end{aligned}$$

When the above inequality is summed up for all possible $1 \le i, j \le 2^m$,

$$V^t(m+1) \le V^{t^0}(m) + V^{t^1}(m) + V^{t^2}(m).$$

Summing up for all $t \in \{0,1,2\}^n$, we have

$$\sum_{t \in \{0,1,2\}^n} V^t(m+1) \le 3 \cdot \sum_{t \in \{0,1,2\}^n} V^t(m).$$

(ii) Using Lemma 5 (ii) on array $\mathcal{A}^t(i,j,m)$ and its four subarrays,

$$D_r\left(\mathcal{A}^t(2i-1,2j-1,m+1),\frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right)+D_r\left(\mathcal{A}^t(2i-1,2j,m+1),\frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right)$$
$$+D_r\left(\mathcal{A}^t(2i,2j-1,m+1),\frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right)+D_r\left(\mathcal{A}^t(2i,2j,m+1),\frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right)$$
$$\leq\frac{1}{2}D_r\left(\mathcal{A}^t(i,j,m),\frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}\right)+\frac{1}{2}Sum(\mathcal{A}^t(i,j,m))+2\frac{2^{2(n-m-1)}}{2^{2n}}.$$

Summing up above inequality for all $1\leq i,j\leq 2^m$, we get

$$\sum_{1\ \leq i,j\leq 2^{m+1}}D_r\left(\mathcal{A}^t(i,j,m+1),\frac{\mathcal{I}_{2^{n-m-1}}}{2^{2n}}\right)$$
$$\leq\frac{1}{2}\sum_{1\leq i,j\leq 2^m}D_r\left(\mathcal{A}^t(i,j,m),\frac{\mathcal{I}_{2^{n-m}}}{2^{2n}}\right)+\frac{1}{2}Sum(\mathcal{A}^t)+\frac{1}{2}.$$

Since $Sum(\mathcal{A}^t)=1$,

$$V^t(m+1)\leq\frac{1}{2}\cdot V^t(m)+1.$$

□

3.4.3 Proof of lower bound

We use the equations and the lemma in the previous subsection to prove the following lower bound on $|S|$. Once it is proved, equivalently it shows our proposed lower bound on the key length.

Theorem 7. $|S|\geq O\left(min\{2^{2n},\frac{2^n}{\epsilon^{log_3 2}}\}\right)$

Proof. Let $W_m=\sum_{t\in\{0,1,2\}^n}V^t(m)$. Then from Lemma 6,

$$W_{m+1}\leq 3W_m \text{ and} \tag{3.8}$$
$$W_{m+1}\leq\tfrac{1}{2}W_m+3^n. \tag{3.9}$$

Note that from Equation (3.6), (3.7),

$$W_0 \leq 3^n \epsilon \text{ and} \tag{3.10}$$

$$W_n = 2\left(1 - \tfrac{|S|}{2^{2n}}\right) 3^n. \tag{3.11}$$

Let $k = log_3(\frac{2}{5\epsilon})$. Consider the following two cases where $k \geq n$ or $k < n$:

(i) When $k \geq n$, using Equations (3.9) and (3.11),

$$W_n \leq 3^n W_0 \leq 3^{2n}\epsilon \leq 3^{n+k}\epsilon = \left(\frac{2}{5}\right) 3^n.$$

Combine above with Equation (3.11) to get

$$|S| \geq \left(\frac{4}{5}\right) 2^{2n}.$$

(ii) When $k < n$, using Equations (3.9) and (3.11),

$$W_k \leq 3^k W_0 \leq 3^{n+k}\epsilon = \left(\frac{2}{5}\right) 3^n.$$

From Equation (3.9)

$$(W_{m+1} - 2 \cdot 3^n) \leq \frac{1}{2}(W_m - 2 \cdot 3^n).$$

Hence,

$$\begin{aligned} W_n - 2 \cdot 3^n &\leq \left(\frac{1}{2}\right)^{n-k} (W_k - 2 \cdot 3^n) \\ &\leq \left(\frac{1}{2^n}\right) 2^k \left(\frac{2}{5}3^n - 2 \cdot 3^n\right) \\ &= \left(\frac{1}{2^n}\right) 2^{log_3 \frac{2}{5\epsilon}} \left(-\frac{8}{5}\right) 3^n \\ &= \frac{1}{2^n}\left(\frac{2}{5\epsilon}\right)^{\log_3 2} \left(-\frac{8}{5}\right) 3^n. \end{aligned}$$

Plugging above in Equation (3.11), we have

$$|S| \geq \left(\frac{2}{5\epsilon}\right)^{\log_3 2} \left(\frac{4}{5}\right) 2^n.$$

From (i) and (ii),

$$|S| \geq min \left\{ \frac{4}{5} 2^{2n}, \left(\frac{2}{5\epsilon}\right)^{\log_3 2} \left(\frac{4}{5}\right) 2^n \right\}.$$

□

3.5 Discussion

In the order of n, our lower bound $n + O(1)$ is a vast improvement over the previous best lower bound of $log\ n + O(1)$. Also it matches with the previous best upper bound for the optimal construction of $n + O(1)$.

However, for a non-constant ϵ, there is still a gap between lower bound and the upper bound for the optimal construction. Our lower bound is $min\{2n, n + log_2 3\ log(1/\epsilon)\} + O(1)$, which does not match the upper bound of $n + 2log(1/\epsilon) + O(1)$ by $(2 - log_2 3) log(1/\epsilon)$.

Bibliography

[1] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience, New York, 2000.

[2] A. Ambainis, M. Mosca, A. Tapp, and R. de Wolf. Private quantum channels. *IEEE Symposium on Foundation of Computer Science*, pages 547–553, 2000.

[3] A. Ambainis and A. Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. *Proceedings of RANDOM*, pages 249–260, 2004.

[4] H. Barnum, E. Knill, and M.A.Nielsen. On quantum fidelities and channel capacities. *IEEE Trans. Inf. Theory*, 46, 2000.

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895, 1993.

[6] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin. Capacities of quantum erasure channels. *Phys. Rev. Lett.*, 78:3217, 1997.

[7] 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, 1996.

[8] P. O. Boykin and V. Roychowdhury. Optimal encryption of quantum bits. *Physical Review A.*, 67, 2003.

[9] N. J. Cerf and C. Adami. Negative entropy and information in quantum mechanics. *Phys. Rev. Lett.*, 79:5194–5197, 1997.

[10] Paul A. Dickinson and Anshwin Nayak. Approximate randomization of quantum states with fewer bits of key. *a*, a.

[11] M. Grassl, T. Beth, and T. Pellizzari. Code for the quantum erasure channel. *Phys. Rev. A.*, 56:33, 1997.

[12] A. W. Harrow. Coherent communication of classical messages. *Phys. Rev. Lett.*, 92:097902, 2004.

[13] Patrick Hayden, Debbie Leung, Peter W. Shor, and Andreas Winter. Randomizing quantum states: Constructions and applications. *a*, a.

[14] M.A.Nielsen and I.L.Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[15] B. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A.*, 54:2614–2628, 1996.

[16] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Phys. Rev. A*, 54:2629–2635, 1996.