



Quantum Polar Codes

Ashutosh Goswami

► To cite this version:

Ashutosh Goswami. Quantum Polar Codes. Information Theory [math.IT]. Université Grenoble Alpes [2020-..], 2021. English. NNT: 2021GRALM042 . tel-03588305

HAL Id: tel-03588305

<https://theses.hal.science/tel-03588305>

Submitted on 24 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : Mathématiques et Informatique

Arrêté ministériel : 25 mai 2016

Présentée par

Ashutosh GOSWAMI

Thèse dirigée par **Valentin SAVIN**, Chercheur, CEA, LETI, UGA

Codirigée par **Mehdi MHALLA**, Chargé de Recherche, CNRS, UGA

Préparée au sein du **Laboratoire d'Informatique de Grenoble**

dans l'**École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Codes Polaires Quantiques

Quantum Polar Codes

Thèse soutenue publiquement le **25/10/2021**,
devant le jury composé de :

Monsieur Omar FAWZI

Directeur de Recherche, INRIA Centre-Grenoble-Rhone-Alpes,
Rapporteur

Monsieur Joseph RENES

Professeur, Ecole Polytechnique Fédérale de Zurich, Rapporteur

Madame Cécilia LANCIEN

Chargée de Recherche, CNRS Delegation Alpes, Examinatrice

Monsieur Jean-Pierre TILLICH

Directeur de Recherche, INRIA Centre de Paris, Président



Acknowledgement

I thank my supervisors Mehdi Mhalla and Valentin Savin, for their constant support and encouragement. They have taught me a great deal, not just about the scientific part of the thesis but also about writing articles and making presentations. I am grateful for their guidance, without which this thesis would not have been possible.

I thank my co-author, Frédéric Dupuis, for his many interesting ideas and stimulating discussions. It was a pleasure to work with him.

I thank the jury members for their cooperation, carefully reading the manuscript, and providing many valuable comments, typos, and corrections.

Finally, I thank my family for always believing in me and supporting me through thick and thin.

Abstract

In classical information theory, polar codes are the first explicit construction of a family of codes that provably achieve the channel capacity for any discrete memoryless classical channel. In this thesis, we investigate the generalizations of polar codes to the case of quantum channels with qudit-inputs, where the dimension of the qudit quantum system is $d \geq 2$. We start by reviewing the Calderbank-Shor-Steane (CSS) quantum polar coding, which is proposed for qubit quantum systems ($d = 2$). The CSS quantum polar codes utilize the fact that the recursive construction of polar codes using the quantum CNOT gate yields classical polarization in both the amplitude and phase bases. The first important theorem of this thesis proves a “purely quantum polarization” for qudit-input quantum channels, where synthesized virtual channels tend to be either completely noiseless or noisy as quantum channels, and not merely in a basis. The channel combining operation for purely quantum polarization is randomly chosen from a finite set of two-qudit unitaries. Taking advantage of this purely quantum polarization phenomenon, we construct an efficient quantum code, where the completely noisy channels are frozen by half of a preshared EPR pair between the encoder and decoder. Hence, our quantum polar code is entanglement assisted. Further, it achieves a quantum communication rate equal to half the symmetric mutual information, which is the symmetric channel capacity for the entanglement assisted quantum communication. Moreover, by chaining several quantum polar codes, we provide a coding scheme, which uses the preshared EPR pairs as catalyst, so that the rate of preshared entanglement vanishes asymptotically.

Subsequently, we focus on an important family of quantum channels known as qubit Pauli channels. Given a Pauli channel, we associate to it a classical channel with a four symbol set as the input alphabet, and show that the former polarizes quantumly if and only if the latter polarizes classically. Based on the classical counterpart channel, we provide an alternative proof of quantum polarization for the Pauli channel. More importantly, we also provide an efficient way to decode Pauli errors by decoding the polar code on the classical counterpart channel. We also consider a multilevel polarization for Pauli channels, where polarization happens in multi-levels, such that the synthesized virtual channels can also be “half-noisy” instead of being completely noiseless or noisy. This construction does not use randomization of the channel combining operation. Further, we show that half-noisy channels can be frozen in either the amplitude or the phase basis. Hence, multilevel polarization can be effectively used to construct a quantum polar coding scheme.

Finally, we report on our ongoing work on CSS quantum polar codes in the context of fault tolerant quantum computing. We provide fault tolerant procedures for preparing logical encoded quantum states and for error syndrome extraction. Hence, we can protect the encoded logical states for arbitrarily long time in the low noise limit. Further, for quantum information processing, we provide fault tolerant procedures to implement the logical Pauli, the Hadamard, and the CNOT gates. Therefore, the only thing missing to have universal fault tolerant quantum computing with polar codes is a fault tolerant procedure for the implementation of the T gate.

Contents

List of Publications	9
Introduction	12
1 Polar Coding: From Classical to Quantum	16
1.1 Classical Information Theory	16
1.2 Classical Polar Coding	19
1.2.1 Polar Code Construction	22
1.2.2 Encoding	23
1.2.3 Decoding	23
1.2.4 Fast Polarization	26
1.3 Quantum Information Theory	26
1.3.1 Quantum States	27
1.3.2 Unitary Operators	28
1.3.3 Quantum Measurement	29
1.3.4 Quantum Channels	29
1.3.5 Unitary 2-Designs	31
1.3.6 Distinguishing Quantum States	31
1.3.7 Measures of Quantum Information	32
1.3.8 Qubit Quantum Systems	37
1.4 Quantum Channel Coding	39
1.4.1 Quantum Communication without Entanglement Assistance	40
1.4.2 Quantum Communication with Entanglement Assistance	41
1.4.3 Stabilizer and CSS Quantum Codes	41
1.5 Quantum Polar Coding	44
1.5.1 Quantum CSS Polar Code	45
1.5.2 Induced Amplitude and Phase Channels	47
1.5.3 Encoding	48
1.5.4 Decoding	50
2 Purely Quantum Polar codes	54
2.1 A General Set of Conditions for Stochastic Process Polarization	55
2.2 Purely Quantum Polarization	56
2.2.1 Channel Combining and Splitting Procedure	56
2.2.2 Properties of the Channel Combining and Splitting	57
2.2.3 Rényi-Bhattacharyya Parameter	58
2.2.4 Quantum Channel Polarization	59
2.3 A Channel Combining Set	62
2.4 Reduction of the Channel Combining Set	66
2.4.1 A Channel Combining Set with $d^4 + d^2 - 2$ Elements for Qudit Channels	66

2.4.2	A Channel Combining Set with Nine Elements for Qubit Channels	70
2.5	Quantum Polar Coding	71
2.5.1	Quantum Polar Codes	71
2.5.2	Quantum Polar Codes as Entanglement-Assisted Stabilizer Codes	72
2.6	Quantum Polar Codes with Vanishing Rate of Preshared Entanglement	73
3	Purely Quantum Polar Codes for Qubit Pauli Channels	76
3.1	Classical Counterpart of a Pauli channel	76
3.1.1	Classical Mixture of Pauli Channels	76
3.1.2	Classical Counterpart of a CMP channel	77
3.1.3	Channel Combining and Splitting Procedure for the Classical Counterpart Channel	80
3.2	Proof of Quantum Polarization Using the Classical Counterpart Channel	84
3.3	Further Reducing the Channel Combining Set	87
3.3.1	Polarization with a Set Containing Three Two-Qubit Clifford Unitaries	88
3.3.2	Polarization with Only One Two-Qubit Clifford Unitary	88
3.4	Fast Polarization	94
3.5	Decoding the Quantum Polar Code Using its Classical Counterpart	102
3.6	Purely Quantum vs. CSS-based Polarization	103
4	Multilevel Polarization of Pauli Channels	104
4.1	Noiseless, Half-noisy and Noisy channels	105
4.2	Multilevel Polarization	110
4.2.1	Several Inequalities for the Good and Bad Channels	110
4.2.2	Proof of Multilevel Polarization	114
4.3	Quantum Coding Scheme	118
4.3.1	Encoding	118
4.3.2	Decoding	120
4.3.3	Number of Preshared EPR Pairs	120
4.3.4	Fast Polarization	122
4.4	An Alternative Construction	123
4.5	Multilevel Polarization for the Quantum Erasure Channel	124
4.5.1	Bit-level Erasure Channel	124
4.5.2	Bhattacharyya Parameters for the First Construction	125
4.5.3	Bhattacharyya Parameters for the Second construction	126
4.5.4	A Comparison of the Speed of Polarization between the Two Constructions	126
5	Towards Fault Tolerant Quantum Computing using Quantum Polar Codes	129
5.1	Quantum Circuits, Noise Model and Fault tolerant Procedures	131
5.2	Fault Tolerant Syndrome Extraction	132
5.3	Preparation of Encoded Logical States	137
5.3.1	Removing X errors from the polar code state $ 0_L\rangle$	138
5.3.2	Removing Z errors from the polar code state $ 0_L\rangle$	140
5.3.3	Numerical Results	141
5.4	Fault Tolerant Logical Gates on Encoded Quantum States	143
5.4.1	Fault Tolerant Procedure for encoded Pauli X and Z Gates	144
5.4.2	Fault Tolerant Procedure for the encoded CNOT Gate	144
5.4.3	Fault Tolerant Procedure for the encoded Hadamard Gate	144
	Conclusion and Perspectives	147

Appendix	151
Bibliography	155

List of Publications

International Journals

- [A1] Frédéric Dupuis, Ashutosh Goswami, Mehdi Mhalla, and Valentin Savin. Polarization of quantum channels using clifford-based channel combining. *IEEE Transactions on Information Theory* 67.5 (2021), pp. 2857–2877. DOI: [10.1109/TIT.2021.3063093](#). arXiv: [1904.04713](#).
- [A2] Ashutosh Goswami, Mehdi Mhalla, and Valentin Savin. Multilevel polarization for quantum channels. *Quantum Information & Computation*, 21.7 & 8 (2021), pp. 577–606. DOI: [10.26421/QIC21.7-8-4](#). arXiv: [2006.12652](#)

Peer Reviewed International Conferences

- [A3] Frédéric Dupuis, Ashutosh Goswami, Mehdi Mhalla, and Valentin Savin. Purely quantum polar codes. *2019 IEEE Information Theory Workshop (ITW)*, (August 2019), pp. 1–5. DOI: [10.1109/ITW44776.2019.8989387](#).
- [A4] Frédéric Dupuis, Ashutosh Goswami, Mehdi Mhalla, and Valentin Savin. Purely quantum polar codes. *23rd Annual Conference on Quantum Information Processing (QIP 2020)* (January 2020), Shenzhen, China.
- [A5] Ashutosh Goswami, Mehdi Mhalla, and Valentin Savin. Quantum polarization of qudit channels. *2021 IEEE International Symposium on Information Theory (ISIT)*, (July 2021). arXiv: [2101.10194](#)

Poster Presentations

- [1] Purely Quantum Polar Codes, European Quantum Technology Conference (EQTC'19), February 2019, Grenoble, France.
- [2] Purely Quantum Polar Codes, GDR IQFA's 10th Colloquium - IQFA'X, November 2019, Paris, France.
- [3] Quantum polarization of qudit channels, 16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021), July 2021, Riga, Latvia.

Introduction

In 1948, Shannon laid down the foundation of information theory in his seminal paper “A mathematical theory of communication” [1]. One of the aspects of Shannon’s theory deals with the problem of transmitting information over noisy channels, where the sender has access to a noisy channel to transmit information to the receiver, and the goal is to reproduce the transmitted information at the receiver end despite being corrupted by the noisy channel. Shannon’s noisy channel coding theorem states that when one can use a noisy channel arbitrarily many times, it is possible to transmit information reliably using error correcting codes if the rate of transmission is below a limit, referred to as the channel capacity. Further, the channel capacity is given by the mutual information of the channel. The channel coding theorem led to the development of the theory of error correcting codes, which aims at finding effective constructions of codes and decoding algorithms, achieving the channel capacity.

In 2009, 60 years after Shannon’s noisy coding theorem, Arikan proposed polar codes, which is the first explicit construction of a family of codes, achieving the channel capacity with efficient encoding and decoding algorithms [2]. The construction of polar codes is based on the recursive application of a channel combining and splitting procedure. It first combines two instances of a noisy channel, using a controlled-NOT gate as channel combining operation, and then splits the combined channel into two virtual channels, referred to as good and bad channels. When applied recursively n times, the above procedure yields $N = 2^n$ virtual channels. These virtual channels exhibit a polarization property, in the sense that they tend to become either completely noisy or noiseless, as N goes to infinity. A channel parameter known as the Bhattacharyya parameter plays an important role in Arikan’s proof of channel polarization. Polar coding consists of efficient encoding and decoding algorithms that take effective advantage of the channel polarization property.

As information needs to be stored in physical systems, it becomes important to take into account the physics of the underlying system. Shannon’s information theory assumes that information is stored in classical systems, which obey the laws of classical mechanics. Considering quantum systems that obey the laws of quantum mechanics to store and process information, initialized the study of quantum information and computation around the 1970s and 1980s [3, 4, 5, 6, 7]. Quantum information, that is the information represented by the state of a quantum system, is fundamentally different from classical information as it has features, such as, quantum superposition and entanglement, which do not have any classical counterparts. Despite of these fundamental differences, many of Shannon’s ideas have been generalized to quantum information, leading to the development of quantum Shannon theory [8]. Quantum Shannon theory is important from both theoretical and practical perspectives, as it provides operational ways to interpret many measures of quantum information.

The aspect of quantum Shannon theory that we consider in this thesis is the problem of transmitting quantum information reliably through quantum channels. Two scenarios are possible in this regard, quantum communication with or without entanglement assistance. For quantum communication without entanglement assistance, Llyod-Shor-

Devetak (LSD) theorem [9, 10, 11] states that the channel capacity is given by the regularized coherent information. However, this regularized expression is not easy to evaluate for general quantum channels because of superadditivity. Hence, channel capacity is still an open problem for general quantum channels. In the entanglement assisted scenario, channel capacity has been established in [12], and it is easy to compute in this case as it is given by half the quantum mutual information of the quantum channel.

Motivated by the fact that classical polar codes achieve the symmetric capacity of classical channels, we investigate polar coding in the context of quantum communication over quantum channels. The first known construction of quantum polar codes is the Calderbank-Shor-Steane (CSS) quantum polar codes, which basically utilize classical channel polarization in the amplitude and phase bases [13]. One of the main contributions of this thesis is to introduce a new family of purely quantum polar codes, which are entanglement assisted and achieve the symmetric channel capacity for entanglement assisted quantum communication. For purely quantum polar codes, we utilize a polarization phenomenon happening at the quantum level, and not merely in a basis. To prove this polarization, using quantum Rényi entropies, we introduce a new channel parameter that plays the role of the Bhattacharyya parameter from the classical channel polarization. We refer to this channel parameter as *Rényi-Bhattacharyya parameter*. We first provide the purely quantum polar codes for general quantum channels with qudit-input, and further extensively explore them for qubit Pauli channels, for which we also provide an efficient decoding algorithm.

Quantum Shannon theory assumes that quantum operations required for encoding and decoding procedures can be performed perfectly and quantum information is only corrupted during the transmission through noisy channels. In the last chapter of the thesis, we relax this assumption and investigate quantum polar codes in a broader framework of fault tolerant quantum computing.

The thesis is organized as follows.

Chapter 2 (Polar Coding: From Classical to Quantum): The first part of the chapter is devoted to classical polar coding, where we review the construction and decoding of classical polar codes. In the second part, we provide basic definitions in quantum information theory that will be used throughout this thesis. Further, we describe, from a slightly different perspective, the quantum CSS polar codes from [13], whose construction is based on classical polar codes in amplitude and phase bases.

Chapter 3 (Purely Quantum Polar codes): In this chapter, we introduce a purely quantum polarization for quantum channels with qudit input, where polarization happens at the quantum level, not merely in a basis. Our purely quantum polarization is based on a quantum channel combining and splitting procedure, where a two-qudit unitary, randomly chosen from a unitary 2-design is used as a channel combining operation. Under the recursive application of this channel combining and splitting procedure, we show that synthesized virtual channels tend to be either completely noisy or completely noiseless as quantum channels.

Using the fact that the generalized two-qudit Clifford group is a unitary 2-design, we conclude that the channel combining operation can be randomly chosen from this set. Further, we show that purely quantum polarization also happens for a subset of two-qudit generalized Clifford group, which is not a unitary 2-design. We exploit the purely quantum polarization to construct a quantum coding scheme, in which good virtual channels are used for quantum communication, while bad virtual channels are frozen using pre-shared EPR pairs. Hence, our coding scheme is entanglement-assisted, and we show that it achieves half the symmetric mutual information of the channel. Moreover, by chaining

several quantum polar codes, we provide a coding scheme for which the rate of pre-shared entanglement vanishes asymptotically and the resulting quantum code achieves the symmetric coherent information.

This chapter is based on our published works [A1, A3, A5].

Chapter 4 (Purely Quantum Polar Codes for Qubit Pauli Channels): In this chapter, we further investigate the purely quantum polarization for the particular case of qubit Pauli channels. To a Pauli channel, we associate a classical non-binary input symmetric channel, referred as the classical counterpart of the Pauli channel. We show that a Pauli channel polarizes quantumly if and only if its classical counterpart polarizes classically. Finally, we exploit this equivalence to provide an alternative proof of the quantum polarization of a Pauli channel. Based on this equivalence, we further devise an effective method to decode the quantum polar code on a Pauli channel, by decoding its classical counterpart. We also provide a fast polarization property ensuring the reliability of decoding.

This chapter is based on our published work [A1].

Chapter 5 (Multilevel Polarization of Pauli Channels): In this chapter, we investigate purely quantum polarization, using a fixed channel combining operation instead of a randomized one as in Chapter 3. For a fixed channel combining operation, we show that polarization happens in a different way for Pauli channels, where polarization happens in multilevels instead of two levels. In particular, synthesized virtual channels can also be “half-noisy” except being completely noisy or noiseless. The half-noisy channels need to be frozen by fixing their inputs in either the amplitude or the phase basis, while preshared EPR pairs are required for the completely noisy channels as before. This allows reducing the number of preshared EPR pairs compared to Chapter 3.

This chapter is based on our published work [A2].

Chapter 6 (Towards Fault Tolerant Quantum Computing using Quantum Polar Codes): In this chapter, we report on our ongoing work on using CSS quantum polar codes for fault tolerant quantum computing. The reason to choose CSS quantum polar codes for fault tolerance instead of purely quantum polar codes is that entanglement assistance goes to zero for CSS quantum polar codes, if a low noise condition is satisfied. For fault tolerant quantum computing, an efficient quantum code must be accompanied with fault tolerant procedures for preparing encoded (logical) states, operating on encoded states, and extracting information about the error that has happened in the form of an error syndrome. We provide a procedure to prepare encoded logical state that works for finite codelengths if the failure probability of the CNOT gate used in the preparation is sufficiently small. We also provide fault tolerant procedures to implement the logical Pauli, the Hadamard, and the CNOT gates and also a fault tolerant procedure for error syndrome extraction. Therefore, we only need a fault tolerant procedure for the T gate for universal fault tolerant quantum computing.

1

Polar Coding: From Classical to Quantum

The chapter consists of two parts. In the first part (up to Section 1.2.4), we provide basic definitions in classical information theory, and describe the construction and the decoding of Arikan's classical polar codes [2]. In the second part, we provide basic definitions in quantum information theory that shall be used throughout this thesis. Further, we describe, from a slightly different perspective, the quantum CSS polar codes from [13], whose construction is based on two classical polar codes in amplitude and phase bases.

1.1 Classical Information Theory

In this section, we briefly explain the basic concepts in classical information theory and coding to motivate polar coding. The reader may refer to [14] for an extensive review of the field.

One of the aspects of Shannon's information theory is concerned with the reliable transmission of information over noisy channels. Noisy channels affect the transmitted information in an uncontrollable and undesirable way, that may lead to the loss of information at the receiver end. The sender has access to an information source, which produces a message to be transmitted to the receiver. The information source is modeled as a random variable, which selects letters from an alphabet \mathcal{X} , according to some probability distribution. Suppose that the message received at the receiver end contains letters from another alphabet \mathcal{Y} .

The noisy channel W with the input alphabet \mathcal{X} and the output alphabet \mathcal{Y} is defined by a set of transition probabilities,

$$\{W(y|x) | x \in \mathcal{X}, y \in \mathcal{Y}\},$$

where $W(y|x)$ is the conditional probability of y given x . In the above description, the transition probabilities $W(y|x)$ do not depend on previously transmitted symbols, hence, W is called a memoryless channel. In the following, we shall restrict our attention to a binary-input, discrete output, memoryless channel (B-DMC) W from \mathcal{X} to \mathcal{Y} , where $\mathcal{X} = \{0, 1\}$ is the input alphabet, and $\mathcal{Y} := \{y_1, y_2, \dots\}$ is the output alphabet.

Channel coding is a way to transmit information reliably through a noisy channel so that it can be recovered at the receiver end. A channel coding scheme, as depicted in Figure 1.1, consists of two main procedures, namely encoding and decoding.



Figure 1.1: Channel coding

Let \mathbf{u} be a K -bit source message, generated using a random variable \mathbf{U} . The encoding procedure maps \mathbf{u} to a N -bit message \mathbf{x} , where $N \geq K$. Let $\mathcal{E}^c : \{0, 1\}^K \rightarrow \{0, 1\}^N$ be the encoding operation. The *code* \mathcal{C} , generated by \mathcal{E}^c , is defined as,

$$\mathcal{C} := \{\mathcal{E}^c(\mathbf{u}) \mid \mathbf{u} \in \{0, 1\}^K\}. \quad (1.1)$$

Instead of sending $\mathbf{u} \in \{0, 1\}^K$ directly, the encoded N -bit message $\mathbf{x} := \mathcal{E}^c(\mathbf{u}) \in \{0, 1\}^N$ is transmitted using N times the channel W . Suppose $\mathbf{y} \in \mathcal{Y}^N$ is received as the channel output of N instances. The decoding procedure is applied on \mathbf{y} in order to generate an estimate $\hat{\mathbf{u}}$ of the source message. Let $\mathcal{D}^c : \mathcal{Y}^N \rightarrow \{0, 1\}^K$ be the decoding procedure, then the estimate is given by,

$$\hat{\mathbf{u}} := \mathcal{D}^c(\mathbf{y}) \in \{0, 1\}^K. \quad (1.2)$$

The coding scheme is said to be ϵ -reliable if the probability of error is at most ϵ , that is,

$$\Pr(\hat{\mathbf{U}} \neq \mathbf{U}) \leq \epsilon, \quad (1.3)$$

where $\hat{\mathbf{U}}$ is the random variable corresponding to the output $\hat{\mathbf{u}}$ of the decoder.

As $N \geq K$, channel coding improves the reliability of transmission by making use of redundancy. The rate of transmission for the code \mathcal{C} is defined as the number of information bits transmitted per channel use, that is,

$$R := \frac{K}{N} \in [0, 1]. \quad (1.4)$$

A rate R is achievable for W , if for any $\epsilon, \delta > 0$, there exists a ϵ -reliable code with rate $R - \delta$. The *capacity* of a channel W is defined as the tightest upper bound on the achievable transmission rate, that is,

$$C(W) := \sup\{R \mid R \text{ is achievable}\}. \quad (1.5)$$

Shannon's coding theorem [1] states that the capacity $C(W)$ is given by the mutual information of the channel W (see Definition 2 below), that is,

$$C(W) = \mathcal{I}_m(W). \quad (1.6)$$

Since Shannon stated the channel coding theorem, a lot of progress has been made in developing explicit coding schemes, as for instance Reed-Muller codes [15, 16], Reed-Solomon codes [17], Bose–Chaudhuri–Hocquenghem (BCH) codes [18, 19], and algebraic geometric codes [20, 21] etc. to name a few. Another important family of codes to be mentioned is low density parity check (LDPC) codes [22], that are capacity approaching. However, an explicit code, that provably achieves the capacity was not constructed for a long time. Arikan finally resolved this problem by proposing polar codes [2]. Shortly after, another capacity achieving construction based on spatially coupled low density parity check (LDPC) codes has been proposed [23]. In what follows, we shall discuss Arikan's polar codes at length.

Shannon's information theory provides two important measures of information, known as the Shannon entropy and the mutual information, which are defined below.

Definition 1 (Shannon Entropy). *Let X be an information source (random variable), which outputs a letter x from an alphabet \mathcal{X} , with probability $p_X(x)$. The Shannon entropy of the source X is defined as,*

$$H(X) := \sum_{x \in \mathcal{X}} -p_X(x) \log_2 p_X(x) \in [0, 1]. \quad (1.7)$$

The Shannon entropy $H(X)$ measures the information content of X . More precisely, Shannon's source coding theorem states that N independent and identically distributed (i.i.d) random variables X_1, \dots, X_N can be compressed into $NH(X)$ bits as $N \rightarrow \infty$, with very low probability of information loss. Further, this compression is optimal in the sense that information will be lost with very high probability if they are compressed further than $NH(X)$ bits.

Definition 2 (Mutual information). *(a) Let X and Y be the input and the output random variables of the channel W , such that X selects a letter x from an alphabet \mathcal{X} , with probability $p_X(x)$, and Y is related to X by transition probabilities $W(y|x), \forall y \in \mathcal{Y}, x \in \mathcal{X}$. Then, the mutual information between X and Y is defined as,*

$$I(X; Y) := \sum_{y \in \mathcal{Y}} \sum_{x \in \{0,1\}} W(y|x) p_X(x) \log_2 \frac{W(y|x)}{\sum_{x \in \{0,1\}} W(y|x) p_X(x)} \in [0, 1]. \quad (1.8)$$

Alternatively, it can be expressed as follows in terms of the Shannon entropy,

$$I(X; Y) = H(X) + H(Y) - H(X, Y), \quad (1.9)$$

where $H(X)$, $H(Y)$ and $H(X, Y)$ are the Shannon entropies of the input X , the output Y , and the input and output jointly, respectively. Therefore, the mutual information is a correlation measure between the input and output of the channel W

(b) The mutual information of the channel W is defined as the mutual information between X and Y , maximized over the input probability distribution $p_X(x)$, that is,

$$I_m(W) := \sup_{p_X(x)} I(X; Y). \quad (1.10)$$

(c) The symmetric mutual information of W is defined as the mutual information when $p_X(x)$ is the uniform distribution, that is,

$$I(W) := \frac{1}{2} \sum_{y \in \mathcal{Y}} \sum_{x \in \{0,1\}} W(y|x) \log_2 \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)} \in [0, 1]. \quad (1.11)$$

Observe that, when W is a symmetric channel, the symmetric mutual information is equal to the capacity, i.e., $I(W) = C(W)$ [14]. Further, if $I(W)$ goes to 0, W tend to be the completely noisy channel, which completely randomizes the one-bit input. If $I(W)$ goes to 1, W tend to be the completely noiseless channel, which perfectly transmits the input bit. In the sequel, we shall assume that input random variable X is uniformly distributed, hence, we mainly consider the symmetric mutual information.

We now define the Bhattacharyya parameter of the channel W .

Definition 3 (Bhattacharyya Parameter). *The Bhattacharyya parameter of W is defined as,*

$$Z(W) := \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)} \in [0, 1]. \quad (1.12)$$

For uncoded transmission over W , given $y \in \mathcal{Y}$ as the channel output, the maximum a posteriori estimate of the input x is given by,

$$\hat{x} = \operatorname{argmax}_{x=0,1} W(y|x). \quad (1.13)$$

The Bhattacharyya parameter $Z(W)$ gives an upper bound on the error probability of the maximum a posteriori decoding [24], that is,

$$\Pr(\hat{X} \neq X) \leq Z(W), \quad (1.14)$$

where \hat{X} is the random variable corresponding to \hat{x} . Further, it can be seen that when W is an identity channel ($W(y|x)$ is equal to 1 if $x = y$, equal to 0 otherwise), $Z(W)$ is equal to 0. Also, when W is completely randomizing channel ($W(y|0) = W(y|1)$, for all y), $Z(W)$ is equal to 1. Therefore, intuitively, we would expect that $I(W)$ approaches to 0 iff $Z(W)$ approaches to 1, and $I(W)$ approaches to 1 iff $Z(W)$ approaches to 0. This has been made precise in the following lemma.

Lemma 4 ([2]). *For any B-DMC W , we have that*

$$I(W) \geq \log_2 \frac{2}{1 + Z(W)}. \quad (1.15)$$

$$I(W) \leq \sqrt{1 - Z(W)^2}. \quad (1.16)$$

1.2 Classical Polar Coding

Polar codes proposed by Arikan [2] are the first explicit construction of channel codes, endowed with efficient encoding and decoding, provably achieving the capacity of any B-DMC. The construction of polar codes is based on a recursive channel combining and splitting procedure. It first combines two instances of a channel using the controlled-NOT gate as channel combining operation. Then, the combined channel is split into two so-called virtual channels, referred to as the good and the bad channel. Applying the channel combining and splitting procedure recursively n times yields a set of $N = 2^n$ virtual channels. These virtual channels tend to become either completely noisy or completely noiseless, as N goes to infinity, which is known as *channel polarization*. Polar codes are constructed taking advantage of the channel polarization phenomenon. Intuitively, the sender freezes the inputs to the virtual channels that are close to completely noisy to the values known to the receiver, and transmits information bits using only virtual channels that are close to completely noiseless. Moreover, polar codes are equipped with an efficient decoding algorithm, known as successive cancellation (SC) decoding.

In the following, we describe the channel polarization, the code construction, and the decoding of polar codes in more detail. Unless otherwise stated, we shall assume column vectors throughout this thesis and denote them by bold letters such as \mathbf{u}, \mathbf{v} . Hence, for a vector $\mathbf{v} \in \{0, 1\}^N$, we may simply write $\mathbf{v} = (v_0, v_1, \dots, v_{N-1})$, where $v_0, v_1, \dots, v_{N-1} \in \{0, 1\}$.

We first define the channel combining and splitting procedure from [2]. The channel combining of two instances of W yields a channel W_2 as depicted in Figure 1.2.

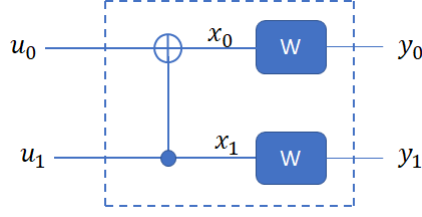


Figure 1.2: Combined channel $W_2(y_0, y_1|u_0, u_1)$. Here, $x_0 = u_0 \oplus u_1$, $x_1 = u_1$.

Given vectors $\mathbf{u} := (u_0, u_1)$ and $\mathbf{x} := (x_0, x_1)$, we have the following equality,

$$\mathbf{x} = P_2 \mathbf{u}, \quad (1.17)$$

where $P_2 := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The transition probability of W_2 is given by,

$$W_2(y_0, y_1|u_0, u_1) = W(y_1|u_0 \oplus u_1)W(y_1|u_1). \quad (1.18)$$

The combined channel is split into two virtual channels $W^{(0)}$ and $W^{(1)}$ as illustrated in Figure 1.3.

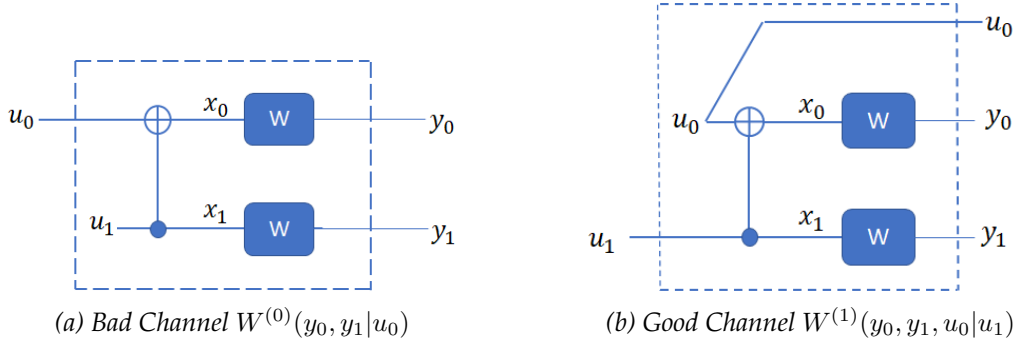


Figure 1.3: Channel splitting

The channel $W^{(0)}$ has input u_0 and output y_0, y_1 , and the channel $W^{(1)}$ has input u_1 and output y_0, y_1, u_0 . Transition probabilities for $W^{(0)}$ and $W^{(1)}$ are given below,

$$W^{(0)}(y_0, y_1|u_0) = \frac{1}{2} \sum_{u_1} W(y_1|u_0 \oplus u_1)W(y_1|u_1). \quad (1.19)$$

$$W^{(1)}(y_0, y_1, u_0|u_1) = \frac{1}{2} W(y_1|u_0 \oplus u_1)W(y_1|u_1). \quad (1.20)$$

Lemma 5 ([2]). We have the following for the symmetric mutual informations of $W^{(0)}$ and $W^{(1)}$,

$$\mathcal{I}(W^{(0)}) + \mathcal{I}(W^{(1)}) = 2\mathcal{I}(W) \quad (1.21)$$

$$\mathcal{I}(W^{(0)}) \leq \mathcal{I}(W) \leq \mathcal{I}(W^{(1)}), \quad (1.22)$$

where $\mathcal{I}(W^{(0)}) = \mathcal{I}(W^{(1)}) = \mathcal{I}(W)$ if and only if $\mathcal{I}(W) \in \{0, 1\}$.

Therefore, the channel combining and splitting procedure redistributes the symmetric mutual information between $W^{(0)}$ and $W^{(1)}$, while preserving the total amount. The channels $W^{(0)}$ and $W^{(1)}$ are called good and bad channel due to (1.22).

Lemma 6 ([2]). *The following relations hold for the Bhattacharyya parameters of $W^{(0)}$ and $W^{(1)}$,*

$$Z(W^{(0)}) \leq 2Z(W) - Z(W)^2, \quad (1.23)$$

$$Z(W^{(1)}) = Z(W)^2, \quad (1.24)$$

where the inequality for the bad channel is an equality if and only if W is an erasure channel.

Applying two times the transform $W \mapsto (W^{(0)}, W^{(1)})$ gives four virtual channels $\{(W^{(i_1)})^{(i_2)} : i_1, i_2 \in \{0, 1\}\}$. Thus, applied recursively n times, we obtain the following 2^n virtual channels,

$$(W^{(i_1 \dots i_n)}) := (W^{(i_1 \dots i_{n-1})})^{(i_n)}, \quad (i_1 \dots i_n) \in \{0, 1\}^n. \quad (1.25)$$

We are now in a position to state Arikan's main polarization theorem.

Theorem 7 ([2]). *For any B-DMC W , let $\{W^{(i_1 \dots i_n)} : (i_1 \dots i_n) \in \{0, 1\}^n\}$ be the set of channels defined in (1.25). Then, for any $\delta > 0$,*

$$\lim_{n \rightarrow \infty} \frac{\#\{(i_1 \dots i_n) \in \{0, 1\}^n : I(W^{(i_1 \dots i_n)}) \in (\delta, 1 - \delta)\}}{2^n} = 0. \quad (1.26)$$

and furthermore,

$$\lim_{n \rightarrow \infty} \frac{\#\{(i_1 \dots i_n) \in \{0, 1\}^n : I(W^{(i_1 \dots i_n)}) \geq 1 - \delta\}}{2^n} = I(W). \quad (1.27)$$

Here, we briefly discuss the main steps of the proof (for details we refer to [2]).

- (a) **I is preserved:** First, the recursive application of the channel combining and splitting procedure can be modeled as a discrete time stochastic process $\{I_n : n \geq 0\}$, with $I_0 = I(W)$ and $I_n = I(W^{(i_1 \dots i_n)})$ for $n > 0$, where i_1, \dots, i_n, \dots are randomly and independently chosen from $\{0, 1\}$. From (1.21), the mutual information is preserved under channel combining and splitting, therefore, it follows that $\{I_n : n \geq 0\}$ is a martingale. Thus, it converges almost everywhere to a random variable I_∞ , whose expectation satisfies $E[I_\infty] = I_0$.
- (b) **I and Z polarize simultaneously:** Equation (1.26) states that I_∞ takes values in $\{0, 1\}$. This can be proven by taking an indirect approach. Indeed, by Lemma 4, it is enough to prove (1.26) while substituting the Bhattacharyya parameter to the mutual information, since they polarize simultaneously.
- (c) **Guaranteed improvement of $Z(W^{(1)})$:** To prove polarization of the Bhattacharyya parameter, the main ingredient is its *guaranteed improvement*, when taking the good channel. Here, guaranteed improvement means that there exists a continuous function $f : [0, 1] \rightarrow [0, 1]$, satisfying $f(0) = 0$, $f(1) = 1$, and $f(z) < z, \forall z \in (0, 1)$, such that $Z(W^{(1)}) \leq f(Z(W))$, for any B-DMC W . In our case, according to (1.24), we may take $f(z) = z^2$. It is worth noticing that for the mutual information, we have $I(W^{(1)}) > I(W)$, for $I(W) \in (0, 1)$, but this inequality is weaker than guaranteed improvement condition. This explains the need of the indirect approach.
- (d) **Expected value argument:** Eventually, (1.27) follows by recognizing that the left hand side limit is equal to $\Pr(I_\infty = 1)$, which is the same as the expectation $E[I_\infty]$. Since I_∞ takes values in $\{0, 1\}$, we have $\Pr(I_\infty = 1) = I_0 = I(W)$, as desired.

1.2.1 Polar Code Construction

We first describe the recursive construction of polar code for $N = 2^n$ copies of W (see Figure 1.6 from right to left). We divide the N copies of W into $\frac{N}{2}$ pairs and then apply the channel combining and splitting procedure on each pair. This gives us $\frac{N}{2}$ copies of both $W^{(0)}$ and $W^{(1)}$. In the next step, for all $i_1 \in \{0, 1\}$, we group together the $\frac{N}{2}$ copies of $W^{(i_1)}$, and divide them into $\frac{N}{4}$ pairs. After applying the channel combining and splitting procedure again on each pair, we get $\frac{N}{4}$ copies of $W^{(i_1 i_2)}$, $\forall i_1, i_2 \in \{0, 1\}$. Similarly, at any level of recursion $k \leq n$, we group copies of a virtual channel $W^{(i_1 \dots i_k)}$ together and divide them in pairs, and subsequently apply the channel combining and splitting procedure on each pair. The recursion stops at $k = n$, as we have only one copy of each virtual channel $W^{(i_1 \dots i_n)}$, $(i_1 \dots i_n) \in \{0, 1\}^n$.

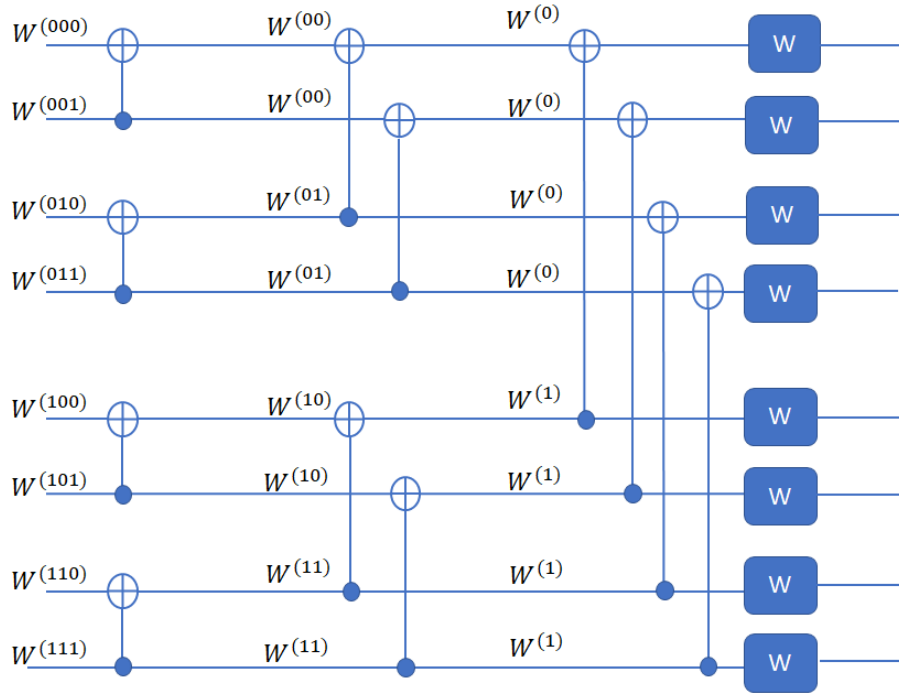


Figure 1.4: Polar code construction for $N = 2^3$.

We shall denote $W^{(i)} := W^{(i_1 \dots i_n)}$, where $(i_1 \dots i_n)$ is the binary representation of $i \in \{0, \dots, N - 1\}$. By the polarization theorem, $I(W^{(i)})$ goes to 0 or 1, as N goes to infinity, except for a vanishing fraction of channels. However, for finite N , it is more difficult to distinguish between good and bad virtual channels, since the polarization is incomplete. Nevertheless, it is always possible to determine the best virtual channels, say according to the value of the Bhattacharyya parameter¹. Thus, we shall fix some subset $\mathcal{I} \subset \{0, \dots, N - 1\}$, referred to as the information set, such that $Z(W^{(i)}) \leq Z(W^{(j)})$, for any $i \in \mathcal{I}$ and $j \in \mathcal{J} := \mathcal{I}^c$. Inputs of the virtual channels $W^{(i)}$, $i \in \mathcal{I}$, are set to information bits, while inputs of $W^{(j)}$, $j \in \mathcal{J}$, are frozen, *i.e.*, set to the values known to both the sender and the receiver (see Section 1.2.2). Hence, the transmission rate is given by $R = \frac{|\mathcal{I}|}{N}$. If N is large enough, the information set \mathcal{I} may be chosen such that $|\mathcal{I}|$ is close to $I(W)N$. For finite N , the size of the information set may be determined according to the

¹One may use either the mutual information or the Bhattacharyya parameter. The latter is generally preferred due to the fact that it can be used to upper-bound the decoding error probability, as discussed in Section 1.2.4.

desired error probability. We will see in Section 1.2.4 that the decoding error probability is upper-bounded by $\sum_{i \in \mathcal{I}} Z(W^{(i)})$. Thus, in practice, \mathcal{I} can be determined such that the above sum is less than some desired probability value.

1.2.2 Encoding

The polar transform after n steps, *i.e.*, the transform obtained by recursively applying the channel combining operation for n levels, as depicted in Figure 1.4, shall be denoted by P_N . In the matrix form, we have that

$$P_N = P_2^{\otimes n}, \quad (1.28)$$

where P_2 is the polar transform for $N = 2$ defined in (1.17).

The input vector $\mathbf{u} := (u_0, \dots, u_{N-1}) \in \{0, 1\}^N$ is composed as follows,

- For an index i belonging to the information set \mathcal{I} , u_i can be randomly chosen from $\{0, 1\}$. The value of u_i is not known to the decoder.
- For an index j belonging to the frozen set \mathcal{J} , u_j is frozen (fixed) and known to both the encoder and decoder. The bit u_j can be fixed to either 0 or 1, only it should be known to the receiver. Hence, we may set $u_j = 0, \forall j \in \mathcal{J}$.

The vector \mathbf{u} is encoded into $\mathbf{x} := (x_0, \dots, x_{N-1}) \in \{0, 1\}^N$ using the classical polar transform P_N . Using (1.28), we may write

$$\mathbf{x} = P_N \mathbf{u}. \quad (1.29)$$

1.2.3 Decoding

Let the vector \mathbf{x} be transmitted through N copies of W and let the channel output vector be $\mathbf{y} := (y_0, \dots, y_{N-1}) \in \mathcal{Y}^N$. The task of decoder is to output estimates \hat{u}_i corresponding to the information bits $u_i, i \in \mathcal{I}$, using the channel output vector \mathbf{y} and the frozen bits $u_j, j \in \mathcal{J}$. Arikan's successive cancellation (SC) decoder generates estimates \hat{u}_i one by one, from $i = 0$ to $i = N - 1$. To do this, observe first that u_i is the input of the virtual channel $W^{(i)}$, whose output consists of the vector \mathbf{y} , together with the inputs to all the previous virtual channels $\mathbf{u}_0^{i-1} := (u_0, \dots, u_{i-1})$. Hence, the maximum a posteriori estimate of u_i is given by

$$\hat{u}_i = \operatorname{argmax}_{u_i=0,1} W^{(i)} \left(\mathbf{y}, \mathbf{u}_0^{i-1} \mid u_i \right) \quad (1.30)$$

Since the inputs to the previous virtual channels are unknown, the SC decoder substitutes the previously estimated inputs $\hat{\mathbf{u}}_0^{i-1}$ to \mathbf{u}_0^{i-1} , in the above equation. It is also more convenient to rewrite the above argmax computation in the form of a log-likelihood-ratio (LLR) computation. Hence, for $i = 0, \dots, N - 1$, we define recursively

$$\lambda_i^{(\mathbf{u})} := \log \frac{W^{(i)} \left(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} \mid u_i = 0 \right)}{W^{(i)} \left(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} \mid u_i = 1 \right)}, \quad (1.31)$$

$$\hat{u}_i := \begin{cases} 0, & \text{if } i \in \mathcal{J}, \\ \frac{1 - \operatorname{sign}(\lambda_i^{(\mathbf{u})})}{2}, & \text{if } i \in \mathcal{I}, \end{cases} \quad (1.32)$$

where $\operatorname{sign}(x) \in \{\pm 1\}$, and $\operatorname{sign}(0)$ is set to either -1 or $+1$ with equal probability.

Now, we explain how the LLR values $\lambda_i^{(u)}$ can be computed in an efficient way. Remember that the actual transmission consists of sending the encoded vector $\mathbf{x} := (x_0, x_1, \dots, x_{N-1})$ using N instances of the W channel, producing the observed channel output $\mathbf{y} := (y_0, y_1, \dots, y_{N-1})$. Hence, one may compute the LLR of each x_i , conditioned on the observed channel output y_i , that is,

$$\lambda_i^{(x)} := \log \frac{W(y_i | x_i = 0)}{W(y_i | x_i = 1)}. \quad (1.33)$$

These LLR values only depend on the W channel and the observed channel output. To compute the LLR values $\lambda_i^{(u)}$ from (1.31), one may actually *propagate* the $\lambda_i^{(x)}$ values through the polar transform (see Figure 1.4), from the right to the left side.

For $N = 2$, the propagation rules are illustrated in Figure 1.5, and given as follows:

$$\begin{aligned} \text{(bad channel)} \quad \lambda_0^{(u)} &= \log \frac{1 + e^{\lambda_0^{(x)} \lambda_1^{(x)}}}{e^{\lambda_0^{(x)}} + e^{\lambda_1^{(x)}}} \end{aligned} \quad (1.34)$$

$$\text{(good channel)} \quad \lambda_1^{(u)} = (-1)^{\hat{u}_0} \lambda_0^{(x)} + \lambda_1^{(x)} \quad (1.35)$$

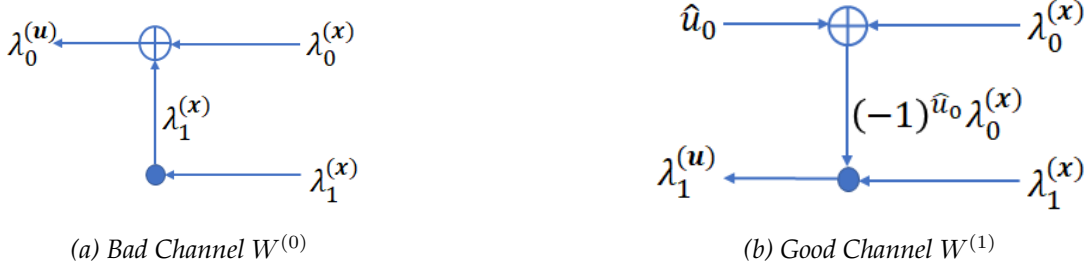


Figure 1.5: Propagation of LLR for $N = 2$.

For $N > 2$, the above propagation rules are used to propagate the LLR values through the different *kernels* (i.e., 2 subblocks) composing the polar transform. When both LLR values at the right hand side of a kernel are known, they may be propagated to the corresponding bad channel. Propagation to the good channel is delayed until the estimate of the bad channel is known. When the propagation reaches some position, say i , on the left hand side of the polar transform, the propagated value is nothing else but the LLR $\lambda_i^{(u)}$ in (1.31). The estimate \hat{u}_i may then be computed by using (1.32) and propagated forward (from left to right) together with previous estimates \hat{u}_0^{i-1} , so as to determine the estimates needed by the delayed propagations.

To illustrate the above procedure, we describe the SC decoding for two steps of polarization, i.e., $N = 2^2$, using the binary tree graph of depth two given in Figure 1.6.

Each node in the graph may store two messages at a given time; a list of LLRs and a list of estimates \hat{u}_i . The list of LLRs are propagated (represented by blue arrows) from the root node (node on the top) to the leaf nodes, in such a way that its size becomes half after each increment in depth. In Figure 1.6, a list of LLRs corresponding to a node (except for the root node) represents the list that has been propagated from its parent node. Therefore, any leaf node receives a list containing only one LLR, based on which the estimate is generated for the corresponding virtual channel using (1.32). The estimates are stored as a list of length 1 at the leaf node, which are propagated towards the root node (represented by orange arrows) in such a way that its size doubles after each decrement in depth. Therefore, the root node must receive a list of estimates of size 4. In Figure 1.6, a list of

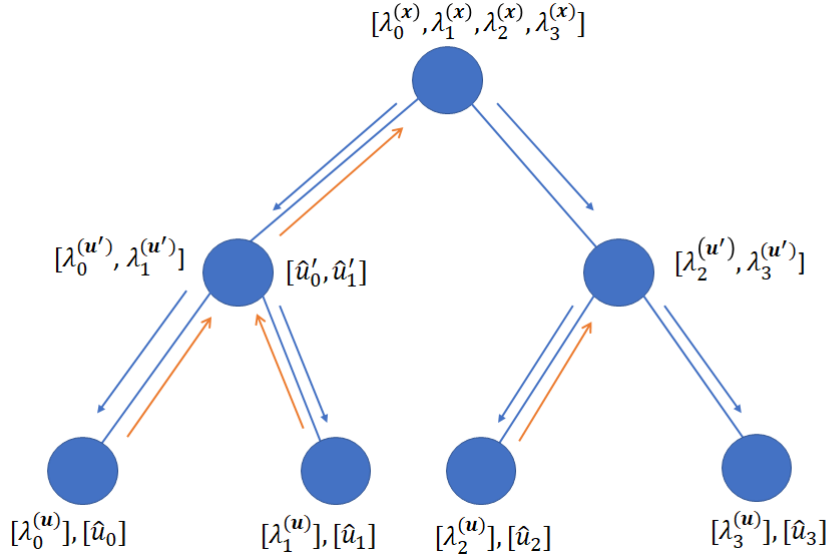


Figure 1.6: Successive cancellation decoder for $N = 2^2$. Here, \mathbf{u}' represents the vector \mathbf{u} after the first step of polarization.

estimates corresponding to a node represents the list of estimates that will be propagated to its parent node. This list is not represented for the root node as it is the last node. The propagation of LLRs on the binary tree graph is done as follows,

1. The computation is initialized by giving the list of LLRs $L = [\lambda_0^{(x)}, \lambda_1^{(x)}, \lambda_2^{(x)}, \lambda_3^{(x)}]$ as input to the root node.
2. We first make two lists $L_1 = [\lambda_0^{(x)}, \lambda_1^{(x)}]$ and $L_2 = [\lambda_2^{(x)}, \lambda_3^{(x)}]$ by dividing the input list down the middle. Then, we pair the elements of L_1 with the corresponding element of L_2 , which gives us two pairs $(\lambda_0^{(x)}, \lambda_2^{(x)})$ and $(\lambda_1^{(x)}, \lambda_3^{(x)})$. Using (1.34), the LLR of the bad channel is computed for each pair, i.e., $\lambda_0^{(u')}$ and $\lambda_1^{(u')}$, where \mathbf{u}' represents the vector \mathbf{u} after the first step of polarization. The list $[\lambda_0^{(u')}, \lambda_1^{(u')}]$ is propagated to the left child of the root node, which is then activated. Meanwhile the root node waits for the list of estimates $[\hat{u}_0', \hat{u}_1']$ to arrive from its left child. Once it receives the estimates, using (1.35), LLRs of the good channels are computed for pairs $(\lambda_0^{(x)}, \lambda_2^{(x)})$ and $(\lambda_1^{(x)}, \lambda_3^{(x)})$, i.e., $\lambda_2^{(u')}$ and $\lambda_3^{(u')}$, respectively. Subsequently, the list $[\lambda_2^{(u')}, \lambda_3^{(u')}]$ is sent to the right child of the root node, which is then activated.
3. Whenever a node apart from one of the leaf nodes is activated after receiving a list of LLRs from its parent node, the above process is repeated. In other words, the received list is divided into two lists down the middle, corresponding elements of two lists are paired together and LLRs of the bad channels are computed for each pair using (1.34) and stored as a list. This list is then propagated to the left child of the node, which is then activated. Meanwhile the node waits for the list of estimates to arrive from its left child in order to calculate the list for the LLRs of the good channels, which are to be sent to its right child.
4. When one of the leaf nodes is reached, the estimate \hat{u}_i is generated using (1.32). If the leaf node is the last one (the rightmost leaf node in the figure), the decoding procedure ends as all the estimates \hat{u}_i has been generated. Otherwise, the generated estimate is sent back to the parent node of the corresponding leaf.

5. When a node has received a list of estimates from both its right and left child, denoted by $\hat{\mathbf{u}}_L, \hat{\mathbf{u}}_R$, it combines them by applying bitwise XOR and makes a new list $\hat{\mathbf{u}}_L \oplus \hat{\mathbf{u}}_R$. The new list $\hat{\mathbf{u}}_L \oplus \hat{\mathbf{u}}_R$ is merged with $\hat{\mathbf{u}}_R$ and sent upward to the parent node. As mentioned before, the waiting parent node, then uses this estimate to compute the list of LLRs for the good channel.

1.2.4 Fast Polarization

Let $P_{\text{err}}(W^{(i)}) := \Pr(\hat{u}_i \neq u_i)$ be the decoding error probability on the i -th virtual channel². As mentioned before, the Bhattacharyya parameter gives an upper bound on the error probability of maximum a posteriori decoding for uncoded transmission through a channel. Therefore,

$$P_{\text{err}}(W^{(i)}) \leq Z(W^{(i)}). \quad (1.36)$$

The SC decoding fails if any one of the virtual channels $W^{(i)}$ for $i \in \mathcal{I}$ has been decoded incorrectly. Hence, we have the following for the block error probability of SC decoding P_{err}^B ,

$$\begin{aligned} P_{\text{err}}^B &= 1 - \prod_{i \in \mathcal{I}} (1 - P_{\text{err}}(W^{(i)})) \\ &\leq \sum_{i \in \mathcal{I}} P_{\text{err}}(W^{(i)}) \\ &\leq \sum_{i \in \mathcal{I}} Z(W^{(i)}). \end{aligned} \quad (1.37)$$

Therefore, in order to ensure that P_{err}^B goes to zero, as N goes to infinity, it is sufficient to prove that the Bhattacharyya parameter scales as $Z(W^{(i)}) = \mathcal{O}(N^{-1-\theta})$, for some $\theta > 0, \forall i \in \mathcal{I}$. This property is called *fast polarization*. The following stronger fast polarization result is given in [25], where it is shown that $Z(W^{(i)}), i \in \mathcal{I}$ scales exponentially with respect to N .

Proposition 8 ([25]). *For any B-DMC W , and for any $\beta < \frac{1}{2}$, following holds,*

$$Z(W^{(i)}) = \mathcal{O}(2^{-N^\beta}), \forall i \in \mathcal{I}. \quad (1.38)$$

Hence, the above proposition ensures the reliability of the SC decoding.

1.3 Quantum Information Theory

In this section, without going into too much detail, we provide some definitions in quantum information theory, that shall be used throughout this thesis. For more comprehensive review of the field of quantum information, the reader may refer to [8, 26].

The following notation is used:

1. Any quantum system is associated with a complex vector space with inner product, referred to as the Hilbert space. We denote by \mathcal{H}_A the Hilbert space associated with a quantum system A .

²Here, in $\Pr(\hat{u}_i \neq u_i)$, \hat{u}_i and u_i are understood as random variables.

2. Throughout this thesis, we consider finite dimensional Hilbert spaces. A quantum system with Hilbert space dimension $d \geq 2$ is referred to as a *qudit*, and for the particular case $d = 2$, it is called a *qubit*.
3. The Hilbert space corresponding to a multipartite system $A_1 \cdots A_N$ is given by the tensor product of the individual Hilbert spaces, *i.e.*, $\mathcal{H}_{A_1 \cdots A_N} = \mathcal{H}_{A_1} \otimes \cdots \otimes \mathcal{H}_{A_N}$.
4. Dirac's bra-ket notation is used to represent a vector, and its dual (transpose of the complex conjugate of the vector). A vector is denoted by $|\phi\rangle_A \in \mathcal{H}_A$ (column vector), and its dual by $\langle\phi|_A$ (row vector). The inner product of any two vectors $|\phi\rangle_A, |\psi\rangle_A \in \mathcal{H}_A$ is denoted by $\langle\psi|\phi\rangle := \langle\psi|_A|\phi\rangle_A$. Therefore, norm of a vector is given by, $|||\psi\rangle|| = \sqrt{\langle\psi|\psi\rangle}$.
5. The set of all linear operators from \mathcal{H}_A to \mathcal{H}_B is denoted by $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$. For the linear operators on \mathcal{H}_A , we simply write $\mathcal{L}(\mathcal{H}_A) := \mathcal{L}(\mathcal{H}_A, \mathcal{H}_A)$. Further, let $\text{Pos}(\mathcal{H}_A) \subset \mathcal{L}(\mathcal{H}_A)$ be the set of all positive semidefinite operators on A .
6. For any operator $L \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$, L^\dagger is defined as the transpose of the complex conjugate of L_A .
7. Throughout this thesis, logarithm is taken in base d , the dimension of the qudit.

We will also need the definitions of the trace and partial trace of a linear operator.

Definition 9 (Trace and partial trace). Let d_A be the dimension of \mathcal{H}_A and let $\{|i\rangle_A \mid i \in \{0, \dots, d_A - 1\}\}$ be an orthonormal basis for \mathcal{H}_A . The trace is a map $\text{Tr} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathbb{C}$, where \mathbb{C} is the set of complex numbers, such that

$$\text{Tr}[L_A] := \sum_{i=0}^{d_A-1} \langle i|_A L_A |i\rangle_A, \quad L_A \in \mathcal{L}(\mathcal{H}_A). \quad (1.39)$$

Consider a bipartite quantum system AB . The partial trace with respect to A is a map $\text{Tr}_A : \mathcal{L}(\mathcal{H}_{AB}) \rightarrow \mathcal{L}(\mathcal{H}_B)$ such that

$$\text{Tr}_A[L_{AB}] := \sum_{i=0}^{d_A-1} (\langle i|_A \otimes I_B) L_{AB} (|i\rangle_A \otimes I_B), \quad L_{AB} \in \mathcal{L}(\mathcal{H}_{AB}). \quad (1.40)$$

It's important to mention that the trace and the partial trace are independent of the chosen basis.

1.3.1 Quantum States

A quantum system is described by its quantum state, which is defined below.

Definition 10 (Quantum states). (a) Quantum state of a system A , also known as density matrix, is given by a positive semidefinite operator $\rho_A \in \text{Pos}(\mathcal{H}_A)$ such that $\text{Tr}[\rho_A] = 1$. We denote by $\mathcal{D}(\mathcal{H}_A) \subset \text{Pos}(\mathcal{H}_A)$ the set of all quantum states.

(b) A quantum state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ is called a pure quantum state if and only if it can be written as $\rho_A = |\psi\rangle_A \langle\psi|_A$, where $|\psi\rangle_A \in \mathcal{H}_A$ is a unit vector, *i.e.*, $|||\psi\rangle|| = 1$. We will often represent a pure quantum state by the corresponding vector $|\psi\rangle_A$.

(c) A mixed quantum state is a convex combination of pure states, *i.e.*, $\rho_A = \sum_i p_i |\psi_i\rangle_A \langle\psi_i|_A$, such that $\sum_i p_i = 1$. The quantum state $\frac{\mathbb{1}_A}{d_A}$, where $\mathbb{1}_A \in \text{Pos}(\mathcal{H}_A)$ is the identity matrix, is called the maximally mixed quantum state.

(d) The support of a quantum state ρ is defined as the set of eigenstates of ρ with non-zero eigenvalues,

$$\text{supp}(\rho) := \{|\psi\rangle \mid \rho|\psi\rangle = \lambda|\psi\rangle, \text{ with } \lambda > 0\}.$$

Given a bipartite quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, one obtains the quantum state of A or B , separately, by doing the partial trace with respect to B or A , respectively. For example, the quantum state corresponding to A is given by, $\rho_A = \text{Tr}_B[\rho_{AB}]$.

One of the peculiar features of quantum systems is quantum entanglement that do not have a parallel in the classical systems. Quantum entanglement refers to correlation between quantum systems, as defined below.

Definition 11 (Entangled quantum states). (a) A pure bipartite quantum state $|\psi\rangle_{AB}$ is called entangled if it is not a product state, that is, $|\psi\rangle_{AB} = |\phi_1\rangle_A \otimes |\phi_2\rangle_B$, for some pure quantum states $|\phi_1\rangle_A$ and $|\phi_2\rangle_B$.

(b) A mixed bipartite state ρ_{AB} is said to be entangled if it can not be written as a convex combination of the product states, that is, $\sum_i p_i \psi_A^i \otimes \phi_B^i$, where $p_i \geq 0$, $\sum_i p_i = 1$ and $\psi_A^i \in \mathcal{D}(\mathcal{H}_A)$, $\phi_B^i \in \mathcal{D}(\mathcal{H}_B)$, $\forall i$.

(c) An Einstein-Podolsky-Rosen (EPR) pair on quantum systems A and A' , each of dimension d_A , is a maximally entangled quantum state defined as, $\Phi_{AA'} := |\Phi\rangle_{AA'} \langle \Phi|_{AA'}$, where $|\Phi\rangle_{AA'} := \frac{1}{\sqrt{d_A}} \sum_i |i\rangle_A |i\rangle_{A'}$.

1.3.2 Unitary Operators

Here, we provide the definition of unitary operators, which are used to describe the dynamics of closed (without any interaction with its environment) quantum systems. More general quantum maps are presented in Section 1.3.4 below.

Definition 12 (Unitary operators). An operator $U_A \in \mathcal{L}(\mathcal{H}_A)$ is said to be unitary if the following holds,

$$U_A^\dagger U_A = U_A U_A^\dagger = I. \quad (1.41)$$

Note that for a unitary operator U_A , we have that $\|U_A|\psi\rangle_A\| = 1$, hence a unitary operator maps a pure quantum state to another pure quantum state. We now provide two important class of unitaries for qudit quantum systems. They are groups in the mathematical sense, and are referred to as the generalized Pauli and Clifford group³.

Definition 13 (Generalized Pauli Group). (a) The generalized Pauli operators X and Z for a qudit quantum system are defined as $X = \sum_{j=0}^{d-1} |j\rangle\langle j \oplus 1|$, and $Z = \sum_{j=0}^{d-1} \omega^j |j\rangle\langle j|$, where \oplus denotes the sum modulo d , and $\omega = e^{\frac{2\pi i}{d}}$.

(b) The generalized Pauli group on one qudit is defined as $\mathcal{P}_d^1 := \{\omega^\lambda P_{r,s} \mid \lambda, r, s = 0, \dots, d-1\}$, where $P_{r,s} := X^r Z^s$.

(c) The generalized Pauli group on n qudits is defined as $\mathcal{P}_d^n := \mathcal{P}_d^1 \otimes \mathcal{P}_d^1 \otimes \dots \otimes \mathcal{P}_d^1$.

It is easily seen that $X^d = Z^d = I$ and $XZ = \omega ZX$, hence \mathcal{P}_d^1 is indeed a group. Applying the commutation relation $XZ = \omega ZX$ appropriately many times, we have that

$$P_{r,s} P_{t,u} = \omega^{ru-st} P_{t,u} P_{r,s}. \quad (1.42)$$

³They are generalizations of the Pauli and Clifford group from the qubit case. Here, we first present them for qudit quantum systems and for the particular case of qubit quantum systems, they are presented later in Section 1.3.8.

Definition 14 (Generalized Clifford Group). *The Clifford group \mathcal{C}_d^n is the unitary group on n qudits that takes \mathcal{P}_d^n to \mathcal{P}_d^n by conjugation,*

$$\mathcal{C}_d^n = \{U \in U(d^n) \mid UPU^\dagger \in \mathcal{P}_d^n, \forall P \in \mathcal{P}_d^n\}. \quad (1.43)$$

1.3.3 Quantum Measurement

Quantum measurement is performed to extract classical information from a quantum system. The quantum state of the system changes according to the classical output. This phenomenon is known as the *collapse of the quantum state*, which is another peculiar feature of quantum systems.

Definition 15 (Quantum Measurement). (a) *A quantum measurement M on a quantum system A with n classical outputs is specified by n operators $\{M_i \in \mathcal{L}(\mathcal{H}_A) \mid i = 1, \dots, n\}$ satisfying $\sum_i M_i^\dagger M_i = I_A$.*

(b) *When performed on a quantum state ρ_A , M outputs a classical output $i \in \{1, \dots, n\}$ randomly with probability $p_i = \text{Tr}(M_i^\dagger M_i \rho)$ and the state of the system after measurement is given by,*

$$\frac{M_i \rho M_i^\dagger}{\text{Tr}(M_i^\dagger M_i \rho_A)}. \quad (1.44)$$

(c) *A quantum measurement M is called projective if each M_i is a projector, that is, $M_i^2 = M_i$.*

(d) *If one is only interested in the classical output (not in the quantum state after the measurement), a quantum measurement with n outcomes $\{1, \dots, n\}$ can be specified by a set of positive operators E_1, E_2, \dots, E_n , such that $\sum_i E_i = I$, and the probability of getting i as the outcome is given by $p_i = \text{Tr}[E_i \rho]$. This is called the positive operator valued measurement (POVM).*

An example of quantum measurement is the measurement of an observable, which is defined below.

Definition 16 (Measurement of an Observable). (a) *An observable $L_A \in \mathcal{L}(\mathcal{H}_A)$ is a Hermitian operator, i.e., $L^\dagger = L$.*

(b) *Measurement of an observable L_A corresponds to the projective measurement defined by the operators $M_\alpha = \sum_i |i\rangle\langle i|$, for each eigenvalue α of L_A , where the states $|i\rangle$ are eigenvectors of L_A with eigenvalue α ⁴. The outcome of the measurement corresponding to M_α is given by the eigenvalue α .*

(c) *A Pauli observable is a Hermitian Pauli operator (Definition 13), and a Pauli measurement is the measurement corresponding to a Pauli observable.*

1.3.4 Quantum Channels

Quantum channels are used to describe the dynamics of open (interacting with its environment) quantum systems. We will need the definition of superoperators to define quantum channels.

Definition 17 (Superoperators). (a) *A superoperator $\mathcal{T}_{A \rightarrow B} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ maps an operator $L_A \in \mathcal{L}(\mathcal{H}_A)$ to another operator $L_B \in \mathcal{L}(\mathcal{H}_B)$.*

⁴ $|\psi\rangle_A \in \mathcal{H}_A$ is an eigenvector of L_A with eigenvalue α if $L_A |\psi\rangle_A = \alpha |\psi\rangle_A$.

(b) A superoperator $\mathcal{T}_{A \rightarrow B}$ is completely positive if for any bipartite positive semidefinite operator $P_{AA'} \in \text{Pos}(\mathcal{H}_{AA'})$, we have that $(\mathcal{T}_{A \rightarrow B} \otimes \mathbb{I}_{A'})(P_{AA'}) \in \text{Pos}(\mathcal{H}_{AA'})$, where \mathbb{I}_A is the identity superoperator.

(c) A superoperator $\mathcal{T}_{A \rightarrow B}$ is trace preserving if for any $L_A \in \mathcal{L}(\mathcal{H}_A)$, we have that $\text{Tr}[L_A] = \text{Tr}[\mathcal{T}_{A \rightarrow B}(L_A)]$.

Definition 18 (Quantum channels). A quantum channel $\mathcal{N}_{A \rightarrow B}$, from A to B , is a superoperator from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$, which is a completely positive and trace preserving (CPTP) map. As $\mathcal{N}_{A \rightarrow B}$ is CPTP, it follows that $\mathcal{N}_{A \rightarrow B}$ maps a quantum state on A to a quantum state on B .

Remark 19. Given a bipartite quantum state $\psi_{AA'}$, to simplify the notation, we often use $\mathcal{N}_{A \rightarrow B}(\psi_{AA'}) := (\mathcal{N}_{A \rightarrow B} \otimes \mathbb{I}_{A'})(\psi_{AA'})$ to indicate that \mathcal{N} is applied only on part A of $\psi_{AA'}$.

Any quantum channel $\mathcal{N}_{A \rightarrow B}$ can be written as $\mathcal{N}_{A \rightarrow B}(L_A) = \sum_i N_i L_A N_i^\dagger$, for a set of operators $\{N_i | N_i \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)\}$, which satisfy $\sum_i N_i^\dagger N_i = I_A$. This is called the *Kraus representation* of quantum channels and operators N_i are called the *Kraus operators*. Quantum channels can be interpreted as a quantum measurement, defined in a slightly more general way than in Definition 15, with measurement operators $M_i \in \mathcal{L}(\mathcal{H}_A)$ replaced by Kraus operators $N_i \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$, where the classical outcome is not known.

Another interpretation of quantum channels is given in terms of the isometry transformations. Isometry transformations are a subset of quantum channels, that are completely noiseless, as defined below.

Definition 20 (Isometry transformation). (a) An isometry operator $U_{A \rightarrow B} \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ is an operator that satisfies $U_{B \rightarrow A}^\dagger U_{A \rightarrow B} = I_A$. If $B = A$, the operator $U_A := U_{A \rightarrow B}$ is a unitary (Definition 12).

(b) An isometry transformation $\mathcal{U}_{A \rightarrow B} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ is a quantum channel induced by the conjugate action of an isometry, that is,

$$\mathcal{U}_{A \rightarrow B}(\rho_A) = U_{A \rightarrow B} \rho_A U_{B \rightarrow A}^\dagger, \quad (1.45)$$

where $U_{A \rightarrow B}$ is an isometry.

Note that an isometry operator $U_{A \rightarrow B} \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ maps a pure quantum state $|\psi\rangle_A \in \mathcal{H}_A$ to another pure quantum state $|\phi\rangle_B = U_{A \rightarrow B}|\psi\rangle_A \in \mathcal{H}_B$. It is easily seen that if $|\psi\rangle_A$ is a unit vector, $|\phi\rangle_B$ is also a unit vector.

Quantum channels excluding isometry transformations are noisy as they are irreversible. The *Stinespring's dilation* theorem states that any quantum channel $\mathcal{N}_{A \rightarrow B}(\rho_A)$ can be obtained from an isometry transformation as follows,

$$\mathcal{N}_{A \rightarrow B}(\rho_A) = \text{Tr}_E[U_{A \rightarrow BE} \rho_A U_{BE \rightarrow A}^\dagger] \quad (1.46)$$

where $U_{A \rightarrow BE}$ is an isometry, whose output contains an extra system E , referred to as the environment. Therefore, the noisy effect of the channel can be considered to arise due to interaction with the environment, which is not accessible. Further, using the Stinespring dialation, any quantum channel can be associated with a complementary channel, as follows.

Definition 21 (Complementary Channel). Let $U_{A \rightarrow BE}$ be a Strinspring dialation of the quantum channel $\mathcal{N}_{A \rightarrow B}$. The complementary channel of $\mathcal{N}_{A \rightarrow B}$, denoted by $\mathcal{N}_{A \rightarrow E}^c$ is defined as,

$$\mathcal{N}_{A \rightarrow E}^c(\rho_A) := \text{Tr}_B[U_{A \rightarrow BE} \rho_A U_{BE \rightarrow A}^\dagger]. \quad (1.47)$$

1.3.5 Unitary 2-Designs

Let $\mathcal{U}(\mathcal{H}_A)$ be the set of unitary operators on \mathcal{H}_A . In many quantum information theoretic tasks, one needs to sample unitaries from the set $\mathcal{U}(\mathcal{H}_A)$ according to the Haar measure. As it contains infinitely many elements, in practice it's not efficient to sample from it. However, in certain situations, this problem is avoided by sampling from a unitary design, which is finite. In this thesis, we would need unitary 2-designs, which is defined below.

Let $\mathcal{W}_A := \mathcal{W}_{A \rightarrow A}$ be a quantum channel. The twirling of \mathcal{W}_A with respect to $\mathcal{U}(\mathcal{H}_A)$ is defined as the quantum channel that maps a ρ_A as follows,

$$\rho_A \mapsto \int U_A^\dagger \mathcal{W}_A(U_A \rho_A U_A^\dagger) U_A d\eta, \quad (1.48)$$

where $U_A \in \mathcal{U}(\mathcal{H}_A)$ is randomly chosen according to the Haar measure η . The twirling of \mathcal{W}_n with respect to a finite subset $\mathcal{U} \subset \mathcal{U}(\mathcal{H}_A)$ is defined as the quantum channel acting as,

$$\rho \mapsto \frac{1}{|\mathcal{U}|} \sum_{U_A \in \mathcal{U}} U_A^\dagger \mathcal{W}_A(U_A \rho U_A^\dagger) U_A. \quad (1.49)$$

Definition 22 (Unitary 2-design). *A finite subset $U_A \in \mathcal{U}(\mathcal{H}_A)$ is said to form a unitary 2-design if it satisfies the following, for all quantum channels \mathcal{W}_A , and quantum states ρ_A :*

$$\frac{1}{|\mathcal{U}|} \sum_{U_A \in \mathcal{U}} U_A^\dagger \mathcal{W}_A(U_A \rho_A U_A^\dagger) U_A = \int U_A^\dagger \mathcal{W}_A(U_A \rho_A U_A^\dagger) U_A d\eta. \quad (1.50)$$

1.3.6 Distinguishing Quantum States

In many situations, it's useful to compare quantum states with each other. In this section, two measures known as the fidelity and the trace distance are provided, which quantify the degree to which two quantum states are different.

Definition 23 (Fidelity). *Given any two quantum states, $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$, the fidelity is defined as,*

$$F(\rho, \sigma) := \|\sqrt{\rho} \sqrt{\sigma}\|_1 \in [0, 1], \quad (1.51)$$

where $\|\cdot\|_1$ is the trace norm defined as, $\|L\|_1 = \text{Tr}[\sqrt{L^\dagger L}]$, for any $L \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$.

For pure quantum states, the expression of fidelity simplifies as, $F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|$. It can be seen that $F(\rho, \sigma) = 1$, if and only if $\rho = \sigma$, and $F(\rho, \sigma) = 0$, if and only if matrices ρ and σ are orthogonal to each other, that is, $\rho\sigma = 0$. Therefore, fidelity is a measure of closeness between quantum states.

Definition 24 (Trace Distance). *Trace distance between quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$ is defined as,*

$$D(\rho, \sigma) := \frac{1}{2} \|\rho - \sigma\|_1. \quad (1.52)$$

The trace distance is a metric on the set of quantum states in the mathematical sense. Further, it has operational significance in the task of discriminating quantum states ρ and σ , as explained below.

Suppose we are given a quantum system with the guarantee that its quantum state is either ρ or σ , and our task is to guess which quantum state has been given. Consider a POVM measurement $E := \{E_0, E_1\}$. Suppose that if the measurement outcome is 0, we conclude that the given state is ρ and if the outcome is 1, we conclude that the given state

is σ . Then, we incorrectly guess the state if ρ was given and the measurement outcome is 1, and if σ was given and the measurement outcome is 0. Hence, the error probability p_e^E is given by,

$$p_e^E = \frac{1}{2} \text{Tr}[E_1 \rho] + \frac{1}{2} \text{Tr}[E_0 \sigma]. \quad (1.53)$$

The error probability p_e of discriminating quantum states ρ and σ is defined as the minimum of p_e^E over all the POVM measurements E , that is,

$$p_e := \min_E p_e^E. \quad (1.54)$$

We have the following relation between p_e and $D(\rho, \sigma)$,

$$p_e = \frac{1}{2} - \frac{1}{2} D(\rho, \sigma). \quad (1.55)$$

The trace distance and fidelity are related as follows.

Proposition 25 (Fuchs-van de Graaf inequality [27]). *For any two quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$, we have the following,*

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}. \quad (1.56)$$

1.3.7 Measures of Quantum Information

Many measures of information from classical information theory are generalized to quantum information. Here, we review some of the quantum information measures that we shall need in this thesis. In the following, when no confusion is possible, we shall drop the quantum systems from the notations of the quantum channels and quantum states.

von-Neumann Entropy

We first define the von-Neumann entropy, which is the quantum analog of the Shannon entropy.

Definition 26 (von-Neumann Entropy). (a) *The von-Neumann entropy of a quantum state ρ_A is defined as,*

$$H(\rho_A) := -\text{Tr}(\rho_A \log \rho_A) \in [0, 1].$$

(b) *The conditional von-Neumann entropy of a bipartite quantum state ρ_{AB} is defined as,*

$$H(A|B)_{\rho_{AB}} := H(\rho_{AB}) - H(\rho_B) \in [-1, 1].$$

The von-Neumann entropy $H(\rho_A)$ measures the information content of the quantum state ρ_A . More precisely, Schumacher's quantum source coding theorem [28] states that N i.i.d quantum states $\rho^{\otimes N}$ can be compressed into $NH(\rho)$ qubits, with very low probability of information loss and also information will be lost with very high probability if they are compressed further than $NH(\rho)$ qubits. Further, $H(\rho) = 0$ if and only if ρ_A is a pure quantum state and $H(\rho) = 1$ if and only if ρ_A is the maximally mixed state, that is, $\rho_A = \frac{\mathbb{1}_A}{d}$.

The conditional quantum entropy $H(A|B)_{\rho_{AB}}$ measures the information content of the system A , given access to the system B , when the quantum state of the system AB is ρ_{AB} . This interpretation is somewhat non-intuitive as $H(A|B)$ can take negative values for entangled quantum states. The negative values can be better understood in the context of

the quantum state merging protocol [29]. The conditional entropy gives the entanglement cost of quantum state merging, that is, if $H(A|B) > 0$ means entanglement needs to be consumed, and $H(A|B) < 0$ means entanglement is gained [29]. Further, for a tripartite pure quantum state ρ_{ABC} , the conditional entropy satisfies the following duality relation,

$$H(A|B) + H(A|C) = 0. \quad (1.57)$$

The above equation follows from the fact that for a pure quantum state ρ_{ABC} , we have $\text{supp}(\rho_{AB}) = \text{supp}(\rho_C)$, and $\text{supp}(\rho_{AC}) = \text{supp}(\rho_B)$, where $\text{supp}(\rho)$ is the support of ρ (Definition 10).

Using the von-Neumann entropy, we define below two information measures of quantum channels, namely the coherent information and the mutual information.

Definition 27 (Coherent Information). (a) The coherent information of a bipartite quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$ is defined as,

$$I(A)B := -H(A|B)_{\rho_{AB}}. \quad (1.58)$$

(b) Consider the action of a quantum channel $\mathcal{N}_{A' \rightarrow B}$ on A' -half of a bipartite quantum state $\psi_{AA'} \in \mathcal{D}(\mathcal{H}_{AA'})$, i.e., $\mathcal{N}(\psi_{AA'}) := (\mathbb{I}_A \otimes \mathcal{N}_{A' \rightarrow B})(\psi_{AA'})$. The coherent information of the quantum channel $\mathcal{N}_{A' \rightarrow B}$ is defined as the maximum coherent information over all the input quantum states $\psi_{AA'}$,

$$Q_1(\mathcal{N}_{A' \rightarrow B}) := \max_{\psi_{AA'}} I(A)B_{\mathcal{N}(\psi_{AA'})} \in [0, 1]. \quad (1.59)$$

(c) The symmetric coherent information of a quantum channel $\mathcal{N}_{A' \rightarrow B}$ is defined as the coherent information when half of an EPR pair $\Phi_{AA'}$ is given as the input,

$$I(\mathcal{N}_{A' \rightarrow B}) := I(A)B_{\mathcal{N}(\Phi_{AA'})} \in [-1, 1]. \quad (1.60)$$

When there is no entanglement assistance, the regularized coherent information is equal to the capacity of the quantum channel, as discussed in Section 1.4.1. Further, the coherent information can be strictly superadditive, i.e., $Q_1(\mathcal{N} \otimes \mathcal{M}) > Q_1(\mathcal{N}) + Q_1(\mathcal{M})$, hence it becomes intractable to compute it for tensor product channels $\mathcal{N}^{\otimes N}$ for large N , as one needs to maximize over the input states of the channel $\mathcal{N}^{\otimes N}$ [30, 31, 32]. It can be seen that the symmetric coherent information is a lower bound on the coherent information of a channel from their definitions. Further, the symmetric coherent information is easier to compute on tensor product channels as it is additive, i.e., $I(\mathcal{N} \otimes \mathcal{M}) = I(\mathcal{N}) + I(\mathcal{M})$. For this reason, it is often used instead of the coherent information.

The symmetric coherent information $I(\mathcal{N})$ is equal to 1 if and only if \mathcal{N} is the identity channel⁵, i.e., $\mathcal{N}(\rho) = \rho, \forall \rho \in \mathcal{D}(\mathcal{H}_A)$. Further, if $I(\mathcal{N})$ is equal to -1 , from the duality relation of the conditional entropy in (1.57), it follows that $I(\mathcal{N}^c)$ is equal to 1, where \mathcal{N}^c is a complementary channel of \mathcal{N} (Definition 21). Therefore, \mathcal{N}^c is the identity channel. This implies that the channel \mathcal{N} is a useless channel that outputs a fixed state ρ_f regardless of the channel input, i.e., $\mathcal{N}_{A' \rightarrow B}(\rho_{A'}) = \rho_f, \forall \rho \in \mathcal{D}(\mathcal{H}_{A'})$.

Definition 28 (Quantum mutual information). (a) The quantum mutual information of a bipartite quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$ is defined as,

$$I(A; B) := H(\rho_A) + H(\rho_B) - H(\rho_{AB}). \quad (1.61)$$

⁵an isometry transformation (Definition 20) instead of the identity channel to be precise. As isometry transformations are reversible, we consider the identity channel for the sake of simplicity.

- (b) The mutual information of a quantum channel $\mathcal{N}_{A' \rightarrow B}$ is defined as the maximum mutual information over all the pure input quantum states $\psi_{AA'} \in \mathcal{D}(\mathcal{H}_{AA'})$,

$$I_m(\mathcal{N}_{A' \rightarrow B}) := \max_{\psi_{AA'}} I(A; B)_{\mathcal{N}(\psi_{AA'})} \in [0, 1]. \quad (1.62)$$

- (c) The symmetric mutual information of a quantum channel $\mathcal{N}_{A' \rightarrow B}$ is defined as the mutual information when half of an EPR pair $\Phi_{AA'}$ is given as the input,

$$I_{ms}(\mathcal{N}_{A' \rightarrow B}) := I(A; B)_{\mathcal{N}(\Phi_{AA'})} \in [0, 1]. \quad (1.63)$$

The mutual informations of the quantum and classical channels are analogous in the sense that both are given by the entropy of the input plus output minus the joint entropy of the input and output, maximized over the input. In the classical case, entropy is the Shannon entropy and in the quantum case, it is the von-Neumann entropy. The quantum mutual information gives the capacity of a quantum channel in the entanglement assisted scenario, as discussed in Section 1.4.2.

Rényi Entropies

We first briefly discuss the classical Rényi entropies and divergences to provide the context and then define their quantum counterparts.

Based on an axiomatic approach, Rényi in his seminal work [33] provided a family of information measures generalizing the Shannon's entropy, now known as Rényi entropies.

Definition 29 (Rényi Entropies [33]). Let X be a random variable defined on an alphabet \mathcal{X} , with probability distribution $p_X(x)$, where $x \in \mathcal{X}$. Then, the Rényi entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ is defined as follows,

$$H_\alpha(X) := \frac{1}{1 - \alpha} \log \left(\sum_{x \in \mathcal{X}} (p_X(x))^\alpha \right). \quad (1.64)$$

The Shannon entropy is given by the limit value of $H_\alpha(X)$ as α approaches 1, that is,

$$H(X) = \lim_{\alpha \rightarrow 1} H_\alpha(X). \quad (1.65)$$

The relative entropy is an important information measure, which measures how far two random variables are from each other. It is defined using the Shannon entropy as follows.

Definition 30 (Relative entropy [34]). Let X and X' be two random variables defined on the same alphabet \mathcal{X} , with probability distributions $p_X(x)$ and $p_{X'}(x)$, respectively. The relative entropy (also known as Kullback–Leibler divergence) of X with respect to X' is defined as,

$$D(X||X') := \begin{cases} \sum_x p_X(x) (\log p_X(x) - \log p_{X'}(x)) & \text{if } X' \gg X, \\ \infty & \text{otherwise,} \end{cases}$$

where $X' \gg X$ means if $p_{X'}(x) = 0$ for some $x \in \mathcal{X}$, we also have $p_X(x) = 0$.

The relative entropy $D(X||X')$ is zero if and only if X and X' are identical. Hence, it is sort of a distance measure between X and X' . However, it's not a metric in the mathematical sense as it is not symmetric under the exchange of X and X' , that is, $D(X||X') \neq D(X'||X)$.

Rényi further provided a family of divergences generalizing the relative entropy, which are known as Rényi divergences.

Definition 31 (Rényi divergences [33]). *The Rényi divergence of order $\alpha \in (0, 1) \cup (1, \infty)$ is defined as follows,*

$$D_\alpha(X||X') := \begin{cases} \frac{1}{\alpha-1} \log \left(\sum_{x \in \mathcal{X}} \frac{(p_X(x))^\alpha}{(p_{X'}(x))^{(1-\alpha)}} \right) & \text{if } X' \gg X. \\ \infty & \text{otherwise.} \end{cases}$$

The relative entropy $D(X||X')$ is recovered from $D_\alpha(X||X')$ as α approaches 1, that is,

$$D(X||X') = \lim_{\alpha \rightarrow 1} D_\alpha(X||X'). \quad (1.66)$$

The other important cases of Rényi divergences, which have several applications in information theory, are $\alpha = 0, \frac{1}{2}, 2, \infty$. The reader may refer to [35] for more details.

The Rényi entropies and divergences have also been generalized for quantum states, which have found several applications in quantum information theory.

Definition 32 (Quantum Rényi Entropies [36]). *The generalization of Rényi entropies to the quantum case is straightforward and given as follows for any quantum state $\rho \in \mathcal{D}(\mathcal{H})$ and order $\alpha \in (0, 1) \cup (1, \infty)$,*

$$H_\alpha(\rho) = \frac{1}{1-\alpha} \log (\text{Tr}[\rho^\alpha]). \quad (1.67)$$

The von-Neumann entropy is recovered from H_α as α approaches 1, that is,

$$H(\rho) = \lim_{\alpha \rightarrow 1} H_\alpha(\rho). \quad (1.68)$$

Definition 33 (Quantum relative entropy [8, 26]). *The quantum relative entropy or quantum Kullback–Leibler divergence of a quantum state ρ with respect to a quantum state σ is defined as,*

$$D(\rho||\sigma) := \begin{cases} \text{Tr} [\rho (\log \rho - \log \sigma)] & \text{if } \sigma \gg \rho, \\ \infty & \text{otherwise,} \end{cases}$$

where the symbol $\sigma \gg \rho$ means that $\text{supp}(\rho)$ is included in $\text{supp}(\sigma)$.

The classical relative entropy can be recovered from quantum relative entropy when quantum states ρ and σ commute with each other, that is, $[\rho, \sigma] = 0$. Several properties from the classical case are preserved, for example, the quantum relative entropy $D(\rho||\sigma)$ is zero if and only if $\rho = \sigma$. Further, it's not symmetric in general under the exchange of ρ and σ , that is, $D(\rho||\sigma) \neq D(\sigma||\rho)$.

There are many generalizations of Rényi divergences to the quantum case. Here, we mention two important generalizations that we shall use later; the first quantum Petz-Rényi divergences [37, 38] and the second quantum sandwiched Rényi divergences [36, 39].

Definition 34 (Quantum Petz-Rényi divergences [37, 38]). *Quantum Rényi relative entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ of the quantum state ρ with respect to the quantum state σ is defined as,*

$$D_\alpha(\rho||\sigma) := \begin{cases} \frac{1}{\alpha-1} \text{Tr}[\rho^\alpha \sigma^{1-\alpha}] & \text{if } \sigma \gg \rho. \\ \infty & \text{otherwise.} \end{cases}$$

Definition 35 (Quantum sandwiched Rényi divergences [36, 39]). *Sandwiched quantum Rényi relative entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ of the quantum state ρ with respect to the quantum state σ is defined as,*

$$\tilde{D}_\alpha(\rho||\sigma) := \begin{cases} \frac{1}{\alpha-1} \text{Tr} \left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho^\alpha \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right] & \text{if } \sigma \gg \rho. \\ \infty & \text{otherwise.} \end{cases}$$

When ρ and σ commute, both generalizations $D_\alpha(\rho||\sigma)$ and $\tilde{D}_\alpha(\rho||\sigma)$ yield the classical Rényi divergence of order α . Further, as α approaches 1, $D_\alpha(\rho||\sigma)$ and $\tilde{D}_\alpha(\rho||\sigma)$ approach to the quantum relative entropy $D(\rho||\sigma)$. It is shown in [39], using the Araki-Lieb-Thirring trace inequality [40, 41] that

$$D_\alpha(\rho||\sigma) \geq \tilde{D}_\alpha(\rho||\sigma). \quad (1.69)$$

Moreover, functions $\alpha \rightarrow D_\alpha(\rho||\sigma)$ and $\alpha \rightarrow \tilde{D}_\alpha(\rho||\sigma)$ are monotonically increasing [36, 42].

Remark 36. For $\alpha > 1$, a new generalization of quantum Rényi divergence $D_\alpha^\#(\rho||\sigma)$ is proposed recently in [43] via convex optimization, whose regularization gives the quantum sandwiched Rényi divergence $\tilde{D}_\alpha(\rho||\sigma)$.

We now turn to the quantum conditional Rényi entropies, which are defined using quantum Rényi divergences D_α and \tilde{D}_α [44]. First note that the quantum conditional entropy $H(A|B)$ (Definition 26) can be derived from the quantum relative entropy as follows,

$$\begin{aligned} H(A|B)_{\rho_{AB}} &= -D(\rho_{AB}||\mathbb{1} \otimes \rho_B). \\ &= \sup_{\sigma_B} -D(\rho_{AB}||\mathbb{1} \otimes \sigma_B), \end{aligned}$$

where $\rho_B = \text{Tr}_A[\rho_{AB}]$. The second equality follows from the fact that $\sigma_B = \rho_B$ maximizes the quantity $D(\rho_{AB}||\mathbb{1} \otimes \sigma_B)$. However, this is not true in general for Rényi relative entropies D_α and \tilde{D}_α . Similarly to the above, using D_α and \tilde{D}_α , quantum conditional Rényi entropies are defined in [44], as follows.

Definition 37 (Quantum conditional Petz-Rényi entropies). For a bipartite quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$ and $\alpha \in (0, 1) \cup (1, \infty)$,

$$H_\alpha^\downarrow(A|B)_{\rho_{AB}} := -D_\alpha(\rho_{AB}||\mathbb{1}_A \otimes \rho_B). \quad (1.70)$$

$$H_\alpha^\uparrow(A|B)_{\rho_{AB}} := \sup_{\sigma_B} -D_\alpha(\rho_{AB}||\mathbb{1}_A \otimes \sigma_B). \quad (1.71)$$

Definition 38 (Quantum conditional sandwiched Rényi entropies). For a bipartite quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$ and $\alpha \in (0, 1) \cup (1, \infty)$,

$$\tilde{H}_\alpha^\downarrow(A|B)_\rho := -\tilde{D}_\alpha(\rho_{AB}||\mathbb{1}_A \otimes \rho_B). \quad (1.72)$$

$$\tilde{H}_\alpha^\uparrow(A|B)_\rho := \sup_{\sigma_B} -\tilde{D}_\alpha(\rho_{AB}||\mathbb{1}_A \otimes \sigma_B). \quad (1.73)$$

In the above definitions \uparrow is used for maximization and \downarrow is used when there is no maximization. Note that when α approaches 1, we have the following for any quantum state ρ_{AB} ,

$$\lim_{\alpha \rightarrow 1} H_\alpha^\downarrow(A|B) = \lim_{\alpha \rightarrow 1} H_\alpha^\uparrow(A|B) = \lim_{\alpha \rightarrow 1} \tilde{H}_\alpha^\downarrow(A|B) = \lim_{\alpha \rightarrow 1} \tilde{H}_\alpha^\uparrow(A|B) = H(A|B). \quad (1.74)$$

The duality relation in (1.57) also holds for conditional Rényi entropies, as below.

Theorem 39 ([44]). For any $\alpha, \beta \in (0, 1) \cup (1, \infty)$, satisfying $\alpha.\beta = 1$, and any pure quantum state $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_{ABC})$, the following relation holds,

$$H_\alpha^\uparrow(A|B)_\rho + \tilde{H}_\beta^\downarrow(A|B)_\rho = 0. \quad (1.75)$$

We will make use of the conditional Petz-Rényi entropy $\tilde{H}_\alpha^\uparrow$ of order $\alpha = \frac{1}{2}$, and the conditional sandwiched Rényi entropy $\tilde{H}_\alpha^\downarrow(A|B)_\rho$ of order $\alpha = 2$ for the proof of our quantum polarization theorem in Chapter 2. The quantity \tilde{H}_2^\downarrow is first introduced by Renner [45], and it is also known as *quantum conditional collision entropy*. Note that $H_{\frac{1}{2}}^\uparrow$ and \tilde{H}_2^\downarrow satisfy the duality relation in Theorem 39. The explicit expressions of $H_{\frac{1}{2}}^\uparrow$ and \tilde{H}_2^\downarrow are given below for a bipartite quantum state ρ_{AB} .

$$\tilde{H}_2^\downarrow(A|B)_\rho = -\log \text{Tr} \left[(\mathbb{1}_A \otimes \rho_B^{-\frac{1}{2}}) \rho_{AB} (\mathbb{1}_A \otimes \rho_B^{-\frac{1}{2}}) \rho_{AB} \right]. \quad (1.76)$$

$$H_{\frac{1}{2}}^\uparrow(A|B)_\rho = 2 \log \sup_{\sigma_B} \text{Tr} \left[\rho_{AB}^{\frac{1}{2}} (\mathbb{1}_A \otimes \sigma_B^{\frac{1}{2}}) \right]. \quad (1.77)$$

1.3.8 Qubit Quantum Systems

In this section, we focus on qubit quantum systems. As mentioned before, the dimension of the associated Hilbert space is two for a qubit quantum system. The orthonormal basis for a single qubit system, defined by the set of vectors $\{|0\rangle, |1\rangle\}$, where $|0\rangle := \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $|1\rangle := \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, is called the *amplitude or computational* basis. Another important basis for a single qubit system is the *phase* basis, which is defined by the set $\{|+\rangle, |-\rangle\}$, where $|+\rangle := \frac{|0\rangle + |1\rangle}{2}$, and $|-\rangle := \frac{|0\rangle - |1\rangle}{2}$.

Qubit Gates

Qubit gates are unitary transformations on qubits. In this section, we give some examples of single and two-qubit gates.

Pauli gates: Pauli gates or matrices are single qubit gates consisting of the identity gate I , and three single qubit gates X, Y, Z , which act as follows in the amplitude basis,

$$X|x\rangle = |x \oplus 1\rangle, \quad (1.78)$$

$$Z|x\rangle = (-1)^x |x\rangle, \quad (1.79)$$

$$Y|x\rangle = (-1)^x i |x \oplus 1\rangle, \quad (1.80)$$

where $x \in \{0, 1\}$. Note that the Pauli X acts like the classical NOT gate in the amplitude basis, hence, referred to as the *bit-flip* gate. Also, it can be easily seen that the Pauli Z acts as the classical NOT gate in the phase basis, hence, referred to as the *phase-flip* gate. Further, It can be seen that all the Pauli gates are Hermitian, i.e., $T^\dagger = T, \forall T \in \{I, X, Y, Z\}$, and they satisfy $XY = iZ$.

Hadamard gate: The Hadamard gate is a single qubit gate, which takes an amplitude basis state to a phase basis state as below,

$$H|x\rangle = \frac{|0\rangle + (-1)^x |1\rangle}{\sqrt{2}}, x \in \{0, 1\}. \quad (1.81)$$

Phase gate: The phase gate corresponding to a $\theta \in [0, 2\pi]$ is a single qubit gate, which acts as follows in the amplitude basis,

$$R_\theta|x\rangle = e^{i2\theta x} |x\rangle, x \in \{0, 1\}. \quad (1.82)$$

CNOT gate: The quantum Controlled-NOT (CNOT) gate is a two-qubit gate. When the first qubit is target and the second qubit is control, as depicted in Figure 1.7, it acts as follows,

$$C_{2 \rightarrow 1}|x\rangle|y\rangle = |x \oplus y\rangle|y\rangle, \quad x, y \in \{0, 1\}. \quad (1.83)$$

Therefore, the quantum CNOT gate acts like the classical CNOT in the amplitude basis. This explains the same circuit being used for both the classical and the quantum CNOT.



Figure 1.7: Quantum CNOT

Swap gate: The swap gate is a two-qubit gate, which acts as follows in the amplitude basis,

$$S|x\rangle|y\rangle = |y\rangle|x\rangle, \quad x, y \in \{0, 1\}. \quad (1.84)$$

Pauli and Clifford Group

Here, we present the Pauli and Clifford groups for qubit quantum systems.

Definition 40 (Pauli Group). (a) The Pauli group on one qubit \mathcal{G}_1 consists of Pauli matrices, with phase factors $\{\pm 1, \pm i\}$, as defined below,

$$\mathcal{G}_1 := \{\pm I, \pm iI, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\}. \quad (1.85)$$

(b) The Pauli group \mathcal{G}_N on N -qubits is defined as

$$\mathcal{G}_N := \mathcal{G}_1^{\otimes N}. \quad (1.86)$$

As product of any two single qubit Pauli matrices yields another Pauli matrix upto a phase factor in $\{\pm 1, \pm i\}$, and $I^2 = X^2 = Y^2 = Z^2 = I$, therefore, \mathcal{G}_N is indeed a group. Further, any $g_1, g_2 \in \mathcal{G}_N$ either commute or anti-commute, that is,

$$g_1 \text{ and } g_2 \text{ commute if } [g_1, g_2] := g_1 g_2 - g_2 g_1 = 0. \quad (1.87)$$

$$g_1 \text{ and } g_2 \text{ anti-commute if } \{g_1, g_2\} := g_1 g_2 + g_2 g_1 = 0. \quad (1.88)$$

Definition 41 (Clifford group). The Clifford Group on N -qubits \mathcal{C}_N is a set of unitaries that takes \mathcal{G}_N to \mathcal{G}_N by conjugation,

$$\mathcal{C}_N = \{U \in U(2^N) \mid U g U^\dagger \in \mathcal{G}_N, \forall g \in \mathcal{G}_N\}. \quad (1.89)$$

Universal Set of Qubit Gates

We first define the operator norm.

Definition 42 (Operator norm). For any $L \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$, the operator norm is defined as,

$$\|L\| = \sup_{\psi} \|L|\psi\rangle\|, \quad (1.90)$$

where maximization is taken over all the normalized vectors (pure quantum states) in \mathcal{H}_A .

A unitary operator V is an ϵ -approximation of a unitary operator U if the following holds,

$$\|U - V\| \leq \epsilon. \quad (1.91)$$

When ϵ is close to zero, for any quantum measurement M , we have almost the same output probabilities for quantum states $U|\psi\rangle$ and $V|\psi\rangle$ [26, Chapter 4].

Definition 43 (Universal set of gates). *A finite set of gates \mathbb{G} is said to be universal if for any n -qubit unitary U and $\epsilon > 0$, there exists a sequence of gates $G_1, G_2, \dots, G_N \in \mathbb{G}$, which gives an ϵ -approximation of U .*

Remark 44. *The gate set $\{C, H, T\}$, where C is the CNOT gate, H is the Hadamard gate, and T is the $\frac{\pi}{8}$ -phase gate $R_{\frac{\pi}{8}}$, is an example of universal set [26, Chapter 4].*

Qubit Channels

Here, we give some examples of the qubit quantum channels, whose inputs and outputs are qubit quantum states. We denote $\mathcal{N}_A := \mathcal{N}_{A \rightarrow A}$.

Definition 45 (Pauli channels). *A quantum channel \mathcal{N}_A , acting on a single qubit quantum state, is called a Pauli channel, if there exists a Kraus representation of the following form,*

$$\mathcal{N}_A(\rho_A) := p_I \rho_A + p_X X \rho_A X + p_Y Y \rho_A Y + p_Z Z \rho_A Z, \quad (1.92)$$

where $p_I, p_X, p_Y, p_Z \geq 0$, such that $p_I + p_X + p_Y + p_Z = 1$.

A Pauli channel can be interpreted as randomly applying a single qubit Pauli gate from I, X, Y, Z on the qubit-input ρ_A , with probabilities p_I, p_X, p_Y, p_Z , respectively. Note that, when $p_I = 1$, the Pauli channel is the identity channel. Further, when $p_I = p_X = p_Y = p_Z = \frac{1}{4}$, we have $\mathcal{N}_A(\rho_A) = \frac{\mathbb{1}_A}{2}, \forall \rho_A \in \mathcal{D}(\mathcal{H}_A)$. Hence, the Pauli channel is completely noisy, as it outputs the maximally mixed state regardless of the channel input.

Definition 46 (Depolarizing channels). *A depolarizing channel is a Pauli channel, which acts as follows on a single qubit quantum state,*

$$\mathcal{N}_A(\rho_A) := (1 - p)\rho_A + p \frac{\mathbb{1}_A}{2}, \quad p \geq 0. \quad (1.93)$$

In other words, depolarizing channel outputs a convex combination of the input state ρ_A and the maximally mixed state $\frac{\mathbb{1}_A}{2}$.

Definition 47 (Qubit erasure channels). *A qubit erasure channel is defined as,*

$$\mathcal{N}_{A \rightarrow FA}(\rho_A) := (1 - p)|0\rangle\langle 0|_F \otimes \rho_A + p|1\rangle\langle 1|_F \otimes \frac{\mathbb{1}_A}{2}, \quad p \geq 0. \quad (1.94)$$

In other words, the qubit erasure channel outputs a classical flag F along with the quantum output A . If the flag is found to be in $|0\rangle$ state, the output quantum state is equal to the input state ρ_A and if it is in $|1\rangle$ state, the output state is the maximally mixed quantum state $\frac{\mathbb{1}_A}{2}$.

1.4 Quantum Channel Coding

A quantum channel (Definition 18) can be used to transmit either classical or quantum information (quantum states). Here, we shall consider the transmission of quantum information through quantum channels. The following two scenarios are discussed; the first quantum communication without entanglement assistance and the second quantum communication with entanglement assistance.

1.4.1 Quantum Communication without Entanglement Assistance

As discussed in Section 1.3.4, quantum channels are in general irreversible, therefore, they corrupt the input quantum state. The goal of quantum channel coding is to transmit quantum information reliably through a noisy quantum channel so that it can be recovered at the receiver end. Similarly to classical channel coding, a quantum channel coding scheme consists of encoding and decoding procedures as illustrated in Figure 1.8.

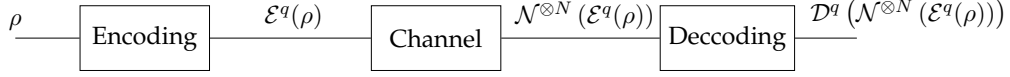


Figure 1.8: Quantum channel coding

We consider qudit quantum systems for which the dimension of the associated Hilbert space is $d \geq 2$. The encoding procedure is given by an isometry transformation (Definition 20) $\mathcal{E}_{(A_1 \dots A_K) \rightarrow (A_1 \dots A_N)}^q$, where $N \geq K$, mapping a K -qudit quantum state $\rho_{A_1 \dots A_K} \in \mathcal{D}(\mathcal{H}_{A_1 \dots A_K})$ to a N -qudit quantum state $\mathcal{E}^q(\rho_{A_1 \dots A_K}) \in \mathcal{D}(\mathcal{H}_{A_1 \dots A_N})$. Let $\{|i_1, \dots, i_K\rangle \mid i_1, \dots, i_K \in \{0, 1\}\}$ be a basis of $\mathcal{H}_{A_1 \dots A_K}$, and $E_{(A_1 \dots A_K) \rightarrow (A_1 \dots A_N)} \in \mathcal{L}(\mathcal{H}_{A_1 \dots A_K}, \mathcal{H}_{A_1 \dots A_N})$ be the isometry operator associated with \mathcal{E}^q , such that,

$$\mathcal{E}^q(\rho_{A_1 \dots A_K}) = E \rho_{A_1 \dots A_K} E^\dagger. \quad (1.95)$$

Then, the quantum code \mathcal{C}_Q , generated by the encoding \mathcal{E}^q , is the subspace of $\mathcal{H}_{A_1 \dots A_N}$ corresponding to the basis,

$$\{E|i_1, \dots, i_K\rangle \mid i_1, \dots, i_K \in \{0, 1\}\}. \quad (1.96)$$

The rate of transmission for the code \mathcal{C}_Q is defined as,

$$R := \frac{K}{N} \in [0, 1]. \quad (1.97)$$

The encoded quantum state $\mathcal{E}^q(\rho_{A_1 \dots A_K})$ is transmitted over N instances of a quantum channel $\mathcal{N}_{A \rightarrow B}$, which output the quantum state $\mathcal{N}^{\otimes N}(\mathcal{E}^q(\rho_{A_1 \dots A_K}))$. Upon receiving the channel output, the receiver applies the decoding procedure to generate an estimate $\hat{\rho}_{A_1 \dots A_K}$ of the input quantum state. The decoding procedure is given by an isometry $\mathcal{D}_{(B_1 \dots B_N) \rightarrow (A_1 \dots A_K)}^q$ such that

$$\hat{\rho}_{A_1 \dots A_K} := \mathcal{D}^q(\mathcal{N}^{\otimes N}(\mathcal{E}^q(\rho_{A_1 \dots A_K}))). \quad (1.98)$$

A code is ϵ -reliable, if the following holds for the trace distance $D(\rho, \hat{\rho})$,

$$D(\rho, \hat{\rho}) \leq \epsilon, \forall \rho \in \mathcal{D}(\mathcal{H}_{A_1 \dots A_K}). \quad (1.99)$$

A rate R is achievable for \mathcal{N} , if for any $\epsilon, \delta > 0$, there exists a ϵ -reliable quantum code with rate $R - \delta$. *Quantum capacity* $Q(\mathcal{N})$ is defined as the tightest upper bound on the achievable rate of transmission of quantum information (see (1.5)). The Llyod-Shor-Devetak quantum capacity theorem [9, 10, 11] states that the quantum capacity is characterized by the following regularized expression,

$$Q(\mathcal{N}) = \lim_{N \rightarrow \infty} \frac{1}{N} Q_1(\mathcal{N}^{\otimes N}), \quad (1.100)$$

where $Q_1(\mathcal{N}^{\otimes N})$ is the coherent information, maximized over all the input states of $\mathcal{N}^{\otimes N}$ (Definition 27). As the coherent information is superadditive, it follows that the coherent information of single channel use $Q_1(\mathcal{N})$ only gives a lower bound on the capacity $Q(\mathcal{N})$ [30, 31, 32]. Further, the superadditive effect makes it hard to evaluate the value of the above limit for general quantum channels. For this reason, unlike the classical capacity, we do not yet have a closed formula for the capacity of general quantum channels. It is worth pointing out here that there are a subset of quantum channels known as *degradable* channels for which the coherent information is additive, therefore, the coherent information of the single channel use $Q_1(\mathcal{N})$ is the channel capacity for degradable channels [46].

1.4.2 Quantum Communication with Entanglement Assistance

A quantum channel coding scheme is said to be entanglement assisted if the coding scheme uses entangled quantum states that have been priorly shared between the sender and receiver. The encoding and decoding for an entanglement assisted coding scheme is done as follows.

A K qudit quantum state is encoded into a N qudit quantum state, using an isometry from $K + T$ qudit system to N qudit system for some $0 < T \leq N - K$, as below,

$$\mathcal{E}_{A_1 \dots A_K \mathcal{A}' \rightarrow A_1 \dots A_N}^q (\rho_{A_1 \dots A_K} \otimes \psi_{\mathcal{A}' \mathcal{B}'}), \quad (1.101)$$

where $\psi_{\mathcal{A}' \mathcal{B}'}$ is an entangled quantum state on T qudit quantum systems $\mathcal{A}' := A'_1 \dots A'_T$ and $\mathcal{B}' := B'_1 \dots B'_T$ shared between the sender and receiver, such that the sender has \mathcal{A}' and the receiver \mathcal{B}' . Note that the encoding isometry only acts on the system \mathcal{A}' of the shared entangled state $\psi_{\mathcal{A}' \mathcal{B}'}$. The sender transmits the encoded N -qudit system $A_1 \dots A_N$, using N times a quantum channel $\mathcal{N}_{A \rightarrow B}$, which output the following quantum state,

$$\mathcal{N}^{\otimes N} (\mathcal{E}^q (\rho_{A_1 \dots A_K} \otimes \psi_{\mathcal{A}' \mathcal{B}'}))_{B_1 \dots B_N \mathcal{B}'}. \quad (1.102)$$

The decoding is done by applying an isometry $\mathcal{D}_{B_1 \dots B_N \mathcal{B}' \rightarrow A_1 \dots A_K}^q$, which gives the following quantum state,

$$\mathcal{D}^q (\mathcal{N}^{\otimes N} (\mathcal{E}^q (\rho_{A_1 \dots A_K} \otimes \psi_{\mathcal{A}' \mathcal{B}'})))_{A_1 \dots A_K}. \quad (1.103)$$

It is shown in [12] that regularization is not necessary for the channel capacity in the case of the entanglement assisted quantum communication, and it is given by half the mutual information of the channel $\mathcal{N}_{A \rightarrow B}$ (Definition 28),

$$Q_{\text{ent}}(\mathcal{N}) = \frac{I_m(\mathcal{N})}{2}. \quad (1.104)$$

1.4.3 Stabilizer and CSS Quantum Codes

Here, we provide efficient constructions of quantum codes for qubit quantum systems ($d = 2$). We first consider well-known stabilizer codes [47]. Then, we consider a special case of the stabilizer codes, known as the Calderbank-Shor-Steane (CSS) codes [48, 49], for which we also briefly discuss the extension to the entanglement assisted communication scenario.

Consider an Abelian subgroup of the N -qubit Pauli group $\mathcal{S} \subset \mathcal{G}_N$ (Definition 40). The subgroup \mathcal{S} stabilizes a pure quantum state $|\psi\rangle$ if the following holds,

$$S|\psi\rangle = |\psi\rangle, \forall S \in \mathcal{S}. \quad (1.105)$$

In other words, $|\psi\rangle$ is fixed by every $S \in \mathcal{S}$. The above can be generalized for mixed state ρ in a straightforward way as follows. The subgroup \mathcal{S} stabilizes a mixed quantum state ρ , with a pure state decomposition $\sum_i p_i |\psi_i\rangle\langle\psi_i|$, if all the pure states $|\psi_i\rangle$ in the decomposition are stabilized by \mathcal{S} . For the sake of clarity, we will consider only pure states in this section. However, everything described here remains valid for mixed states.

Given an Abelian subgroup $\mathcal{S} \subset \mathcal{G}_N$, define a subspace $\mathcal{C}(\mathcal{S})$ of the N qubit Hilbert space as follows,

$$\mathcal{C}(\mathcal{S}) := \{|\psi\rangle \mid S|\psi\rangle = |\psi\rangle, \forall S \in \mathcal{S}\}. \quad (1.106)$$

It can be easily seen that $-I \in \mathcal{S}$ implies that $\mathcal{C}(\mathcal{S})$ is an empty set.

Definition 48. A stabilizer group \mathcal{S} is an Abelian subgroup of \mathcal{G}_N , which does not contain $-I$. In this case, $\mathcal{C}(\mathcal{S})$ is called the stabilizer code generated by the stabilizer group \mathcal{S} .

Any stabilizer group $\mathcal{S} \subset \mathcal{G}_N$ has 2^{N-K} elements, for some $0 \leq K \leq N$ [26, Chapter 10, Section 10.5]. Further, a subgroup with 2^{N-K} elements stabilizes a subspace of dimension 2^K of the N -qubit Hilbert space. Therefore, a subgroup with 2^N elements stabilizes only one pure quantum state, which is called a *stabilizer state*.

A subset $\{S_1, \dots, S_n\} \subset \mathcal{S}$ is said to be a generating set of \mathcal{S} if every element in \mathcal{S} can be written as a product of elements from $\{S_1, \dots, S_n\}$. Further, a generating set is said to be independent if any $S_i \in \{S_1, \dots, S_n\}$ is not a product of elements from $\{S_1, \dots, S_n\} \setminus S_i$. Any independent generating set of \mathcal{S} has cardinality $\log_2 |\mathcal{S}|$, where $|\mathcal{S}|$ is the cardinality of \mathcal{S} . In the following, unless otherwise mentioned, we shall always consider independent generating sets of a stabilizer group.

The stabilizer codes are defined in a way that they are suitable for correcting Pauli errors. Suppose a (unknown) Pauli error $E \in \mathcal{G}_N$ occurs on a quantum state $|\psi\rangle \in \mathcal{C}(\mathcal{S})$. Then, the corrupted quantum state is given by, $|\psi'\rangle := E|\psi\rangle$. First, note that if $E \in \mathcal{S}$, the error does not affect the state as $|\psi'\rangle = |\psi\rangle$, hence, there is no need of error correction. In general, we need to identify the error E up to a stabilizer, i.e., we can identify E with E' such that $E = E'S$, for any $S \in \mathcal{S}$. Then, the initial state can be recovered by applying the operator E'^\dagger back on the corrupted state. We now briefly give the intuition behind the error correction properties of the stabilizer codes.

As elements in \mathcal{G}_N either commute or anti-commute, we have the following for any $S \in \mathcal{S}$,

$$\begin{aligned} S|\psi'\rangle &= SE|\psi\rangle \\ &= (-1)^a E|\psi\rangle, \end{aligned} \quad (1.107)$$

where $a = 0$, if $[S, E] = 0$, and $a = 1$ if $\{S, E\} = 0$. We now consider the following two cases,

1. The error E anti-commutes with at least one $S \in \mathcal{S}$: From (1.107), it follows that the corrupted state $|\psi'\rangle$ is orthogonal to the code $\mathcal{C}(\mathcal{S})$. Hence, E can be detected by performing a projective measurement corresponding to the generator S , which anticommutes with E . As \mathcal{S} is an Abelian group, all the generators $\{S_1, \dots, S_n\}$ can be measured simultaneously. The *error syndrome* is defined as the set of outcomes of the measurements corresponding to all the generators $\{S_1, \dots, S_n\}$.

2. The error E commutes with every element in the stabilizer group \mathcal{S} , and $E \notin \mathcal{S}$: Then $E \in N(\mathcal{S}) \setminus \mathcal{S}$, where $N(\mathcal{S})$ is the centralizer of the stabilizer group \mathcal{S} . From (1.107), the commutation condition implies that the corrupted state $|\psi'\rangle$ belongs to the code $\mathcal{C}(\mathcal{S})$. Further, as $E \notin \mathcal{S}$, $|\psi'\rangle$ is not equal to $|\psi\rangle$. In this case, E can not be detected by performing measurements corresponding to the generators.

Using the above two conditions, the set of correctable errors for any stabilizer code $\mathcal{C}(\mathcal{S})$ can be determined as given in the following lemma.

Lemma 49 ([47]). *A set of errors \mathcal{E} is correctable if for any $E_i, E_j \in \mathcal{E}$, the condition $E_i^\dagger E_j \notin N(\mathcal{S}) \setminus \mathcal{S}$ holds.*

We briefly give the intuition behind the proof of the above lemma. From Point (1) of the above discussion, it follows that if $E_i^\dagger E_j \notin N(\mathcal{S}) \setminus \mathcal{S}$, then $\langle \psi | E_i^\dagger E_j | \psi \rangle = 0$, for any $|\psi\rangle \in \mathcal{C}(\mathcal{S})$. Therefore, errors E_i and E_j take the state $|\psi\rangle$ to two different orthogonal spaces to the code $\mathcal{C}(\mathcal{S})$, implying that they produce different error syndromes. Hence, we can distinguish them from one another.

Note that the above lemma does not explicitly provide a decoding scheme to identify the errors given the error syndrome. Lastly it's important to mention that using a stabilizer code, the symmetric coherent information can be achieved for Pauli channels (see [8, Chapter 24, Section 6]).

Definition 50 (Calderbank-Shor-Steane (CSS) code [48, 49]). *An stabilizer code $\mathcal{C}(\mathcal{S})$ is called a CSS code if there exists a generating set $G := G_X \cup G_Z$, such that any $g_x \in G_X$ consists of Pauli matrices I and X , i.e., $g_x = X^{u_0} \otimes \cdots \otimes X^{u_{N-1}}$, for some $u_0, \dots, u_{N-1} \in \{0, 1\}$, and any $g_z \in G_Z$ consists of Pauli matrices I and Z , i.e., $g_z = Z^{v_0} \otimes \cdots \otimes Z^{v_{N-1}}$, for some $v_0, \dots, v_{N-1} \in \{0, 1\}$. We shall refer G_X as the X type generating set and G_Z as the Z type generating set.*

The problem of constructing a CSS code is equivalent to the construction of two orthogonal classical binary linear codes as explained below.

Given a CSS code \mathcal{C} , consider the X and Z part of the generators G_X and G_Z . One may associate a $K_X \times N$ matrix H_X with G_X , where K_X is the cardinality of G_X , such that the rows of H_X are vectors (u_0, \dots, u_{N-1}) corresponding to $g_x = X^{u_0} \otimes \cdots \otimes X^{u_{N-1}}$, $\forall g_x \in G_X$. Similarly, a $K_Z \times N$ matrix H_Z is associated with G_Z , where K_Z is the cardinality of G_Z , such that the rows of H_Z are vectors (v_0, \dots, v_{N-1}) corresponding to $g_z = Z^{v_0} \otimes \cdots \otimes Z^{v_{N-1}}$, $\forall g_z \in G_Z$.

Consider an error $E \in \mathcal{G}_N$ occurs on a quantum state $|\psi\rangle$ belonging to the CSS code \mathcal{C} . We only need to correct errors up to the phase factor, hence we may take $E \in \mathcal{G}_N \setminus \{\pm 1, \pm i\}$, i.e., the centralizer of the Pauli group \mathcal{G}_N with respect to its quotient. As any element in $\bar{\mathcal{G}}_1$ can be written as $X^u Z^v$ for some $u, v \in \{0, 1\}$, it follows that $E = X^{u_0} Z^{v_0} \otimes \cdots \otimes X^{u_N} Z^{v_N}$ for some vectors $(u'_0, \dots, u'_{N-1}), (v'_0, \dots, v'_{N-1}) \in \{0, 1\}^N$.

It can be seen that the measurement of a generator $g_x = X^{u_0} \otimes \cdots \otimes X^{u_{N-1}} \in G_X$ on the error corrupted quantum state $E|\psi\rangle$ outputs $\sum_{i=0}^{N-1} u_i v'_i \in \{0, 1\}$, where the sum is XOR. Therefore, it follows that the measurement outcome of all the generators $g_x \in G_X$ is given by the following vector,

$$H_X(v'_0, \dots, v'_{N-1}) \in \{0, 1\}^{K_X}. \quad (1.108)$$

Similarly, the measurement outcome of all the generators $g_z \in G_Z$ is given by the vector,

$$H_Z(u'_0, \dots, u'_{N-1}) \in \{0, 1\}^{K_Z}. \quad (1.109)$$

Hence, a quantum CSS code yields two classical binary linear codes defined by the parity check matrices H_X and H_Z (see [26, Chapter 10] and [14, Chapter 13] for the Parity check matrices). Further, the commutativity of generators $g_x, g_z, \forall g_X \in G_X, g_Z \in G_Z$ imposes the following orthogonality constraint on matrices H_X and H_Z ,

$$H_X H_Z^\top = 0, \quad (1.110)$$

where H_Z^\top is the transpose of H_Z . Therefore, two classical codes associated with a CSS codes are orthogonal to each other. Moreover, given two classical codes with check matrices H_1 and H_2 satisfying $H_1 H_2^\top = 0$, one can obtain a quantum CSS code by defining the X and Z type generating sets G_X and G_Z , using the rows of H_1 and H_2 , respectively. Hence, the problem of constructing a CSS code is reduced to constructing two orthogonal classical binary linear codes.

Entanglement assisted CSS code: Given the entanglement assistance, a quantum CSS code can be built using two classical codes even if they are not orthogonal [50, 51, 52]. Consider parity check matrices H_1 and H_2 associated with two classical codes, and assume $H_1 H_2^\top \neq 0$. Using the rows of H_1 , define the X type generator G_X and using the rows of H_2 , define the Z type generator G_Z . Let \mathcal{S} be the group generated by $G_X \cup G_Z$. The group \mathcal{S} is not a stabilizer group as there are elements in $G_X \cup G_Z$, that anti-commute with each other. However, given any two anti-commuting operators $g_x = X^{u_0} \otimes \dots \otimes X^{u_{N-1}} \in G_X$ and $g_z = X^{v_0} \otimes \dots \otimes X^{v_{N-1}} \in G_Z$, one can modify the operators g_x and g_z by adding an extra noiseless system, so that they commute, as follows,

$$\begin{aligned} g'_x &= X \otimes X^{u_0} \otimes \dots \otimes X^{u_{N-1}}. \\ g'_z &= Z \otimes X^{v_0} \otimes \dots \otimes X^{v_{N-1}}. \end{aligned}$$

It can be seen that $[g'_x, g'_z] = 0$ if $\{g_X, g_Z\} = 0$. The above modification is basically done by sharing an EPR pair between the sender and receiver, such that the receiver's half of the preshared EPR pair is noiseless (see [50] for more details).

1.5 Quantum Polar Coding

As polar coding for classical channels achieves the channel capacity, it is desirable to extend it for quantum channels. Polar codes are first generalized for sending classical information through qubit quantum channels, that is, classical-quantum (cq) channels in [53], achieving the Holevo information as the rate of transmission. The encoding, in this case, is the same as the classical polar codes, hence, it is efficient. For decoding, a quantum analog of SC decoding has been proposed, for which the error probability approaches to zero, exponentially in the codelength. Although, this decoding is not efficient in the sense that it requires collective measurements. Moreover, it has been shown in [54] that one necessarily needs to perform a certain number of collective measurements to achieve the Holevo information.

A generalization of polar codes for quantum communication over qubit-input Pauli and erasure channels is given in [13], which subsumes efficient encoding and decoding, and achieves the symmetric coherent information as the rate of transmission. The quantum polar coding scheme in [13] first associates two classical B-DMCs with a given Pauli or erasure quantum channel, referred to as the *induced* amplitude and phase channels. Then,

it is shown that the recursive construction of polar codes, using a quantum CNOT gate, yields classical channel polarization for both induced channels. Finally, a Calderbank-Shor-Steane (CSS) quantum code is constructed using the classical polar codes on the induced amplitude and phase channels. This construction requires a small number of EPR pairs to be shared between the sender and the receiver, thus making the resulting code entanglement-assisted. However, in the low noise limit, the number of EPR pairs goes to zero. A refined CSS construction is proposed in [55], where preshared entanglement is completely suppressed at the cost of a more complicated multilevel coding scheme. The quantum CSS polar codes are further extended to general quantum channels in [56, 53], which also achieves the symmetric coherent information. For general quantum channels, the associated induced amplitude and phase channels are cq channels instead of classical channels.

In the following, we restrict our attention to Pauli channels and present the CSS construction of polar codes from [13] in more detail, and from a slightly different perspective.

1.5.1 Quantum CSS Polar Code

Notation: We denote $\bar{0} := +$, and $\bar{1} := -$. For any vector $\mathbf{u} = (u_0, \dots, u_{N-1}) \in \{0, 1\}^N$, $\bar{\mathbf{u}}$ is obtained by substituting its components $u_k \in \{0, 1\}$ with \bar{u}_k . Therefore, the N qubit phase basis can be represented as, $\{|\bar{\mathbf{u}}\rangle \mid \mathbf{u} \in \{0, 1\}^N\}$.

We define the quantum polar transform Q_N , as the unitary matrix obtained by replacing the classical CNOT with the quantum one in the classical polar transform P_N . From the definition of the quantum CNOT gate (see Section 1.3.8), it follows that the quantum polar transform Q_N acts like the classical polar transform P_N on a N -qubit amplitude basis state, that is,

$$Q_N|\mathbf{u}\rangle = |P_N\mathbf{u}\rangle, \forall \mathbf{u} \in \{0, 1\}^N. \quad (1.111)$$

The quantum CNOT gate acts as below in the phase basis,

$$C_{2 \rightarrow 1}|\overline{(x, y)}\rangle = |\overline{(x, x \oplus y)}\rangle, \forall (x, y) \in \{0, 1\}^2. \quad (1.112)$$

The above can be easily seen from the circuit equivalence in Figure 1.9,

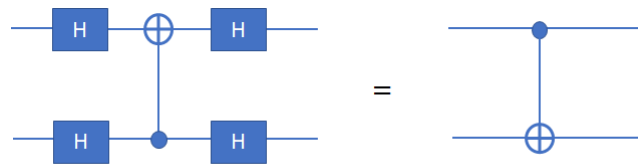


Figure 1.9: The two circuits produce the same output for a given two-qubit input.

Note that the quantum CNOT gate also yields the classical CNOT in the phase basis, only the inputs and outputs are arranged in the reversed order compared to the amplitude basis. Therefore, the quantum polar transform Q_N acts like the classical polar transform on a N -qubit phase basis state, with inputs and outputs arranged in the reversed order compared to the amplitude basis, as illustrated in Figure 1.10. Let P_N^r be the reversed classical polar transform defined as,

$$P_N^r(u_0, \dots, u_i, \dots, u_{N-1}) := RP_N R(u_0, \dots, u_i, \dots, u_{N-1}), \quad (1.113)$$

where R is a permutation matrix defined as

$$R(u_0, \dots, u_i, \dots, u_{N-1}) := (u_{N-1}, \dots, u_{N-1-i}, \dots, u_0). \quad (1.114)$$

Then, we have that

$$Q_N|\bar{\mathbf{u}}\rangle = |\overline{P_N^r \mathbf{u}}\rangle, \forall \mathbf{u} \in \{0, 1\}^N. \quad (1.115)$$

Proposition 51. *We have the following identity*

$$P_N^r = P_N^\top, \quad (1.116)$$

where P_N^\top is the transpose of P_N .

Proof. The equation (1.112) can be written as following,

$$C_{2 \rightarrow 1}|\overline{(x, y)}\rangle = |\overline{P_2^r(x, y)}\rangle, \quad (1.117)$$

where $P_2^r := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = P_2^\top$. Since $P_N^r = P_2^{r \otimes n}$, it follows that $P_N^r = P_N^\top$. \square

We show in Section 1.5.2 that two classical channels, namely the induced amplitude and phase channels, can be associated with a Pauli channel. As the quantum polar transform Q_N acts like the classical polar transforms P_N and P_N^r in amplitude and phase bases, respectively, it implies that Q_N yields the classical channel polarization for both the induced amplitude and phase channel. The CSS quantum polar construction relies on the polarization of the induced amplitude and phase channels, the encoding and decoding for which is given in Sections 1.5.3 and 1.5.4, respectively.

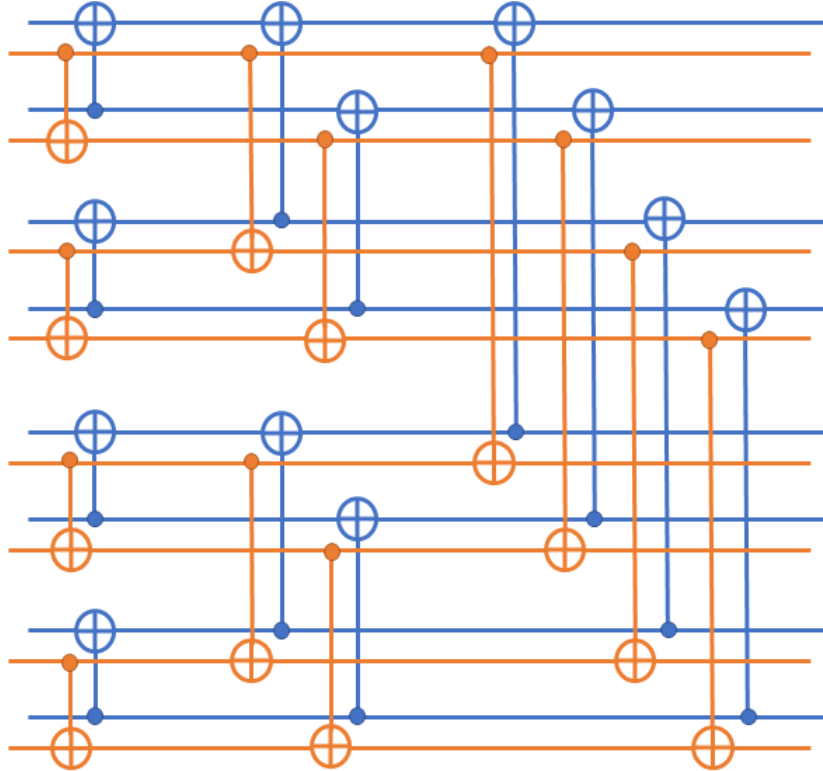


Figure 1.10: Quantum Polar transform for $N = 2^3$ as two classical polar transforms. The blue wires represent quantum information in the amplitude basis and the orange wires in the phase basis.

1.5.2 Induced Amplitude and Phase Channels

Here, we associate two classical channels, induced amplitude and phase channels, with a Pauli channel. Consider a Pauli channel $\mathcal{W}(\rho) := p_I \rho + p_X X \rho X + p_Y Y \rho Y + p_Z Z \rho Z$. In the amplitude basis, only X and Y errors matter, therefore, \mathcal{W} is equal to a channel \mathcal{W}_A , such that

$$\mathcal{W}_A(\rho) = (p_I + p_Z)\rho + (p_X + p_Y)X\rho X. \quad (1.118)$$

Since X corresponds to the bit flip, the above channel acts like a binary symmetric classical channel (BSC) W_A in the amplitude basis, with crossover probability $p_A = p_X + p_Y$, as depicted in Figure 1.11. We shall refer to W_A as the induced amplitude channel.

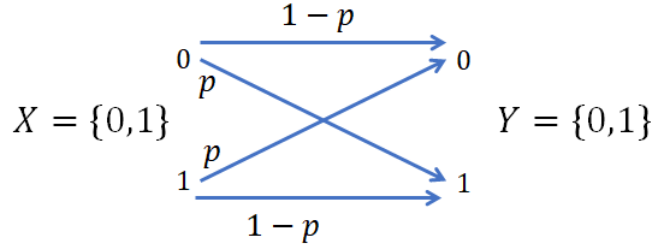


Figure 1.11: A BSC with crossover probability p .

In the phase basis, only Z and Y errors matter. Therefore, \mathcal{W} is equal to a channel \mathcal{W}'_P , such that

$$\mathcal{W}'_P(\rho) = (p_I + p_X)\rho + (p_Z + p_Y)Z\rho Z. \quad (1.119)$$

Since Z corresponds to the phase flip, the above channel acts like a BSC W'_P in the phase basis, with crossover probability $p_P = p_Z + p_Y$.

The classical channels W_A and W'_P do not take into account the correlation between X and Z errors due to Y error. To accomodate the correlation, we assume the knowledge whether X error has occurred or not, and modify the quantum channel \mathcal{W}'_P as below,

$$\mathcal{W}_P(|\bar{x}\rangle\langle\bar{x}|) = (P_I + P_Z)|0\rangle\langle 0|_F \otimes \mathcal{W}_0(|\bar{x}\rangle\langle\bar{x}|) + (P_X + P_Y)|1\rangle\langle 1|_F \otimes \mathcal{W}_1(|\bar{x}\rangle\langle\bar{x}|), \quad (1.120)$$

where the classical flag F indicates whether X error has occurred or not, and

$$\begin{aligned} \mathcal{W}_0(|\bar{x}\rangle\langle\bar{x}|) &:= \frac{p_I}{p_I + p_Z}|\bar{x}\rangle\langle\bar{x}| + \frac{p_Z}{p_I + p_Z}Z|\bar{x}\rangle\langle\bar{x}|Z \\ \mathcal{W}_1(|\bar{x}\rangle\langle\bar{x}|) &:= \frac{p_X}{p_X + p_Y}|\bar{x}\rangle\langle\bar{x}| + \frac{p_Y}{p_X + p_Y}Z|\bar{x}\rangle\langle\bar{x}|Z. \end{aligned} \quad (1.121)$$

The above channels \mathcal{W}_0 and \mathcal{W}_1 act like BSCs W_0 and W_1 , with crossover probabilities $\frac{p_Z}{p_I + p_Z}$ and $\frac{p_Y}{p_X + p_Y}$, respectively. Hence, \mathcal{W}_P acts like a mixture of two BSCs $W_P := \{\mathcal{W}_0, \mathcal{W}_1\}$, chosen with probabilities $P_I + P_Z$ and $P_X + P_Y$, respectively. We shall refer to W_P as the induced phase channel. The transition probabilities of W_P are given by,

$$W_P(u, y|x) = p_u W_u(y|x), \quad u \in \{0, 1\}, \quad (1.122)$$

where $p_0 = p_I + p_Z$ and $p_1 = p_X + p_Y$.

Lemma 52. *The following equality holds,*

$$I(\mathcal{W}) = I(W_A) + I(W_P) - 1, \quad (1.123)$$

where $I(\mathcal{W})$ is the symmetric coherent information of \mathcal{W} and $I(W_A)(I(W_P))$ is the symmetric mutual information of $W_A(W_P)$.

Proof. The symmetric coherent information of the Pauli channel \mathcal{W} is given by,

$$I(\mathcal{W}) = 1 - H(\mathbf{p}), \quad (1.124)$$

where $\mathbf{p} := (p_I, p_X, p_Y, p_Z)$ and $H(\mathbf{p})$ is the Shannon entropy of the probability vector \mathbf{p} , that is, $H(\mathbf{p}) = -p_I \log p_I - p_X \log p_X - p_Y \log p_Y - p_Z \log p_Z$.

It can be seen that the mutual information of a BSC channel with crossover probability ϵ is equal to $1 - H(\epsilon)$, where $H(\epsilon)$ represents the Shannon entropy of a binary probability vector $(\epsilon, 1 - \epsilon)$. Hence, the symmetric mutual information of W_A is given by,

$$I(W_A) = 1 - H(p_X + p_Y), \quad (1.125)$$

Similarly, the mutual information of W_P is given by,

$$\begin{aligned} I(W_P) &= (p_I + p_Z)I(W_0) + (p_X + p_Y)I(W_1) \\ &= (p_I + p_Z)H\left(\frac{p_Z}{p_I + p_Z}\right) + (p_X + p_Y)H\left(\frac{p_Y}{p_X + p_Y}\right) \\ &= 1 - H(\mathbf{p}) + H(p_X + p_Y). \end{aligned} \quad (1.126)$$

From (1.125) and (1.126), we have that

$$I(W_A) + I(W_P) - 1 = 1 - H(\mathbf{p}). \quad (1.127)$$

□

1.5.3 Encoding

To an index $i \in \{0, \dots, N-1\}$, we associate two classical virtual channels as follows,

- The virtual channel $W_A^{(i)}$, obtained by channel combining and splitting procedure on W_A , using the classical polar transform P_N .
- The virtual channel $W_P^{(i^c)}$, where $i^c := N - i - 1$, obtained by channel combining and splitting procedure on W_P , using the reversed classical polar transform P_N^r .

As channel polarization happens for both W_A and W_P , therefore, their respective virtual channels tend to be either completely noiseless or completely noisy as $N \rightarrow \infty$. We call a binary-input classical channel W δ -noiseless if $Z(W) < \delta$, and δ -noisy if $Z(W) > 1 - \delta$. Hence, for any $\delta < \frac{1}{2}$ and sufficiently large N , all but a vanishing fraction of indices from the set $\mathcal{S} := \{0, 1, \dots, N-1\}$ can be grouped in the following four disjoint subsets

- $\mathcal{A} \subseteq \mathcal{S}$ such that $\forall j \in \mathcal{A}$, both $W_A^{(j)}$ and $W_P^{(j^c)}$ are δ -noiseless.
- $\mathcal{B} \subseteq \mathcal{S}$ such that $\forall j \in \mathcal{B}$, $W_A^{(j)}$ is δ -noiseless and $W_P^{(j^c)}$ is δ -noisy.
- $\mathcal{C} \subseteq \mathcal{S}$ such that $\forall j \in \mathcal{C}$, $W_A^{(j)}$ is δ -noisy and $W_P^{(j^c)}$ is δ -noiseless.
- $\mathcal{D} \subseteq \mathcal{S}$ such that $\forall j \in \mathcal{D}$ both $W_A^{(j)}$ and $W_P^{(j^c)}$ are δ -noisy.

Let $\bar{\mathcal{D}}$ be the complement of the set $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$. Note that $|\bar{\mathcal{D}}|$ is almost equal to $|\mathcal{D}|$ for sufficiently large N . With a slight abuse of notation, we denote $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\bar{\mathcal{D}}$ as quantum systems composed of $|\mathcal{A}|$, $|\mathcal{B}|$, $|\mathcal{C}|$, and $|\bar{\mathcal{D}}|$ qubits, respectively. It will be clear from the context when they represent quantum systems or when sets of indices.

We first set the quantum state of systems $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\bar{\mathcal{D}}$ as follows,

- The quantum state corresponding to the system \mathcal{A} is an arbitrary quantum state $\rho_{\mathcal{A}} \in \mathcal{D}(\mathcal{H}_{\mathcal{A}})$, which is to be transmitted to the receiver.
- The quantum state of the system \mathcal{B} is frozen (fixed) by the quantum state $\rho_{\mathcal{B}}^+ := \otimes_{b \in \mathcal{B}} |+\rangle\langle+|_b$.
- The quantum state of the system \mathcal{C} is frozen by the quantum state $\rho_{\mathcal{C}}^0 := \otimes_{c \in \mathcal{C}} |0\rangle\langle 0|_c$.
- Let $\Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}$ be the maximally entangled quantum state, defined on isomorphic quantum systems $\bar{\mathcal{D}}$ and $\bar{\mathcal{D}}'$ each containing $|\bar{\mathcal{D}}|$ qubits, as follows,

$$\Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'} := \otimes_{d \in \bar{\mathcal{D}}} \Phi_{dd'}, \quad (1.128)$$

where d and d' are the d th qubit of the quantum systems $\bar{\mathcal{D}}$ and $\bar{\mathcal{D}}'$. The quantum state of the system $\bar{\mathcal{D}}$ is frozen by the $\bar{\mathcal{D}}$ part of $\Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}$. The other part $\bar{\mathcal{D}}'$ is directly given to the decoder.

Then, the quantum polar transform Q_N is applied on the $\mathcal{ABC}\bar{\mathcal{D}}$ part, which gives the following quantum state,

$$\varphi_{\mathcal{ABC}\bar{\mathcal{D}}\bar{\mathcal{D}}'} = Q_N \otimes I_{\bar{\mathcal{D}}'} (\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}^+ \otimes \rho_{\mathcal{C}}^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) Q_N^\dagger \otimes I_{\bar{\mathcal{D}}'}. \quad (1.129)$$

Generators of the quantum polar code stabilizer group: The quantum polar code can be seen as a stabilizer code as follows. The quantum state $(\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}^+ \otimes \rho_{\mathcal{C}}^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'})$ for any $\rho_{\mathcal{A}} \in \mathcal{D}(\mathcal{H}_{\mathcal{A}})$ is stabilized by the following Pauli operators,

$$X_b \otimes_{k \neq b} I_k, \quad \forall b \in \mathcal{B}. \quad (1.130)$$

$$Z_c \otimes_{k \neq c} I_k, \quad \forall c \in \mathcal{C}. \quad (1.131)$$

$$X_d \otimes X_{d'} \otimes_{k \neq d, d'} I_k, \quad \forall d \in \bar{\mathcal{D}}. \quad (1.132)$$

$$Z_d \otimes Z_{d'} \otimes_{k \neq d, d'} I_k, \quad \forall d \in \bar{\mathcal{D}}. \quad (1.133)$$

Therefore, the stabilizer group of the polar code is generated by,

$$Q_N \otimes I_{\bar{\mathcal{D}}'} (X_b \otimes_{k \neq b} I_k) Q_N^\dagger \otimes I_{\bar{\mathcal{D}}'}, \quad \forall b \in \mathcal{B}. \quad (1.134)$$

$$Q_N \otimes I_{\bar{\mathcal{D}}'} (Z_c \otimes_{k \neq c} I_k) Q_N^\dagger \otimes I_{\bar{\mathcal{D}}'}, \quad \forall c \in \mathcal{C}. \quad (1.135)$$

$$Q_N \otimes I_{\bar{\mathcal{D}}'} (X_d \otimes X_{d'} \otimes_{k \neq d, d'} I_k) Q_N^\dagger \otimes I_{\bar{\mathcal{D}}'}, \quad \forall d \in \bar{\mathcal{D}}. \quad (1.136)$$

$$Q_N \otimes I_{\bar{\mathcal{D}}'} (Z_d \otimes Z_{d'} \otimes_{k \neq d, d'} I_k) Q_N^\dagger \otimes I_{\bar{\mathcal{D}}'}, \quad \forall d \in \bar{\mathcal{D}}. \quad (1.137)$$

Note that the following holds for the CNOT gate,

$$C_{2 \rightarrow 1}(X \otimes I)C_{2 \rightarrow 1} = X \otimes I \quad (1.138)$$

$$C_{2 \rightarrow 1}(I \otimes X)C_{2 \rightarrow 1} = X \otimes X \quad (1.139)$$

$$C_{2 \rightarrow 1}(Z \otimes I)C_{2 \rightarrow 1} = Z \otimes Z \quad (1.140)$$

$$C_{2 \rightarrow 1}(I \otimes Z)C_{2 \rightarrow 1} = I \otimes Z \quad (1.141)$$

From the above four equations, it follows that any stabilizer generator from (1.134)-(1.137) consists of either X and I , or Z and I . Therefore, from Definition 50, the set of stabilizer generators of the quantum polar code gives a CSS code. However, as the preshared EPR pairs are required, it is an entanglement assisted CSS code.

Rate of the quantum CSS polar code: The rate of the quantum CSS polar code is given by,

$$R = \frac{|\mathcal{A}|}{N}. \quad (1.142)$$

As $N \rightarrow \infty$, from Theorem 7, we have that

$$|\mathcal{A}| + |\mathcal{B}| \rightarrow NI(W_A). \quad (1.143)$$

$$|\mathcal{A}| + |\mathcal{C}| \rightarrow NI(W_P). \quad (1.144)$$

$$|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}| + |\mathcal{D}| \rightarrow N. \quad (1.145)$$

From the above three equations, we have that

$$\frac{|\mathcal{A}|}{N} \rightarrow (I(W_A) + I(W_P) - 1) + \frac{|\mathcal{D}|}{N}. \quad (1.146)$$

Hence, the quantum CSS polar code achieves the symmetric coherent information plus the fraction of preshared EPR pairs. Hence, the *net communication rate*, that is, the rate less than the fraction of preshared EPR pair, approaches the symmetric coherent information $I(\mathcal{W})$.

Entanglement free condition: As mentioned above, preshared EPR pairs are needed for the indices in the set $\bar{\mathcal{D}}$. In the following lemma, we give a condition on the Bhattacharyya parameters of W_A, W_P , which guarantees that $|\mathcal{D}| \rightarrow 0$ as $N \rightarrow \infty$. In other words, the number of preshared EPR pairs goes to zero for sufficiently large N .

Lemma 53 ([13]). *If the constraint $Z(W_A) + Z(W_P) \leq 1$ holds, then $|\mathcal{D}| \rightarrow 0$ as $N \rightarrow \infty$.*

Proof. From (1.23) and (1.24), we have the following after the first polarization step,

$$\begin{aligned} Z(W_A^{(0)}) + Z(W_P^{(1)}) &\leq 2Z(W_A) - Z(W_A)^2 + Z(W_P)^2 \\ &\leq 2Z(W_A) - Z(W_A)^2 + (1 - Z(W_A))^2 \\ &= 1. \end{aligned} \quad (1.147)$$

Applying the above recursively, it can be seen that following holds for any $i \in S$,

$$Z(W_A^{(i)}) + Z(W_P^{(ic)}) \leq 1. \quad (1.148)$$

As $N \rightarrow \infty$, from the polarization theorem (Theorem 7), we have that the parameters $Z(W_A^{(i)})$ and $Z(W_P^{(ic)})$ approach to either 0 or 1. The above equation implies that both $Z(W_A^{(i)})$ and $Z(W_P^{(ic)})$ can't approach 1 together. Hence, $|\mathcal{D}| \rightarrow 0$ as $N \rightarrow \infty$. \square

1.5.4 Decoding

The quantum system $ABC\bar{\mathcal{D}}$ is sent using N instances of the Pauli channel \mathcal{W} . As no errors occur on the system $\bar{\mathcal{D}}'$, the following is the channel output,

$$\psi_{ABC\bar{\mathcal{D}}\bar{\mathcal{D}}'} := (\mathcal{W}^{\otimes N} \otimes I_{\bar{\mathcal{D}}'}) (\varphi_{ABC\bar{\mathcal{D}}\bar{\mathcal{D}}'}). \quad (1.149)$$

Since \mathcal{W} is a Pauli channel, we have that

$$\psi_{ABC\bar{\mathcal{D}}\bar{\mathcal{D}}'} = (E_{ABC\bar{\mathcal{D}}} Q_N \otimes I_{\bar{\mathcal{D}}'}) (\rho_A \otimes \rho_B^+ \otimes \rho_C^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) (Q_N^\dagger E_{ABC\bar{\mathcal{D}}}^\dagger \otimes I_{\bar{\mathcal{D}}'}), \quad (1.150)$$

for a random N -qubit Pauli error $E_{ABC\bar{\mathcal{D}}} \in \mathcal{G}_N \setminus \{\pm 1, \pm i\}$. The decoding is performed in the following steps,

Step 1: Apply the inverse quantum polar transform on the channel output state. Applying Q_N^\dagger on the output state $\psi_{ABC\bar{\mathcal{D}}\bar{\mathcal{D}}'}$, we have that

$$\begin{aligned} Q_N^\dagger \psi_{ABC\bar{\mathcal{D}}\bar{\mathcal{D}}'} Q_N &= (Q_N^\dagger E_{ABC\bar{\mathcal{D}}} Q_N \otimes I_{\bar{\mathcal{D}}'}) (\rho_A \otimes \rho_B^+ \otimes \rho_C^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) (Q_N^\dagger E_{ABC\bar{\mathcal{D}}}^\dagger Q_N \otimes I_{\bar{\mathcal{D}}'}) \\ &= (E'_{ABC\bar{\mathcal{D}}} \otimes I_{\bar{\mathcal{D}}'}) (\rho_A \otimes \rho_B^+ \otimes \rho_C^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) (E'_{ABC\bar{\mathcal{D}}} \otimes I_{\bar{\mathcal{D}}'}), \end{aligned}$$

where $E'_{ABCD} := Q_N^\dagger E_{ABCD} Q_N$. From (1.138)-(1.141), it follows that $E'_{ABCD} \in \mathcal{G}_N \setminus \{\pm 1, \pm i\}$ is also a Pauli error.

Step 2: Quantum measurement. Let $E'_{ABCD} = X^{u'_0} Z^{v'_0} \otimes \dots \otimes X^{u'_N} Z^{v'_N}$ and let \mathbf{u}'_X and \mathbf{v}'_Z be the X and Z error vectors of E'_{ABCD} , respectively, that is,

$$\mathbf{u}'_X := (u'_0, \dots, u'_{N-1}).$$

$$\mathbf{v}'_Z := (v'_0, \dots, v'_{N-1}).$$

The receiver performs the Pauli X measurement on each $b \in \mathcal{B} \subset \mathcal{S}$, which determines the components of \mathbf{v}'_Z corresponding to the set \mathcal{B} . Further, the Pauli Z measurement on each $c \in \mathcal{C}$, which determines the components of \mathbf{u}'_X corresponding to the set \mathcal{C} . Finally, the receiver performs the *Bell measurement*, that is, the measurement corresponding to the Pauli operators $X \otimes X$ and $Z \otimes Z$, on the two-qubit system dd' for each $d \in \bar{\mathcal{D}} \subset \mathcal{S}$, which determines the components of both \mathbf{u}'_X and \mathbf{v}'_Z corresponding to $\bar{\mathcal{D}}$.

Step 3: Decode the classical polar codes on W_A and W_P . Let \mathbf{u}_X and \mathbf{v}_Z be X and Z error vectors, respectively, corresponding to the error $E_{ABCD} \in \bar{\mathcal{G}}_N$. Since $E'_{ABCD} = Q_N^\dagger E_{ABCD} Q_N$, from (1.138)-(1.141), we have the following equalities,

$$\mathbf{u}'_X = P_N \mathbf{u}_X. \quad (1.151)$$

$$\mathbf{v}'_Z = P_N^r \mathbf{v}_Z. \quad (1.152)$$

Further, as $P_N^2 = P_N^{r^2} = I$, we also have the following,

$$\mathbf{u}_X = P_N \mathbf{u}'_X. \quad (1.153)$$

$$\mathbf{v}_Z = P_N^r \mathbf{v}'_Z. \quad (1.154)$$

The vectors \mathbf{u}_X and \mathbf{v}_Z are decoded as follows,

(i) **Decoding of \mathbf{u}'_X :** note that when the *all-zero vector* $0^N := (0, \dots, 0)$ is input to the N instances of W_A , denoted by W_A^N , then the vector \mathbf{u}_X can be considered as a channel output of W_A^N . As the channel W_A is a BSC, we have that

$$W_A^N(\mathbf{u}_X | 0^N) = W_A^N(0^N | \mathbf{u}_X). \quad (1.155)$$

Therefore, we can equivalently consider 0^N as the observed channel output, and \mathbf{u}_X (unknown) as the channel input. We have been given,

- The components of the vector \mathbf{u}'_X corresponding to the set $\mathcal{C} \cup \bar{\mathcal{D}} \subset \mathcal{S}$.
- A noisy observation of the channel input $\mathbf{u}_X = P_N \mathbf{u}'_X$ (namely 0^N).

Based on the above, we can determine the components of the vector \mathbf{u}'_X corresponding to the set $(\mathcal{A} \cup \mathcal{B})$, using the SC decoding of polar codes as described in Section 1.2.3.

(ii) **Decoding of \mathbf{v}'_Z :** Recall the W_P is a mixture of two BSCs W_0 and W_1 , which are chosen with some probability (see (1.122) for the transition probability of W_P). Note that when 0^N is input to W_P^N (N instances of W_P), the pair $(\mathbf{u}_x, \mathbf{v}_Z)$ can be considered as a channel output, where the vector \mathbf{u}_x is known and the vector \mathbf{v}_Z is unknown. As channels W_0 and W_1 are BSCs, from (1.122), we have that

$$W_P^N(\mathbf{u}_x, \mathbf{v}_Z | 0^N) = W_P^N(\mathbf{u}_x, 0^N | \mathbf{v}_Z). \quad (1.156)$$

Therefore, we can equivalently consider $(\mathbf{u}_x, 0^N)$ as the channel output and \mathbf{v}_Z as the channel input. We have been given,

- The components of the vector v'_Z corresponding to the set $\mathcal{B} \cup \mathcal{D} \subset \mathcal{S}$.
- A noisy observation of the channel input $v_Z = F_N^r v'_Z$ (namely $(u_X, 0^N)$).

Based on the above, we can determine the components of the vector v'_Z corresponding to the set $(\mathcal{A} \cup \mathcal{C})$, using the SC decoding of polar codes.

2

Purely Quantum Polar codes

Quantum CSS polar code (for qubit-input quantum channels) presented in the previous chapter takes advantage of the fact that the recursive construction with the CNOT gate (channel combining operation) yields channel polarization for both induced amplitude and phase channels, which are associated with a quantum channel. The induced amplitude and phase channels associated with a Pauli channel are classical channels and for general quantum channels, they are classical-quantum channels. Therefore, in general, quantum CSS polar coding is based on the polarization of two classical-quantum channels.

In this chapter, we take a different approach to construct quantum polar codes. We prove a channel polarization phenomenon for quantum channels, where synthesized virtual channels tend to be completely noisy or noiseless as quantum channels, not just in one basis. We refer to this phenomenon as *purely quantum polarization*. Here, we prove the purely quantum polarization for quantum channels with qudit-input of dimension $d \geq 2$.

Our purely quantum polarization is based on a quantum channel combining and splitting procedure, where a two-qudit unitary, randomly chosen from a unitary 2-design (see Section 1.3.5) is used as a channel combining operation. Further, using the fact that the generalized two-qudit Clifford group (Definition 14) is a unitary 2-design [57] (we also provide a simple proof of this fact), we conclude that the channel combining operation can be randomly chosen from this set.

Using a symmetry argument, we reduce the set from which the channel combining operation is chosen. Precisely, when the qudit dimension d is a prime, we show that polarization happens for a subset of two-qudit Clifford unitaries containing only $d^4 + d^2 - 2$ elements, which is not a unitary 2-design. Hence, unitary 2-designs are not necessary for the quantum polarization of qudit-input channels. When $d = 2$, the channel combining set can be further reduced to $\frac{d^4 + d^2 - 2}{2} = 9$ elements.

We exploit the purely quantum polarization to construct a quantum coding scheme, where the virtual channels that are completely noiseless as quantum channels are used for quantum communication, while the virtual channels that are completely noisy are frozen using preshared EPR pairs. Hence, our coding scheme is entanglement-assisted. Further, it achieves half the symmetric mutual information of the channel.

2.1 A General Set of Conditions for Stochastic Process Polarization

Recall from the discussion regarding the proof of classical polarization (Theorem 7) that the recursive application of the channel combining and splitting procedure can be modeled as a discrete time stochastic process $\{I_n : n \geq 0\}$, where $I_n = I(W^{i_1 \dots i_n})$ is the mutual information of the virtual channel $W^{i_1 \dots i_n}$. Further, the proof of polarization of I_n therein is done indirectly using the Bhattacharyya parameter.

The following lemma provides a general set of conditions for polarization of an information measure I , using a Bhattacharyya like parameter T , under the recursive application of a channel combining and splitting procedure. The lemma below is a slightly modified version of [24, Lemma 2], so as to meet our specific needs.

Lemma 54 ([24, Lemma 2]). *Suppose B_i , $i = 1, 2, \dots$ are independent and identically distributed (i.i.d.), $\{0, 1\}$ -valued random variables with $P(B_1 = 0) = P(B_1 = 1) = 1/2$, defined on a probability space (Ω, \mathcal{F}, P) . Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as the trivial σ -algebra and set \mathcal{F}_n , $n \geq 1$, to be the σ -field generated by $(B_1 \dots B_n)$. Suppose further that two stochastic processes $\{I_n : n \geq 0\}$ and $\{T_n : n \geq 0\}$ are defined on this probability space with the following properties:*

- (i.1) I_n takes values in $[\iota_0, \iota_1]$ and is measurable with respect to \mathcal{F}_n . That is, I_0 is a constant, and I_n is a function of $B_1 \dots B_n$.
- (i.2) $\{(I_n, \mathcal{F}_n) : n \geq 0\}$ is a martingale, i.e., $\mathbb{E}_{B_{n+1}}[I_{n+1} \mid I_n = i_n, I_{n-1} = i_{n-1}, \dots, I_1 = i_1, I_0] = i_n$, for any $n \geq 0$ and all possible i_1, \dots, i_n .
- (t.1) T_n takes values in the interval $[\theta_0, \theta_1]$ and is measurable with respect to \mathcal{F}_n .
- (i&t.1) For any $\epsilon > 0$ there exists $\delta > 0$, such that $I_n \in (\iota_0 + \epsilon, \iota_1 - \epsilon)$ implies $T_n \in (\theta_0 + \delta, \theta_1 - \delta)$.
- (t.2) **Guaranteed improvement:** $T_{n+1} \leq f(T_n)$ when $B_{n+1} = 1$, where $f : [\theta_0, \theta_1] \rightarrow [\theta_0, \theta_1]$ is a continuous function, such that $f(\theta) < \theta, \forall \theta \in (\theta_0, \theta_1)$.

Then, $I_\infty := \lim_{n \rightarrow \infty} I_n$ exists with probability 1, I_∞ takes values in $\{\iota_0, \iota_1\}$, and $\mathbb{E}(I_\infty) := \iota_0 P(I_\infty = \iota_0) + \iota_1 P(I_\infty = \iota_1) = I_0$.

Proof. As the process $\{(I_n, \mathcal{F}_n) : n \geq 0\}$ is a martingale, it follows that almost surely I_n converges to a limit value. This means that for any $\epsilon > 0$, there exists a n_0 such that for all $n > n_0$, we have that $|I_n - \iota| \leq \epsilon$, for some $\iota \in [\iota_0, \iota_1]$. It can be seen that $\iota \in \{\iota_0, \iota_1\}$ as follows,

Suppose for $\delta > 0$, $\iota \in (\iota_0 + \delta, \iota_1 - \delta)$ holds, then it follows that there exists a n_0 , such that for all $n > n_0$, I_n takes values in $(\iota_0 + \delta, \iota_1 - \delta)$. From point (i&t.1), we have that $I_n \in (\iota_0 + \delta, \iota_1 - \delta) \implies T_n \in (\theta_0 + \delta', \theta_1 - \delta')$ for some $\delta' > 0$. For some $n > n_0$, consider an event E_n , such that $B_n = B_{n+1} = \dots = B_{n+k} = 1$ for $k > 0$. We have the following,

- (i) $B_i, i = 0, 1, \dots$ are i.i.d random variables.
- (ii) The probability of E_n for any n is equal to $2^{-k} > 0$, therefore we have $\sum_n \Pr(E_n) = \infty$.

Hence, from the Borel-Cantelli lemma, it follows that E_n occurs infinitely often for any finite $k > 0$. From point (t.2), when E_n occurs for some $n > n_0$, we have that

$$T_{n+k} \leq f^k(T_n), \quad (2.1)$$

where f^k is the k -fold composition of f . As f is a continuous function and $f(T_n) < T_n$, we have that for any $\delta' > 0$ and n such that $T_n \in (\theta_0 + \delta', \theta_1 - \delta')$, there exists a k such that $f^k(T_n) < \theta_0 + \delta'$. Hence, it follows that I_n does not converge to any value in $(\iota_0 + \delta, \iota_1 - \delta)$, therefore, we have that $\iota \in \{\iota_0, \iota_1\}$. Finally, using the martingale property (i.2), we have that

$$\mathbb{E}(I_\infty) := \iota_0 P(I_\infty = \iota_0) + \iota_1 P(I_\infty = \iota_1) = I_0. \quad (2.2)$$

□

2.2 Purely Quantum Polarization

In this section, we show our purely quantum polarization phenomenon for quantum channels. To do so, we introduce a channel combining and splitting procedure for quantum channels. We show that all the conditions of polarization from Section 2.1 are satisfied for the stochastic process obtained by recursively applying this channel combining and splitting procedure, hence polarization happens.

2.2.1 Channel Combining and Splitting Procedure

Consider two quantum channels $\mathcal{N}_{A' \rightarrow B}$ and $\mathcal{M}_{A' \rightarrow B}$, where A' is a qudit quantum system of dimension $d \geq 2$, and B is a qudit quantum system of arbitrary dimension. Our quantum polarization scheme relies on the channel combining and splitting procedures depicted in Figure 2.1.

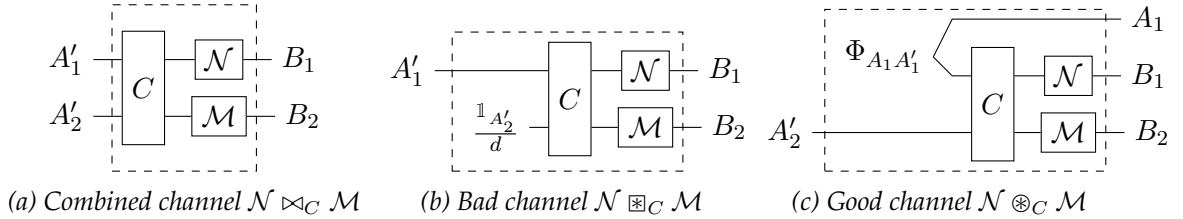


Figure 2.1: Channel combining and splitting. (a) combined channel: a two-qudit unitary C is applied on the two inputs. (b) bad channel: we input a totally mixed state into the second input. (c) good channel: we input half of an EPR pair into the first input, and the other half is given as the output A_1 .

First, quantum channels $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$ are combined, using a two-qudit unitary C , which yields the quantum channel $\mathcal{N} \bowtie_C \mathcal{M}$, as depicted in Figure 2.1a. The combined channel $\mathcal{N} \bowtie_C \mathcal{M}$ is a quantum channel from $A'_1 A'_2$ to $B_1 B_2$, which acts as follows on a two-qudit quantum state $\phi_{AA'} \in \mathcal{D}(\mathcal{H}_{AA'})$,

$$(\mathcal{N} \bowtie_C \mathcal{M})_{A'_1 A'_2 \rightarrow B_1 B_2} \left(\phi_{A'_1 A'_2} \right) := \mathcal{N}_{A'_1 \rightarrow B_1} \otimes \mathcal{M}_{A'_2 \rightarrow B_2} \left(C \phi_{A'_1 A'_2} C^\dagger \right). \quad (2.3)$$

The combined channel is then split into two quantum virtual channels, the bad channel $\mathcal{N} \boxtimes_C \mathcal{M}$, and the good channel $\mathcal{N} \otimes_C \mathcal{M}$, as depicted in Figures 2.1b and 2.1c, respectively.

The bad channel $\mathcal{N} \boxtimes_C \mathcal{M}$ is a channel from A'_1 to $B_1 B_2$, which acts as follows on a quantum state $\rho_{A'_1} \in \mathcal{D}(\mathcal{H}_{A'_1})$,

$$(\mathcal{N} \boxtimes_C \mathcal{M})_{A'_1 \rightarrow B_1 B_2} (\rho_{A'_1}) := (\mathcal{N} \bowtie_C \mathcal{M})_{A'_1 A'_2 \rightarrow B_1 B_2} \left(\rho_{A'_1} \otimes \frac{\mathbb{1}_{A'_2}}{d} \right). \quad (2.4)$$

The good channel $\mathcal{N} \otimes_C \mathcal{M}$ is a quantum channel from A'_2 to $A_1 B_1 B_2$, which acts as follows on a quantum state $\rho_{A'_2} \in \mathcal{D}(\mathcal{H}_{A'_2})$,

$$(\mathcal{N} \otimes_C \mathcal{M})_{A'_2 \rightarrow A_1 B_1 B_2}(\rho_{A'_2}) := (\mathcal{N} \boxtimes_C \mathcal{M})_{A'_1 A'_2 \rightarrow B_1 B_2} \left(\Phi_{A_1 A'_1} \otimes \rho_{A'_2} \right), \quad (2.5)$$

where $\Phi_{A_1 A'_1}$ is an EPR pair (Definition 11).

2.2.2 Properties of the Channel Combining and Splitting

In the following lemma, we show that the complements of the good and bad channels yield the bad and good channels, respectively on the complements of $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$.

Lemma 55. *For quantum channels $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$, we have that*

$$\begin{aligned} (a) \quad (\mathcal{N} \boxtimes_C \mathcal{M})_{A'_1 \rightarrow E_1 E_2 A_2}^c(\rho_{A'_1}) &= \mathcal{N}_{A'_1 \rightarrow E_1}^c \otimes \mathcal{M}_{A'_2 \rightarrow E_2}^c \left(C \left(\rho_{A'_1} \otimes \Phi_{A'_2 A_2} \right) C^\dagger \right), \\ (b) \quad (\mathcal{N} \otimes_C \mathcal{M})_{A'_2 \rightarrow E_1 E_2}^c(\rho_{A'_2}) &= \mathcal{N}_{A'_1 \rightarrow E_1}^c \otimes \mathcal{M}_{A'_2 \rightarrow E_2}^c \left(C \left(\frac{1_{A'_1}}{d} \otimes \rho_{A'_2} \right) C^\dagger \right), \end{aligned}$$

where $\mathcal{N}_{A'_1 \rightarrow E_1}^c$ and $\mathcal{M}_{A'_2 \rightarrow E_2}^c$ are complementary channels of $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$, respectively.

Proof. Let $U_{A'_1 \rightarrow B_1 E_1}$ and $V_{A'_2 \rightarrow B_2 E_2}$ be the Stinespring dilations of $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$, respectively. Define isometries $W_{A'_1 \rightarrow B_1 B_2 E_1 E_2 A_2}$ and $W'_{A'_2 \rightarrow B_1 B_2 E_1 E_2 A_1}$ as following,

$$W(\rho_{A'_1})W^\dagger := U \otimes V \left(C \left(\rho_{A'_1} \otimes \Phi_{A'_2 A_2} \right) C^\dagger \right) U^\dagger \otimes V^\dagger. \quad (2.6)$$

$$W'(\rho_{A'_2})W'^\dagger := U \otimes V \left(C \left(\Phi_{A_1 A'_1} \otimes \rho_{A'_2} \right) C^\dagger \right) U^\dagger \otimes V^\dagger. \quad (2.7)$$

It can be seen that

$$\text{tr}_{E_1 E_2 A_2}(W(\rho_{A'_1})W^\dagger) = \mathcal{N} \otimes \mathcal{M} \left(C \left(\rho_{A'_1} \otimes \frac{1_{A'_2}}{d} \right) C^\dagger \right) = (\mathcal{N} \boxtimes_C \mathcal{M})(\rho_{A'_1}). \quad (2.8)$$

$$\text{tr}_{E_1 E_2}(W'(\rho_{A'_2})W'^\dagger) = \mathcal{N} \otimes \mathcal{M} \left(C \left(\Phi_{A_1 A'_1} \otimes \rho_{A'_2} \right) C^\dagger \right) = (\mathcal{N} \otimes_C \mathcal{M})(\rho_{A'_2}). \quad (2.9)$$

Therefore, isometries $W_{A'_1 \rightarrow B_1 B_2 E_1 E_2 A_2}$ and $W'_{A'_2 \rightarrow B_1 B_2 E_1 E_2 A_1}$ are Stinespring dilations of the channels $(\mathcal{N} \boxtimes_C \mathcal{M})$ and $(\mathcal{N} \otimes_C \mathcal{M})$, respectively (see also [58, Theorem 1]). Therefore, by tracing out channel outputs of $(\mathcal{N} \boxtimes_C \mathcal{M})$ and $(\mathcal{N} \otimes_C \mathcal{M})$, we get their respective complementary channels,

$$(\mathcal{N} \boxtimes_C \mathcal{M})^c(\rho_{A'_1}) = \text{tr}_{B_1 B_2}(W(\rho_{A'_1})W^\dagger) = \mathcal{N}^c \otimes \mathcal{M}^c \left(C \left(\rho_{A'_1} \otimes \Phi_{A'_2 A_1} \right) C^\dagger \right). \quad (2.10)$$

$$(\mathcal{N} \otimes_C \mathcal{M})^c(\rho_{A'_2}) = \text{tr}_{A_1 B_1 B_2}(W'(\rho_{A'_2})W'^\dagger) = \mathcal{N}^c \otimes \mathcal{M}^c \left(C \left(\frac{1_{A'_1}}{d} \otimes \rho_{A'_2} \right) C^\dagger \right). \quad (2.11)$$

□

In the following lemma, we show that the channel combining and splitting procedure preserves the total symmetric coherent information.

Lemma 56. *Given two channels $\mathcal{N}_{A'_1 \rightarrow B_1}$ and $\mathcal{M}_{A'_2 \rightarrow B_2}$ with qudit inputs, then*

$$I(\mathcal{N} \otimes_C \mathcal{M}) + I(\mathcal{N} \boxtimes_C \mathcal{M}) = I(\mathcal{N}) + I(\mathcal{M}),$$

and this holds for all choices of C .

Proof. Consider the following state,

$$\rho_{A_1 A_2 B_1 B_2} = (\mathcal{N}_{A'_1 \rightarrow B_1} \otimes \mathcal{M}_{A'_2 \rightarrow B_2})(C(\Phi_{A_1 A'_1} \otimes \Phi_{A_2 A'_2})C^\dagger). \quad (2.12)$$

Then, we have that

$$I(\mathcal{N} \boxtimes_C \mathcal{M}) = -H(A_1|B_1 B_2)_\rho \quad (2.13)$$

$$I(\mathcal{N} \otimes_C \mathcal{M}) = -H(A_2|A_1 B_1 B_2)_\rho. \quad (2.14)$$

Therefore, by the chain rule,

$$\begin{aligned} I(\mathcal{N} \boxtimes_C \mathcal{M}) + I(\mathcal{N} \otimes_C \mathcal{M}) &= -H(A_1|B_1 B_2)_\rho - H(A_2|A_1 B_1 B_2)_\rho \\ &= -H(A_1 A_2|B_1 B_2)_\rho. \end{aligned} \quad (2.15)$$

Now, recall that the EPR pair has the property that $(Z \otimes \mathbb{1})|\Phi\rangle = (\mathbb{1} \otimes Z^\top)|\Phi\rangle$ for any matrix Z . Using this, we can move C from the input systems A'_1 and A'_2 to the purifying systems $A_1 A_2$: $\rho = C^\top(\mathcal{N} \otimes \mathcal{M})(\Phi_{A_1 A'_1} \otimes \Phi_{A_2 A'_2})\bar{C}$, where \bar{C} is the complex conjugate of C . Hence, we have

$$\begin{aligned} -H(A_1 A_2|B_1 B_2)_\rho &= -H(A_1 A_2|B_1 B_2)_{(\mathcal{N} \otimes \mathcal{M})(\Phi_{A_1 A'_1} \otimes \Phi_{A_2 A'_2})} \\ &= -H(A_1|B_1)_{\mathcal{N}(\Phi_{A_1 A'_1})} - H(A_2|B_2)_{\mathcal{M}(\Phi_{A_2 A'_2})} \\ &= I(\mathcal{N}) + I(\mathcal{M}). \end{aligned}$$

□

2.2.3 Rényi-Bhattacharyya Parameter

Here, we introduce a new channel parameter, based on the quantum conditional Rényi entropies $H_{\frac{1}{2}}^\uparrow$ and \tilde{H}_2^\downarrow (see Definitions 37, 38, and also (1.76) and (1.77)). This parameter will play the role of the Bhattacharyya parameter in the proof of our polarization Theorem 60 below. For this reason, we call it the Rényi-Bhattacharyya parameter.

Definition 57 (Rényi-Bhattacharyya parameter). *For any quantum channel $\mathcal{N}_{A' \rightarrow B}$ with qudit input A' , the Rényi-Bhattacharyya parameter is defined as,*

$$R(\mathcal{N}) := d^{\frac{H_{\frac{1}{2}}^\uparrow(A|B)_{\mathcal{N}(\Phi_{AA'})}}{2}} = d^{-\tilde{H}_2^\downarrow(A|E)_{\mathcal{N}^c(\Phi_{AA'})}} \in \left[\frac{1}{d}, d\right], \quad (2.16)$$

where $\Phi_{AA'}$ is an EPR pair and $\mathcal{N}_{A' \rightarrow E}^c$ is a complementary channel of \mathcal{N} .

Note that the second equality in (2.16) follows from the duality relation of the conditional quantum entropies in Theorem 39. Further, as mentioned before, the choice of the complementary channel is not unique. However, the value of $\tilde{H}_2^\downarrow(A|E)_{\mathcal{N}^c(\Phi_{AA'})}$ is independent of the choice.

We now give the following relation between the Rényi-Bhattacharyya parameter and the coherent information.

Lemma 58. *Let $\mathcal{W}_{A' \rightarrow B}$ be a channel with qudit input. Then, for any $\delta > 0$, we have that*

$$(a) \quad R(\mathcal{W}) \leq \frac{1}{d} + \delta \Rightarrow I(\mathcal{W}) \geq 1 - \log(1 + d\delta).$$

$$(b) \quad R(\mathcal{W}) \geq d - \delta \Rightarrow I(\mathcal{W}) \leq -1 + 2\sqrt{\frac{\delta}{d}} + \frac{\sqrt{d} + \sqrt{\delta}}{\sqrt{d}} H\left(\frac{\sqrt{\delta}}{\sqrt{d} + \sqrt{\delta}}\right), \text{ where } H(\cdot) \text{ denotes the binary entropy function.}$$

Proof. **Point (a).** For $\rho_{AB} = \mathcal{W}(\Phi_{AA'})$, we have that

$$\frac{1}{d} + \delta \geq R(\mathcal{W}) = d^{\frac{H_{\frac{1}{2}}^{\uparrow}(A|B)_{\rho}}{2}} \geq d^{H(A|B)_{\rho}} = d^{-I(\mathcal{W})},$$

where we have used $H_{\frac{1}{2}}^{\uparrow}(A|B)_{\rho} \geq H(A|B)_{\rho}$ for the second inequality, which follows from the monotonically decreasing property of the conditional Petz-Rényi entropy with respect to its order [36, Theorem 7]. Hence, $I(\mathcal{W}) \geq 1 - \log(1 + d\delta)$.

Point (b). We have that

$$\begin{aligned} d - \delta &\leq R(\mathcal{W}) \leq R(\mathcal{W}) \\ &= \max_{\sigma_B} \text{Tr} \left[\rho_{AB}^{\frac{1}{2}} \sigma_B^{\frac{1}{2}} \right]^2 \\ &= d \max_{\sigma_B} \text{Tr} \left[\sqrt{\rho_{AB}} \sqrt{\frac{\mathbb{1}_A}{d} \otimes \sigma_B} \right]^2 \\ &\leq d \max_{\sigma_B} \left\| \sqrt{\rho_{AB}} \sqrt{\frac{\mathbb{1}_A}{d} \otimes \sigma_B} \right\|_1^2 \end{aligned} \tag{2.17}$$

$$= d \max_{\sigma_B} F \left(\rho_{AB}, \frac{\mathbb{1}_A}{d} \otimes \sigma_B \right)^2 \tag{2.18}$$

Using the Fuchs-van de Graaf inequalities [27], we get that there exists a σ_B such that $\frac{1}{2} \|\rho_{AB} - \frac{\mathbb{1}_A}{d} \otimes \sigma_B\|_1 \leq \sqrt{\frac{\delta}{d}}$. We are now in a position to use the Alicki-Fannes-Winter [59, Lemma 2] inequality, which states that

$$|H(A|B)_{\rho} - 1| \leq 2\sqrt{\frac{\delta}{d}} + \frac{\sqrt{d} + \sqrt{\delta}}{\sqrt{d}} H \left(\frac{\sqrt{\delta}}{\sqrt{d} + \sqrt{\delta}} \right).$$

□

2.2.4 Quantum Channel Polarization

We now consider the channel combining and splitting procedure for two copies of a quantum channel $\mathcal{W}_{A' \rightarrow B}$, and we will use the following notation,

$$\begin{aligned} \mathcal{W}_C^{(0)} &:= \mathcal{W} \boxtimes_C \mathcal{W} \\ \mathcal{W}_C^{(1)} &:= \mathcal{W} \otimes_C \mathcal{W} \end{aligned}$$

The polarization construction is obtained by recursively applying the channel combining and splitting procedure, while choosing C randomly from some finite set of two-qudit unitaries, denoted by $\mathcal{U} \subset \mathcal{U}(d^2)$. To accommodate the random choice of $C \in \mathcal{U}$, a classical description of C is included as part of the output of the bad and good channels. Hence, for $i = 0, 1$, we define,

$$\mathcal{W}^{(i)}(\rho) = \frac{1}{|\mathcal{U}|} \sum_{C \in \mathcal{U}} |C\rangle\langle C| \otimes \mathcal{W}_C^{(i)}(\rho), \tag{2.19}$$

where $\{|C\rangle\}_{C \in \mathcal{U}}$ is an orthogonal basis of some auxiliary system. Applying twice the transformation $\mathcal{W} \mapsto (\mathcal{W}^{(0)}, \mathcal{W}^{(1)})$, we get four virtual channels $\mathcal{W}^{(i_1 i_2)} := (\mathcal{W}^{(i_1)})^{(i_2)}$,

where $(i_1 i_2) \in \{00, 01, 10, 11\}$. In general, after n levels or recursion, we obtain 2^n channels:

$$\mathcal{W}^{(i_1 \dots i_n)} := \left(\mathcal{W}^{(i_1 \dots i_{n-1})} \right)^{(i_n)}, \quad \forall (i_1 \dots i_n) \in \{0, 1\}^n. \quad (2.20)$$

Our quantum polarization theorem below (Theorem 60) states that the symmetric coherent information of the synthesized virtual channels $\mathcal{W}^{(i_1 \dots i_n)}$ polarizes, meaning that it goes to either -1 or $+1$ as n goes to infinity (except possibly for a vanishing fraction of channels), provided that \mathcal{U} is a unitary 2-design.

Before stating the polarization theorem, we first provide the following lemma on the symmetric coherent information I and the Rényi-Bhattacharyya parameter R of a classical mixture of quantum channels. It will allow us to derive the main steps in the proof of the polarization theorem, by conveniently working with the $\mathcal{W}_C^{(0)}(\rho) / \mathcal{W}_C^{(1)}(\rho)$ construction, rather than the $\mathcal{W}^{(0)}(\rho) / \mathcal{W}^{(1)}(\rho)$ mixture (in which a classical description of C is included in the output). The proof is omitted, since part (a) is trivial, and part (b) follows easily from [36, Section B.2].

Lemma 59. *Let $\mathcal{N}(\rho) = \sum_{x \in X} \lambda_x |x\rangle\langle x| \otimes \mathcal{N}_x(\rho)$, be a classical mixture of quantum channels \mathcal{N}_x , where $\{|x\rangle\}_{x \in X}$ is some orthonormal basis of an auxiliary system, and $\sum_{x \in X} \lambda_x = 1$. Then,*

- (a) $I(\mathcal{N}) = \mathbb{E}_X I(\mathcal{N}_x) := \sum_{x \in X} \lambda_x I(\mathcal{N}_x)$.
- (b) $R(\mathcal{N}) = \mathbb{E}_X R(\mathcal{N}_x) := \sum_{x \in X} \lambda_x R(\mathcal{N}_x)$.

We are now in a position to state our polarization theorem.

Theorem 60. *Let \mathcal{U} be a unitary 2-design. For any qudit-input quantum channel \mathcal{W} , let $\{\mathcal{W}^{(i_1 \dots i_n)} : (i_1 \dots i_n) \in \{0, 1\}^n\}$ be the set of channels defined in (2.20), with channel combining unitary C randomly chosen from \mathcal{U} . Then, for any $\delta > 0$,*

$$\lim_{n \rightarrow \infty} \frac{\#\{(i_1 \dots i_n) \in \{0, 1\}^n : I(\mathcal{W}^{(i_1 \dots i_n)}) \in (-1 + \delta, 1 - \delta)\}}{2^n} = 0$$

and furthermore,

$$\lim_{n \rightarrow \infty} \frac{\#\{(i_1 \dots i_n) \in \{0, 1\}^n : I(\mathcal{W}^{(i_1 \dots i_n)}) \geq 1 - \delta\}}{2^n} = \frac{I(\mathcal{W}) + 1}{2}$$

Proof. We will utilize Lemma 54 for the proof. This basically requires us to find two parameters I and T that respectively play the roles of the symmetric mutual information and the Bhattacharyya parameter from the classical case. The parameter I for us is the symmetric coherent information. The choice of T is crucial to prove polarization as all the constraints in Lemma 54 should be satisfied for I and T . We shall consider the Rényi-Bhattacharyya parameter as our T .

Let $\{B_n : n \geq 1\}$ be a sequence of i.i.d., $\{0, 1\}$ -valued random variables with $P(B_n = 0) = P(B_n = 1) = 1/2$, as in Lemma 54. Let $\{I_n : n \geq 0\}$ and $\{R_n : n \geq 0\}$ be the stochastic processes defined by $I_n := I(\mathcal{W}^{(B_1 \dots B_n)})$ and $R_n := R(\mathcal{W}^{(B_1 \dots B_n)})$. By convention, $\mathcal{W}^{(\emptyset)} := \mathcal{W}$, thus $I_0 = I(\mathcal{W})$ and $R_0 = R(\mathcal{W})$. For the first part, we prove that all the conditions of Lemma 54 hold for I_n and $T_n := R_n$.

(i.1) Straightforward (with $[\iota_0, \iota_1] = [-1, 1]$).

(i.2) We must show that I_n forms a martingale. In other words, that the channel combining and splitting transformation does not change the total coherent information, i.e., $I(\mathcal{W}^{(0)}) + I(\mathcal{W}^{(1)}) = 2I(\mathcal{W})$. This follows from Lemma 56, and Lemma 59 (a), given before.

(t.1) Straightforward (with $[\theta_0, \theta_1] = [\frac{1}{d}, d]$).

(i&t.1) For any $\varepsilon > 0$, there exists a $\delta > 0$ such that $I_n \in (-1 + \varepsilon, 1 - \varepsilon)$ implies that $R_n \in (\frac{1}{d} + \delta, d - \delta)$. In other words, we need to show that if R polarizes, then so does I . This holds for any choice of the Clifford unitary in the channel combining operation, and is proven in Lemma 58 before.

(t.2) We will show the *guaranteed improvement* of the good channel such that

$$R_{n+1} = \frac{d}{d^2 + 1} (1 + R_n^2), \text{ when } B_{n+1} = 1.$$

It is enough to prove the above for $n = 0$ (i.e., the first step of recursion), since in the general case the proof is obtained simply by replacing \mathcal{W} with $\mathcal{W}^{(B_1 \dots B_n)}$. First, by using Lemma 59 (b), and assuming $B_1 = 1$, we get

$$R_1 := R(\mathcal{W}^{(1)}) = \mathbb{E}_C R(\mathcal{W}_C^{(1)}) = \mathbb{E}_C R(\mathcal{W} \otimes_C \mathcal{W}),$$

where the last equality is simply a reminder of our notation $\mathcal{W}_C^{(1)} := \mathcal{W} \otimes_C \mathcal{W}$. Then, from Lemma 61 below, it follows that $\mathbb{E}_C R(\mathcal{W}_C^{(1)}) = \frac{d}{d^2 + 1} (1 + R(\mathcal{W})^2)$.

Lemma 61. *Let $\mathcal{N}_{A' \rightarrow B}$ and $\mathcal{M}_{A' \rightarrow B}$ be two quantum channels with qudit input. Then,*

$$\mathbb{E}_C R(\mathcal{N} \otimes_C \mathcal{M}) = \frac{d}{d^2 + 1} (1 + R(\mathcal{N})R(\mathcal{M}))$$

where \mathbb{E}_C denotes the expectation operator, C is the channel combining unitary, chosen uniformly at random from a unitary 2-design \mathcal{U} .

Proof. Let $\mathcal{N}_{A'_1 \rightarrow E_1}^c$ and $\mathcal{M}_{A'_2 \rightarrow E_2}^c$ be the complementary channels of \mathcal{N} and \mathcal{M} respectively. From Lemma 55, we have that

$$(\mathcal{N} \otimes_C \mathcal{M})^c(\rho_{A'_2}) = (\mathcal{N}^c \otimes \mathcal{M}^c) \left(C \left(\frac{\mathbb{1}_{A'_1}}{d} \otimes \rho_{A'_2} \right) C^\dagger \right).$$

Therefore, $R(\mathcal{N} \otimes_C \mathcal{M}) = d^{-\tilde{H}_2^{\downarrow}(A_2|E_1E_2)_\rho}$, where $\rho_{A_2E_1E_2} = (\mathcal{N} \otimes_C \mathcal{M})^c(\Phi_{A_2A'_2})$. Note that $\rho_{E_1E_2} = \mathcal{N}^c(\frac{\mathbb{1}}{d}) \otimes \mathcal{M}^c(\frac{\mathbb{1}}{d})$, which is independent of C . Now, to compute the expected value of this for a random choice of C , we proceed as follows:

$$\begin{aligned} \mathbb{E}_C d^{-\tilde{H}_2^{\downarrow}(A_2|E_1E_2)_\rho} &= \mathbb{E}_C \text{Tr} \left[\left(\rho_{E_1E_2}^{-\frac{1}{4}} \rho_{A_2E_1E_2} \rho_{E_1E_2}^{-\frac{1}{4}} \right)^2 \right] \\ &= \mathbb{E}_C \text{Tr} \left[\left(\rho_{E_1E_2}^{-\frac{1}{4}} (\mathcal{N}^c \otimes \mathcal{M}^c) \left(C \left(\frac{\mathbb{1}_{A'_1}}{d} \otimes \Phi_{A_2A'_2} \right) C^\dagger \right) \rho_{E_1E_2}^{-\frac{1}{4}} \right)^2 \right]. \end{aligned}$$

Note that this is basically the same calculation as in [60, Equation (3.32)] (there, U is chosen according to the Haar measure over the full unitary group, but all that is required is a unitary 2-design). However, we will not make the simplifications after (3.44) and (3.45) in [60], but will instead keep all the terms. We therefore get

$$\begin{aligned} \mathbb{E}_C d^{-\tilde{H}_2^{\downarrow}(A_2|E_1E_2)_\rho} &= \alpha \text{Tr} \left[\left(\frac{\mathbb{1}_{A_2}}{d} \right)^2 \right] + \beta \text{Tr} \left[\left(\frac{\mathbb{1}_{A'_1}}{d} \otimes \Phi_{A_2A'_2} \right)^2 \right] \\ &= \frac{1}{d} \alpha + \frac{1}{d} \beta, \end{aligned}$$

where

$$\alpha = \frac{d^4}{d^4 - 1} - \frac{d^2}{d^4 - 1} d^{-\tilde{H}_2^\downarrow(A_1 A_2 | E_1 E_2)_\omega},$$

$$\beta = \frac{d^4}{d^4 - 1} d^{-\tilde{H}_2^\downarrow(A_1 A_2 | E_1 E_2)_\omega} - \frac{d^2}{d^4 - 1},$$

$$\text{and } \omega_{A_1 A_2 E_1 E_2} := (\mathcal{N}^c \otimes \mathcal{M}^c)(\Phi_{A_1 A'_1} \otimes \Phi_{A_2 A'_2}).$$

Hence,

$$\begin{aligned} \mathbb{E}_C d^{-\tilde{H}_2^\downarrow(A_2 | E_1 E_2)_\rho} &= \frac{d}{d^2 + 1} + \frac{d}{d^2 + 1} d^{-\tilde{H}_2^\downarrow(A_1 A_2 | E_1 E_2)_\omega} \\ &= \frac{d}{d^2 + 1} + \frac{d}{d^2 + 1} d^{-\tilde{H}_2^\downarrow(A_1 | E_1)_{\mathcal{N}^c(\Phi_{A_1 A'_1})}} d^{-\tilde{H}_2^\downarrow(A_2 | E_2)_{\mathcal{M}^c(\Phi_{A_2 A'_2})}} \\ &= \frac{d}{d^2 + 1} (1 + R(\mathcal{N})R(\mathcal{M})), \end{aligned}$$

where we have used that the conditional sandwiched Rényi entropy of order 2 is additive with respect to tensor-product states, which follows easily from the definition of \tilde{H}_2^\downarrow . \square

The second part of the theorem follows from the martingale property (i.2). From Lemma 54, we have that

$$\iota_0 P(I_\infty = \iota_0) + \iota_1 P(I_\infty = \iota_1) = I_0,$$

here, $\iota_0 = -1$, $\iota_1 = 1$ and, $I_0 = I(\mathcal{W})$. Since I_∞ takes values in $\{-1, 1\}$, $P(I_\infty = -1) = 1 - P(I_\infty = 1)$. Therefore, we have,

$$P(I_\infty = 1) = \frac{1 + I(\mathcal{W})}{2},$$

Note that the fraction of noiseless channels, that is, $P(I_\infty = 1)$ is equal to half the symmetric mutual information of the channel \mathcal{W} . \square

2.3 A Channel Combining Set

As proven in Section 2.2, the purely quantum polarization phenomenon happens by choosing the channel combining operation from a unitary 2-design. Here, we provide a simple proof of the fact that the generalized Clifford group (Definition 14) on two-qudits is a unitary 2-design. Hence, one may sample the channel combining operation from this set.

Theorem 62. *The generalized Clifford group on two qudits, \mathcal{C}_d^2 , is a unitary 2-design¹.*

¹We note that it may be inferred from Lemmas 1, 2 and 3 in [57]. We give an alternative and more elementary proof here, by generalizing the proof from [61] to the qudit case.

Proof. It is shown in [61, Theorem 1] (see also [62]) that the Clifford group on n -qubits forms a unitary 2-design for any $n \geq 1$. Here, we generalize the proof from [61] to the qudit case, and for $n = 2$. We need to prove that the Clifford group \mathcal{C}_d^2 satisfies the Definition 22. For this, it is sufficient to prove (1.50), with $\mathcal{U} = \mathcal{C}_d^2$ for two-qudit input quantum channels of the form $\mathcal{W}_2(\rho) := A\rho B$ (since any quantum channel is a convex combination of quantum channels of this form).

We first consider the twirling of \mathcal{W}_2 with respect to the Clifford group \mathcal{C}_d^2 . Since the Pauli group \mathcal{P}_d^2 is a normal subgroup of \mathcal{C}_d^2 , we may choose a subset $\bar{\mathcal{C}}_d^2 \subset \mathcal{C}_d^2$ containing one representative for each equivalence class in the quotient group $\mathcal{C}_d^2/\mathcal{P}_d^2$. Thus, any element of \mathcal{C}_d^2 can be uniquely written as a product CP , where $C \in \bar{\mathcal{C}}_d^2$ and $P \in \mathcal{P}_d^2$. Therefore, in order to twirl \mathcal{W}_2 with respect to \mathcal{C}_d^2 , we may first twirl it with respect to \mathcal{P}_d^2 , then twirl again the obtained channel with respect to $\bar{\mathcal{C}}_d^2$.

The elements of \mathcal{P}_d^2 have the form $\omega^\lambda P_{r,s} \otimes P_{r',s'}$, with $\lambda, r, s, r', s' = 0, \dots, d-1$. Hence, twirling \mathcal{W}_2 with respect to \mathcal{P}_d^2 gives a quantum channel, denoted \mathcal{W}'_2 , defined below

$$\begin{aligned} \mathcal{W}'_2(\rho) &:= \frac{1}{d^5} \sum_{\lambda, r, s, r', s'} (\omega^\lambda P_{r,s} \otimes P_{r',s'})^\dagger A (\omega^\lambda P_{r,s} \otimes P_{r',s'}) \rho (\omega^\lambda P_{r,s} \otimes P_{r',s'})^\dagger B (\omega^\lambda P_{r,s} \otimes P_{r',s'}), \\ &= \frac{1}{d^4} \sum_{r, s, r', s'} (P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) A (P_{r,s} \otimes P_{r',s'}) \rho (P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) B (P_{r,s} \otimes P_{r',s'}). \end{aligned} \quad (2.21)$$

The last equality from the above shows that it is actually enough to twirl \mathcal{W}_2 with respect to the subset $\bar{\mathcal{P}}_d^2 := \{P_{r,s} \otimes P_{r',s'} \mid r, s, r', s' = 0, \dots, d-1\}$, obtained by omitting phase factors. Since $\bar{\mathcal{P}}_d^2$ forms an operator basis (for two-qudit operators), we may write

$$A = \sum_{r, s, r', s'} \alpha(r, s, r', s') P_{r,s} \otimes P_{r',s'} \quad (2.22)$$

$$B = \sum_{r, s, r', s'} \beta(r, s, r', s') P_{r,s} \otimes P_{r',s'}. \quad (2.23)$$

We now prove two lemmas 63 and 64, which imply Theorem 62.

Lemma 63. *The quantum channel \mathcal{W}'_2 , obtained by twirling \mathcal{W}_2 with respect to $\bar{\mathcal{P}}_d^2$, is a Pauli channel satisfying the following*

$$\mathcal{W}'_2(\rho) = \sum_{r, s, r', s'} \gamma_{r, s, r', s'} (P_{r,s} \otimes P_{r',s'}) \rho (P_{r,s}^\dagger \otimes P_{r',s'}^\dagger), \quad (2.24)$$

where $\gamma_{r, s, r', s'} := \omega^{rs+rs'} \alpha(r, s, r', s') \beta(-r, -s, -r', -s')$ and $-x$ denotes the additive inverse of x modulo d .

Proof. Recall that $\bar{\mathcal{P}}_d^2 = \{P_{r,s} \otimes P_{r',s'} \mid r, s, r', s' = 0, \dots, d-1\}$ is the subset of two-qudit Pauli, without phase factors. Hence, twirling of \mathcal{W}_2 with respect to $\bar{\mathcal{P}}_d^2$ gives

$$\mathcal{W}'_2(\rho) = \frac{1}{d^4} \sum_{r, s, r', s'} (P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) A (P_{r,s} \otimes P_{r',s'}) \rho (P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) B (P_{r,s} \otimes P_{r',s'}) \quad (2.25)$$

Since $\bar{\mathcal{P}}_d^2$ forms an operator basis, we may write

$$A = \sum_{r, s, r', s'} \alpha(r, s, r', s') P_{r,s} \otimes P_{r',s'}, \quad (2.26)$$

$$B = \sum_{r, s, r', s'} \beta(r, s, r', s') P_{r,s} \otimes P_{r',s'} \quad (2.27)$$

Substituting A and B in the above equation, we get

$$\mathcal{W}'_2(\rho) = \frac{1}{d^4} \sum_{t,u,t',u'} \sum_{v,w,v',w'} \alpha(t,u,t',u') \beta(v,w,v',w') \kappa, \quad (2.28)$$

where κ is defined as,

$$\kappa := \sum_{r,r',s,s'} (P_{r,s}^\dagger P_{t,u} P_{r,s}) \otimes (P_{r',s'}^\dagger P_{t',u'} P_{r',s'}) \rho(P_{r,s}^\dagger P_{v,w} P_{r,s}) \otimes (P_{r',s'}^\dagger P_{v',w'} P_{r',s'}). \quad (2.29)$$

From (1.42), we have that $P_{t,u} P_{r,s} = \omega^{-ru+st} P_{r,s} P_{t,u}$. Then, we may write

$$\kappa = k(P_{t,u} \otimes P_{t',u'}) \rho(P_{v,w} \otimes P_{v',w'}), \quad (2.30)$$

where k is defined as,

$$k := \sum_{r,s} \omega^{-r(u+w)+s(v+t)} \sum_{r',s'} \omega^{-r'(u'+w')+s'(v'+t')}. \quad (2.31)$$

When $u + w = v + t = 0 \pmod{d}$, we have $\sum_{r,s} \omega^{-r(u+w)+s(v+t)} = d^2$. When either $u + v \neq 0 \pmod{d}$ or $t + w \neq 0 \pmod{d}$, we have $\sum_{r,s} \omega^{-r(u+w)+s(v+t)} = \frac{(\omega^{-d}-1)(\omega^d-1)}{(\omega^{-1}-1)(\omega-1)} = 0$. Therefore,

$$k = \begin{cases} d^4, & \text{when } u + w = v + t = u' + w' = v' + t' = 0 \pmod{d} \\ 0, & \text{otherwise} \end{cases} \quad (2.32)$$

The condition $u + w = v + t = 0 \pmod{d}$ implies that $P_{t,u} P_{v,w} = X^t Z^u X^v Z^w = \omega^{-uv} I$. Using $t = -v \pmod{d}$, we have that $P_{v,w} = \omega^{tv} P_{t,u}^\dagger$. Plugging κ into (2.28), we get

$$\mathcal{W}'_2(\rho) = \sum_{t,u,t',u'} \gamma_{t,u,t',u'} (P_{t,u} \otimes P_{t',u'}) \rho(P_{t,u}^\dagger \otimes P_{t',u'}^\dagger), \quad (2.33)$$

where $\gamma_{t,u,t',u'}$ is defined as,

$$\gamma_{t,u,t',u'} := \omega^{tu+t'u'} \alpha(t,u,t',u') \beta(-t,-u,-t',-u'). \quad (2.34)$$

Hence, \mathcal{W}'_2 is a qudit Pauli channel, as desired. \square

Lemma 64. *The quantum channel obtained by twirling \mathcal{W}'_2 with respect to $\bar{\mathcal{C}}_d^2$, is the quantum channel \mathcal{W}''_2 acting as*

$$\mathcal{W}''_2(\rho) = \frac{\text{Tr}(AB)}{d^4} \mathbb{1} \otimes \mathbb{1} + \frac{d^2 \text{Tr}(A) \text{Tr}(B) - \text{Tr}(AB)}{d^2(d^4 - 1)} \left(\rho - \frac{1}{d^2} \mathbb{1} \otimes \mathbb{1} \right). \quad (2.35)$$

Proof. Recall that $\bar{\mathcal{C}}_d^2 \subset \mathcal{C}_d^2$ is a subset containing one representative for each equivalence class in the quotient group $\mathcal{C}_d^2 / \mathcal{P}_d^2$. Twirling of \mathcal{W}'_2 with respect to $\bar{\mathcal{C}}_d^2$ gives

$$\mathcal{W}''_2(\rho) = \sum_{t,u,t',u'} \gamma_{t,u,t',u'} \frac{1}{|\bar{\mathcal{C}}_d^2|} \sum_{C \in \bar{\mathcal{C}}_d^2} C^\dagger (P_{t,u} \otimes P_{t',u'}) C \rho C^\dagger (P_{t,u}^\dagger \otimes P_{t',u'}^\dagger) C. \quad (2.36)$$

We know that the conjugate action of the entire set $\bar{\mathcal{C}}_d^2$ maps any $P_{t,u} \otimes P_{t',u'} \neq I \otimes I$ to all $d^4 - 1$ two-qudit Paulis excluding $I \otimes I$, an equal number of times. In other words,

$P_{t,u} \otimes P_{t',u'} \neq I \otimes I$ gets mapped to a Pauli $P_{r,s} \otimes P_{r',s'} \neq I \otimes I$, $\frac{|\tilde{\mathcal{C}}_d^2|}{d^4-1}$ times. Further, $I \otimes I$ is always mapped to $I \otimes I$. Therefore, we have that

$$\mathcal{W}_2''(\rho) = \gamma_{0,0,0,0}\rho + \frac{1}{d^4-1}\gamma' \sum_{(r,s,r',s') \neq (0,0,0,0)} (P_{r,s} \otimes P_{r',s'})\rho(P_{r,s}^\dagger \otimes P_{r',s'}^\dagger), \quad (2.37)$$

where γ' is defined as,

$$\gamma' := \sum_{(t,u,t',u') \neq (0,0,0,0)} \gamma_{t,u,t',u'}. \quad (2.38)$$

Using the following three identities, we can easily transform (2.37) into the form of (2.35).

1. $\gamma_{0,0,0,0} = \frac{\text{Tr}(A)\text{Tr}(B)}{d^4}$.
2. $\sum_{t,u,t',u'} \gamma_{t,u,t',u'} = \frac{\text{Tr}(AB)}{d^2}$.
3. $\sum_{r,s,r',s'} (P_{r,s} \otimes P_{r',s'})\rho(P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) = d^2 I \otimes I$.

Proof of identity 1) We have that $\gamma_{0,0,0,0} = \alpha(0,0,0,0)\beta(0,0,0,0)$. Also,

$$\text{Tr}(P_{r,s}) = \begin{cases} d, & \text{when } P_{r,s} = I \\ 0, & \text{otherwise} \end{cases}$$

Using (2.26) and (2.27), we get $\text{Tr}(A) = \alpha(0,0,0,0)d^2$ and $\text{Tr}(B) = \beta(0,0,0,0)d^2$. Hence, we have $\gamma_{0,0,0,0} = \frac{\text{Tr}(A)\text{Tr}(B)}{d^4}$.

Proof of identity 2) We have,

$$\begin{aligned} \text{Tr}(AB) &= \sum_{t,u,t',u'} \sum_{v,w,v',w'} \alpha(t,u,t',u')\beta(v,w,v',w')\text{Tr}(P_{t,u}P_{v,w})\text{Tr}(P_{t',u'}P_{v',w'}) \\ &= \sum_{t,u,t',u'} d^2 \omega^{tu+t'u'} \alpha(t,u,t',u')\beta(-t,-u,-t',-u') \\ &= d^2 \sum_{t,u,t',u'} \gamma_{t,u,t',u'}. \end{aligned}$$

Proof of identity 3) Let $\rho = \sum_{r,s,r',s'} \rho_{r,s,r',s'} P_{r,s} \otimes P_{r',s'}$. Since ρ is a density matrix, we have $\rho_{0,0,0,0} = \frac{\text{Tr}(\rho)}{d^2} = \frac{1}{d^2}$. Hence,

$$\begin{aligned} \sum_{r,s,r',s'} (P_{r,s} \otimes P_{r',s'})\rho(P_{r,s}^\dagger \otimes P_{r',s'}^\dagger) &= \sum_{r,s,r',s'} \sum_{t,u,t',u'} \rho_{t,u,t',u'} (P_{r,s}P_{t,u}P_{r,s}^\dagger) \otimes (P_{r',s'}P_{t',u'}P_{r',s'}^\dagger) \\ &= \sum_{t,u,t',u'} \rho_{t,u,t',u'} \left(\sum_{r,s,r',s'} \omega^{-st+ru} \omega^{-s't'+r'u'} \right) P_{t,u} \otimes P_{t',u'} \\ &= d^4 \rho_{0,0,0,0} I \otimes I \\ &= d^2 I \otimes I. \end{aligned}$$

We get (2.35) from (2.37) by using the above identities, while also substituting the notation $\mathbb{1}$ for the identity matrix I , as it denotes a quantum state here. □

Now, the quantum channel \mathcal{W}_2'' is the twirling of \mathcal{W}_2 with respect to \mathcal{C}_d^2 . To conclude that \mathcal{C}_d^2 is a unitary 2-design, we need to show that twirling \mathcal{W}_2 with respect to $\mathcal{U}(d^2)$ yields the same channel, which follows from [63]. \square

2.4 Reduction of the Channel Combining Set

In this section, we prove that when the qudit dimension d is a prime number, quantum polarization can be achieved by taking a subset of the two-qudit Clifford group \mathcal{C}_d^2 , containing only $d^4 + d^2 - 2$ elements. For the qubit case, that is, $d = 2$, this set can be further reduced to $\frac{d^4 + d^2 - 2}{2}$ elements. Recall from the proof of Theorem 60 that only condition that depends on the channel combining set is the guaranteed improvement condition (t.2). We show that guaranteed improvement condition (t.2) is still fulfilled, when the channel combining operation is chosen from the reduced set of two-qudit Cliffords, which implies that polarization happens.

2.4.1 A Channel Combining Set with $d^4 + d^2 - 2$ Elements for Qudit Channels

We first define an equivalence relation on \mathcal{C}_2 , whose equivalence classes are the left cosets of $\mathcal{C}_1 \otimes \mathcal{C}_1$ as follows.

Definition 65. We say that C' and $C'' \in \mathcal{C}_2$ are equivalent, and denote it by $C' \sim C''$, if there exist $C_1, C_2 \in \mathcal{C}_1$ such that $C'' = C'(C_1 \otimes C_2)$ (see also Figure 2.2).

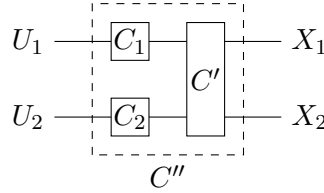


Figure 2.2: Equivalent two-qudit Clifford gates $C' \sim C''$

Now, we observe that two equivalent Clifford gates, used to combine any two quantum channels, yield the same Rényi-Bhattacharyya parameter of the bad/good channels. This is stated in the following lemma.

Lemma 66. Let $C', C'' \in \mathcal{C}_2$. If $C' \sim C''$, then for any two quantum channels \mathcal{N} and \mathcal{M} with qudit inputs, we have the following,

- (a) $R(\mathcal{M} \boxtimes_{C'} \mathcal{N}) = R(\mathcal{M} \boxtimes_{C''} \mathcal{N})$.
- (b) $R(\mathcal{M} \otimes_{C'} \mathcal{N}) = R(\mathcal{M} \otimes_{C''} \mathcal{N})$.

Proof. **Point (i).** Given $C''_{A'_1 A'_2} = C'_{A'_1 A'_2} (C^1_{A'_1} \otimes C^2_{A'_2})$, let

$$\begin{aligned} \rho' &:= (\mathcal{N} \boxtimes_{C'} \mathcal{M})_{A'_1 \rightarrow E_1 E_2 A_2}^c (\Phi_{A_1 A'_1}) \\ \rho'' &:= (\mathcal{N} \boxtimes_{C''} \mathcal{M})_{A'_1 \rightarrow E_1 E_2 A_2}^c (\Phi_{A_1 A'_1}). \end{aligned}$$

Then, we have that

$$\begin{aligned}
 \rho'' &= \mathcal{N}_{A'_1 \rightarrow E_1}^c \otimes \mathcal{M}_{A'_2 \rightarrow E_2}^c \left(C_{A'_1 A'_2}' (C_{A'_1}^1 \otimes C_{A'_2}^2) \left(\Phi_{A_1 A'_1} \otimes \Phi_{A'_2 A_2} \right) (C_{A'_1}^{1\dagger} \otimes C_{A'_2}^{2\dagger}) C_{A'_1 A'_2}'^\dagger \right) \\
 &= \mathcal{N}_{A'_1 \rightarrow E_1}^c \otimes \mathcal{M}_{A'_2 \rightarrow E_2}^c \left(C_{A'_1 A'_2}' (C_{A'_1}^{1\top} \otimes C_{A'_2}^{2\top}) \left(\Phi_{A_1 A'_1} \otimes \Phi_{A'_2 A_2} \right) (\bar{C}_{A'_1}^1 \otimes \bar{C}_{A'_2}^2) C_{A'_1 A'_2}'^\dagger \right) \\
 &= C_{A'_1}^{1\top} \otimes C_{A'_2}^{2\top} \left(\mathcal{N}_{A'_1 \rightarrow E_1}^c \otimes \mathcal{M}_{A'_2 \rightarrow E_2}^c \left(C_{A'_1 A'_2}' \left(\Phi_{A_1 A'_1} \otimes \Phi_{A'_2 A_2} \right) C_{A'_1 A'_2}'^\dagger \right) \bar{C}_{A'_1}^1 \otimes \bar{C}_{A'_2}^2 \right) \\
 &= C_{A'_1}^{1\top} \otimes C_{A'_2}^{2\top} (\rho') \bar{C}_{A'_1}^1 \otimes \bar{C}_{A'_2}^2,
 \end{aligned} \tag{2.39}$$

where the first equality follows from part (a) of Lemma 55 and the second equality follows from the relation $(\mathbb{1} \otimes Z)|\Phi\rangle = (Z^\top \otimes \mathbb{1})|\Phi\rangle$, for any matrix Z .

From (57), we have that

$$R(\mathcal{N} \boxtimes_{C''} \mathcal{M}) = 2^{-\tilde{H}_2^\downarrow(A_1|E_1 E_2 A_2)_{\rho''}},$$

where $\tilde{H}_2^\downarrow(A|B)_\rho = -\tilde{D}_2(\rho_{AB}||I \otimes \rho_B)$ (Definition 38). We have the following equality for $\tilde{D}_2(\rho||\sigma)$,

$$\tilde{D}_2(\rho||\sigma) = \tilde{D}_2(U \rho U^\dagger || U \sigma U^\dagger), \tag{2.40}$$

where U is a unitary operator. Hence,

$$\begin{aligned}
 \tilde{H}_2^\downarrow(A_1|E_1 E_2 A_2)_{\rho''} &= -\tilde{D}_2(\rho''||\mathbb{1} \otimes \text{Tr}_{A_1}(\rho'')) \\
 &= -\tilde{D}_2(C_{A'_1}^{1\top} \otimes C_{A'_2}^{2\top} (\rho') \bar{C}_{A'_1}^1 \otimes \bar{C}_{A'_2}^2 || \mathbb{1} \otimes (C_{A'_2}^{2\top} \text{Tr}_{A_1}(\rho') \bar{C}_{A'_2}^2)) \\
 &= -\tilde{D}_2(\rho' || \mathbb{1} \otimes \text{Tr}_{A_1}(\rho')) \\
 &= \tilde{H}_2^\downarrow(A_1|E_1 E_2 A_2)_{\rho'},
 \end{aligned}$$

where the second equality follows from (2.39) and $\text{Tr}_{A_1}(\rho'') = C_{A'_2}^{2\top} \text{Tr}_{A_1}(\rho') \bar{C}_{A'_2}^2$, and the third equality follows from the unitary equivalence in (2.40). Therefore, we get $R(\mathcal{N} \boxtimes_{C''} \mathcal{M}) = R(\mathcal{N} \boxtimes_{C'} \mathcal{M})$, as desired.

Point (ii). Given $C_{A'_1 A'_2}'' = C_{A'_1 A'_2}' (C_{A'_1}^1 \otimes C_{A'_2}^2)$, let

$$\begin{aligned}
 \rho' &:= (\mathcal{N} \otimes_{C'} \mathcal{M})_{A'_1 \rightarrow E_1 E_2 A_2}^c (\Phi_{A_1 A'_1}) \\
 \rho'' &:= (\mathcal{N} \otimes_{C''} \mathcal{M})_{A'_1 \rightarrow E_1 E_2 A_2}^c (\Phi_{A_1 A'_1}).
 \end{aligned}$$

Then, it can be shown similarly to the point (i) that

$$\rho'' = C_{A'_2}^{2\top} \rho' \bar{C}_{A'_2}^2. \tag{2.41}$$

From the unitary equivalence in (2.40), it follows that $R(\mathcal{M} \otimes_{C'} \mathcal{N}) = R(\mathcal{M} \otimes_{C''} \mathcal{N})$. \square

Using Lemma 66, we reduce the channel combining set as follows.

Theorem 67. *If d is prime, there exists a subset $\mathcal{U} \subset \mathcal{C}_d^2$, of size $|\mathcal{U}| = d^4 + d^2 - 2$, which is not a unitary 2-design, and such that polarization happens when the channel combining unitary C is randomly chosen from \mathcal{U} .*

Proof. We divide the set \mathcal{C}_d^2 into equivalence classes with respect to the equivalence relation in Definition 65. The equivalence classes are left cosets of $\mathcal{C}_d^1 \otimes \mathcal{C}_d^1$ in \mathcal{C}_d^2 . The number of the left cosets are equal to $\frac{|\mathcal{C}_d^2|}{|\mathcal{C}_d^1|^2}$. Consider now a set of representatives \mathcal{R} , containing

only one two-qudit Clifford unitary from each left coset. As a consequence of Lemma 66, we have the following for the Rényi-Bhattacharyya parameter of the good channel $\mathcal{W}_C^{(1)}$,

$$\mathbb{E}_{C \in \mathcal{R}} R(\mathcal{W}_C^{(1)}) = \mathbb{E}_{C \in \mathcal{C}_d^2} R(\mathcal{W}_C^{(1)}). \quad (2.42)$$

Therefore, we need a set containing only $\frac{|\mathcal{C}_d^2|}{|\mathcal{C}_d^1|^2}$ two-qudit Cliffords to achieve the polarization. In the following lemma, we give $|\mathcal{C}_d^2|$ and $|\mathcal{C}_d^1|$.

Lemma 68. *If d is a prime number, $|\mathcal{C}_d^1| = d^3(d^2 - 1)$ and $|\mathcal{C}_d^2| = d^8(d^4 - 1)(d^2 - 1)$.*

Proof. Consider the one-qudit Clifford group \mathcal{C}_d^1 . We count first the permutations generated by \mathcal{C}_d^1 on $\bar{\mathcal{P}}_d^1 := \{P_{r,s} | r, s = 0, \dots, d-1\}$, and later we will accommodate the phase factors. Any Clifford $C \in \mathcal{C}_d^1$ is uniquely determined by its conjugate action on the generators of the Pauli group, X and Z . Suppose that C maps $X \mapsto P_{r,s}$ and $Z \mapsto P_{t,u}$ via its conjugate action, where $P_{r,s}, P_{t,u} \neq I$. On the one hand, since commutation relations are preserved under unitary conjugation, $P_{r,s}$ and $P_{t,u}$ must satisfy $P_{r,s}P_{t,u} = \omega P_{t,u}P_{r,s}$. On the other hand, from (1.42), we have that $P_{r,s}P_{t,u} = \omega^{ru-st}P_{t,u}P_{r,s}$. Therefore, r, u, s, t must be such that $ru - st = 1 \pmod{d}$. We fix r, s and solve for t, u . Since $P_{r,s} \neq I$, it follows that either r or s is non-zero. Without loss of generality, we may assume that $r \neq 0$. Since d is a prime number, r is invertible under multiplication modulo d . Therefore, for any $t \in \{0, \dots, d-1\}$, there exists a unique $u := r^{-1}(1 + st) \pmod{d}$, satisfying $ru - st = 1$. Hence, there are exactly d choices for the t, u pair. Since we have $d^2 - 1$ choices for the r, s pair, it follows that there are $d(d^2 - 1)$ pairs of Paulis, $P_{r,s}$ and $P_{t,u}$, such that $P_{r,s}P_{t,u} = \omega P_{t,u}P_{r,s}$. Taking into account the phase factors, $\omega^\lambda, \lambda \in \{0, \dots, d-1\}$, it follows that \mathcal{C}_d^1 has $d^3(d^2 - 1)$ elements.

We now count the number of elements in \mathcal{C}_d^2 . The two-qudit Pauli group \mathcal{P}_d^2 is generated by a set of four Paulis $I \otimes X, I \otimes Z, X \otimes I$ and $Z \otimes I$, and any Clifford $C \in \mathcal{C}_d^2$ is uniquely determined by its conjugate action on these four generators. The commutation relations between the four generators are illustrated in Fig. 2.3.

$$\begin{array}{cc} I \otimes X & X \otimes I \\ | & | \\ I \otimes Z & Z \otimes I \end{array}$$

Figure 2.3: Connected Paulis satisfy $AB = \omega BA$, with A is the Pauli on the top row, and B the Pauli on the bottom row. Paulis that are not connected commute.

Consider a mapping $I \otimes X \mapsto A, I \otimes Z \mapsto B, X \otimes I \mapsto A', Z \otimes I \mapsto B'$, where $A, B, A', B' \in \bar{\mathcal{P}}_d^2$, that preserves all the commutation relations between generators. Pauli $I \otimes X$ can be mapped to any two-qudit Pauli $A \neq I \otimes I$, so there are $d^4 - 1$ choices for A . It is not very difficult to see that for any $A \neq I \otimes I$ there are d^3 choices for B such that $AB = \omega BA$. Further, there are $d(d^2 - 1)$ pairs of two-qudit Paulis A' and B' , which commute with both A and B , and satisfy $A'B' = \omega B'A'$. Therefore, we have $d^4(d^4 - 1)(d^2 - 1)$ possible permutations on $\bar{\mathcal{P}}_d^2$, which satisfy all the commutation relations. Taking into account the phase factors, it follows that \mathcal{C}_d^2 has $d^8(d^4 - 1)(d^2 - 1)$ elements. \square

Therefore, from the above lemma, we have that $\frac{|\mathcal{C}_d^2|}{|\mathcal{C}_d^1|^2} = d^4 + d^2$. Further, we may choose our representative set \mathcal{R} such that it contains the identity (I) and the swap (S) gates. Note that $R(\mathcal{W}_C^{(1)}) = R(\mathcal{W})$ for $C \in \{I, S\}$. Since $\mathbb{E}_{C \in \mathcal{R}} R(\mathcal{W}_C^{(1)}) \leq R(\mathcal{W})$, it follows

that removing I and S can only decrease the expectation value, that is, $\mathbb{E}_{R \in \mathcal{C} \setminus \{I, S\}} \leq \mathbb{E}_{C \in \mathcal{R}} R(\mathcal{W}_C^{(1)})$. More precisely, we have the following,

$$\mathbb{E}_{C \in \mathcal{R} \setminus \{I, S\}} R(\mathcal{W}_C^{(1)}) = \frac{d^3}{d^4 + d^2 - 2} - \frac{2R(\mathcal{W})}{d^4 + d^2 - 2} + \frac{d^3 R(\mathcal{W})^2}{d^4 + d^2 - 2} \leq \frac{d}{d^2 + 1} (1 + R(\mathcal{W})^2).$$

Hence, condition (t.2) from Theorem 60 remains satisfied even after removing the I and S gates from the set of representatives \mathcal{R} . Therefore, polarization happens by sampling the channel combining operation from a set containing $d^4 + d^2 - 2$ two-qudit Clifford. \square

In Lemma 69 below, we prove that when two-qudit Clifford unitaries C' and C'' , which are connected by the swap gate S , such that $C'' := SC'$, are used to combine two copies of a quantum channel $\mathcal{W}_{A \rightarrow B}$ with qudit input, yield the same Rényi-Bhattacharyya parameter of the bad/good channels. Note that this property is weaker than the one in Lemma 66, which holds for any two quantum channels \mathcal{N} and \mathcal{M} . However, it is sufficient in the context of polarization, where one has many copies of the same channel.

Lemma 69. *Let $C', C'' \in \mathcal{C}_2$, such that $C'' = SC'$, where S is the swap gate. Then, for two copies of a quantum channel \mathcal{W} with qudit input,*

$$R(\mathcal{W} \boxtimes_{C'} \mathcal{W}) = R(\mathcal{W} \boxtimes_{C''} \mathcal{W}) \quad \text{and} \quad R(\mathcal{W} \otimes_{C'} \mathcal{W}) = R(\mathcal{W} \otimes_{C''} \mathcal{W}).$$

Proof. First, we note that by applying a unitary on the output of any quantum channel does not change the Rényi-Bhattacharyya parameter. Precisely, let $\mathcal{N}_{A \rightarrow B}$ be any quantum channel, and $U\mathcal{N}_{A \rightarrow B}U^\dagger$ be the quantum channel² obtained by applying the unitary U on the output system B , that is, $(U\mathcal{N}_{A \rightarrow B}U^\dagger)(\rho_A) := U\mathcal{N}_{A \rightarrow B}(\rho_A)U^\dagger$. Then,

$$R(U\mathcal{N}_{A \rightarrow B}U^\dagger) = R(\mathcal{N}_{A \rightarrow B}). \quad (2.43)$$

Going back to the proof of the Lemma, by the definition of $\mathcal{W} \boxtimes_C \mathcal{W}$ and using that $S^\dagger = S$, we may write:

$$\begin{aligned} (\mathcal{W} \boxtimes_{C''} \mathcal{W})(\rho) &= (\mathcal{W} \otimes \mathcal{W}) \left(C''(\rho \otimes \frac{\mathbb{1}}{d}) C''^\dagger \right) \\ &= (\mathcal{W} \otimes \mathcal{W}) \left(SC'(\rho \otimes \frac{\mathbb{1}}{d}) C'^\dagger S \right). \end{aligned}$$

Now, it is easily seen that the $\mathcal{W} \otimes \mathcal{W}$ channel is covariant with respect to the swap gate, i.e., the swap gate commutes with the action of the channel. Hence we may further write:

$$\begin{aligned} (\mathcal{W} \boxtimes_{C''} \mathcal{W})(\rho) &= S(\mathcal{W} \otimes \mathcal{W}) \left(C'(\rho \otimes \frac{\mathbb{1}}{d}) C'^\dagger \right) S \\ &= S(\mathcal{W} \boxtimes_{C'} \mathcal{W})(\rho) S \\ &= (S(\mathcal{W} \boxtimes_{C'} \mathcal{W})S)(\rho). \end{aligned}$$

Hence, $\mathcal{W} \boxtimes_{C''} \mathcal{W} = S(\mathcal{W} \boxtimes_{C'} \mathcal{W})S$, and using (2.43), with $\mathcal{N} := \mathcal{W} \boxtimes_{C'} \mathcal{W}$ and $U := S$, we get

$$R(\mathcal{W} \boxtimes_{C'} \mathcal{W}) = R(\mathcal{W} \boxtimes_{C''} \mathcal{W}),$$

as desired. The equality $R(\mathcal{W} \otimes_{C'} \mathcal{W}) = R(\mathcal{W} \otimes_{C''} \mathcal{W})$ may be proven in a similar way. \square

²To see that $U\mathcal{N}_{A \rightarrow B}U^\dagger$ is a quantum channel, it is enough to notice that if $\mathcal{N}_{A \rightarrow B}$ is defined by Kraus operators $\{E_k\}$, then $U\mathcal{N}_{A \rightarrow B}U^\dagger$ is defined by Kraus operators $\{UE_k\}$.

We would like to further reduce the channel combining set using the above lemma. Note that if C' and C'' belong to the same equivalence class under the equivalence relation in Definition 65, this is not significant as we are using only one unitary from each equivalence class. However, if C' and C'' belong to different equivalence classes for all C' , the size of the channel combining set can be reduced to $\frac{d^4+d^2-2}{2}$. We show in Lemma 70 by providing a counterexample that this property does not hold for a two-qudit unitary in dimension $d = 5$. However, as shown in Section 2.4.2 below, it holds for all two-qubit clifford unitaries.

Lemma 70. *For $d = 5$, there exists a $C \in \mathcal{C}_d^2$ such that $SC = C(C_1 \otimes C_2)$ for some $C_1, C_2 \in \mathcal{C}_d^1$.*

Proof. We consider $d = 5$. Let $C_1 = I$ be the identity, and $C'_2 \in \mathcal{C}_d^1$ be such that it maps $X \mapsto X^4$ and $Z \mapsto Z^4$, via conjugation. Since $X^4 Z^4 = \omega Z^4 X^4$, C'_2 is indeed a one-qudit Clifford. We define $C_2 = C'_2 X^2 Z^2$. Further, let $C \in \mathcal{C}_d^2$, such that its conjugate action generates the following permutation on the generators of \mathcal{P}_d^2 ,

$$\begin{aligned} I \otimes X &\mapsto X^4 Z \otimes X Z^4, \\ I \otimes Z &\mapsto X Z \otimes X^4 Z^4, \\ X \otimes I &\mapsto X^4 Z \otimes X^4 Z, \\ Z \otimes I &\mapsto X Z \otimes X Z. \end{aligned}$$

Using (1.42), it is easily seen that the above permutation preserves all the commutation relations between the generators. Now, the conjugate actions of SC and $C(C_1 \otimes C_2)$ generate the same permutation on \mathcal{P}_d^2 . Therefore, $SC = C(C_1 \otimes C_2)$. \square

2.4.2 A Channel Combining Set with Nine Elements for Qubit Channels

For the qubit case, $d = 2$, a set of $d^4 + d^2 - 2 = 18$ representatives can be chosen as follows³.

- For nine out of the remaining 18 equivalence classes, one may find representatives of the form $(C_1 \otimes C_2)C_{2 \rightarrow 1}$, where $C_{2 \rightarrow 1}$ is the CNOT gate with control on the second qubit and target on the first qubit (see Section 1.3.8), $C_1 \in \{I, \sqrt{Z}, \sqrt{Y}\}$, $C_2 \in \{I, \sqrt{X}, \sqrt{Y}\}$, and $\sqrt{P} = \frac{(1-i)(1+iP)}{2}$, for any Pauli matrix $P \in \{X, Y, Z\}$. We denote this set by \mathcal{L} , which is further depicted in Figure 2.4.

$$\mathcal{L} := \left\{ (C_1 \otimes C_2)C_{2 \rightarrow 1} \mid C_1 \in \{I, \sqrt{Z}, \sqrt{Y}\}, C_2 \in \{I, \sqrt{X}, \sqrt{Y}\} \right\}.$$

- For the remaining nine equivalence classes, one may find representatives of the form SL , where S is the swap gate and $L \in \mathcal{L}$. We denote this set by \mathcal{L}' ,

$$\mathcal{L}' := \{SL \mid L \in \mathcal{L}\}.$$

Hence, from Theorem 67 and Lemma 69 either the set \mathcal{L} or \mathcal{L}' is sufficient to achieve polarization.

³We used a computer program to determine such a set of representatives.

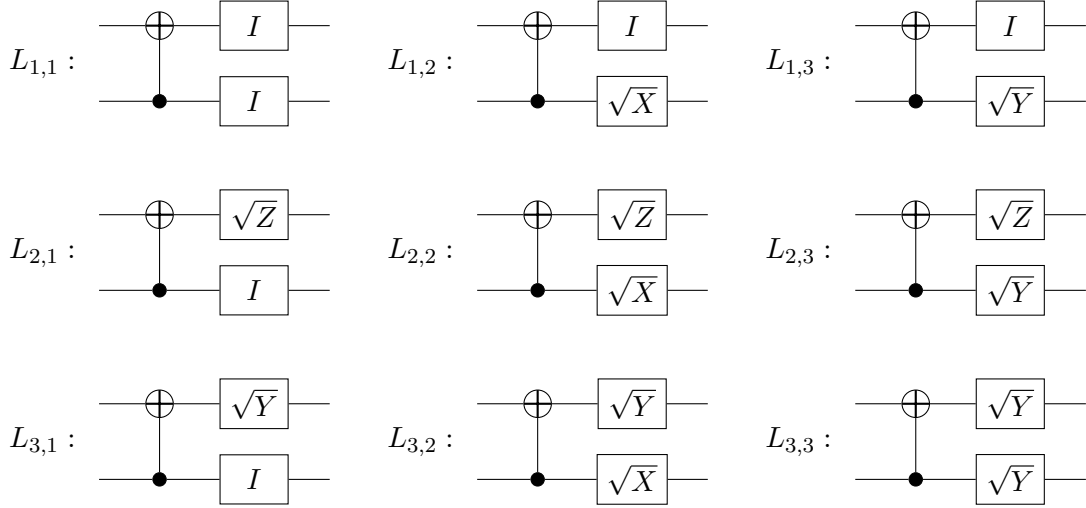


Figure 2.4: The set $\mathcal{L} := \{L_{i,j} \mid 1 \leq i, j \leq 3\}$ containing nine Cliffords.

2.5 Quantum Polar Coding

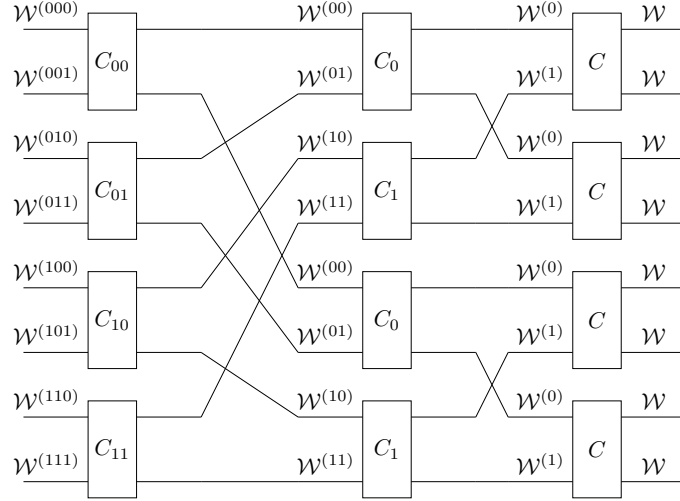
2.5.1 Quantum Polar Codes

We exploit the purely quantum polarization to construct a quantum polar of finite code-length as follows [2] (see also Figure 2.5). For a quantum polar code of length $N = 2^n$, $n > 0$, we start with N copies of the quantum channel \mathcal{W} , pair them in $N/2$ pairs, and apply the channel combining and splitting operation on each pair. The same channel combining Clifford gate is used for each of the $N/2$ pairs, which will be denoted by C . By doing so, we generate $N/2$ copies of the channel $\mathcal{W}^{(0)} := \mathcal{W} \boxtimes_C \mathcal{W}$ and $N/2$ copies of the channel $\mathcal{W}^{(1)} := \mathcal{W} \otimes_C \mathcal{W}$. Hence, for each $i_1 = 0, 1$, we group together the $N/2$ copies of the $\mathcal{W}^{(i_1)}$ channel, pair them in $N/4$ pairs, and apply the channel combining and splitting operation on each pair, by using some channel combining Clifford gate denoted by C_{i_1} . By performing n *polarization steps* (that is, applying the above construction recursively n times), we generate quantum channels $\mathcal{W}^{(i_1 \dots i_n)}$, which can be recursively defined for $n > 0$, as follows:

$$\mathcal{W}^{(i_1 \dots i_n)} := \begin{cases} \mathcal{W}^{(i_1 \dots i_{n-1})} \boxtimes_{C_{i_1 \dots i_{n-1}}} \mathcal{W}^{(i_1 \dots i_{n-1})}, & \text{if } i_n = 0 \\ \mathcal{W}^{(i_1 \dots i_{n-1})} \otimes_{C_{i_1 \dots i_{n-1}}} \mathcal{W}^{(i_1 \dots i_{n-1})}, & \text{if } i_n = 1 \end{cases} \quad (2.44)$$

where, for $n = 1$, in the right hand side term of the above equality, we set by convention $\mathcal{W}^{(\emptyset)} := \mathcal{W}$ and $C_{\emptyset} := C$. Note that, for the sake of simplicity, we have dropped the channel combining Clifford gate from the $\mathcal{W}^{(i_1 \dots i_n)}$ notation. The construction is illustrated in Fig. 2.5, for $N = 8$. Here, horizontal “wires” represent qudits, and for each polarization step, we have indicated on each wire the virtual channel $\mathcal{W}^{(i_1 i_2 \dots)}$ “seen” by the corresponding qudit state.

The above construction synthesizes a set of N channels and, for any $i = 0, \dots, N - 1$, we shall further denote $\mathcal{W}^{(i)} := \mathcal{W}^{(i_1 \dots i_n)}$, where $i_1 \dots i_n$ is the binary decomposition of i . Let $\mathcal{I} \subseteq \{0, 1, \dots, N - 1\}$ denote the set of good channels (*i.e.*, with symmetric coherent information close to 1, or equivalently, Rényi-Bhattacharyya parameter close to $1/2$), and


 Figure 2.5: Quantum polar code of length $N = 8$

let $\mathcal{J} := \{0, 1, \dots, N - 1\} \setminus \mathcal{I}$. With a slight abuse of notation, we shall also denote by \mathcal{I} and \mathcal{J} as quantum systems consisting of $|\mathcal{I}|$ and $|\mathcal{J}|$ qudits, respectively.

A quantum state $\rho_{\mathcal{I}}$ on system \mathcal{I} is encoded by supplying it as input to channels $i \in \mathcal{I}$, while supplying each channel $j \in \mathcal{J}$ with half of an EPR pair, shared between the sender and the receiver. Precisely, let $\Phi_{\mathcal{J}\mathcal{J}'}$ be a maximally entangled state, defined by

$$\Phi_{\mathcal{J}\mathcal{J}'} = \otimes_{j \in \mathcal{J}} \Phi_{jj'}, \quad (2.45)$$

where indices j and j' indicate the j -th qudits of \mathcal{J} and \mathcal{J}' systems, respectively, and $\Phi_{jj'}$ is an EPR pair. Let also Q_N denote the quantum polar transform, that is the unitary operator defined by applying Clifford gates corresponding to the n polarization steps. The encoded state, denoted $\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'}$, is obtained by applying the $Q_N \otimes I_{\mathcal{J}'}$ unitary on the $\mathcal{I}\mathcal{J}\mathcal{J}'$ system, hence:

$$\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'} := (Q_N \otimes I_{\mathcal{J}'})(\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'})(Q_N^\dagger \otimes I_{\mathcal{J}'}). \quad (2.46)$$

Since no errors occur on the \mathcal{J}' system, the channel output state is given by:

$$\psi_{\mathcal{I}\mathcal{J}\mathcal{J}'} := (\mathcal{W}^{\otimes N} \otimes I_{\mathcal{J}'})(\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'}). \quad (2.47)$$

It is worth noticing that randomness is used only at the code construction stage (since Clifford gates used in the n polarization steps are randomly chosen from some predetermined set of gates), but not at the encoding stage. The constructed polar code allows communicating quantum information over a quantum channel \mathcal{W} at a rate $|\mathcal{I}|/N$, which approaches $(1 + I(\mathcal{W}))/2$ (that is, half the symmetric mutual information of the channel), as N goes to infinity. The net communication rate, is given by $(|\mathcal{I}| - |\mathcal{J}|)/N$, and approaches the symmetric coherent information $I(\mathcal{W})$, as N goes to infinity, similarly to the CSS quantum polar code (see discussion on the rate of communication in Section 1.5.3).

2.5.2 Quantum Polar Codes as Entanglement-Assisted Stabilizer Codes

In this section, we consider the purely quantum polar code presented in Section 2.5.1 for qubit quantum systems and observe that it can be considered as an entanglement assisted stabilizer code.

Including all information qubits (\mathcal{I} system) and both systems of the preshared EPR pairs (\mathcal{J} and \mathcal{J}' systems), the quantum polar code from the above section can be described as an entanglement assisted stabilizer code, in the sense of [50]. Precisely, using the notation from the previous section, the quantum state $\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'}$ is stabilized by the set of Pauli operators

$$\mathcal{S}_{\mathcal{I}\mathcal{J}\mathcal{J}'} := \{I_{\mathcal{I}} \otimes X_j X_{j'}, I_{\mathcal{I}} \otimes Z_j Z_{j'} \mid j \in \mathcal{J}\}, \quad (2.48)$$

where $I_{\mathcal{I}}$ denotes the identity on the \mathcal{I} system, and $X_j X_{j'}$ (respectively, $Z_j Z_{j'}$) denotes the tensor product of the Pauli- X (respectively, Pauli- Z) operators of the j -th qubits of systems \mathcal{J} and \mathcal{J}' ⁴. Conversely, any quantum state on the tripartite $\mathcal{I}\mathcal{J}\mathcal{J}'$ system, which is stabilized by Pauli operators in $\mathcal{S}_{\mathcal{I}\mathcal{J}\mathcal{J}'}$, is necessarily of the form $\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'}$ (by a dimension argument). Hence, encoded states ($\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'}$ defined in (2.46)) are stabilized by the set of Pauli operators obtained by *passing the elements of $\mathcal{S}_{\mathcal{I}\mathcal{J}\mathcal{J}'}$ through the polar transform G_q* , that is,

$$\bar{\mathcal{S}}_{\mathcal{I}\mathcal{J}\mathcal{J}'} := (G_q \otimes I_{\mathcal{J}'}) \mathcal{S}_{\mathcal{I}\mathcal{J}\mathcal{J}'} (G_q^\dagger \otimes I_{\mathcal{J}'}). \quad (2.49)$$

For stabilizer codes, the decoding problem for general quantum channels reduces to decoding Pauli errors only, after performing syndrome measurement, *i.e.*, measuring all the generators of the stabilizer group (in our case, the elements of $\bar{\mathcal{S}}_{\mathcal{I}\mathcal{J}\mathcal{J}'}$). Here, the implicit assumption is that syndrome measurement induces appropriate projections, such that it results in the standard Pauli error model.

For qubit Pauli channels, we provide an efficient decoding algorithm (Section 3.5), achieving the symmetric coherent information of the channel. For general quantum channels, syndrome measurement coupled with the above decoding on the induced Pauli error model may yield a practical solution to the decoding problem. However, such a solution is not optimal, due to the loss of information incurred during syndrome measurement. Besides, the polar code should be fitted to (and thus exploit the polarization of) the induced Pauli error model, rather than the quantum channel itself. Devising an efficient decoding algorithm capable of achieving the symmetric coherent information of general quantum channels is an open problem.

2.6 Quantum Polar Codes with Vanishing Rate of Preshared Entanglement

In this section we present a code construction using an asymptotically vanishing rate or preshared entanglement, while achieving a net transmission rate equal to the symmetric coherent information of the channel. In particular, we shall assume that the coherent information of the channel is positive, $I(\mathcal{W}) > 0$. The proposed construction bears similarities to the universal polar code construction in [64, Section V], capable of achieving the compound capacity of a finite set of classical channels.

Let $P_q(N, \mathcal{J}, \mathcal{I})$ denote a quantum polar code of length $N = 2^n$, for some $n > 0$, where \mathcal{I} and \mathcal{J} denote the sets of good and bad channels respectively. By Theorem 60, as n goes to infinity, $|\mathcal{I}|$ approaches $\frac{1+I(\mathcal{W})}{2}N$, and thus $|\mathcal{J}|$ approaches $\frac{1-I(\mathcal{W})}{2}N$. Since $I(\mathcal{W}) > 0$, it follows that $|\mathcal{J}| < |\mathcal{I}|$, provided that n is large enough. Therefore, we may find a subset

⁴Note that the definition of $\mathcal{S}_{\mathcal{I}\mathcal{J}\mathcal{J}'}$ in (2.48) depends only on index $j \in \mathcal{J}$, since $j' \in \mathcal{J}'$ is the counterpart of j (thus, uniquely determined by the latter).

of good channels $\mathcal{I}' \subset \mathcal{I}$, such that $|\mathcal{I}'| = |\mathcal{J}|$. In the sequel, we shall extend the definition of a polar code to include such a subset \mathcal{I}' , and denote it by $P_q(N, \mathcal{J}, \mathcal{I}, \mathcal{I}')$.

Let us now consider k copies of a quantum polar code $P_q(N, \mathcal{J}, \mathcal{I}, \mathcal{I}')$, denoted by $P_q^l(N, \mathcal{J}_l, \mathcal{I}_l, \mathcal{I}'_l)$ or simply by P_q^l , for any $l \in \{0, 1, \dots, k-1\}$. We define a quantum code C_q^k of codelength $|C_q^k| = kN$, by *chaining* them in the following way (see also Fig. 2.6):

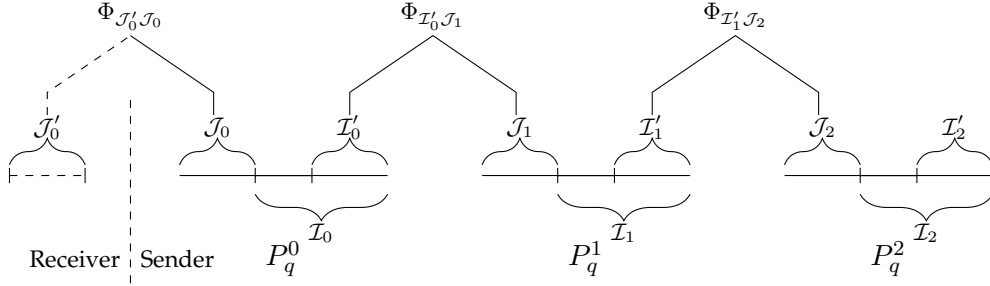


Figure 2.6: C_q^3 : Chaining construction with $k = 3$ copies of a quantum polar codes P_q

- (i) For system \mathcal{J}_0 , the input quantum state before encoding is half of a maximally entangled state $\Phi_{\mathcal{J}_0 \mathcal{J}'_0}$, where system \mathcal{J}'_0 is part of channel output. This is the only preshared entanglement between the sender and the receiver.
- (ii) For systems \mathcal{I}'_{l-1} and \mathcal{J}_l , with $l \neq 0$, the input quantum state before encoding is a maximally entangled state $\Phi_{\mathcal{I}'_{l-1} \mathcal{J}_l}$.
- (iii) Systems $\mathcal{I}_l \setminus \mathcal{I}'_l$, for $l \neq k-1$, and \mathcal{I}_{k-1} are information systems, meaning that the corresponding quantum state is the one that needs to be transmitted from the sender to the receiver.

It can be easily seen that the transmission (coding) rate of the proposed scheme is given by

$$R := \frac{\sum_{l=0}^{k-2} |\mathcal{I}_l \setminus \mathcal{I}'_l| + |\mathcal{I}_{k-1}|}{kN} \xrightarrow{n \rightarrow \infty} \frac{(k-1)I(\mathcal{W}) + \frac{1+I(\mathcal{W})}{2}}{k} \xrightarrow{k \rightarrow \infty} I(\mathcal{W}),$$

while the rate of preshared entanglement is given by

$$E := \frac{|\mathcal{J}_0|}{kN} \xrightarrow{n \rightarrow \infty} \frac{1 - I(\mathcal{W})}{2k} \xrightarrow{k \rightarrow \infty} 0.$$

Decoding C_q^k : We shall assume that we are given an effective decoding algorithm of the quantum polar code P_q , capable of achieving the symmetric coherent information of the channel. In the next chapter, we indeed provide an efficient decoding for qubit Pauli channels (Section 3.5), but it remains an open problem for general quantum channels. In this case, C_q^k can be decoded sequentially, by decoding first P_q^0 , then P_q^1 , P_q^2 , and so on. Indeed, after decoding P_q^0 , thus in particular correcting the state of the \mathcal{I}'_0 system, the EPR pairs $\Phi_{\mathcal{I}'_0 \mathcal{J}_1}$ will play the role of the preshared entanglement required to decode P_q^1 . Therefore, P_q^1 can be decoded once P_q^0 has been decoded, and similarly, P_q^l can be decoded after P_q^{l-1} has been decoded, for any $l \in \{2, \dots, k-1\}$.

Entanglement as a catalyst: Finally, the above coding scheme can be slightly modified, such that preshared entanglement between the sender and the receiver is not consumed. In the above construction, we have considered that for the last P_q^{k-1} polar code, the \mathcal{I}'_{k-1} system is an information system, i.e., used to transmit quantum information from the

sender to the receiver (system \mathcal{I}'_2 in Fig. 2.6). Let us now assume that the input quantum state to the \mathcal{I}'_{k-1} system is half of a maximally entangled state $\Phi_{\mathcal{I}'_{k-1}\mathcal{J}_k}$, where quantum system \mathcal{J}_k is held by the sender. When the receiver completes decoding of the C_q^k code, it restores the initial state of the \mathcal{I}'_{k-1} , thus resulting in a maximally entangled state $\Phi_{\mathcal{I}'_{k-1}\mathcal{J}_k}$ shared between the sender (\mathcal{J}_k system) and the receiver (\mathcal{I}'_{k-1} system). Hence, the initial preshared entanglement $\Phi_{\mathcal{J}_0\mathcal{J}'_0}$ acts as a catalyst, in that it produces a new state $\Phi_{\mathcal{I}'_{k-1}\mathcal{J}_k}$ shared between the sender and the receiver, which can be used for the next transmission.

3

Purely Quantum Polar Codes for Qubit Pauli Channels

In this chapter, we further investigate the quantum polarization from Chapter 2 for the particular case of qubit Pauli channels (Definition 45). First, to a Pauli channel, we associate a classical non-binary symmetric channel, with both input and output alphabets given by a set containing four elements. We refer to it as the classical counterpart of the Pauli channel. We further define a channel combining and splitting procedure for the classical counterpart, using the permutation generated by a two qubit Clifford on the two qubit Pauli group as the channel combining operation. We then show that a Pauli channel and its classical counterpart polarize simultaneously under their respective channel combining and splitting procedure. Therefore, the quantum polarization of a Pauli channel implies the polarization of its classical counterpart and vice-versa. We use this equivalence to provide an alternative proof of the quantum polarization of a Pauli channel, by explicitly showing the polarization of its classical counterpart, using techniques from classical polarization. In particular, we show polarization when the channel combining operation is randomly chosen from the set of nine Clifford unitary \mathcal{L} given in Figure 2.4, and also when it is randomly chosen from a subset $\mathcal{S} \subset \mathcal{L}$ containing three elements. Finally, we provide an effective method to decode the quantum polar code on a Pauli channel, by decoding its classical polar code on its classical counterpart channel. We also show a fast polarization property when the channel combining operation is chosen from \mathcal{S} , ensuring the reliability of the decoding.

3.1 Classical Counterpart of a Pauli channel

3.1.1 Classical Mixture of Pauli Channels

We will consider a slightly more general class of quantum channels, namely classical mixture of Pauli channels. The reason for this is that the virtual channels synthesized during polarization of a Pauli channel are *identifiable* (see Definition 72) to classical mixture of Pauli channels (this will be proved later in Proposition 78).

Definition 71 (Classical Mixture of Pauli (CMP) channels). *A classical mixture of Pauli (CMP) channels is a quantum channel defined as, $\mathcal{N}(\rho) = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{N}_x(\rho)$, where $\{|x\rangle \mid x \in \mathcal{X}\}$ is some orthonormal basis of an auxiliary system, \mathcal{N}_x are Pauli channels, and $p_x \geq 0$ such that $\sum_x p_x = 1$. We will simply refer to \mathcal{N} as a CMP channel.*

Definition 72 (Identifiable Channels). *We say that a quantum channel $\mathcal{N}_{A \rightarrow B_1 B_2}$ is identifiable to a channel $\mathcal{N}'_{A \rightarrow B_1}$, if for some unitary operator C on the $B_1 B_2$ system, we have that*

$$\mathcal{N}(\rho) = C \left(\mathcal{N}'(\rho) \otimes \frac{I_{B_2}}{d_{B_2}} \right) C^\dagger, \quad (3.1)$$

where d_{B_2} denotes the dimension of the B_2 system.

3.1.2 Classical Counterpart of a CMP channel

The following notation will be used,

1. Let $\bar{\mathcal{G}}_N = \mathcal{G}_N / \{\pm 1, \pm i\}$ be the Abelian group obtained by taking the quotient of the N qubit Pauli group \mathcal{G}_N by its centralizer.
2. We write $\bar{\mathcal{G}}_1 = \{\sigma_i \mid i = 0, \dots, 3\}$, with $\sigma_0 = I$, $\sigma_1 = Z$, $\sigma_2 = X$, $\sigma_3 = Y$, and $\bar{\mathcal{G}}_2 = \{\sigma_{i,j} := \sigma_i \otimes \sigma_j \mid i, j = 0, \dots, 3\} \simeq \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$.
3. For any two-qubit Clifford unitary C , we denote by $\Gamma(C)$, or simply Γ when no confusion is possible, the conjugate action of C on $\bar{\mathcal{G}}_2$. In other words, Γ is the automorphism of $\bar{\mathcal{G}}_2$ (or equivalently $\bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$), defined by $\Gamma(\sigma_{i,j}) = C \sigma_{i,j} C^\dagger$.

We will first introduce the classical counterpart of a Pauli channels and then we will extend the definition to CMP channels.

Definition 73 (Classical counterpart of a Pauli channel). *Let \mathcal{N} be a Pauli channel that acts as $\mathcal{N}(\rho) = \sum_i p_i \sigma_i \rho \sigma_i$. The classical counterpart of \mathcal{N} , denoted by $\mathcal{N}^\#$, is the classical channel with the input and output alphabets $\bar{\mathcal{G}}_1$, and transition probabilities $\mathcal{N}^\#(\sigma_i | \sigma_j) = p_k$, where k is such that $\sigma_i \sigma_j = \sigma_k$, where equality is understood as equivalence classes in $\bar{\mathcal{G}}_1$.*

Note that the classical counterpart of a Pauli channel is a memoryless symmetric channel. Hence, its capacity is given by the symmetric mutual information, which is defined below.

Definition 74. *The symmetric mutual information of a classical channel W with input set $\bar{\mathcal{G}}_1$ is given by,*

$$I(W) := \frac{1}{4} \sum_y \sum_{x \in \bar{\mathcal{G}}_1} W(y|x) \log_2 \frac{W(y|x)}{W(y)} \in [0, 2], \quad (3.2)$$

where $W(y) = \frac{1}{4} \sum_{x' \in \bar{\mathcal{G}}_1} W(y|x')$.

From (3.2), the symmetric mutual information of $\mathcal{N}^\#$ is given by,

$$I(\mathcal{N}^\#) = 2 - H(\mathbf{p}) \in [0, 2], \quad (3.3)$$

where $H(\mathbf{p})$ is the Shannon entropy of the probability vector $\mathbf{p} = (p_0, p_1, p_2, p_3)$. From (3.3) and (1.124) it follows that, we have the following relation between the symmetric mutual information of $\mathcal{N}^\#$ and the symmetric coherent information of \mathcal{N} ,

$$I(\mathcal{N}^\#) = 1 + I(\mathcal{N}). \quad (3.4)$$

Note that the right hand side of (3.4) is the symmetric mutual information of the Pauli channel \mathcal{N} . We now extend the definition of the classical counterpart to CMP channels.

Definition 75 (Classical counterpart of a CMP channel). *Given a CMP channel $\mathcal{N}(\rho) = \sum_{x \in \mathcal{X}} p_x |x\rangle\langle x| \otimes \mathcal{N}_x(\rho)$, its classical classical counterpart $\mathcal{N}^\#$ is a channel from $\mathcal{X} \times \bar{\mathcal{G}}_1$ to $\bar{\mathcal{G}}_1$, with transition probability,*

$$\mathcal{N}^\#(x, \sigma_i | \sigma_j) = p_x \mathcal{N}_x^\#(\sigma_i | \sigma_j).$$

It can be seen that the symmetric mutual information of the classical counterpart of a CMP channel \mathcal{N} is given by,

$$\mathbb{I}(\mathcal{N}^\#) = \sum_x p_x \mathbb{I}(\mathcal{N}_x^\#). \quad (3.5)$$

Hence, from the above equation and Lemma 59, it follows that the relation in (3.4) also holds for the CMP channel.

Definition 76 (Equivalent classical channels). *We say two classical channels W and W' are equivalent and denote it by $W \equiv W'$, if they have the same transition probability matrix, modulo a permutation of rows and columns.*

As mentioned before, the virtual channels synthesized during polarization process of a Pauli channel are identifiable to a CMP channel. It can be seen that if a quantum channel is identifiable to two CMP channels $\mathcal{N}_{A \rightarrow B_1}$ and $\mathcal{N}'_{A \rightarrow B_1}$, then $(\mathcal{N}')^\#$ and $(\mathcal{N}'')^\#$ are classically equivalent channels in the sense of Definition 76. This follows from the lemma below.

Lemma 77. *Let \mathcal{N}' and \mathcal{N}'' be two CMP channels, such that the following holds for some unitary C ,*

$$\mathcal{N}'(\rho) \otimes \frac{I_{B_2}}{d_{B_2}} = C \left(\mathcal{N}''(\rho) \otimes \frac{I_{B_2}}{d_{B_2}} \right) C^\dagger, \quad (3.6)$$

where d_{B_2} is dimension of B_2 . Then, $(\mathcal{N}')^\# \equiv (\mathcal{N}'')^\#$ in the sense of Definition 76.

Proof. We restrict ourselves to the case when \mathcal{N}' and \mathcal{N}'' are Pauli channels, since the case of CMP channels follows in a similar manner, by introducing an auxiliary system providing a classical description of the Pauli channel being used. Hence, we may write

$$\mathcal{N}'(\rho) = \sum_{i=0}^3 p'_i \sigma_i \rho \sigma_i^\dagger, \quad (3.7)$$

$$\mathcal{N}''(\rho) = \sum_{i=0}^3 p''_i \sigma_i \rho \sigma_i^\dagger, \quad (3.8)$$

where $\sum_{i=0}^3 p'_i = \sum_{i=0}^3 p''_i = 1$. It follows that $\mathcal{N}'(\sigma_k) = \alpha'_k \sigma_k$ and $\mathcal{N}''(\sigma_k) = \alpha''_k \sigma_k$, where $\alpha'_0 = \alpha''_0 = 1$, and for $k = 1, 2, 3$, $\alpha'_k = p'_0 + p'_k - p'_{k_1} - p'_{k_2}$, $\alpha''_k = p''_0 + p''_k - p''_{k_1} - p''_{k_2}$, with $\{k_1, k_2\} = \{1, 2, 3\} \setminus \{k\}$. Using the vector notation $\mathbf{p}' := (p'_0, p'_1, p'_2, p'_3)$, and similarly $\mathbf{p}'', \alpha', \alpha''$, the above equalities rewrite as

$$\alpha' = A\mathbf{p}' \text{ and } \alpha'' = A\mathbf{p}'', \quad (3.9)$$

$$\text{where } A := \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Now, replacing ρ by σ_k in (3.6), we have that

$$\alpha'_k \sigma_k \otimes I_{B_2} = C (\alpha''_k \sigma_k \otimes I_{B_2}) C^\dagger. \quad (3.10)$$

Since the conjugate action of the unitary C preserves the Hilbert–Schmidt norm of an operator, it follows that $\|\alpha'_k \sigma_k \otimes I_{B_2}\|_{\text{HS}} = \|\alpha''_k \sigma_k \otimes I_{B_2}\|_{\text{HS}}$, and therefore $|\alpha'_k| = |\alpha''_k|$.

Case 1: We first assume that $\alpha'_k = \alpha''_k, \forall k = 1, 2, 3$. In this case, using (3.9), it follows that $\mathbf{p}' = \mathbf{p}''$, and therefore $(\mathcal{N}')^\# = (\mathcal{N}'')^\#$.

Case 2: We consider now the case when $\alpha'_k \neq \alpha''_k$, for some $k = 1, 2, 3$. To address this case, we start by writing $C = \sum_{i=0}^3 \sigma_i \otimes C_i$, where C_i are linear operators on the system B_2 . Hence, (3.6) rewrites as

$$\mathcal{N}'(\rho) \otimes \frac{I_{B_2}}{d_{B_2}} = \sum_{i,j} \left(\sigma_i \mathcal{N}''(\rho) \sigma_j^\dagger \right) \otimes \frac{C_i C_j^\dagger}{d_{B_2}}. \quad (3.11)$$

Tracing out the B_2 system, we have

$$\mathcal{N}'(\rho) = \sum_{i,j} \gamma_{i,j} \sigma_i \mathcal{N}''(\rho) \sigma_j^\dagger, \quad \text{where } \gamma_{i,j} = \frac{\text{Tr}(C_i C_j^\dagger)}{d_{B_2}}. \quad (3.12)$$

We define $\gamma_i := \gamma_{i,i}$, and from (3.12) it follows that $\gamma_i := \gamma_{i,i} \in \mathbb{R}_+$. Replacing $\rho = \sigma_k$ in (3.12), we have that for all $k = 0, \dots, 3$,

$$\alpha'_k \sigma_k = \alpha''_k \sum_i \gamma_i \sigma_i \sigma_k \sigma_i^\dagger + \alpha''_k \sum_{i,j, i \neq j} \gamma_{i,j} \sigma_i \sigma_k \sigma_j^\dagger. \quad (3.13)$$

The left hand side of the above equation has only σ_k term, so only σ_k on the right hand side should survive as Pauli matrices form an orthogonal basis. It follows that either $\alpha'_k = \alpha''_k = 0$, or the terms of the second sum in the right hand side of the above equation necessarily cancel each other. In both cases, we have that

$$\alpha'_k \sigma_k = \alpha''_k \sum_i \gamma_i \sigma_i \sigma_k \sigma_i^\dagger = \alpha''_k \lambda_k \sigma_k, \quad (3.14)$$

$$\text{and thus, } \alpha'_k = \lambda_k \alpha''_k, \quad (3.15)$$

$$\text{where, } \lambda_0 := \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 \quad (3.16)$$

$$\lambda_1 := \gamma_0 + \gamma_1 - \gamma_2 - \gamma_3 \quad (3.17)$$

$$\lambda_2 := \gamma_0 - \gamma_1 + \gamma_2 - \gamma_3 \quad (3.18)$$

$$\lambda_3 := \gamma_0 - \gamma_1 - \gamma_2 + \gamma_3 \quad (3.19)$$

We also note that $\lambda_0 = 1$, since $\alpha'_0 = \alpha''_0 = 1$. We further rewrite (3.15) as

$$\boldsymbol{\alpha}' = \Lambda \boldsymbol{\alpha}''. \quad (3.20)$$

where $\Lambda = \text{diag}(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ is the square diagonal matrix with λ_i 's on the main diagonal. Plugging (3.9) into (3.20), and using $A^2 = 4I$, we get

$$\mathbf{p}' = \frac{1}{4} A \Lambda A \mathbf{p}'' = \Gamma \mathbf{p}'', \quad (3.21)$$

$$\text{where } \Gamma := \frac{1}{4} A \Lambda A = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_1 & \gamma_0 & \gamma_3 & \gamma_2 \\ \gamma_2 & \gamma_3 & \gamma_0 & \gamma_1 \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix}$$

We now come back to our assumption, namely $\alpha'_k \neq \alpha''_k$, for some $k = 1, 2, 3$. Without loss of generality, we may assume that $\alpha'_1 \neq \alpha''_1$. Since $|\alpha'_1| = |\alpha''_1|$ and $\alpha'_1 = \lambda_1 \alpha''_1$, it follows that $\lambda_1 = -1$. Then, using (3.16) and (3.17), we have that $2(\gamma_0 + \gamma_1) = \lambda_0 + \lambda_1 = 0$, which implies

$$\gamma_0 = \gamma_1 = 0, \quad (3.22)$$

since they are non-negative. We proceed now with several sub-cases:

Case 2.1: either $\alpha'_2 \neq \alpha''_2$ or $\alpha'_3 \neq \alpha''_3$. Similarly to the derivation of (3.22), we get either $\gamma_2 = 0$ (in which case $\gamma_3 = 1$) or $\gamma_3 = 0$ (in which case $\gamma_2 = 1$). In either case Λ is a permutation matrix, which implies that $(\mathcal{N}')^\# \equiv (\mathcal{N}'')^\#$, as desired.

Case 2.2: $\alpha'_2 = \alpha''_2$ and $\alpha'_3 = \alpha''_3$, and either $\alpha'_2 = \alpha''_2 \neq 0$ or $\alpha'_3 = \alpha''_3 \neq 0$. Let us assume that $\alpha'_2 = \alpha''_2 \neq 0$. In this case, using (3.15), we have that $\lambda_2 = 1$, and from (3.18) it follows that $\gamma_2 - \gamma_3 = 1$. This implies $\gamma_2 = 1$ and $\gamma_3 = 0$, therefore Λ is a permutation matrix, and thus $(\mathcal{N}')^\# \equiv (\mathcal{N}'')^\#$, as desired.

Case 2.3: $\alpha'_2 = \alpha''_2 = 0$ and $\alpha'_3 = \alpha''_3 = 0$. Using $\alpha'_k = 2(p'_0 + p'_k) - 1, \forall k \neq 0$, we get $p'_2 = p'_3 = \frac{1}{2} - p'_0$, and similarly $p''_2 = p''_3 = \frac{1}{2} - p''_0$. Moreover, using (3.21) and the fact that $\gamma_2 + \gamma_3 = 1$, we get $p'_0 = p'_1 = p''_2 = p''_3$ and $p'_2 = p'_3 = p''_0 = p''_1$. This implies that $(\mathcal{N}')^\# \equiv (\mathcal{N}'')^\#$, as desired.

This concludes the second case, and finishes the proof. \square

3.1.3 Channel Combining and Splitting Procedure for the Classical Counterpart Channel

Simplified notation: To simplify notation, we shall identify $(\bar{\mathcal{G}}_1, \times) \cong (\{0, 1, 2, 3\}, \oplus)$, by identifying $\sigma_u \cong u, \forall u = 0, \dots, 3$, where the additive group operation $u \oplus v$ is given by the bitwise exclusive OR (XOR) between the binary representations of integers u, v . The classical counterpart $\mathcal{N}^\#$ of a Pauli channel $\mathcal{N}(\rho) = \sum_{u=0}^3 p_u \sigma_u \rho \sigma_u^\dagger$ (Definition 73), is therefore identified to a channel with input and output alphabet $\bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, and transition probabilities $\mathcal{N}^\#(u | v) = p_{u \oplus v}$.

Let N and M be two classical channels, both with the input alphabet $\bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, and output alphabets A and B , respectively. Channel transition probabilities are denoted by $N(a | u)$ and $M(b | v)$, for $u, v \in \bar{\mathcal{G}}_1, a \in A$, and $b \in B$. Let $\Gamma : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$ be any permutation, and write $\Gamma = (\Gamma_1, \Gamma_2)$, with $\Gamma_i : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1, i = 1, 2$. The channel combining and splitting procedures are depicted in Figure 3.1 below.

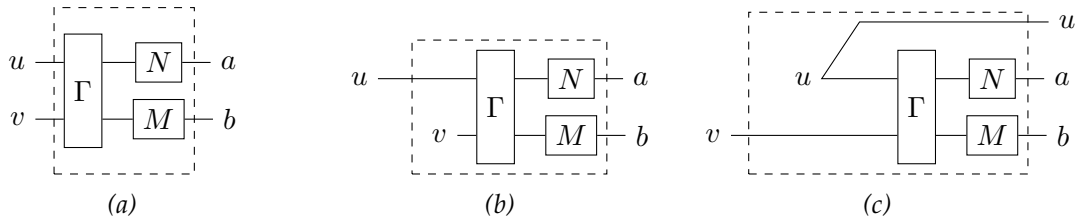


Figure 3.1: Channel combining and splitting. (a) combined channel: $N \bowtie_\Gamma M(a, b | u, v)$. (b) bad channel: $N \boxtimes_\Gamma M(a, b | u)$. (c) good channel: $N \oplus_\Gamma M(u, a, b | v)$.

The combined channel $N \bowtie_\Gamma M$ is defined as,

$$(N \bowtie_\Gamma M)(a, b | u, v) = N(a | \Gamma_1(u, v)) M(b | \Gamma_2(u, v)) \quad (3.23)$$

The combined channel is further *split* into two channels $N \boxtimes_{\Gamma} M$ and $N \otimes_{\Gamma} M$, as follows

$$(N \boxtimes_{\Gamma} M)(a, b | u) = \frac{1}{4} \sum_v (N \bowtie_{\Gamma} M)(a, b | u, v), \quad (3.24)$$

$$(N \otimes_{\Gamma} M)(a, b, u | v) = \frac{1}{4} (N \bowtie_{\Gamma} M)(a, b | u, v). \quad (3.25)$$

From the chain rule of the mutual information, we have that

$$I(N \boxtimes_{\Gamma} M) + I(N \otimes_{\Gamma} M) = I(N) + I(M), \quad (3.26)$$

which means that the symmetric mutual information is preserved under the above channel combining and splitting procedure. Applying the above construction to the classical counterparts of two CMP channels, we prove the following proposition.

Proposition 78. *Let $\mathcal{N}_{U \rightarrow A}$ and $\mathcal{M}_{V \rightarrow B}$ be two CMP channels, and C be any two-qubit Clifford unitary, acting on the two qubit system UV . Let $\mathcal{N}^{\#}$ and $\mathcal{M}^{\#}$ denote the two classical counterparts of the above CMP channels, and $\Gamma := \Gamma(C)$ be the permutation induced by the conjugate action of C on $\bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$. Then $\mathcal{N} \boxtimes_C \mathcal{M}$ and $\mathcal{N} \otimes_C \mathcal{M}$ are identifiable to CMP channels, thus $(\mathcal{N} \boxtimes_C \mathcal{M})^{\#}$ and $(\mathcal{N} \otimes_C \mathcal{M})^{\#}$ are well defined, and the following properties hold:*

$$(a) \quad (\mathcal{N} \boxtimes_C \mathcal{M})^{\#} \equiv \mathcal{N}^{\#} \boxtimes_{\Gamma} \mathcal{M}^{\#}.$$

$$(b) \quad (\mathcal{N} \otimes_C \mathcal{M})^{\#} \equiv \mathcal{N}^{\#} \otimes_{\Gamma} \mathcal{M}^{\#}.$$

Proof. We identify the automorphism $\Gamma = \Gamma(C)$ induced by the conjugate action of a two-qubit Clifford unitary C on $\bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$, to a linear permutation $\Gamma : \{0, 1, 2, 3\}^2 \rightarrow \{0, 1, 2, 3\}^2$, such that $C\sigma_{i,j}C^{\dagger} = \sigma_{\Gamma(i,j)}$. We shall also write $\Gamma = (\Gamma_1, \Gamma_2)$, with $\Gamma_i : \{0, 1, 2, 3\}^2 \rightarrow \{0, 1, 2, 3\}$, $i = 1, 2$.

It can be easily seen that it is enough to prove the statements of the proposition for the case when \mathcal{N} and \mathcal{M} are Pauli channels. Let $\mathcal{N}(\rho) = \sum_{i=0}^3 p_i \sigma_i \rho \sigma_i^{\dagger}$ and $\mathcal{M}(\rho) = \sum_{j=0}^3 q_j \sigma_j \rho \sigma_j^{\dagger}$.

Point (a). We have the following,

$$\begin{aligned} (\mathcal{N} \boxtimes \mathcal{M})(\rho_U) &= (\mathcal{N} \otimes \mathcal{M}) \left(C \left(\rho_U \otimes \frac{I_V}{2} \right) C^{\dagger} \right) \\ &= \sum_{i,j} p_i q_j \sigma_{i,j} C \left(\rho_U \otimes \frac{I_V}{2} \right) C^{\dagger} \sigma_{i,j}^{\dagger} \\ &= \sum_{i,j} r_{i,j} C \sigma_{\Gamma^{-1}(i,j)} \left(\rho_U \otimes \frac{I_V}{2} \right) \sigma_{\Gamma^{-1}(i,j)}^{\dagger} C^{\dagger}, \text{ where } r_{i,j} := p_i q_j \\ &= C \left(\sum_{i,j} r_{\Gamma(i,j)} \sigma_{i,j} \left(\rho_U \otimes \frac{I_V}{2} \right) \sigma_{i,j}^{\dagger} \right) C^{\dagger} \\ &= C \left(\sum_{i,j} r_{\Gamma(i,j)} \sigma_{i,j} \left(\rho_U \otimes \frac{I_V}{2} \right) \sigma_{i,j}^{\dagger} \right) C^{\dagger} \\ &= C \left(\sum_{i,j} r_{\Gamma(i,j)} \sigma_i \rho_U \sigma_i^{\dagger} \otimes \frac{I_V}{2} \right) C^{\dagger} \\ &= C \left(\sum_i s_i \sigma_i \rho_U \sigma_i^{\dagger} \otimes \frac{I_V}{2} \right) C^{\dagger}, \text{ where } s_i := \sum_j r_{\Gamma(i,j)} \end{aligned} \quad (3.27)$$

where the fourth equality follows from the variable change $(i, j) \mapsto \Gamma(i, j)$. Omitting the conjugate action of the unitary C and discarding the V system, we may further identify:

$$(\mathcal{N} \boxtimes \mathcal{M})(\rho_U) = \sum_i s_i \sigma_i \rho_U \sigma_i^\dagger.$$

Hence, the associated classical channel $(\mathcal{N} \boxtimes \mathcal{M})^\#$ is defined by the probability vector $\mathbf{s} = (s_0, s_1, s_2, s_3)$, meaning that

$$(\mathcal{N} \boxtimes \mathcal{M})^\#(i | j) = s_{i \oplus j}. \quad (3.28)$$

On the other hand, we have:

$$\begin{aligned} (\mathcal{N}^\# \boxtimes \mathcal{M}^\#)(a, b | u) &= \frac{1}{4} \sum_v \mathcal{N}^\#(a | \Gamma_1(u, v)) \mathcal{M}^\#(b | \Gamma_2(u, v)) \\ &= \frac{1}{4} \sum_v p_{a \oplus \Gamma_1(u, v)} q_{b \oplus \Gamma_2(u, v)}. \end{aligned} \quad (3.29)$$

Applying Γ^{-1} on the channel output, we may identify $\mathcal{N}^\# \boxtimes \mathcal{M}^\#$ to a channel with output $(a', b') = \Gamma^{-1}(a, b)$, and transition probabilities given by:

$$\begin{aligned} (\mathcal{N}^\# \boxtimes \mathcal{M}^\#)(a', b' | u) &= \frac{1}{4} \sum_v p_{\Gamma_1(a', b') \oplus \Gamma_1(u, v)} q_{\Gamma_2(a', b') \oplus \Gamma_2(u, v)} \\ &= \frac{1}{4} \sum_v p_{\Gamma_1((a', b') \oplus (u, v))} q_{\Gamma_2((a', b') \oplus (u, v))} \\ &= \frac{1}{4} \sum_v p_{\Gamma_1(a' \oplus u, b' \oplus v)} q_{\Gamma_2(a' \oplus u, b' \oplus v)} \\ &= \frac{1}{4} \sum_v p_{\Gamma_1(a' \oplus u, v)} q_{\Gamma_2(a' \oplus u, v)} \\ &= \frac{1}{4} \sum_v r_{\Gamma(a' \oplus u, v)} \\ &= \frac{1}{4} s_{a' \oplus u}. \end{aligned} \quad (3.30)$$

We can then discard the b' output, since the channel transition probabilities do not depend on it, which gives a channel defined by transition probabilities:

$$(\mathcal{N}^\# \boxtimes \mathcal{M}^\#)(a' | u) = s_{a' \oplus u}. \quad (3.31)$$

Finally, using (3.28) and (3.31), and noticing that omitting the conjugate action of the unitary C and discarding the V system in the derivation of (3.28) is equivalent to applying Γ^{-1} on the channel output and discarding the b' output in the derivation of (3.31), we conclude that $(\mathcal{N} \boxtimes \mathcal{M})^\# \equiv \mathcal{N}^\# \boxtimes \mathcal{M}^\#$.

Point (b). Similar to the derivations used for point (i), we get

$$\begin{aligned} (\mathcal{N} \otimes \mathcal{M})(\rho_V) &= C \left(\sum_{i,j} r_{\Gamma(i,j)} \sigma_{i,j} (\Phi_{U'U} \otimes \rho_V) \sigma_{i,j}^\dagger \right) C^\dagger \\ &= C \left(\sum_{i,j} r_{\Gamma(i,j)} \left((I_{U'} \otimes \sigma_i)(\Phi_{U'U})(I_{U'} \otimes \sigma_i^\dagger) \right) \otimes (\sigma_j \rho_V \sigma_j^\dagger) \right) C^\dagger \end{aligned} \quad (3.32)$$

Omitting the conjugate action of the unitary C , and expressing $(I_{U'} \otimes \sigma_i)(\Phi_{U'U})(I_{U'} \otimes \sigma_i^\dagger)$ in the Bell basis, $\{|i\rangle\}_{i=0,\dots,3} := \{\frac{|00\rangle+|11\rangle}{\sqrt{2}}, \frac{|01\rangle+|10\rangle}{\sqrt{2}}, \frac{|01\rangle-|10\rangle}{\sqrt{2}}, \frac{|00\rangle-|11\rangle}{\sqrt{2}}\}$, we get:

$$(\mathcal{N} \otimes \mathcal{M})(\rho_V) = \sum_{i,j} r_{\Gamma(i,j)} |i\rangle\langle i| \otimes (\sigma_j \rho_V \sigma_j^\dagger).$$

Let $\lambda_i := \sum_j r_{\Gamma(i,j)}$ and $s_{i,j} := r_{\Gamma(i,j)}/\lambda_i$ (with $s_{i,j} := 0$ if $\lambda_i = 0$). Denoting by \mathcal{S}_i the Pauli channel defined by $\mathcal{S}(\rho)_i = \sum_j s_{i,j} \sigma_j \rho \sigma_j^\dagger$, we may rewrite:

$$(\mathcal{N} \otimes \mathcal{M})(\rho_V) = \sum_i \lambda_i |i\rangle\langle i| \otimes \mathcal{S}_i(\rho_V).$$

Hence, $(\mathcal{N} \otimes \mathcal{M})^\#$ is the mixture of the channels $\mathcal{S}_i^\#$, with $\mathcal{S}_i^\#$ being used with probability λ_i , whose transition probabilities are given by:

$$(\mathcal{N} \otimes \mathcal{M})^\#(i, j | k) = \lambda_i s_{i,j \oplus k} = r_{\Gamma(i,j \oplus k)}. \quad (3.33)$$

On the other hand, we have:

$$\begin{aligned} (\mathcal{N}^\# \otimes \mathcal{M}^\#)(a, b, u | v) &= \frac{1}{4} \mathcal{N}^\#(a | \Gamma_1(u, v)) \mathcal{M}^\#(b | \Gamma_2(u, v)) \\ &= \frac{1}{4} p_{a \oplus \Gamma_1(u, v)} q_{b \oplus \Gamma_2(u, v)}. \end{aligned} \quad (3.34)$$

We apply Γ^{-1} on the (a, b) output of the channel, which is equivalent to omitting the conjugate action of the unitary C in (3.32), and then identify $\mathcal{N}^\# \otimes \mathcal{M}^\#$ to a channel with output (a', b', u) , where $(a', b') = \Gamma^{-1}(a, b)$, and transition probabilities:

$$\begin{aligned} (\mathcal{N}^\# \otimes \mathcal{M}^\#)(a', b', u | v) &= \frac{1}{4} p_{\Gamma_1(a', b') \oplus \Gamma_1(u, v)} q_{\Gamma_2(a', b') \oplus \Gamma_2(u, v)} \\ &= \frac{1}{4} p_{\Gamma_1(a' \oplus u, b' \oplus v)} q_{\Gamma_2(a' \oplus u, b' \oplus v)} \\ &= \frac{1}{4} r_{\Gamma(a' \oplus u, b' \oplus v)}. \end{aligned} \quad (3.35)$$

We further perform a change of variable, replacing (a', u) by $(a' \oplus u, u)$, which makes the above transition probability independent of u . We may then discard the u output, and thus identify $\mathcal{N}^\# \otimes \mathcal{M}^\#$ to a channel with output (a', b') and transition probabilities:

$$(\mathcal{N}^\# \otimes \mathcal{M}^\#)(a', b' | v) = r_{\Gamma(a', b' \oplus v)}. \quad (3.36)$$

Finally, using (3.33) and (3.36), we conclude that $(\mathcal{N} \otimes \mathcal{M})^\# \equiv \mathcal{N}^\# \otimes \mathcal{M}^\#$. \square

A consequence of the above proposition is that a CMP channel polarizes under the recursive application of the channel combining and splitting rules, if and only if its classical counterpart does so. Moreover, processes of both quantum and classical polarization yield the same set of indices for the good/bad channels. More precisely, we have the following.

Corollary 79. *Let \mathcal{W} be a CMP channel, and $\mathcal{W}^{(i_1 \dots i_n)}$ be defined recursively as in (2.44), $\forall n > 0, \forall i_1 \dots i_n \in \{0, 1\}^n$. Let $\mathcal{W}^\#$ be the classical counterpart of \mathcal{W} , and $(\mathcal{W}^\#)^{(i_1 \dots i_n)}$ be defined recursively, similar to (2.44), while replacing \mathcal{W} by $\mathcal{W}^\#$, and Clifford unitaries $C_{i_1 \dots i_n}$ by the corresponding permutations $\Gamma_{i_1 \dots i_n} := \Gamma(C_{i_1 \dots i_n})$. Then $(\mathcal{W}^{(i_1 \dots i_n)})^\# \equiv (\mathcal{W}^\#)^{(i_1 \dots i_n)}$, $\forall n, \forall i_1 \dots i_n \in \{0, 1\}^n$. In particular:*

$$I((\mathcal{W}^\#)^{(i_1 \dots i_n)}) = 1 + I(\mathcal{W}^{(i_1 \dots i_n)}).$$

As we have already proven in Theorem 60 that as n goes to infinity, $I(\mathcal{W}^{(i_1 \dots i_n)})$ approaches to either -1 or 1 , hence, $I((\mathcal{W}^\#)^{(i_1 \dots i_n)})$ approaches to either 0 or 2 . Therefore, polarization also happens for the classical counterpart.

3.2 Proof of Quantum Polarization Using the Classical Counterpart Channel

In this section, we provide an alternative proof of quantum polarization for the Pauli channel by proving classical polarization for the classical counterpart channel. The channel combining here is chosen randomly from the set of nine two-qubit Clifford unitaries \mathcal{L} given in Figure 2.4. The proof of polarization for the classical counterpart may be obtained by verifying the conditions from Lemma 54, with the stochastic process $\{T_n : n \geq 0\}$ given by the Bhattacharyya parameters Z_n of the synthesized virtual channels. Note that we only need to prove the guaranteed improvement condition under the set \mathcal{L} as all the other conditions are already shown in [24].

We give below the definition of the Bhattacharyya parameter for a classical channel W with non-binary input, as defined in [24]. We shall restrict our attention to classical channels with input alphabet $\bar{\mathcal{G}}_1$.

Definition 80 (Bhattacharyya Parameter [24]). *Let W be a classical channel, with input alphabet $\bar{\mathcal{G}}_1 \cong (\{0, 1, 2, 3\}, \oplus)$ and output alphabet Y . For $u, u', d \in \bar{\mathcal{G}}_1$, we define*

$$Z(W_{u,u'}) := \sum_{y \in Y} \sqrt{W(y|u)W(y|u')} \in [0, 1].$$

$$Z_d(W) := \frac{1}{4} \sum_{u \in \bar{\mathcal{G}}_1} Z(W_{u,u \oplus d}) \in [0, 1].$$

In particular, note that $Z(W_{u,u}) = 1, \forall u \in \bar{\mathcal{G}}_1$, and $Z_0(W) = 1$. The Bhattacharyya parameter of W , denoted $Z(W)$, is then defined as

$$Z(W) := \frac{1}{3} \sum_{d \neq 0} Z_d(W) = \frac{1}{12} \sum_{u \neq u'} Z(W_{u,u'}) \in [0, 1].$$

In the following lemma, we show the guaranteed improvement for the Bhattacharyya parameter, which implies the polarization of the classical counterpart.

Lemma 81. *Let \mathcal{W} be a CMP channel and $W := \mathcal{W}^\#$ its classical counterpart. Given two instances of the channel W , we have that*

$$\mathbb{E}_{\Gamma \in \Gamma(\mathcal{L})} Z(W \otimes_\Gamma W) = \frac{1}{3} Z(W) + \frac{2}{3} Z(W)^2 \leq Z(W). \quad (3.37)$$

where $\Gamma(\mathcal{L})$ denotes the set of permutations generated on the two-qubit Pauli group $\bar{\mathcal{G}}_2$ by the conjugate action of Cliffords in \mathcal{L} . Further, the inequality in the above equation is an equality if and only if $Z(W) \in \{0, 1\}$.

Before proving Lemma 81, we will prove Lemmas 82 and 83.

Lemma 82. *For any two classical channels N, M , with input alphabet $\bar{\mathcal{G}}_1 \cong (\{0, 1, 2, 3\}, \oplus)$, and any linear permutation $\Gamma = (A, B) : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$, the following equality holds for any $d \in \bar{\mathcal{G}}_1$:*

$$Z_d(N \otimes_\Gamma M) = Z_{A(0,d)}(N) Z_{B(0,d)}(M).$$

Proof. According to Definition 80, for the channel $N \otimes_{\Gamma} M$, we have that

$$\begin{aligned}
 Z((N \otimes_{\Gamma} M)_{v,v'}) &= \sum_{u,y_1,y_2} \sqrt{(N \otimes_{\Gamma} M)(y_1, y_2, u | v) (N \otimes_{\Gamma} M)(y_1, y_2, u | v')} \\
 &= \frac{1}{4} \sum_{u,y_1,y_2} \sqrt{N(y_1 | A(u, v)) M(y_2 | B(u, v)) N(y_1 | A(u, v')) M(y_2 | B(u, v'))} \\
 &= \frac{1}{4} \sum_{u,y_1,y_2} \sqrt{N(y_1 | A(u, v)) N(y_1 | A(u, v')) M(y_2 | B(u, v)) M(y_2 | B(u, v'))} \\
 &= \frac{1}{4} \sum_u Z(N_{A(u,v), A(u,v')}) Z(M_{B(u,v), B(u,v')}).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 Z_d(N \otimes_{\Gamma} M) &= \frac{1}{4} \sum_v Z((N \otimes_{\Gamma} M)_{v,v \oplus d}) \\
 &= \frac{1}{16} \sum_{u,v} Z(N_{A(u,v), A(u,v \oplus d)}) Z(M_{B(u,v), B(u,v \oplus d)}) \\
 &= \frac{1}{16} \sum_{u,v} Z(N_{A(u,v), A(u,v) \oplus A(0,d)}) Z(M_{B(u,v), B(u,v) \oplus B(0,d)}) \\
 &= \frac{1}{16} \sum_a Z(N_{a, a \oplus A(0,d)}) \sum_b Z(M_{b, b \oplus B(0,d)}) \\
 &= Z_{A(0,d)}(N) Z_{B(0,d)}(M),
 \end{aligned} \tag{3.38}$$

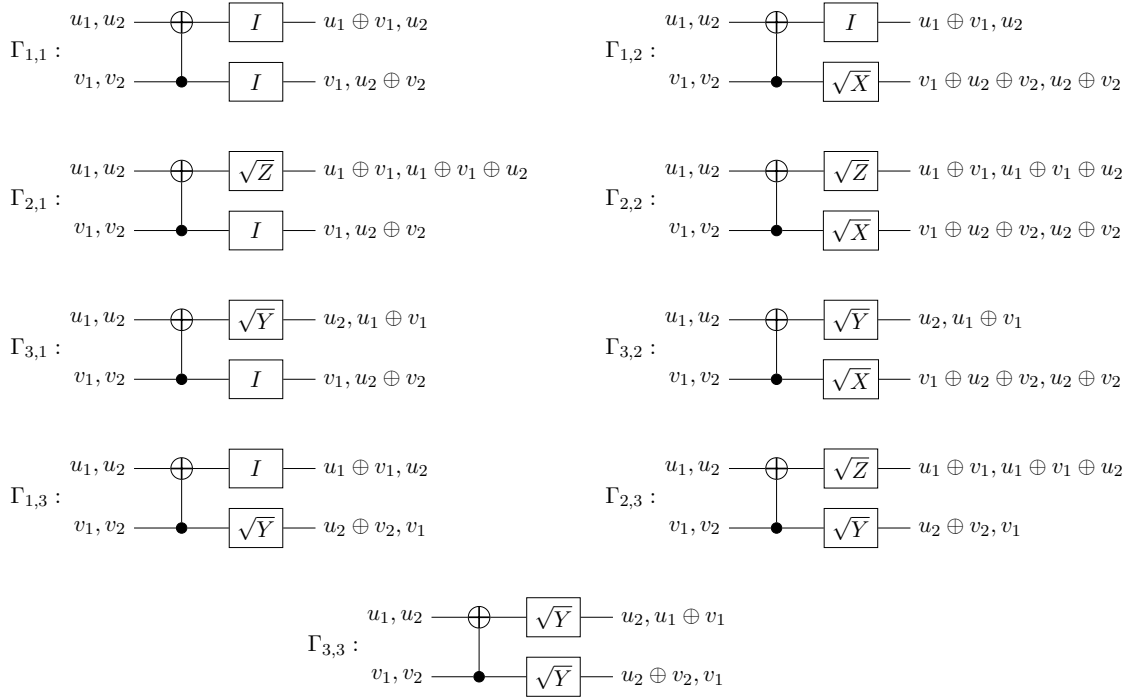
where the third equality follows from the linearity of the permutation $\Gamma = (A, B)$, and the fourth equality follows from the change of basis for the summation from (u, v) to $(a, b) := (A(u, v), B(u, v))$. \square

We denote by $u := [u_1, u_2]$ the binary representation of a given $u \in \bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, where $u_1, u_2 \in \{0, 1\}$ and u_2 is the least significant bit.

Lemma 83. Let $\Gamma_{i,j} := \Gamma(L_{i,j}) : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$ be the permutation defined by the conjugate action of $L_{i,j} \in \mathcal{L}$. Then $\Gamma_{i,j} = (A_i, B_j), \forall 1 \leq i, j \leq 3$, with $A_i, B_j : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1$ given by:

$$\begin{aligned}
 A_1(u, v) &= [u_1 \oplus v_1, u_2], & B_1(u, v) &= [v_1, u_2 \oplus v_2] \\
 A_2(u, v) &= [u_1 \oplus v_1, u_1 \oplus v_1 \oplus u_2], & B_2(u, v) &= [v_1 \oplus u_2 \oplus v_2, u_2 \oplus v_2] \\
 A_3(u, v) &= [u_2, u_1 \oplus v_1] & B_3(u, v) &= [u_2 \oplus v_2, v_1]
 \end{aligned}$$

where inputs u and v are represented in binary form, $u := [u_1, u_2]$ and $v := [v_1, v_2]$, with $u_1, u_2, v_1, v_2 \in \{0, 1\}$ ($\Gamma_{i,j}$ permutations are also depicted in Fig. 3.2).


 Figure 3.2: Elements of the set $\Gamma(\mathcal{L})$

Proof. Recall from Section 2.4.2, that $L_{i,j} = (C' \otimes C'')C_{2 \rightarrow 1}$, where $C' \in \{I, \sqrt{Z}, \sqrt{Y}\}$, $C'' \in \{I, \sqrt{X}, \sqrt{Y}\}$, and $C_{2 \rightarrow 1}$ is the CNOT gate. Recall also that by identifying $\bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, we have $I = \sigma_0 \cong 0$, $Z = \sigma_1 \cong 1$, $X = \sigma_2 \cong 2$, $Y = \sigma_3 \cong 3$. The conjugate action of \sqrt{X} on $\bar{\mathcal{G}}_1$, fixes I and X , and permutes Y and Z . Hence, the corresponding permutation on $\bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, can be written as $(0, 3, 2, 1)$. Similarly, the conjugate action of \sqrt{Y} and \sqrt{Z} induces the permutations $(0, 2, 1, 3)$ and $(0, 1, 3, 2)$, respectively. Replacing $u \in \{0, 1, 2, 3\}$ by its binary representation $[u_1, u_2]$, we may write:

$$\begin{aligned} \sqrt{X} : [u_1, u_2] &\mapsto [u_1 \oplus u_2, u_2] \\ \sqrt{Y} : [u_1, u_2] &\mapsto [u_2, u_1] \\ \sqrt{Z} : [u_1, u_2] &\mapsto [u_1, u_1 \oplus u_2] \end{aligned} \tag{3.39}$$

Moreover, the permutation induced by the conjugate action of the $C_{2 \rightarrow 1}$ gate is the linear permutation on $\bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$ such that:

$$\begin{aligned} C_{2 \rightarrow 1} : (X, I) &\mapsto (X, I), \quad (I, X) \mapsto (X, X) \\ (Z, I) &\mapsto (Z, Z), \quad (I, Z) \mapsto (I, Z) \\ \Rightarrow C_{2 \rightarrow 1} : ([u_1, u_2], [v_1, v_2]) &\mapsto ([u_1 \oplus v_1, u_2], [v_1, u_2 \oplus v_2]). \end{aligned} \tag{3.40}$$

Finally, using (3.39) and (3.40), it can be easily verified that $\Gamma_{i,j} = (A_i, B_j), \forall 1 \leq i, j \leq 3$, with A_i and B_j as given in the lemma. \square

Proof of Lemma 81. Applying Lemma 82 and Lemma 83, we may express $Z_d(W \otimes_{\Gamma_{i,j}} W)$ as a function of the parameters $(Z_1(W), Z_2(W), Z_3(W))$, for any $\Gamma_{i,j} \in \Gamma(\mathcal{L})$ and any $d = 1, 2, 3$ (recall that $Z_0(W) = 1$). The corresponding expressions are given in Table 3.1.

(i, j)	$Z_1(W \otimes_{\Gamma_{i,j}} W)$	$Z_2(W \otimes_{\Gamma_{i,j}} W)$	$Z_3(W \otimes_{\Gamma_{i,j}} W)$
(1, 1)	$Z_1(W)$	$Z_2(W)^2$	$Z_2(W)Z_3(W)$
(1, 2)	$Z_3(W)$	$Z_2(W)^2$	$Z_1(W)Z_2(W)$
(1, 3)	$Z_2(W)$	$Z_1(W)Z_2(W)$	$Z_2(W)Z_3(W)$
(2, 1)	$Z_1(W)$	$Z_2(W)Z_3(W)$	$Z_3(W)^2$
(2, 2)	$Z_3(W)$	$Z_2(W)Z_3(W)$	$Z_1(W)Z_3(W)$
(2, 3)	$Z_2(W)$	$Z_1(W)Z_3(W)$	$Z_3(W)^2$
(3, 1)	$Z_1(W)$	$Z_1(W)Z_2(W)$	$Z_1(W)Z_3(W)$
(3, 2)	$Z_3(W)$	$Z_1(W)Z_2(W)$	$Z_1(W)^2$
(3, 3)	$Z_2(W)$	$Z_1(W)^2$	$Z_1(W)Z_3(W)$

 Table 3.1: $Z_d(W \otimes_{\Gamma_{i,j}} W)$ as a function of $(Z_1(W), Z_2(W), Z_3(W))$

Hence,

$$\begin{aligned}
 \sum_{\Gamma \in \Gamma(\mathcal{L})} Z(W \otimes_{\Gamma} W) &= \frac{1}{3} \sum_{\Gamma \in \Gamma(\mathcal{L})} \sum_{d=1}^3 Z_d(W \otimes_{\Gamma} W) \\
 &= \sum_{d=1}^3 Z_d(W) + \frac{2}{3} \left(\sum_{d=1}^3 Z_d(W) \right)^2 \\
 &= 3Z(W) + 6Z(W)^2,
 \end{aligned} \tag{3.41}$$

and therefore,

$$\begin{aligned}
 \mathbb{E}_{\Gamma \in \Gamma(\mathcal{L})} Z(W \otimes_{\Gamma} W) &= \frac{1}{9} \sum_{\Gamma \in \Gamma(\mathcal{L})} Z(W \otimes_{\Gamma} W) \\
 &= \frac{1}{3} Z(W) + \frac{2}{3} Z(W)^2.
 \end{aligned}$$

□

Remark 84. It can be directly seen that $\mathbb{E}_{\Gamma \in \Gamma(\mathcal{L})} Z(W \otimes_{\Gamma} W) = \mathbb{E}_{\Gamma \in \Gamma(\mathcal{L}')} Z(W \otimes_{\Gamma} W)$, where the elements of the set \mathcal{L}' is connected to the elements of \mathcal{L} by the swap gate, as defined in Section 2.4.2.

3.3 Further Reducing the Channel Combining Set

In this section, we first show polarization, when the channel combining operation is randomly chosen from a set of three two-qubit Cliffords, namely $L_{1,3}, L_{2,2}, L_{3,1}$. Then, we attempt to completely derandomize the channel combining operation, with the help of a numerical simulation. We plot the channel combining operation for two steps of polarization for all the nine permutations, which suggests that polarization may happen for $L_{1,3}, L_{2,2}, L_{3,1}$ individually.

3.3.1 Polarization with a Set Containing Three Two-Qubit Clifford Unitaries

Let \mathcal{S} denote the set containing the Clifford gates $L_{1,3}$, $L_{2,2}$, and $L_{3,1}$ from Fig. 2.4, and $\Gamma(\mathcal{S})$ denote the corresponding set of permutations, namely $\Gamma(L_{1,3})$, $\Gamma(L_{2,2})$ and $\Gamma(L_{3,1})$, generated by the conjugate actions of $L_{1,3}$, $L_{2,2}$, and $L_{3,1}$ on $\bar{\mathcal{G}}_2$.

Lemma 85. *Let \mathcal{W} be a CMP channel and $W := \mathcal{W}^\#$ its classical counterpart. Given two instances of the channel W , then*

$$\mathbb{E}_{\Gamma \in \Gamma(\mathcal{S})} Z(W \otimes_{\Gamma} W) \leq \frac{1}{3} Z(W) + \frac{2}{3} Z(W)^2. \quad (3.42)$$

Proof. Using Table 3.1, for $\Gamma \in \Gamma(\mathcal{S}) = \{\Gamma_{1,3}, \Gamma_{2,2}, \Gamma_{3,1}\}$, we get

$$\begin{aligned} \mathbb{E}_{\Gamma \in \Gamma(\mathcal{S})} Z(W \otimes_{\Gamma} W) &= \frac{1}{3} \sum_{\Gamma \in \Gamma(\mathcal{S})} Z(W \otimes_{\Gamma} W) \\ &= \frac{1}{9} \sum_{\Gamma \in \Gamma(\mathcal{S})} \sum_{1 \leq d \leq 3} Z_d(W \otimes_{\Gamma} W) \\ &= \frac{1}{9} \sum_{1 \leq d \leq 3} Z_d(W) + \frac{2}{9} \sum_{1 \leq d' \neq d'' \leq 3} Z_{d'}(W) Z_{d''}(W) \\ &\leq \frac{1}{3} Z(W) + \frac{2}{3} Z(W)^2, \end{aligned}$$

where, using $Z(W) = (Z_1(W) + Z_2(W) + Z_3(W))/3$, it is easily seen that the last inequality is equivalent to the following inequality,

$$Z_1(W)Z_2(W) + Z_1(W)Z_3(W) + Z_2(W)Z_3(W) \leq Z_1(W)^2 + Z_2(W)^2 + Z_3(W)^2. \quad (3.43)$$

The above inequality follows from $Z_i(W)Z_j(W) \leq (Z_i(W)^2 + Z_j(W)^2)/2$. \square

3.3.2 Polarization with Only One Two-Qubit Clifford Unitary

We note that the main ingredient of proof of polarizations in Sections 3.2 and 3.3.1 is the guaranteed improvement of the good channel. The following two observations are in order,

1. Polarization happens if we have guaranteed degradation¹ instead of guaranteed improvement as Lemma 54 holds with minor modifications for guaranteed degradation.
2. Polarization also happens if guaranteed improvement or degradation happens for a virtual channel after two polarization steps instead of one.

From Table 3.1, it can be verified that none of the permutations yield guaranteed improvement for the good channel after one polarization step. In the remaining section, we investigate with the help of a computer program whether guaranteed improvement or degradation happens after two polarization steps for any of the nine permutations.

We discretize the probability values defining the Pauli channels. In Figures 3.3-3.11 below, each point in the black region corresponds to a different Pauli channel. In the figures, we have plotted the Bhattacharyya parameter of the virtual channels obtained after two

¹For guaranteed degradation, we have $f(\theta) > \theta$, when $B_{n+1} = 0$ (instead of $f(\theta) < \theta$, when $B_{n+1} = 1$) in condition (t.2) of Lemma 54.

steps of polarization, that is, $Z(W^{(i_1 i_2)})$, where W is a Pauli channel and $i_1, i_2 \in \{0, 1\}$, with respect to $Z(W)$, for all 9 permutations. For guaranteed improvement, we require the black region to be always strictly below the red line except at the end values $Z(W) = 0, 1$ and similarly, for guaranteed degradation, we require the black part to be always strictly above the red line.

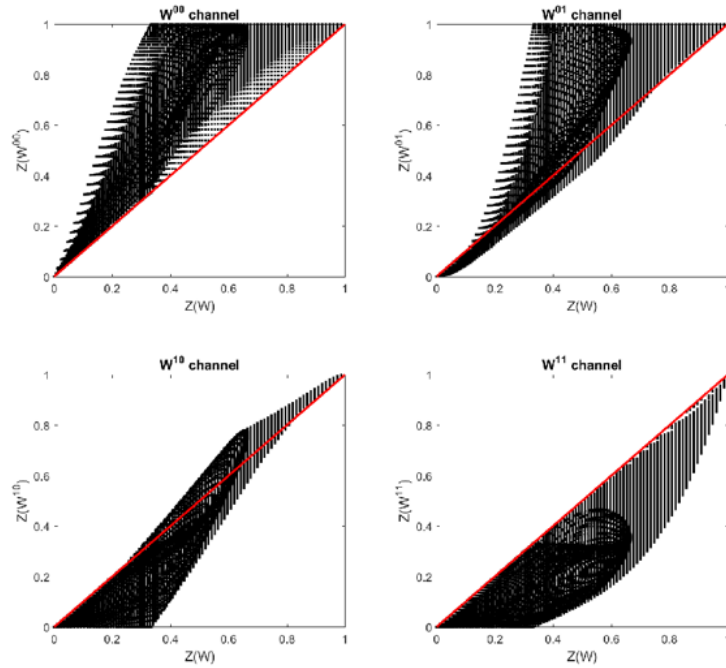


Figure 3.3: Permutation $\Gamma_{1,1}$, no guaranteed improvement or degradation for any virtual channel.

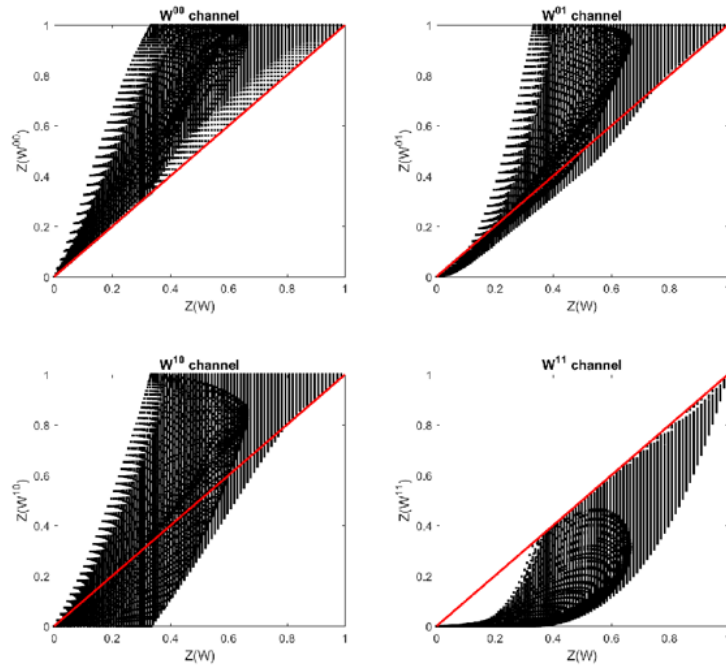


Figure 3.4: Permutation $\Gamma_{1,2}$, no guaranteed improvement or degradation for any virtual channel.

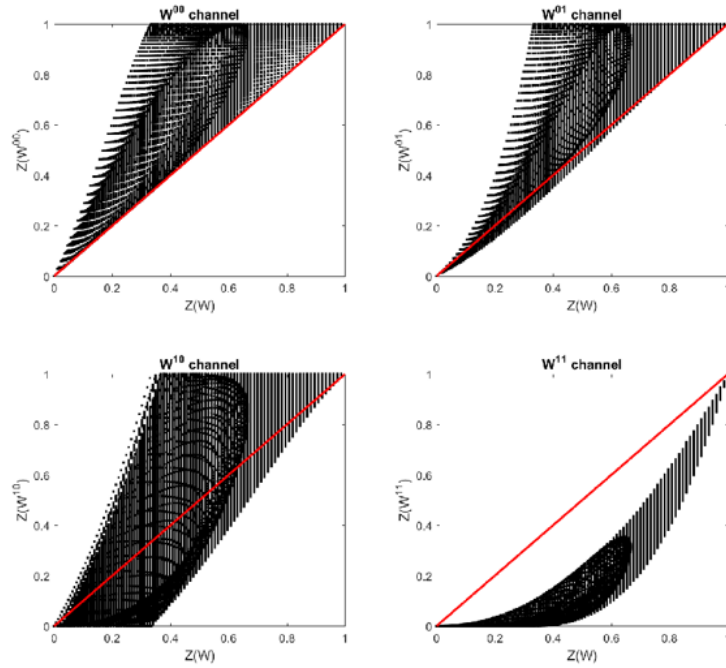


Figure 3.5: Permutation $\Gamma_{1,3}$, guaranteed improvement for the virtual channel $W^{(11)}$.

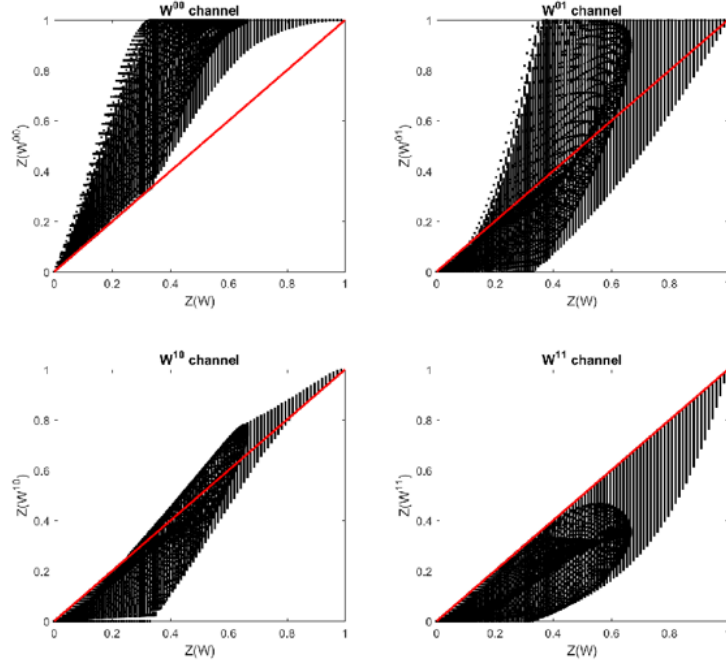


Figure 3.6: Permutation $\Gamma_{2,1}$, no guaranteed improvement or degradation for any virtual channel.

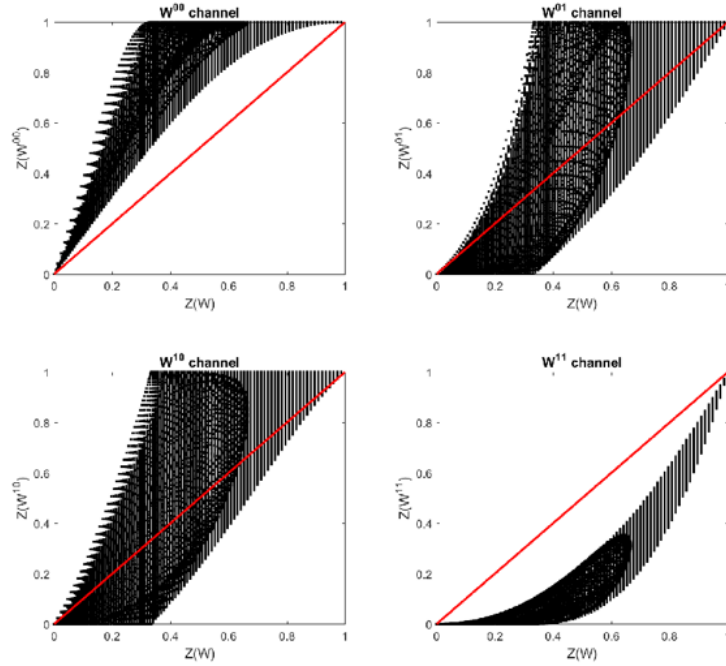


Figure 3.7: Permutation $\Gamma_{2,2}$, guaranteed improvement for the virtual channel $W^{(11)}$ and guaranteed degradation for the virtual channel $W^{(00)}$.

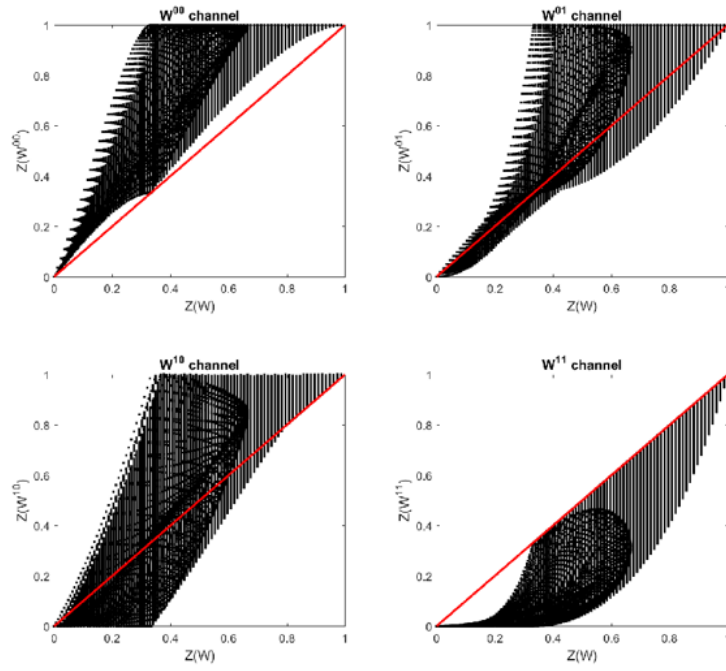


Figure 3.8: Permutation $\Gamma_{2,3}$, no guaranteed improvement or degradation for any virtual channel.

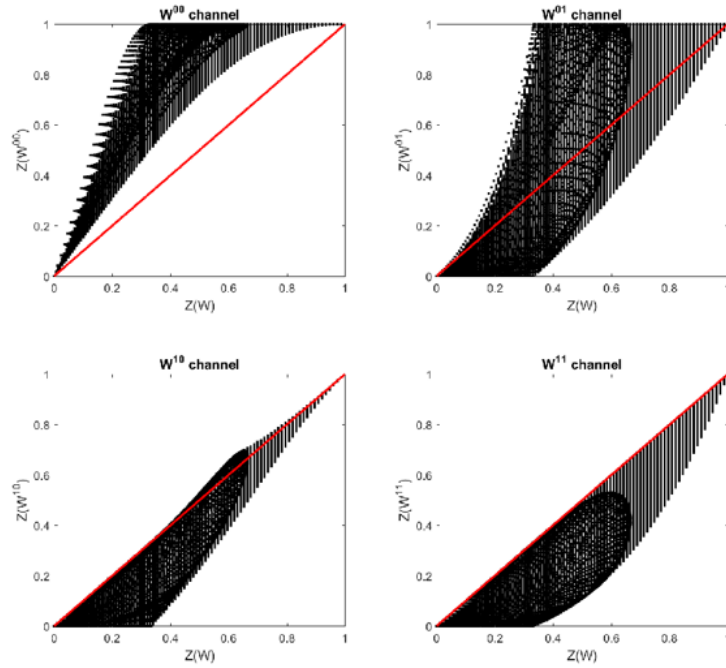


Figure 3.9: Permutation $\Gamma_{3,1}$, guaranteed degradation for the virtual channel $W^{(00)}$.

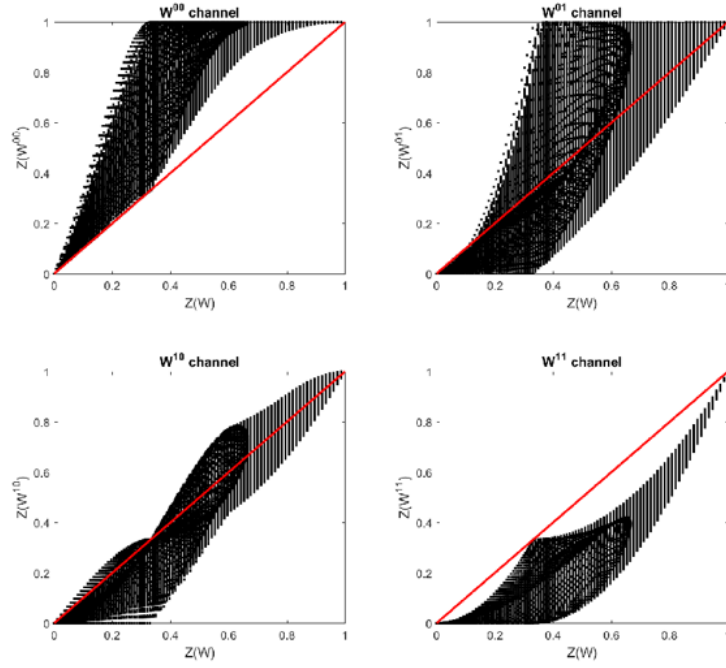


Figure 3.10: Permutation $\Gamma_{3,2}$, no guaranteed improvement or degradation for any virtual channel.

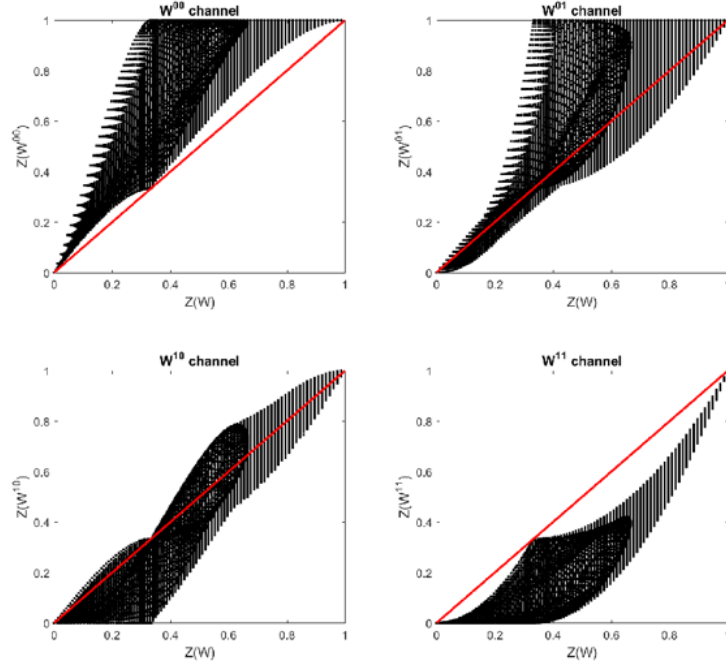


Figure 3.11: Permutation $\Gamma_{3,3}$, no guaranteed improvement or degradation for any virtual channel.

To conclude, the numerical results suggest that polarization might happen for two-qubit Cliffords $\Gamma_{1,3}, \Gamma_{2,2}, \Gamma_{3,1}$ individually. However, to prove polarization, we would need a

rigorous proof of guaranteed improvement or degradation after two polarization steps for CMP channels, which is left as an open problem. Finally, we note the following.

Remark 86. For permutation $\Gamma_{3,3}$, we almost have the guaranteed degradation for $W^{(00)}$ and the guaranteed improvement for $W^{(11)}$ as the black region touches the red line at only one point $Z(W) = \frac{1}{3}$. We show in Chapter 4 that the recursive application of $\Gamma_{3,3}$ yields a multilevel polarization and a quantum polar code can be constructed using this multilevel polarization.

3.4 Fast Polarization

Before discussing decoding of quantum polar codes over Pauli channels (Section 3.5 below), recall that classical polar codes are equipped with a decoding algorithm, known as successive cancellation (SC) [2]. However, the effectiveness of the classical SC decoding, *i.e.*, its capability of successfully decoding at rates close to the capacity, depends on the speed of polarization. The Bhattacharyya parameter of the synthesized channels plays an important role in determining the speed at which polarization takes place. First, we note that for a classical channel W , the Bhattacharyya parameter upper bounds the error probability of uncoded transmission. Precisely, given a classical channel W with input alphabet X , the error probability of the maximum a posteriori estimate for a single channel use, denoted P_e , is upper-bounded as follows ([24, Proposition 2]):

$$P_e \leq (|X| - 1)Z(W).$$

Now, consider a polar code defined by the recursive application of n polarization steps to the classical channel $W := \mathcal{W}^\#$ (the input alphabet is $X := \bar{\mathcal{G}}_1$, of size $|\bar{\mathcal{G}}_1| = 4$). The construction is the same as the one in Section 2.5.1, while replacing the quantum channel \mathcal{W} by its classical counterpart W , and channel combining Clifford gates $C_{i_1 i_2 \dots}$ by the corresponding permutations $\Gamma_{i_1 i_2 \dots} := \Gamma(C_{i_1 i_2 \dots})$. For any $i = 0, \dots, N - 1$, let $W^{(i)} := (\mathcal{W}^\#)^{(i_1 \dots i_n)}$, where $i_1 \dots i_n$ is the binary decomposition of i . For the sake of simplicity, we drop the channel combining permutations Γ 's from the above notation. Let $\mathcal{I} \subset \{0, 1, \dots, N - 1\}$ denote the set of good channels (*i.e.*, channels used to transmit information symbols, as opposed to bad channels, which are frozen to symbol values known to both the encoder and decoder). Since the SC decoding proceeds by decoding successively the synthesized good channels², it can be easily seen that the block error probability of the SC decoder, denoted by $P_e(N, \mathcal{I})$, is upper-bounded by (see also [2, Proposition 2]):

$$P_e(N, \mathcal{I}) \leq 3 \sum_{i \in \mathcal{I}} Z(W^{(i)}). \quad (3.44)$$

If the Bhattacharyya parameters of the $W^{(i)}$ channels, with $i \in \mathcal{I}$, converge sufficiently fast to zero, one can use (3.44) to ensure that $P_e(N, \mathcal{I})$ goes to zero. Since the number of terms in the right hand side of (3.44) is linear in N , it is actually enough to prove that $Z(W^{(i)}) \leq O(N^{-(1+\theta)})$, $\forall i \in \mathcal{I}$, for some $\theta > 0$.

The proof of fast polarization properties given in [25] (see also [24, Lemma 3]) exploits the following two main ingredients:

²An estimate of the input of each good channel is generated using the maximum a posteriori estimate, according to the observed channel output and the previously decoded channels.

- (1) The quadratic improvement of the Bhattacharyya parameter, when taking the good channel, i.e., $Z(W^{(i_1 \dots i_{n-1} i_n)}) \leq Z(W^{(i_1 \dots i_{n-1})})^2$, $\forall i_1 \dots i_{n-1} i_n \in \{0, 1\}^n$, such that $i_n = 1$.
- (2) The linearly upper-bounded degradation of the Bhattacharyya parameter, when taking the bad channel, i.e., $Z(W^{(i_1 \dots i_{n-1} i_n)}) \leq \kappa Z(W^{(i_1 \dots i_{n-1})})$, $\forall i_1 \dots i_{n-1} i_n \in \{0, 1\}^n$, such that $i_n = 0$, for some constant $\kappa > 0$.

Regarding the second condition, in our case we have the following lemma, where for a classical channel W with input alphabet $\bar{\mathcal{G}}_1 \cong \{0, 1, 2, 3\}$, we define

$$\bar{Z}(W) := \max_{d=1,2,3} Z_d(W). \quad (3.45)$$

Lemma 87. *For any classical channel W with input alphabet $\bar{\mathcal{G}}_1$, and any linear permutation $\Gamma : \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1 \rightarrow \bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$, the following inequalities hold:*

$$\begin{aligned} \bar{Z}(W \boxtimes_{\Gamma} W) &\leq 4\bar{Z}(W). \\ Z(W \boxtimes_{\Gamma} W) &\leq 12Z(W). \end{aligned}$$

Proof. According to Definition 80, for the channel $N \boxtimes_{\Gamma} M$, we have:

$$\begin{aligned} Z((N \boxtimes_{\Gamma} M)_{u, u'}) &= \sum_{y_1, y_2} \sqrt{(N \boxtimes_{\Gamma} M)(y_1, y_2 | u) (N \boxtimes_{\Gamma} M)(y_1, y_2 | u')} \\ &= \frac{1}{4} \sum_{y_1, y_2} \sqrt{\sum_v N(y_1 | A(u, v)) M(y_2 | B(u, v)) \sum_{v'} N(y_1 | A(u', v')) M(y_2 | B(u', v'))} \\ &\leq \frac{1}{4} \sum_{v, v'} \sum_{y_1, y_2} \sqrt{N(y_1 | A(u, v)) M(y_2 | B(u, v)) N(y_1 | A(u', v')) M(y_2 | B(u', v'))} \\ &= \frac{1}{4} \sum_{v, v'} \sum_{y_1, y_2} \sqrt{N(y_1 | A(u, v)) N(y_1 | A(u', v')) M(y_2 | B(u, v)) M(y_2 | B(u', v'))} \\ &= \frac{1}{4} \sum_{v, v'} Z(N_{A(u, v), A(u', v')}) Z(M_{B(u, v), B(u', v')}) , \end{aligned}$$

where the inequality above follows from $\sqrt{\sum_v x_v} \leq \sum_v \sqrt{x_v}$. Therefore,

$$\begin{aligned} Z_d(N \boxtimes_{\Gamma} M) &= \frac{1}{4} \sum_u Z((N \boxtimes_{\Gamma} M)_{u, u \oplus d}) \\ &\leq \frac{1}{16} \sum_{u, v, v'} Z(N_{A(u, v), A(u \oplus d, v')}) Z(M_{B(u, v), B(u \oplus d, v')}) \\ &= \frac{1}{16} \sum_{u, v, d'} Z(N_{A(u, v), A(u \oplus d, v \oplus d')}) Z(M_{B(u, v), B(u \oplus d, v \oplus d')}) \\ &= \frac{1}{16} \sum_{u, v, d'} Z(N_{A(u, v), A(u, v) \oplus A(d, d')}) Z(M_{B(u, v), B(u, v) \oplus B(d, d')}) \\ &= \frac{1}{16} \sum_{d'} \sum_a Z(N_{a, a \oplus A(d, d')}) \sum_b Z(M_{b, b \oplus B(d, d')}) \\ &= \sum_{d'} Z_{A(d, d')}(N) Z_{B(d, d')}(M), \end{aligned} \quad (3.46)$$

where the third to last equality follows from the linearity of the permutation $\Gamma = (A, B)$, and the second to last follows from the change of basis for the summation from (u, v) to $(a, b) := (A(u, v), B(u, v))$.

Now, using (3.46), we have that

$$Z_d(W \boxtimes_{\Gamma} W) \leq \sum_{d' \in \bar{\mathcal{G}}_1} Z_{A(d, d')}(W) Z_{B(d, d')}(W).$$

For $d \neq 0$, $A(d, d')$ and $B(d, d')$ cannot be simultaneously zero (recall that $Z_0(W) = 1$), and therefore we get $Z_{A(d, d')}(W) Z_{B(d, d')}(W) \leq \bar{Z}(W)$. Hence, $Z_d(W \boxtimes_{\Gamma} W) \leq 4\bar{Z}(W)$, $\forall d = 1, 2, 3$, which implies $\bar{Z}(W \boxtimes_{\Gamma} W) \leq 4\bar{Z}(W)$, as desired. Finally, we have

$$Z(W \boxtimes_{\Gamma} W) \leq \bar{Z}(W \boxtimes_{\Gamma} W) \leq 4\bar{Z}(W) \leq 12Z(W),$$

which proves the second inequality of the lemma. \square

The condition (1) above – quadratic improvement of the Bhattacharyya parameter, when taking the good channel – is more problematic, due to the linear term in the right hand side of (3.37) and (3.42). In particular, we can not apply [24, Lemma 3] to derive fast polarization properties in our case. Instead, we will prove fast polarization properties by drawing upon arguments similar to those in the proof of [2, Theorem 2]. First, we need the following definition.

Definition 88. Let W be a classical channel with input alphabet $\bar{\mathcal{G}}_1$, and $\Gamma = \{\Gamma, \Gamma_{i_1 \dots i_n} \mid n > 0, i_1 \dots i_n \in \{0, 1\}^n\}$ be an infinite sequence of permutations. For $n > 0$, let

$$W^{(i_1 \dots i_n)} := \begin{cases} W^{(i_1 \dots i_{n-1})} \boxtimes_{\Gamma_{i_1 \dots i_{n-1}}} W^{(i_1 \dots i_{n-1})}, & \text{if } i_n = 0 \\ W^{(i_1 \dots i_{n-1})} \oplus_{\Gamma_{i_1 \dots i_{n-1}}} W^{(i_1 \dots i_{n-1})}, & \text{if } i_n = 1 \end{cases} \quad (3.47)$$

where, for $n = 1$, in the right hand side term of the above equality, we set by convention $W^{(\emptyset)} := W$ and $\Gamma_{\emptyset} := \Gamma$. We say that Γ is a polarizing sequence (or that polarization happens for Γ), if for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \frac{\#\{(i_1 \dots i_n) \in \{0, 1\}^n : I(W^{(i_1 \dots i_n)}) \in (\delta, 1 - \delta)\}}{2^n} = 0.$$

Note that contrary to Lemma 81 and Lemma 85, we consider here a given sequence of permutations, instead of averaging over some set of sequences. If $W = \mathcal{W}^{\#}$ is the classical counterpart of a CMP channel \mathcal{W} , by Lemma 85, we know that polarization happens when averaging over all the sequences $\Gamma \in \Gamma(\mathcal{S})^{\infty}$. As a consequence, there exists a subset $\Gamma(\mathcal{S})_{\text{pol}}^{\infty} \subset \Gamma(\mathcal{S})^{\infty}$ of positive probability³, such that polarization happens for any $\Gamma \in \Gamma(\mathcal{S})_{\text{pol}}^{\infty}$. We are now ready to state the following fast polarization result.

Proposition 89. Let \mathcal{W} be a CMP channel, $W := \mathcal{W}^{\#}$ its classical counterpart, and \mathcal{S} the set of three Clifford gates from Section 3.3.1. Then the following fast polarization property holds for almost all Γ sequences in $\Gamma(\mathcal{S})_{\text{pol}}^{\infty}$:

³Note that $\Gamma(\mathcal{S})^{\infty}$ is the infinite product space of countable many copies of $\Gamma(\mathcal{S})$, and it is endowed with the infinite product probability measure, taking the uniform probability measure on each copy of $\Gamma(\mathcal{S})$. See [65] for infinite product probability measures.

For any $\theta > 0$ and $R < \mathbb{I}(W)$, there exists a sequence of sets $\mathcal{I}_N \subset \{0, \dots, N-1\}$, $N \in \{1, 2, \dots, 2^n, \dots\}$, such that $|\mathcal{I}_N| \geq NR$ and $Z(W^{(i)}) \leq O(N^{-(1+\theta)})$, $\forall i \in \mathcal{I}_N$. In particular, the block error probability of polar coding under SC decoding satisfies

$$P_e(N, \mathcal{I}_N) \leq O(N^{-\theta}).$$

We proceed first with several lemmas, then prove the above proposition. In the following, the notation $x = x(\cdot)$ means that the value of x depends only on the list of variables (\cdot) enclosed between parentheses.

Lemma 90. (i) For any permutation $\Gamma \in \Gamma(\mathcal{S})$, there exist $\delta_1 = \delta_1(\Gamma)$, $\delta_2 = \delta_2(\Gamma)$, $\delta_3 = \delta_3(\Gamma)$, such that $\{\delta_1, \delta_2, \delta_3\} = \{1, 2, 3\}$, and

$$\begin{aligned} Z_1(W \otimes_{\Gamma} W) &= Z_{\delta_1}(W) \\ Z_2(W \otimes_{\Gamma} W) &= Z_{\delta_1}(W)Z_{\delta_2}(W) \\ Z_3(W \otimes_{\Gamma} W) &= Z_{\delta_1}(W)Z_{\delta_3}(W), \end{aligned}$$

and the above equalities hold for any channel W .

(ii) For any $d \in \{1, 2, 3\}$, there exists exactly one permutation $\Gamma \in \Gamma(\mathcal{S})$, such that $\delta_1(\Gamma) = d$.

Proof. Follows from Table 3.1, wherein $\Gamma(\mathcal{S}) = \{\Gamma_{1,3}, \Gamma_{2,2}, \Gamma_{3,1}\}$. Precisely, we have $\delta_1(\Gamma_{1,3}) = 2$, $\delta_1(\Gamma_{2,2}) = 3$, $\delta_1(\Gamma_{3,1}) = 1$. \square

Lemma 91. There exist a constant $\kappa > 1$ and $\delta = \delta(W) \in \{1, 2, 3\}$, such that for any $\Gamma \in \Gamma(\mathcal{S})$ and any $d \in \{1, 2, 3\}$, the following equality holds

$$Z_d(W \boxtimes_{\Gamma} W) \leq \kappa Z_{\delta}(W).$$

Proof. Follows from Lemma 87, for $\kappa = 4$ and $\delta = \delta(W) := \arg\max_{d=1,2,3} Z_d(W)$. \square

We shall also use the following lemma (known as Hoeffding's inequality) providing an upper bound for the probability that the mean of n independent random variables falls below its expected value mean by a positive number.

Lemma 92 ([66, Theorem 1]). Let X_1, X_2, \dots, X_n be independent random variables such that $0 \leq X_i \leq 1$, for any $i = 1, \dots, n$. Let $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$, and $\mu = \mathbb{E}(\bar{X})$. Then, for any $0 < t < \mu$,

$$\Pr \{ \bar{X} \leq \mu - t \} \leq e^{-2nt^2}.$$

Now, let $\Gamma(\mathcal{S})^{\infty}$ be the infinite Cartesian product of countable many copies of $\Gamma(\mathcal{S})$. It is endowed with an infinite product probability measure [65], denoted by P , where the uniform probability measure is taken on each copy of $\Gamma(\mathcal{S})$. For our purposes, an infinite sequence $\Gamma \in \Gamma(\mathcal{S})^{\infty}$ should be written as $\Gamma := \{\Gamma, \Gamma_{i_1 \dots i_n} \mid n > 0, i_1 \dots i_n \in \{0, 1\}^n\}$ (this is always possible, since the set of indices is countable). We further define a sequence of independent and identically distributed (i.i.d) Bernoulli random variables on $\Gamma(\mathcal{S})^{\infty}$, denoted $\Delta^{i_1 \dots i_n}$, $n \geq 0$, $i_1 \dots i_n \in \{0, 1\}^n$,

$$\Delta^{i_1 \dots i_n}(\Gamma) := \mathbf{1}_{\{\delta_1(\Gamma_{i_1 \dots i_n}) \in \{2, 3\}\}},$$

that is, $\Delta^{i_1 \dots i_n}(\Gamma)$ is equal to 1, if $\delta_1(\Gamma_{i_1 \dots i_n}) \in \{2, 3\}$, and equal to 0, if $\delta_1(\Gamma_{i_1 \dots i_n}) = 1$. Note that $\Delta^{i_1 \dots i_n}(\Gamma)$ does actually only depend on the $\Gamma_{i_1 \dots i_n}$ element of Γ (here, n and $i_1 \dots i_n$ are fixed). From Lemma 90 (ii), it follows that $\mathbb{E}(\Delta^{i_1 \dots i_n}) = 2/3$, $\forall n \geq 0$, $\forall i_1 \dots i_n \in \{0, 1\}^n$.

For $0 < \gamma < 2/3$ and $m > 0$, we define

$$\Pi_m(\gamma) := \left\{ \Gamma \in \Gamma(\mathcal{S})^\infty \mid \sum_{i_1 \cdots i_{m-1}} \Delta^{i_1 \cdots i_{m-1} 1}(\Gamma) \geq \left(\frac{2}{3} - \gamma\right) 2^{m-1} \right\} \quad (3.48)$$

$$\bar{\Pi}_m(\gamma) := \bigcap_{n \geq m} \Pi_n(\gamma) \quad (3.49)$$

The sum in (3.48) comprises all the terms $\Delta^{i_1 \cdots i_{m-1} i_m}(\Gamma)$, with $i_1 \cdots i_{m-1} \in \{0, 1\}^{m-1}$ and $i_m = 1$ (here, m is fixed). Thus, $\Pi_m(\gamma)$ is defined by requiring that at least a fraction of $(2/3 - \gamma)$ of $\Delta^{i_1 \cdots i_{m-1} i_m}$ variables are equal to 1, where $i_m = 1$. In (3.49), the above condition must hold for any $n \geq m$.

Lemma 93. For any $0 < \gamma < 2/3$ and $m > 0$,

$$P(\bar{\Pi}_m(\gamma)) \geq 2 - \frac{1}{1 - e^{-\gamma^2 2^m}}. \quad (3.50)$$

Proof. By Lemma 92, $P(\Pi_m(\gamma)) \geq 1 - e^{-\gamma^2 2^m}$. Therefore, we have

$$\begin{aligned} P(\bar{\Pi}_m(\gamma)) &\geq 1 - \sum_{n \geq m} e^{-\gamma^2 2^n} \\ &= 1 - \sum_{n \geq 0} \left(e^{-\gamma^2 2^m}\right)^{2^n} \\ &\geq 1 - \sum_{n \geq 1} \left(e^{-\gamma^2 2^m}\right)^n \\ &= 1 - \left(\frac{1}{1 - e^{-\gamma^2 2^m}} - 1\right) \\ &= 2 - \frac{1}{1 - e^{-\gamma^2 2^m}}. \end{aligned}$$

□

Note that the right hand side term in (3.50) converges to 1 as m goes to infinity. Hence, for $\varepsilon > 0$, we denote by $m(\gamma, \varepsilon)$ the smallest m value, such that $2 - \frac{1}{1 - e^{-\gamma^2 2^m}} \geq 1 - \varepsilon$. It follows that $P(\bar{\Pi}_{m(\gamma, \varepsilon)}(\gamma)) \geq 1 - \varepsilon$.

In the following, we fix once for all some γ value, such that $0 < \gamma < 2/3$. The value of γ will not matter for any of what we do here, we only need $(2/3 - \gamma)$ to be positive. We proceed now with the proof of Proposition 89.

Proof of Proposition 89. Let $\Omega := \{0, 1\}^\infty$ denote the set of infinite binary sequences $\omega := (\omega_1 \omega_2 \cdots) \in \{0, 1\}^\infty$. Hence, Ω can be endowed with an infinite product probability measure, by taking the uniform probability measure on each ω_n component. We denote this probability measure by P (the notation is the same as for the probability measure on $\Gamma(\mathcal{S})^\infty$, but no confusion should arise, since the sample spaces are different).

Let $\varepsilon > 0$ and fix any $\Gamma \in \Gamma(\mathcal{S})_{\text{pol}}^\infty \cap \bar{\Pi}_{m(\gamma, \varepsilon)}(\gamma)$. Given Γ , the polarization process can be formally described as a random process on the probability space Ω [2]. Precisely, for any $\omega = (\omega_1 \omega_2 \cdots) \in \Omega$ and $n > 0$, we define

$$\begin{aligned} Z^{[n]}(\omega) &:= Z\left(W^{(\omega_1 \cdots \omega_n)}\right) \\ Z_d^{[n]}(\omega) &:= Z_d\left(W^{(\omega_1 \cdots \omega_n)}\right), \forall d \in \{1, 2, 3\} \end{aligned}$$

Note that $W^{(\omega_1 \dots \omega_n)}$ is recursively defined as in (3.47), through the implicit assumption of using the channel combining permutations in the given sequence Γ . For $n = 0$, we set $Z^{[0]}(\omega) := Z(W)$ and $Z_d^{[0]}(\omega) := Z_d(W)$.

For $\zeta > 0$ and $m \geq 0$, we define

$$T_m(\zeta) := \left\{ \omega \in \Omega \mid Z_d^{[n]}(\omega) \leq \zeta, \forall d = 1, 2, 3, \forall n \geq m \right\}.$$

Hence, for $\omega \in T_m(\zeta)$, $d \in \{1, 2, 3\}$, and $n > m$, we may write

$$Z_d^{[n]}(\omega) = \frac{Z_{d_n}^{[n]}(\omega)}{Z_{d_{n-1}}^{[n-1]}(\omega)} \frac{Z_{d_{n-1}}^{[n-1]}(\omega)}{Z_{d_{n-2}}^{[n-2]}(\omega)} \dots \frac{Z_{d_{m+1}}^{[m+1]}(\omega)}{Z_{d_m}^{[m]}(\omega)} Z_{d_m}^{[m]}(\omega), \quad (3.51)$$

where $d_n := d$, and d_{n-1}, \dots, d_m are defined as explained below. Recall that $Z_d^{[k]}(\omega) := Z_d(W^{(\omega_1 \dots \omega_k)})$, and for $k \in \{n, n-1, \dots, m+1\}$, we have

$$W^{(\omega_1 \dots \omega_k)} = \begin{cases} W^{(\omega_1 \dots \omega_{k-1})} \boxtimes_{\Gamma_{\omega_1 \dots \omega_{k-1}}} W^{(\omega_1 \dots \omega_{k-1})}, & \text{if } \omega_k = 0 \\ W^{(\omega_1 \dots \omega_{k-1})} \boxplus_{\Gamma_{\omega_1 \dots \omega_{k-1}}} W^{(\omega_1 \dots \omega_{k-1})}, & \text{if } \omega_k = 1 \end{cases}$$

Hence, if $\omega_k = 0$, we set $d_{k-1} := \delta(W^{(\omega_1 \dots \omega_{k-1})})$ from Lemma 91, such that we have

$$\frac{Z_{d_k}^{[k]}(\omega)}{Z_{d_{k-1}}^{[k-1]}(\omega)} \leq \kappa, \quad \text{if } \omega_k = 0. \quad (3.52)$$

If $\omega_k = 1$, we set $d_{k-1} := \delta_1(\Gamma_{\omega_1 \dots \omega_{k-1}})$ from Lemma 90, such that we have

$$\frac{Z_{d_k}^{[k]}(\omega)}{Z_{d_{k-1}}^{[k-1]}(\omega)} = 1, \quad \text{if } \omega_k = 1 \text{ and } d_k = 1. \quad (3.53)$$

$$\frac{Z_{d_k}^{[k]}(\omega)}{Z_{d_{k-1}}^{[k-1]}(\omega)} \leq \zeta, \quad \text{if } \omega_k = 1 \text{ and } d_k \in \{2, 3\}. \quad (3.54)$$

Let $A_{m,n}(\omega) := \{k \in \{m+1, \dots, n\} \mid \omega_k = 1\}$, and $B_{m,n}(\omega) := \{k \in \{m+1, \dots, n\} \mid \omega_k = 1 \text{ and } d_k \in \{2, 3\}\}$. Using (3.51), (3.52)–(3.54), for $\omega \in T_m(\zeta)$ and $n > m$, we get:

$$Z_d^{[n]}(\omega) \leq \kappa^{(n-m)-|A_{m,n}(\omega)|} \zeta^{|B_{m,n}(\omega)|} \zeta. \quad (3.55)$$

Now, we want to upper-bound the right hand side term of the above inequality, by providing lower-bounds for the $|A_{m,n}(\omega)|$ and $|B_{m,n}(\omega)|$ values.

$|A_{m,n}(\omega)|$ lower-bound: Let $A^{[k]}(\omega) := \omega_k$, hence $|A_{m,n}(\omega)| = \sum_{k=m+1}^n A^{[k]}(\omega)$. Fix any $\alpha \in (0, 1/2)$, and let

$$\mathcal{A}_{m,n}(\alpha) := \left\{ \omega \in \Omega \mid \sum_{k=m+1}^n A^{[k]}(\omega) \geq \left(\frac{1}{2} - \alpha \right) (n - m) \right\}.$$

Hence, for any $\omega \in \mathcal{A}_{m,n}(\alpha)$,

$$|A_{m,n}(\omega)| \geq (1/2 - \alpha)(n - m). \quad (3.56)$$

Moreover, by Lemma 92, $P(\mathcal{A}_{m,n}(\alpha)) \geq 1 - e^{-2\alpha^2(n-m)}$.

$|B_{m,n}(\omega)|$ *lower-bound*: First, note that d_k is defined depending on ω_{k+1} value. Hence, we may write

$$\begin{aligned} B_{m,n}(\omega) &= \{k \in \{m+1, \dots, n\} \mid \omega_k = 1 \text{ and } d_k \in \{2, 3\}\} \\ &\supseteq \{k \in \{m+1, \dots, n-1\} \mid \omega_k = 1, \omega_{k+1} = 1, \text{ and } d_k \in \{2, 3\}\} \\ &= \{k \in \{m+1, \dots, n-1\} \mid \omega_k = 1, \omega_{k+1} = 1, \text{ and } \delta_1(\Gamma_{\omega_1 \dots \omega_k}) \in \{2, 3\}\}. \end{aligned}$$

Let $B^{[k]}$ be the Bernoulli random variable on Ω , defined by

$$B^{[k]}(\omega) := \mathbf{1}_{\{\omega_{k+1}=1\}} \mathbf{1}_{\{\omega_k=1\}} \mathbf{1}_{\{\delta_1(\Gamma_{\omega_1 \dots \omega_k}) \in \{2, 3\}\}}.$$

The expected value of $B^{[k]}$ is given by

$$\begin{aligned} \mathbb{E}B^{[k]} &= \frac{1}{2^{k+1}} \sum_{i_1 \dots i_k i_{k+1}} \mathbf{1}_{\{i_{k+1}=1\}} \mathbf{1}_{\{i_k=1\}} \mathbf{1}_{\{\delta_1(\Gamma_{i_1 \dots i_k}) \in \{2, 3\}\}} \\ &= \frac{1}{2^{k+1}} \sum_{i_1 \dots i_{k-1}} \mathbf{1}_{\{\delta_1(\Gamma_{i_1 \dots i_{k-1} 1}) \in \{2, 3\}\}} \\ &= \frac{1}{2^{k+1}} \sum_{i_1 \dots i_{k-1}} \Delta^{i_1 \dots i_{k-1} 1}(\Gamma). \end{aligned}$$

Since $\Gamma \in \bar{\Pi}_{m(\gamma, \epsilon)}(\gamma)$, for $k > m \geq m(\gamma, \epsilon)$, we get

$$\mathbb{E}B^{[k]} \geq \gamma_0 := \frac{1}{4} \left(\frac{2}{3} - \gamma \right).$$

Let $\mathcal{K}(m, n) := \{k \in m+1, \dots, n-1 \mid k = m+1 \pmod{2}\}$, the set of integers $m+1, m+3, \dots$ comprised between $m+1$ and $n-1$. Random variables $B^{[k]}$, $k \in \mathcal{K}(m, n)$, are independent, and the expected value of their mean, denoted $\mathbb{E}B_{\mathcal{K}(m, n)} := \frac{1}{|\mathcal{K}(m, n)|} \mathbb{E}B^{[k]}$, satisfies $\mathbb{E}B_{\mathcal{K}(m, n)} \geq \gamma_0$. Fix any $\beta \in (0, \gamma_0)$, and let

$$\mathcal{B}_{m,n}(\beta) := \left\{ \omega \in \Omega \mid \sum_{k \in \mathcal{K}(m, n)} B^{[k]}(\omega) \geq (\gamma_0 - \beta) |\mathcal{K}(m, n)| \right\}.$$

Hence, for $m \geq m(\gamma, \epsilon)$ and $\omega \in \mathcal{B}_{m,n}(\beta)$, we have⁴

$$|B_{m,n}(\omega)| \geq \sum_{k=m+1}^{n-1} B^{[k]}(\omega) \geq \sum_{k \in \mathcal{K}(m, n)} B^{[k]}(\omega) \geq (\gamma_0 - \beta) |\mathcal{K}(m, n)| \geq (\gamma_0 - \beta) \frac{n-m}{3}. \quad (3.57)$$

Moreover, by applying Lemma 92, we have

$$\begin{aligned} P(\mathcal{B}_{m,n}(\beta)) &\geq P\left(\sum_{k \in \mathcal{K}(m, n)} B^{[k]}(\omega) \geq (\mathbb{E}B_{\mathcal{K}(m, n)} - \beta) |\mathcal{K}(m, n)|\right) \\ &\geq 1 - e^{-2\beta^2 |\mathcal{K}(m, n)|} \\ &\geq 1 - e^{-2\beta^2 \frac{n-m}{3}}. \end{aligned}$$

⁴The last inequality could be tighten, but we only need a non-zero fraction of $n-m$.

We define $\mathcal{U}_{m,n}(\zeta, \alpha, \beta) := T_m(\zeta) \cap \mathcal{A}_{m,n}(\alpha) \cap \mathcal{B}_{m,n}(\beta)$. Using (3.55), (3.56), and (3.57), for $n > m \geq m(\gamma, \epsilon)$ and $\omega \in \mathcal{U}_{m,n}(\zeta, \alpha, \beta)$, we have

$$Z_d^{[n]}(\omega) \leq \kappa^{(\alpha+\frac{1}{2})(n-m)} \zeta^{\frac{\gamma_0-\beta}{3}(n-m)} \zeta = \left(\kappa^{\alpha+\frac{1}{2}} \zeta^{\frac{\gamma_0-\beta}{3}} \right)^{n-m} \zeta.$$

Note that α, β , and γ (thus, γ_0) are some fixed constants. Hence, for any $\theta > 0$ (as in the fast polarization property), we may choose $\zeta > 0$, such that $\kappa^{\alpha+\frac{1}{2}} \zeta^{\frac{\gamma_0-\beta}{3}} \leq 2^{-(1+\theta)}$. Using $Z^{[n]}(\omega) \leq \max_{d=1,2,3} Z_d^{[n]}(\omega)$, we get the following inequality, that holds for any $n > m \geq m(\gamma, \epsilon)$ and any $\omega \in \mathcal{U}_{m,n}(\zeta, \alpha, \beta)$:

$$Z^{[n]}(\omega) \leq c 2^{-n(1+\theta)} = c N^{-(1+\theta)}.$$

where $c = c(m, \alpha, \beta, \gamma, \zeta) := \left(\kappa^{\alpha+\frac{1}{2}} \zeta^{\frac{\gamma_0-\beta}{3}} \right)^{-m} \zeta$, and $N = 2^n$. Note that α, β, γ , and ζ have been fixed at this point, and only the value of m can still be varied.

To complete the proof, we need to show that $\mathcal{U}_{m,n}(\zeta, \alpha, \beta)$ is sufficiently large (for some m , and large enough $n > m$), so that we may find information sets \mathcal{I}_N of size $|\mathcal{I}_N| \geq RN$, for $R < \mathbf{I}(W)$. For this, we need the following lemma, which is essentially the same as Lemma 1 in [2], and the proof follows using exactly the same arguments as in *loc. cit.* (and also using the fact that Γ is a polarizing sequence).

Lemma 94. *For any fixed $\zeta > 0$ and any $0 \leq \delta < \mathbf{I}(W)$, there exists an integer $m_0(\zeta, \delta)$, such that*

$$P(T_{m_0}(\zeta)) \geq \mathbf{I}(W) - \delta.$$

Therefore, $P(T_m(\zeta))$ can be made arbitrarily close to $\mathbf{I}(W)$, by taking m large enough, and once we have made $P(T_m(\zeta))$ as close as desired to $\mathbf{I}(W)$, we can make $P(\mathcal{A}_{m,n}(\alpha))$ and $P(\mathcal{B}_{m,n}(\beta))$ arbitrarily close to 1, by taking $n > m$ large enough. Hence, for any $R < \mathbf{I}(W)$, we may find $m_0 = m_0(\zeta, R)$ and $n_0 = n_0(m_0, \alpha, \beta, \gamma) > m_0$, such that

$$P(\mathcal{U}_{m_0,n}(\zeta, \alpha, \beta)) > R, \quad \forall n \geq n_0,$$

and since we may assume that $m_0 \geq m(\gamma, \epsilon)$, we also have

$$Z^{[n]}(\omega) \leq c_0 N^{-(1+\theta)}, \quad \forall n \geq n_0, \quad \forall \omega \in \mathcal{U}_{m_0,n}(\zeta, \alpha, \beta) \quad (3.58)$$

where $c_0 := c_0(m_0, \alpha, \beta, \gamma, \zeta)$.

Now, for $n > 0$, let $\mathcal{V}_n := \{\omega \in \Omega \mid Z^{[n]}(\omega) \leq c_0 N^{-(1+\theta)}\}$. Using (3.58), we have that $\mathcal{U}_{m_0,n}(\zeta, \alpha, \beta) \subseteq \mathcal{V}_n$, for any $n \geq n_0$, and therefore $P[\mathcal{V}_n] \geq R$. On the other hand,

$$\begin{aligned} P[\mathcal{V}_n] &= \sum_{i_1 \dots i_n \in \{0,1\}^n} \frac{1}{2^n} \mathbf{1} \left\{ Z(W^{(i_1 \dots i_n)}) \leq c_0 N^{-(1+\theta)} \right\} \\ &= \frac{1}{N} |\mathcal{I}_N|, \end{aligned}$$

where $\mathcal{I}_N := \{i \in \{0, \dots, N-1\} \mid Z(W^{(i)}) \leq c_0 N^{-(1+\theta)}\}$. It follows that $|\mathcal{I}_N| \geq RN$, for $n \geq n_0$.

We have shown that, given $\epsilon > 0$, the fast polarization property holds for any $\Gamma \in \Gamma(\mathcal{S})_{\text{pol}}^\infty \cap \bar{\Pi}_{m(\gamma, \epsilon)}(\gamma)$, with $P(\bar{\Pi}_{m(\gamma, \epsilon)}(\gamma)) \geq 1 - \epsilon$. We then conclude that it holds for any $\Gamma \in \Gamma(\mathcal{S})_{\text{pol}}^\infty \cap (\bigcup_{\epsilon > 0} \bar{\Pi}_{m(\gamma, \epsilon)}(\gamma))$, which is a measurable subset of $\Gamma(\mathcal{S})_{\text{pol}}^\infty$, of same probability. \square

3.5 Decoding the Quantum Polar Code Using its Classical Counterpart

Let \mathcal{W} be a CMP channel and $\mathcal{W}^\#$ its classical counterpart. Let Q_N denote the unitary operator corresponding to the quantum polar code (defined by the recursive application of n polarization steps, see Section 2.5.1), and P_N denote the linear transformation corresponding to the classical polar code. Let \mathcal{I} and \mathcal{J} be the set of indices corresponding to the good and bad channels, respectively, with $|\mathcal{I}| + |\mathcal{J}| = N := 2^n$. We shall use the following notation from Section 2.5.1:

- $\rho_{\mathcal{I}}$ denotes the original state of system \mathcal{I} ,
- $\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'} := (Q_N \otimes I_{\mathcal{J}'})(\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'})(Q_N^\dagger \otimes I_{\mathcal{J}'})$ denotes the *encoded state*, where $\Phi_{\mathcal{J}\mathcal{J}'}$ is a maximally entangled state, as defined in (2.45).
- $\psi_{\mathcal{I}\mathcal{J}\mathcal{J}'} := (\mathcal{W}^{\otimes N} \otimes I_{\mathcal{J}'})(\varphi_{\mathcal{I}\mathcal{J}\mathcal{J}'})$ denotes the *channel output state*.

Since \mathcal{W} is a CMP channel, it follows that:

$$\psi_{\mathcal{I}\mathcal{J}\mathcal{J}'} = (E_{\mathcal{I}\mathcal{J}} Q_N \otimes I_{\mathcal{J}'})(\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'})(Q_N^\dagger E_{\mathcal{I}\mathcal{J}}^\dagger \otimes I_{\mathcal{J}'}).$$

for a random N -qubit Pauli error $E_{\mathcal{I}\mathcal{J}} \in \bar{\mathcal{G}}_N$. Hence, quantum polar code decoding can be performed in the 4 steps described below.

Step 1: Apply the inverse quantum polar transform on the channel output state. Applying Q_N^\dagger on the output state $\psi_{\mathcal{I}\mathcal{J}\mathcal{J}'}$, leaves the $\mathcal{I}\mathcal{J}\mathcal{J}'$ system in the following state:

$$\begin{aligned} \psi'_{\mathcal{I}\mathcal{J}\mathcal{J}'} &= (Q_N^\dagger E_{\mathcal{I}\mathcal{J}} Q_N \otimes I_{\mathcal{J}'})(\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'})(Q_N^\dagger E_{\mathcal{I}\mathcal{J}}^\dagger Q_N \otimes I_{\mathcal{J}'}) \\ &= (E'_{\mathcal{I}\mathcal{J}} \otimes I_{\mathcal{J}'})(\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'})(E_{\mathcal{I}\mathcal{J}}'^\dagger \otimes I_{\mathcal{J}'}). \end{aligned}$$

where $E'_{\mathcal{I}\mathcal{J}} := Q_N^\dagger E_{\mathcal{I}\mathcal{J}} Q_N$. As Q_N is a N qubit Clifford, we have that $E'_{\mathcal{I}\mathcal{J}} \in \bar{\mathcal{G}}_N \simeq \bar{\mathcal{G}}_1^N$, and thus write $E'_{\mathcal{I}\mathcal{J}} = P_N^{-1} E_{\mathcal{I}\mathcal{J}}$, or equivalently:

$$E_{\mathcal{I}\mathcal{J}} = P_N E'_{\mathcal{I}\mathcal{J}}.$$

Put differently, $E_{\mathcal{I}\mathcal{J}}$ is the classical polar encoded version of $E'_{\mathcal{I}\mathcal{J}}$.

Step 2: Quantum measurement.⁵ Let $E'_{\mathcal{I}\mathcal{J}} = \bigotimes_{i \in \mathcal{I}} E'_i \otimes \bigotimes_{j \in \mathcal{J}} E'_j$, with $E'_i, E'_j \in \bar{\mathcal{G}}_1$. Measuring $X_j X_{j'}$ and $Z_j Z_{j'}$ observables⁶, allows determining the value of E'_j , for any $j \in \mathcal{J}$, since no errors occurred on the \mathcal{J}' system.

Step 3: Decode the classical polar code counterpart. We note that the error $E_{\mathcal{I}\mathcal{J}}$ can be seen as the output of the classical vector channel $(\mathcal{W}^\#)^N$, when the “all-identity vector” $\sigma_0^N \in \bar{\mathcal{G}}_1^N$ is applied at the channel input. However, by the definition of the classical channel $\mathcal{W}^\#$, we have $(\mathcal{W}^\#)^N(E_{\mathcal{I}\mathcal{J}} \mid \sigma_0^N) = (\mathcal{W}^\#)^N(\sigma_0^N \mid E_{\mathcal{I}\mathcal{J}})$, meaning that we can equivalently consider σ_0^N as being the observed channel output, and $E_{\mathcal{I}\mathcal{J}}$ the (unknown) channel input. Hence, we have given (i) the value of $E'_{\mathcal{J}} := \bigotimes_{j \in \mathcal{J}} E'_j$, and (ii) a noisy observation (namely σ_0^N) of $E_{\mathcal{I}\mathcal{J}} = P_N E'_{\mathcal{I}\mathcal{J}}$. We can then use classical polar code decoding to recover the value of $E'_{\mathcal{I}} := \bigotimes_{i \in \mathcal{I}} E'_i$.

Step 4: Error correction. Once we have recovered the $E'_{\mathcal{J}}$ (step 2) and $E'_{\mathcal{I}}$ (step 3) values, we can apply the $E'_{\mathcal{I}\mathcal{J}} \otimes I_{\mathcal{J}'}$ operator on $\psi'_{\mathcal{I}\mathcal{J}\mathcal{J}'}$, thus leaving the $\mathcal{I}\mathcal{J}\mathcal{J}'$ system in the state $\rho_{\mathcal{I}} \otimes \Phi_{\mathcal{J}\mathcal{J}'}$.

⁵Steps (1) and (2) together perform a set of measurements that are equivalent to measuring the elements of the stabilizer set $\bar{\mathcal{S}}_{\mathcal{I}\mathcal{J}\mathcal{J}'}$ defined in (2.49).

⁶Here, indices j and j' indicate the j -th qubits of \mathcal{J} and \mathcal{J}' systems.

3.6 Purely Quantum vs. CSS-based Polarization

In this section, using a numerical simulation, we compare the speed of polarizations of purely quantum construction presented in this chapter and the quantum CSS construction from Section 1.5.1, for a finite length code. We consider a quantum erasure channel (Definition 47) \mathcal{W} , with erasure probability $\epsilon = 0.25$. The symmetric coherent information of \mathcal{W} is given by,

$$I(\mathcal{W}) = 1 - 2\epsilon = 0.5. \quad (3.59)$$

In Figure 3.12 below, we plot the Bhattacharyya parameter of the synthesized virtual channels, for $n = 16$ polarization steps, for both the purely quantum and the CSS-based polarization. Almost noiseless virtual channels are those with Bhattacharyya parameter approaching 0, while almost noisy virtual channels are those with Bhattacharyya parameter approaching 1.

- For the purely quantum polarization (solid black curve), the Bhattacharyya parameter is given by Definition 80. The fraction of good channels is almost equal to $(1 + I(\mathcal{W})) / 2 = 0.75$ and the fraction of preshared EPR pairs is almost equal to $(1 - I(\mathcal{W})) / 2 = 0.25$.
- For the CSS-based polarization (solid red curve), the Bhattacharyya parameter is the sum of the two classical Bhattacharyya parameters obtained by polarizing induced amplitude and phase channels W_A and W_P . The fraction of good channels approaches $I(\mathcal{W}) = 0.5$.

As expected, our scheme achieves a higher communication rate than the CSS-based one, since it uses entanglement. To compare the two schemes in terms of net rates, that is, the rate minus the fraction of preshared EPR pairs, we shift the purely quantum polarization curve to the left by $(1 - I(\mathcal{W})) / 2 = 0.25$, thus obtaining the dashed black curve in Figure 3.12, which is virtually superimposed on the CSS-based polarization curve. We conclude that our construction and the CSS construction exhibit the same polarization speed, which is given by the slope of the two curves.

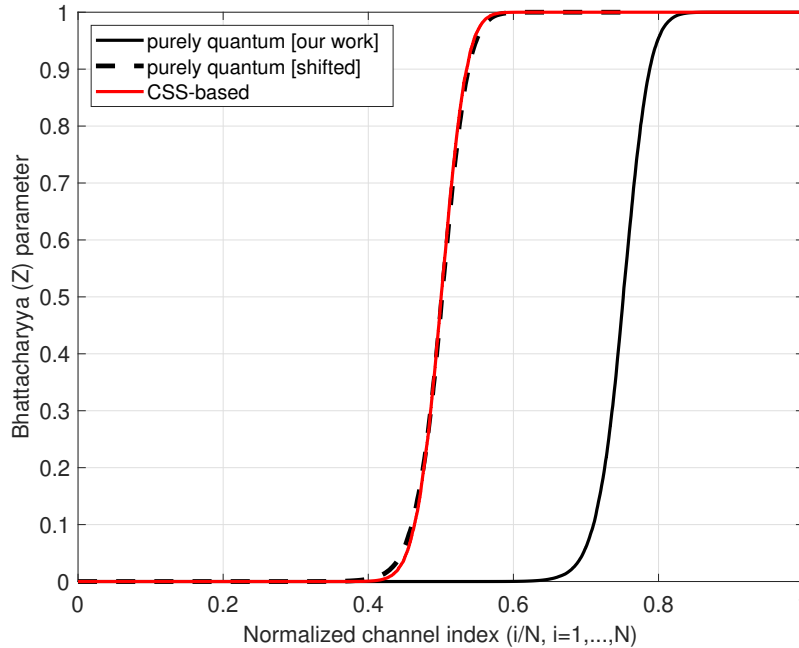


Figure 3.12: Bhattacharyya parameter of the synthesized virtual channels ($n = 16$)

4

Multilevel Polarization of Pauli Channels

Recall that in Chapter 3, for Pauli channels, we first show polarization, using the set of 9 two-qubit Cliffords \mathcal{L} from Figure 2.4 (see Section 3.2). We then show polarization, using a subset $\mathcal{S} \subset \mathcal{L}$ containing three two-qubit Cliffords (see Section 3.3.1). For the proofs, we rely on guaranteed improvement of the good channel after one polarization step. Further, using a computer simulation, we look for guaranteed improvement or degradation after two polarization steps in order to completely derandomize the channel combining operation. Moreover, we note that for $L_{3,3} \in \mathcal{L}$, we almost have the guaranteed improvement for a virtual channel after two polarization steps (see Remark 86).

In this chapter, we consider the polarization construction, with the gate $L_{3,3} \in \mathcal{L}$, for Pauli channels. We show a different polarization phenomenon, where polarization happens in multi-levels instead of two levels in the sense of [67, 68]. More precisely, the synthesized virtual channels also tend to be “half-noisy” except being completely noisy or noiseless. The half-noisy channels need to be frozen by fixing their inputs in either the amplitude or the phase basis, while preshared EPR pairs are required for the completely noisy channels as before.

As some of the bad channels are frozen in either the amplitude or the phase basis, this reduces the number of preshared EPR pairs compared to the construction in Chapter 3. We also give an upper bound on the number of preshared EPR pairs, which is an equality for the quantum erasure channel. In particular, for a quantum erasure channel with erasure probability ϵ , the fraction of preshared EPR pairs is ϵ^2 , while it is ϵ for the construction proposed in Chapter 3 (see Section 3.6). Therefore, the number of preshared EPR pairs is significantly reduced, taking advantage of the multilevel nature of polarization. We construct a quantum polar code based on the multilevel polarization for which the decoding can also be efficiently performed by decoding the classical counterpart.

Finally, we present a slightly different construction utilizing a quantum circuit equivalence, the goal of which is to improve the speed of multilevel polarization. For a quantum erasure channel, we show with the help of a computer program that the speed of multilevel polarization improves significantly for this alternative construction compared to the first construction.

4.1 Noiseless, Half-noisy and Noisy channels

Recall from Chapter 3 that quantum polarization on a Pauli channel happens if and only if the classical polarization on its classical counterpart happens (see Proposition 78 and Corollary 79). This allowed us to prove the quantum polarization by showing polarization for the classical counterpart. In what follows, we prove multilevel polarization for the CMP channel, using the polar code construction on its classical counterpart channel. In this section, using the Bhattacharyya parameter from Definition 80, we provide conditions for the classical counterpart channel to be completely noiseless, half-noisy or completely noisy. We shall use the following notation from Chapter 3.

1. Let $\bar{\mathcal{G}}_N := \mathcal{G}_N / \{\pm 1, \pm i\}$ be the Abelian group obtained by taking the quotient of the N qubit Pauli group \mathcal{G}_N by its centralizer.
2. For the classical counterpart of a CMP channel \mathcal{W} , we shall use $W := \mathcal{W}^\#$. Recall that W is a classical channel with the input alphabet $\bar{\mathcal{G}}_1 := \mathcal{G}_1 / \{\pm 1, \pm i\}$.
3. We shall identify $I \equiv 00$, $Z \equiv 01$, $X \equiv 10$, and $Y \equiv 11$. Using this identification, we may write $\bar{\mathcal{G}}_1 = \{00, 01, 10, 11\}$, or sometimes $\bar{\mathcal{G}}_1 = \{0, 1, 2, 3\}$, which will be clear from the context.

For any $x, x', d \in \bar{\mathcal{G}}_1$, we first define two information measures $I(W_{x,x'})$ and $I_d(W)$ as follows

$$I(W_{x,x'}) := \sum_y \frac{1}{2} \left[W(y|x) \log_2 \frac{W(y|x)}{\frac{1}{2}[W(y|x) + W(y|x')]} + W(y|x') \log_2 \frac{W(y|x')}{\frac{1}{2}[W(y|x) + W(y|x')]} \right]. \quad (4.1)$$

$$I_d(W) := \frac{1}{4} \sum_x I(W_{x,x \oplus d}). \quad (4.2)$$

Note that $I(W_{x,x'})$ is the symmetric mutual information of the binary-input channel (Definition 2) obtained by restricting the input alphabet of W to $\{x, x'\} \subseteq \bar{\mathcal{G}}_1$.

Further note that the parameter $Z_d(W)$ from Definition 80 can be written as follows, for $x, x', x'', d \in \bar{\mathcal{G}}_1$,

$$Z(W_{x,x'}) := \sum_y \sqrt{W(y|x)W(y|x')}. \quad (4.3)$$

$$Z_d(W) := \frac{1}{4} \sum_{x \in \bar{\mathcal{G}}_1} Z(W_{x,x \oplus d}), \quad (4.4)$$

$$= \frac{1}{2} [Z(W_{0,d}) + Z(W_{x'',x'' \oplus d})], \text{ for any } x'' \neq 0, d, \quad (4.5)$$

where (4.5) follows from $Z(W_{x,x'}) = Z(W_{x',x})$. Also recall from Definition 80 that the Bhattacharyya parameter of W is given by,

$$Z(W) := \frac{1}{3} \sum_{d \in \bar{\mathcal{G}}_1: d \neq 0} Z_d(W). \quad (4.6)$$

We now proceed with several lemmas. First, we give the following lemma for the symmetric mutual information $I(W)$ and the Bhattacharyya parameter $Z(W)$ for the non-binary input channel W , analogous to Lemma 4.

Lemma 95 ([24]). *For the classical counterpart channel W , we have*

$$I(W) \geq \log_2 \frac{4}{1 + 3Z(W)}. \quad (4.7)$$

$$I(W) \leq 6(\log_2 e) \sqrt{1 - Z(W)^2}. \quad (4.8)$$

The first inequality in the above Lemma implies that $I(W)$ approaches to 2 if $Z(W)$ approaches to 0, and the second inequality implies that $I(W)$ approaches to 0 if $Z(W)$ approaches to 1.

In the lemma below, we show that if any two parameters from the set $\{Z_1(W), Z_2(W), Z_3(W)\}$, defined in (4.5), approach 1, the remaining third parameter will also approach 1.

Lemma 96. *For any $\{d_1, d_2, d_3\} = \{1, 2, 3\}$, if $Z_{d_1}(W) \geq 1 - \epsilon_1$, and $Z_{d_2}(W) \geq 1 - \epsilon_2$, then,*

$$Z_{d_3}(W) \geq 1 - \epsilon_3, \text{ where } \epsilon_3 = 4(\sqrt{\epsilon_1} + \sqrt{\epsilon_2})^2. \quad (4.9)$$

Proof. For $x \in \bar{\mathcal{G}}_1$, consider a vector $\vec{A}(x)$ such that $\vec{A}(x) = (\sqrt{W(y|x)}, y \in Y)$. It follows that $\vec{A}(x) \cdot \vec{A}(x') = Z(W_{x,x'})$ and $|\vec{A}(x) - \vec{A}(x')| = \sqrt{2(1 - Z(W_{x,x'}))}$, where $|\vec{A}(x) - \vec{A}(x')|$ is the Euclidean distance between the vectors $\vec{A}(x)$ and $\vec{A}(x')$. Using the triangle inequality and $d_1 \oplus d_2 = d_3$, we have that

$$\sqrt{(1 - Z(W_{x,x \oplus d_3}))} \leq \sqrt{(1 - Z(W_{x,x \oplus d_1}))} + \sqrt{(1 - Z(W_{x \oplus d_1, x \oplus d_1 \oplus d_2}))}. \quad (4.10)$$

For $d \in \{d_1, d_2\}$, we have that $Z_d(W) \geq 1 - \epsilon \implies (1 - Z(W_{x,x \oplus d})) \leq 4\epsilon, \forall x$. Then, from (4.10),

$$\begin{aligned} \sqrt{(1 - Z(W_{x,x \oplus d_3}))} &\leq 2(\sqrt{\epsilon_1} + \sqrt{\epsilon_2}), \forall x \\ \implies Z_{d_3}(W) &\geq 1 - 4(\sqrt{\epsilon_1} + \sqrt{\epsilon_2})^2. \end{aligned} \quad \square$$

We now introduce the partial channels of the non-binary input channel W .

Definition 97. (Partial channels). *Consider $x = x_1 x_2 \in \bar{\mathcal{G}}_1 = \{00, 01, 10, 11\}$ is given as the channel input of W . We define the following three binary-input channels that are obtained by randomizing one bit of information from x ,*

$$W^{[1]} : x_1 \rightarrow y; W^{[1]}(y|0) = \frac{W(y|00) + W(y|01)}{2}, W^{[1]}(y|1) = \frac{W(y|10) + W(y|11)}{2}. \quad (4.11)$$

$$W^{[2]} : x_2 \rightarrow y; W^{[2]}(y|0) = \frac{W(y|00) + W(y|10)}{2}, W^{[2]}(y|1) = \frac{W(y|01) + W(y|11)}{2}. \quad (4.12)$$

$$W^{[3]} : x_1 \oplus x_2 \rightarrow y; W^{[3]}(y|0) = \frac{W(y|00) + W(y|11)}{2}, W^{[3]}(y|1) = \frac{W(y|01) + W(y|10)}{2}. \quad (4.13)$$

In particular, the partial channel $W^{[1]}$ takes x_1 as input and randomizes x_2 , the partial channel $W^{[2]}$ takes x_2 as input and randomizes x_1 , and the partial channel $W^{[3]}$ takes $x_1 \oplus x_2$ as input and randomizes both x_1 and x_2 , individually. For $\{d_1, d_2, d_3\} = \{1, 2, 3\}$, the above three definitions can be merged into the following

$$W^{[d_1]}(y|0) = \frac{W(y|0) + W(y|d_1)}{2}, \text{ and } W^{[d_1]}(y|1) = \frac{W(y|d_2) + W(y|d_3)}{2}. \quad (4.14)$$

We now prove several bounds relating $Z_d(W)$, $Z(W^{[d]})$ and $I(W)$.

Lemma 98. Given $\{d_1, d_2, d_3\} = \{1, 2, 3\}$, we have the following inequalities, which bear similarities to Lemmas 9 and 10 from [67]:

- (a) $Z(W^{[d_1]}) \leq Z_{d_2}(W) + Z_{d_3}(W)$.
- (b) $Z(W^{[d_1]}) \geq Z_{d_i}(W)$, where $Z_{d_i}(W) = \max(Z_{d_2}(W), Z_{d_3}(W))$.
- (c) $I(W) \leq \frac{1}{3} \sum_{d \in \{1,2,3\}} \sqrt{1 - Z_d(W)^2} + \frac{1}{3} \sum_{d \in \{1,2,3\}} \sqrt{1 - Z(W^{[d]})^2}$.

Proof. **Point (a):** The Bhattacharyya parameter of the partial channel $W^{[d_1]}$ is given by,

$$\begin{aligned} Z(W^{[d_1]}) &= \sum_y \sqrt{W^{[d_1]}(y|0)W^{[d_1]}(y|1)} \\ &= \frac{1}{2} \sum_y \sqrt{\sum_{l \in \{0, d_1\}} \sum_{m \in \{d_2, d_3\}} W(y|l)W(y|m)} \\ &\leq \frac{1}{2} \sum_{l \in \{0, d_1\}} \sum_{m \in \{d_2, d_3\}} \sum_y \sqrt{W(y|l)W(y|m)} \\ &= Z_{d_2}(W) + Z_{d_3}(W), \end{aligned}$$

where the second equality follows from (4.14), the third inequality follows from $\sqrt{\sum_x a_x} \leq \sum_x \sqrt{a_x}$, and the fourth equality follows from $l \oplus m \in \{d_2, d_3\}, \forall l, m$ as $d_3 = d_1 \oplus d_2$ and (4.5).

Point (b): For $W^{[d_1]}$, we consider the following two-dimensional vectors:

$$\begin{aligned} \vec{B}_0(y) &= (\sqrt{W(y|0)}, \sqrt{W(y|d_1)}). \\ \vec{B}_1(y) &= (\sqrt{W(y|d_2)}, \sqrt{W(y|d_1 \oplus d_2)}). \\ \vec{B}_2(y) &= (\sqrt{W(y|d_1 \oplus d_2)}, \sqrt{W(y|d_2)}). \end{aligned}$$

Then, we have that

$$\begin{aligned} |\vec{B}_0(y)| &= \sqrt{W(y|0) + W(y|d_1)}. \\ |\vec{B}_1(y)| &= |\vec{B}_2(y)| = \sqrt{W(y|d_2) + W(y|d_1 \oplus d_2)}. \\ \vec{B}_0(y) \cdot \vec{B}_1(y) &= \sqrt{W(y|0)}\sqrt{W(y|d_2)} + \sqrt{W(y|d_1)}\sqrt{W(y|d_1 \oplus d_2)}. \\ \vec{B}_0(y) \cdot \vec{B}_2(y) &= \sqrt{W(y|0)}\sqrt{W(y|d_1 \oplus d_2)} + \sqrt{W(y|d_1)}\sqrt{W(y|d_2)}. \end{aligned}$$

From the definitions of $Z(W^{[i]})$ and $Z_d(W)$, it follows:

$$Z(W^{[d_1]}) = \frac{1}{2} \sum_y |\vec{B}_0(y)| |\vec{B}_1(y)| = \frac{1}{2} \sum_y |\vec{B}_0(y)| |\vec{B}_2(y)|. \quad (4.15)$$

$$Z_{d_2}(W) = \frac{1}{2} \sum_y \vec{B}_0(y) \cdot \vec{B}_1(y). \quad (4.16)$$

$$Z_{d_3}(W) = Z_{d_1 \oplus d_2}(W) = \frac{1}{2} \sum_y \vec{B}_0(y) \cdot \vec{B}_2(y). \quad (4.17)$$

Then, from the Cauchy-Schwartz inequality, we have that

$$Z_d(W) \leq Z(W^{[d_1]}), \text{ for } d = d_2, d_3. \quad (4.18)$$

Point (c): $I(W)$ can be written as following [67, Lemma 10],

$$\begin{aligned}
 I(W) &= \frac{1}{4} \sum_y \sum_{x \in \mathcal{G}_1} W(y|x) \log_2 \frac{W(y|x)}{P(y)} \\
 &= \frac{1}{4} \sum_y \frac{1}{6} \sum_{d \in \{d_1, d_2, d_3\}} \sum_x [W(y|x) \log_2 \frac{W(y|x)}{P(y)} + W(y|x \oplus d) \log_2 \frac{W(y|x \oplus d)}{P(y)}] \\
 &= \frac{1}{24} \sum_{d \in \{d_1, d_2, d_3\}} \sum_x W(y|x) \log_2 \frac{W(y|x)}{\frac{1}{2}[W(y|x) + W(y|x \oplus d)]} + W(y|x \oplus d) \log_2 \frac{W(y|x \oplus d)}{\frac{1}{2}[W(y|x) + W(y|x \oplus d)]} \\
 &\quad + \frac{1}{12} \sum_y \sum_{d \in \{d_1, d_2, d_3\}} \sum_x \frac{W(y|x) + W(y|x \oplus d)}{2} \log_2 \frac{\frac{1}{2}[W(y|x) + W(y|x \oplus d)]}{P(y)} \\
 &= \frac{1}{12} \sum_{d \in \{d_1, d_2, d_3\}} \sum_x I(W_{x, x \oplus d}) + \frac{1}{6} \sum_y \sum_{d \in \{d_1, d_2, d_3\}} \frac{W(y|0) + W(y|d)}{2} \log_2 \frac{\frac{1}{2}[W(y|0) + W(y|d)]}{P(y)} \\
 &\quad + \frac{1}{12} \sum_y \sum_{d \in \{d_1, d_2, d_3\}} \sum_{x \neq 0, d} \frac{W(y|x) + W(y|x \oplus d)}{2} \log_2 \frac{\frac{1}{2}[W(y|x) + W(y|x \oplus d)]}{P(y)} \\
 &= \frac{1}{3} \sum_{d \neq 0} I_d(W) + \frac{1}{6} \sum_{d \in \{d_1, d_2, d_3\}} \sum_y [W^{[d]}(y|0) \log_2 \frac{W^{[d]}(y|0)}{P(y)} + W^{[d]}(y|1) \log_2 \frac{W^{[d]}(y|1)}{P(y)}] \\
 &= \frac{1}{3} \sum_{d \neq 0} I_d(W) + \frac{1}{3} \sum_{d \in \{1, 2, 3\}} I(W^{[d]}),
 \end{aligned}$$

where the first equality follows from Definition 74, and parameters $I(W_{x, x'})$, $I_d(W)$ are defined in (4.1) and (4.2), respectively. Also, $I(W^{[d]})$ is the symmetric mutual information of the binary-input partial channel $W^{[d]}$. Using $I(W_{x, x'}) \leq \sqrt{1 - Z(W_{x, x'})^2}$ from [2], and concavity of the function $f(x) = \sqrt{1 - x^2}$, we have that

$$I(W) \leq \frac{1}{3} \sum_{d \in \{1, 2, 3\}} \sqrt{1 - Z_d(W)^2} + \frac{1}{3} \sum_{i \in \{1, 2, 3\}} \sqrt{1 - Z(W^{[i]})^2}. \quad (4.19)$$

□

Finally, we prove the following lemma that will be used to define half-noisy channels.

Lemma 99. *Given $Z_{d_1}(W) \leq \epsilon$, $Z_{d_2}(W) \leq \epsilon$, and $Z_{d_3}(W) \geq 1 - \epsilon$, with $\epsilon > 0$, and $\{d_1, d_2, d_3\} = \{1, 2, 3\}$, then*

$$(a) \quad I(W^{[d_3]}) \in [1 - \log_2(1 + 2\epsilon), 1].$$

$$(b) \quad |I(W) - I(W^{[d_3]})| \leq \Delta, \text{ where } \Delta = \sqrt{2\epsilon} + \log_2(1 + 2\epsilon).$$

Proof. **Point (a):** Since $W^{[d_3]}$ is a binary-input channel, $I(W^{[d_3]}) \leq 1$. From point (a) of Lemma 98, we have that

$$0 \leq Z(W^{[d_3]}) \leq Z_{d_1}(W) + Z_{d_2}(W) \leq 2\epsilon. \quad (4.20)$$

Using the inequality $I(W_b) \geq 1 - \log_2(1 + Z(W_b))$ for any binary-input channel W_b from [2], we can lower bound $I(W^{[d_3]})$ as follows

$$I(W^{[d_3]}) \geq 1 - \log_2(1 + 2\epsilon). \quad (4.21)$$

Hence,

$$I(W^{[d_3]}) \in [1 - \log_2(1 + 2\epsilon), 1]. \quad (4.22)$$

Point (b): From Point (b) of Lemma 98, we have that

$$Z(W^{[d_i]}) \geq Z_{d_3}(W) \geq 1 - \epsilon, \forall d_i = d_1, d_2. \quad (4.23)$$

In point 3 of Lemma 98, substituting the lower bound on $Z_d(W)$, i.e., $Z_{d_1}(W) = Z_{d_2}(W) = 0$, $Z_{d_3}(W) = 1 - \epsilon$, and the lower bound on $Z(W^{[d]})$ from (4.20) and (4.23), i.e., $Z(W^{[d_1]}) = Z(W^{[d_2]}) = 1 - \epsilon$, $Z(W^{[d_3]}) = 0$, we have the following upper bound on $I(W)$,

$$I(W) \leq 1 + \sqrt{2\epsilon}. \quad (4.24)$$

From inequality in (4.7), $I(W)$ can also be lower bounded, as below

$$\begin{aligned} I(W) &\geq \log_2 \frac{4}{2 + \epsilon} \\ &= 1 - \log_2(1 + \frac{\epsilon}{2}). \end{aligned} \quad (4.25)$$

From (4.24) and (4.25), we have that

$$I(W) \in [1 - \log_2(1 + \frac{\epsilon}{2}), 1 + \sqrt{2\epsilon}]. \quad (4.26)$$

From (4.22) and (4.26), we have that

$$|I(W) - I(W^{[d_3]})| \leq \Delta, \quad (4.27)$$

where $\Delta = \max(\sqrt{2\epsilon} + \log_2(1 + 2\epsilon), \log_2(1 + \frac{\epsilon}{2})) = \sqrt{2\epsilon} + \log_2(1 + 2\epsilon)$. \square

We are now in a position to define the noiseless, half-noisy, and noisy channels.

Definition 100. Given $\delta > 0$, a channel W is said to be:

- (a) δ -noiseless if $Z_1(W) < \delta$, $Z_2(W) < \delta$, and $Z_3(W) < \delta$.
- (b) δ -noisy if $Z_1(W) > 1 - \delta$, and $Z_2(W) > 1 - \delta$.
- (c) δ -half-noisy of type d_3 , if $Z_{d_1}(W) < \delta$, $Z_{d_2}(W) < \delta$, and $Z_{d_3}(W) > 1 - \delta$, with $\{d_1, d_2, d_3\} = \{1, 2, 3\}$.

Recall that W takes as input two bits $x_1 x_2$, where x_1, x_2 , and $x_1 \oplus x_2$ are inputs to the partial channels $W^{[1]}$, $W^{[2]}$, and $W^{[3]}$, respectively.

If W is such that $Z_1(W) < \delta$, $Z_2(W) < \delta$, and $Z_3(W) < \delta$, using (4.7), we have that $I(W) \rightarrow 0$ as $\delta \rightarrow 0$. Therefore, we call W , δ -noiseless.

If W is such that $Z_1(W) > 1 - \delta$, $Z_2(W) > 1 - \delta$, using (4.8) and Lemma 96, we have that $I(W) \rightarrow 0$ as $\delta \rightarrow 0$. Therefore, we call W , δ -noisy.

If W is such that $Z_{d_1}(W) < \delta$, $Z_{d_2}(W) < \delta$, and $Z_{d_3}(W) > 1 - \delta$, with $\{d_1, d_2, d_3\} = \{1, 2, 3\}$ and $\delta \rightarrow 0$, from point (a) of Lemma 99, the binary-input partial channel $W^{[d_3]}$ tends to be noiseless, that is, $I(W^{[d_3]}) \rightarrow 1$. We may take $d_3 = 1$, without the loss of generality. Then, we can reliably transmit one bit of information, namely x_1 , the input to the partial channel $W^{[1]}$, using W . Moreover, from point (b) of Lemma 99, $I(W^{[d_3]}) \rightarrow I(W)$. Thus, the remaining one bit from the input of W , namely x_2 , the input to the partial channel $W^{[2]}$, is completely randomized or erased. Therefore, we call W , “ δ -half-noisy of type d_3 ”.

4.2 Multilevel Polarization

In this section, we show that the CMP channel polarizes into completely noiseless, half-noisy or completely noisy channels, under the recursive channel combining and splitting procedure, using the two qubit Clifford $L_{3,3} \in \mathcal{L}$ (see Figure 2.4) as the channel combining operation. We utilize the channel combining and splitting procedure on the classical counterpart to prove the polarization, where the permutation corresponding to $L_{3,3}$, that is, $\Gamma_{3,3}$ from Figure 3.2, is taken as the channel combining operation. Since the single qubit gate $L_{3,3}$ generates the same permutation on the set $\bar{\mathcal{G}}_1$ as the Hadamard gate H by the conjugate action, for our purposes the following two-qubit Clifford L is equivalent to $L_{3,3}$.

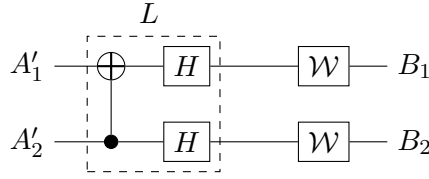


Figure 4.1: Two-qubit Clifford gate L . Here H is the Hadamard gate.

The permutation generated by the conjugate action of L on $\bar{\mathcal{G}}_1 \times \bar{\mathcal{G}}_1$ is depicted in the following figure, which is the same as $\Gamma_{3,3}$ ¹.

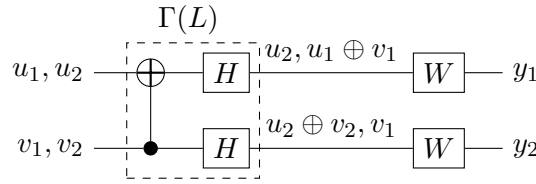


Figure 4.2: The permutation $\Gamma(L)$. To avoid any possible confusion, two bits of the input and output symbols are separated here by a comma.

From (3.24) and (3.25), the virtual channels obtained after the channel combining and splitting procedure on two copies of W , using $\Gamma(L)$ as channel combining operation, are given by

$$(W \boxtimes W)(y_1, y_2 | u_1, u_2) = \frac{1}{4} \sum_{v_1, v_2} W(y_1 | u_2, u_1 \oplus v_1) W(y_2 | u_2 \oplus v_2, v_1), \quad (4.28)$$

$$(W \otimes W)(y_1, y_2, u_1, u_2 | v_1, v_2) = \frac{1}{4} W(y_1 | u_2, u_1 \oplus v_1) W(y_2 | u_2 \oplus v_2, v_1), \quad (4.29)$$

where $u_1, u_2, v_1, v_2 \in \{0, 1\}$.

4.2.1 Several Inequalities for the Good and Bad Channels

Here, we shall provide many inequalities for the good and bad channel synthesized after one polarization step, which we shall use later for the proof of multilevel polarization.

¹Recall that $00 \equiv I, 01 \equiv Z, 10 \equiv X, 11 \equiv Y$.

Lemma 101. *The following equalities hold for the good channel $W \otimes W$,*

$$Z_1(W \otimes W) = Z_2(W). \quad (4.30)$$

$$Z_2(W \otimes W) = Z_1(W)^2. \quad (4.31)$$

Proof. **Proof of (4.30):** We have that

$$\begin{aligned} Z_1(W \otimes W) &= \frac{1}{4} \sum_{\substack{y_1, y_2, u_1, u_2 \\ v_1, v_2}} \sqrt{(W \otimes W)(y_1, y_2, u_1, u_2 | v_1, v_2)(W \otimes W)(y_1, y_2, u_1, u_2 | v_1 + 1, v_2 + 1)} \\ &= \frac{1}{16} \sum_{\substack{y_1, y_2, u_1, u_2 \\ v_1, v_2}} W(y_1 | u_2, u_1 + v_1) \sqrt{W(y_2 | u_2 + v_2, v_1) W(y_2 | u_2 + v_2 + 1, v_1)} \\ &= \frac{1}{16} \sum_{\substack{y_2, u_1, u_2 \\ v_1, v_2}} \sqrt{W(y_2 | u_2 + v_2, v_1) W(y_2 | u_2 + v_2 + 1, v_1)} \\ &= \frac{1}{4} \sum_{u_1, u_2} Z_2(W) = Z_2(W). \end{aligned} \quad (4.32)$$

Proof of (4.31): We have that

$$\begin{aligned} Z_2(W \otimes W) &= \frac{1}{4} \sum_{\substack{y_1, y_2, u_1, u_2 \\ v_1, v_2}} \sqrt{(W \otimes W)(y_1, y_2, u_1, u_2 | v_1, v_2)(W \otimes W)(y_1, y_2, u_1, u_2 | v_1 + 1, v_2)} \\ &= \frac{1}{16} \sum_{\substack{y_1, y_2, u_1, u_2 \\ v_1, v_2}} \sqrt{W(y_1 | u_2, u_1 + v_1) W(y_1 | u_2, u_1 + v_1 + 1)} \\ &\quad \cdot \sqrt{W(y_2 | u_2 + v_2, v_1) W(y_2 | u_2 + v_2, v_1 + 1)} \\ &= \frac{1}{16} \sum_{y_1, u_1, u_2} \sqrt{W(y_1 | u_2, u_1) W(y_1 | u_2, u_1 + 1)} \\ &\quad \cdot \sum_{y_2, v_1, v_2} \sqrt{W(y_2 | v_2, v_1) W(y_2 | v_2, v_1 + 1)} \\ &= Z_1(W)^2. \end{aligned} \quad (4.33)$$

□

Lemma 102. *The following inequalities hold for the partial channels, $(W \boxtimes W)^{[i]}$ and $(W \otimes W)^{[i]}$, for all $i \in \{1, 2\}$,*

$$Z((W \boxtimes W)^{[1]}) \leq 2Z(W^{[2]}) - Z(W^{[2]})^2. \quad (4.34)$$

$$Z((W \boxtimes W)^{[2]}) = Z(W^{[1]}). \quad (4.35)$$

$$Z((W \otimes W)^{[1]}) = Z_1(W)Z(W^{[2]}). \quad (4.36)$$

$$Z((W \otimes W)^{[2]}) \leq Z(W^{[1]}). \quad (4.37)$$

Proof. The transition probabilities of the partial channels (see (4.11) and (4.12)) $(W \otimes W)^{[i]}$ and $(W \boxtimes W)^{[j]}$ for $i, j \in \{1, 2\}$ is given by

$$(W \boxtimes W)^{[1]}(y_1, y_2 | u_1) = \frac{(W \boxtimes W)(y_1, y_2 | u_1, 0) + (W \boxtimes W)(y_1, y_2 | u_1, 1)}{2}$$

$$\begin{aligned}
 &= \frac{1}{8} \sum_{v_1, v_2} [W(y_1|0, u_1 + v_1)W(y_2|v_2, v_1) + W(y_1|1, u_1 + v_1)W(y_2|v_2 + 1, v_1)] \\
 &= \frac{1}{8} \sum_{v_1} [W(y_1|0, u_1 + v_1) \sum_{v_2} W(y_2|v_2, v_1) + W(y_1|1, u_1 + v_1) \sum_{v_2} W(y_2|v_2 + 1, v_1)] \\
 &= \frac{1}{4} \sum_{v_1} [W(y_1|0, u_1 + v_1)W^{[2]}(y_2|v_1) + W(y_1|1, u_1 + v_1)W^{[2]}(y_2|v_1)] \\
 &= \frac{1}{2} \sum_{v_1} W^{[2]}(y_1|u_1 + v_1)W^{[2]}(y_2|v_1). \tag{4.38}
 \end{aligned}$$

$$\begin{aligned}
 (W \boxtimes W)^{[2]}(y_1, y_2|u_2) &= \frac{(W \boxtimes W)(y_1, y_2|0, u_2) + (W \boxtimes W)(y_1, y_2|1, u_2)}{2} \\
 &= \frac{1}{8} \sum_{v_1, v_2} [W(y_1|u_2, v_1)W(y_2|u_2 + v_2, v_1) + W(y_1|u_2, v_1 + 1)W(y_2|u_2 + v_2, v_1)] \\
 &= \frac{1}{8} \sum_{v_1} [W(y_1|u_2, v_1) \sum_{v_2} W(y_2|u_2 + v_2, v_1) + W(y_1|u_2, v_1 + 1) \sum_{v_2} W(y_2|u_2 + v_2, v_1)] \\
 &= \frac{1}{4} \sum_{v_1} [W(y_1|u_2, v_1) + W(y_1|u_2, v_1 + 1)]W^{[2]}(y_2|v_1) \\
 &= \frac{1}{2} W^{[1]}(y_1|u_2) \sum_{v_1} W^{[2]}(y_2|v_1). \tag{4.39}
 \end{aligned}$$

$$\begin{aligned}
 (W \otimes W)^{[1]}(y_1, y_2, u_1, u_2|v_1) &= \frac{(W \otimes W)^{[1]}(y_1, y_2, u_1, u_2|v_1, 0) + (W \otimes W)^{[1]}(y_1, y_2|v_1, 1)}{2} \\
 &= \frac{1}{8} [W(y_1|u_2, u_1 + v_1)W(y_2|u_2, v_1) + W(y_1|u_2, u_1 + v_1)W(y_2|u_2 + 1, v_1)] \\
 &= \frac{1}{4} W(y_1|u_2, u_1 + v_1)W^{[2]}(y_2|v_1). \tag{4.40}
 \end{aligned}$$

$$\begin{aligned}
 (W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|v_2) &= \frac{(W \otimes W)^{[1]}(y_1, y_2, u_1, u_2|0, v_2) + (W \otimes W)^{[1]}(y_1, y_2|1, v_2)}{2} \\
 &= \frac{W(y_1|u_2, u_1)W(y_2|u_2 + v_2, 0) + W(y_1|u_2, u_1 + 1)W(y_2|u_2 + v_2, 1)}{8}. \tag{4.41}
 \end{aligned}$$

Proof of (4.34): From (4.38), we have that

$$\begin{aligned}
 (W \boxtimes W)^{[1]}(y_1, y_2|0) &= \frac{W^{[2]}(y_1|0)W^{[2]}(y_2|0) + W^{[2]}(y_1|1)W^{[2]}(y_2|1)}{2} \\
 (W \boxtimes W)^{[1]}(y_1, y_2|1) &= \frac{W^{[2]}(y_1|0)W^{[2]}(y_2|1) + W^{[2]}(y_1|1)W^{[2]}(y_2|0)}{2}
 \end{aligned}$$

Define $\alpha(y_1) = W^{[2]}(y_1|0)$, $\beta(y_2) = W^{[2]}(y_2|0)$, $\delta(y_1) = W^{[2]}(y_1|1)$ and $\gamma(y_2) = W^{[2]}(y_2|1)$. Then, the following equalities hold,

$$Z(W^{[2]}) = \sum_{y_1} \sqrt{\alpha(y_1)\delta(y_1)} = \sum_{y_2} \sqrt{\beta(y_2)\gamma(y_2)} \tag{4.42}$$

$$\sum_{y_1} \alpha(y_1) = \sum_{y_1} \delta(y_1) = \sum_{y_2} \beta(y_2) = \sum_{y_2} \gamma(y_2) = 1 \tag{4.43}$$

The Bhattacharyya parameter of the partial channel $(W \boxtimes W)^{[1]}$ is given by

$$\begin{aligned}
 Z((W \boxtimes W)^{[1]}) &= \sum_{y_1, y_2} \sqrt{(W \boxtimes W)^{[1]}(y_1, y_2|0)(W \boxtimes W)^{[1]}(y_1, y_2|1)} \\
 &= \frac{1}{2} \sum_{y_1, y_2} \sqrt{\alpha(y_1)\beta(y_2) + \delta(y_1)\gamma(y_2)} \sqrt{\alpha(y_1)\gamma(y_2) + \delta(y_1)\beta(y_2)} \\
 &\leq \frac{1}{2} \sum_{y_1} [\alpha(y_1) + \delta(y_1)] \sum_{y_2} \sqrt{\beta(y_2)\gamma(y_2)} + \frac{1}{2} \sum_{y_1} \sqrt{\alpha(y_1)\delta(y_1)} \sum_{y_2} [\beta(y_2) + \gamma(y_2)] \\
 &\quad - \sum_{y_1, y_2} \sqrt{\alpha(y_1)\delta(y_1)\beta(y_2)\gamma(y_2)} \\
 &= 2Z(W^{[2]}) - Z(W^{[2]})^2,
 \end{aligned}$$

where for the third inequality, we have used the following inequality from [2]

$$\sqrt{(\alpha\beta + \delta\gamma)(\alpha\gamma + \delta\beta)} \leq (\sqrt{\alpha\beta} + \sqrt{\gamma\delta})(\sqrt{\alpha\gamma} + \sqrt{\delta\beta}) - 2\sqrt{\alpha\beta\gamma\delta}. \quad (4.44)$$

and the fourth equality follows from (4.42) and (4.43).

Proof of (4.35): The Bhattacharyya parameter of the partial channel $(W \boxtimes W)^{[2]}$ is given by

$$\begin{aligned}
 Z((W \boxtimes W)^{[2]}) &= \frac{1}{2} \sum_{y_1, y_2} \sqrt{(W \boxtimes W)^{[2]}(y_1, y_2|0)} \sqrt{(W \boxtimes W)^{[2]}(y_1, y_2|1)} \\
 &= \frac{1}{2} \sum_{y_1} \sqrt{W^{[1]}(y_1|0)W^{[1]}(y_1|1)} \sum_{y_2} \sum_{v_1} W^{[1]}(y_2|v_1) \\
 &= Z(W^{[1]}),
 \end{aligned}$$

where the second equality follows from (4.39).

Proof of (4.36): The Bhattacharyya parameter of the partial channel $(W \otimes W)^{[1]}$ is given by

$$\begin{aligned}
 Z((W \otimes W)^{[1]}) &= \sum_{y_1, y_2, u_1, u_2} \sqrt{(W \otimes W)^{[1]}(y_1, y_2, u_1, u_2|0)(W \otimes W)^{[1]}(y_1, y_2, u_1, u_2|1)} \\
 &= \frac{1}{4} \sum_{y_1} \sum_{u_1, u_2} \sqrt{W(y_1|u_2, u_1)W(y_1|u_2, u_1 + 1)} \sum_{y_2} \sqrt{W^{[2]}(y_2|0)W^{[2]}(y_2|1)} \\
 &= Z_1(W)Z(W^{[2]}), \quad (4.45)
 \end{aligned}$$

where the second equality follows from (4.40).

Proof of (4.37): The Bhattacharyya parameter of the partial channel $(W \otimes W)^{[2]}$ is given

by

$$\begin{aligned}
 Z((W \otimes W)^{[2]}) &= \sum_{y_1, y_2, u_1, u_2} \sqrt{(W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|0)(W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|1)} \\
 &\leq \sum_{y_1, y_2, u_2} \sqrt{\sum_{u_1} (W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|0) \sum_{u'_1} (W \otimes W)^{[2]}(y_1, y_2, u'_1, u_2|1)} \\
 &= \frac{1}{4} \sum_{y_1, y_2, u_2} \sqrt{\frac{1}{2} \left(\sum_{u_1} W(y_1|u_2, u_1) \right) W(y_2|u_2, 0) + \frac{1}{2} \left(\sum_{u_1} W(y_1|u_2, u_1 + 1) \right) W(y_2|u_2, 1)} \\
 &\quad \cdot \sqrt{\frac{1}{2} \left(\sum_{u'_1} W(y_1|u_2, u'_1) \right) W(y_2|u_2 + 1, 0) + \frac{1}{2} \left(\sum_{u'_1} W(y_1|u_2, u'_1 + 1) \right) W(y_2|u_2 + 1, 1)} \\
 &= \frac{1}{2} \sum_{y_1, y_2, u_2} W^{[1]}(y_1|u_2) \sqrt{\frac{W(y_2|u_2, 0) + W(y_2|u_2, 1)}{2}} \sqrt{\frac{W(y_2|u_2 + 1, 0) + W(y_2|u_2 + 1, 1)}{2}} \\
 &= \frac{1}{2} \sum_{y_2, u_2} \sqrt{W^{[1]}(y_2|u_2) W^{[1]}(y_2|u_2 + 1)} \\
 &= Z(W^{[1]}),
 \end{aligned}$$

where for the second inequality, consider vectors $\vec{A}(y_1, y_2, u_2) = (\sqrt{(W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|0)})_{u_1}$ and $\vec{B}(y_1, y_2, u_2) = (\sqrt{(W \otimes W)^{[2]}(y_1, y_2, u_1, u_2|1)})_{u_1}$. Then, it follows from the Cauchy-Schwartz inequality, $|\vec{A}(y_1, y_2, u_2) \cdot \vec{B}(y_1, y_2, u_2)| \leq |\vec{A}(y_1, y_2, u_2)| |\vec{B}(y_1, y_2, u_2)|$. The third equality follows from (4.41). \square

4.2.2 Proof of Multilevel Polarization

We define $W^{(0)} := W \boxtimes W$ and $W^{(1)} := W \otimes W$, and consider the recursive application of channel combining and splitting procedure, $W \mapsto (W^{(0)}, W^{(1)})$. After two steps of polarization, we have a set of four virtual channels, $(W^{(i_1)})^{(i_2)}, \forall i_1 i_2 \in \{0, 1\}^2$. Similarly, after n polarization steps, we have the following set of 2^n virtual channels,

$$W^{(i_1 \dots i_n)} := (W^{(i_1 \dots i_{n-1})})^{(i_n)}, \forall i_1 \dots i_n \in \{0, 1\}^n. \quad (4.46)$$

We now state the multilevel polarization theorem.

Theorem 103. Let $\{W^{(i_1 \dots i_n)} | i_1 \dots i_n \in \{0, 1\}^n\}$ be the set of virtual channels defined in (4.46), when the permutation $\Gamma(L)$ is used as channel combining operation. Then, for any $0 < \delta < \frac{1}{2}$,

$$\lim_{n \rightarrow \infty} \frac{\#\{i_1 \dots i_n \in \{0, 1\}^n \mid W^{(i_1 \dots i_n)} \text{ is either } \delta\text{-noiseless, } \delta\text{-half-noisy of type 1 or 2, or } \delta\text{-noisy}\}}{2^n} = 1.$$

Remark 104. The end result of the multilevel polarization may appear to be similar to the quantum CSS construction presented in Section 1.5.1. However, note that the multilevel polarization is fundamentally different from the quantum CSS construction, as the synthesized virtual channels here are quantum channels at every polarization step, which are identifiable to a classical channel with non-binary input. However, for the CSS construction, we have two classical virtual channels with binary input, obtained by two different classical polar code constructions (see Figure 1.10).

Remark 105. It is easily seen that the multilevel polarization does not happen when the channel combining operation is fixed to the CNOT gate (without Hadamard afterward). Taking CNOT

gate as the channel combining operation, we have the following,

$$Z_1(W \otimes W) = Z_1(W). \quad (4.47)$$

$$Z_2(W \otimes W) = Z_2(W)^2. \quad (4.48)$$

Hence, from (4.47), the parameter Z_1 remains fixed, implying that the multilevel polarization into completely noiseless, half-noisy, and completely noisy channels does not happen. Note that in Lemma 101, where the channel combining operation is fixed to $(H \otimes H)CNOT$, $Z_1(W \otimes W)$ depends on $Z_2(W)$ and $Z_2(W \otimes W)$ depends on $Z_1(W)$. Hence, there is improvement for both parameters Z_1 and Z_2 , when we take many good channels in succession, which is one of the key arguments in the proof of multilevel polarization theorem, below.

Note that it is sufficient to prove the above theorem assuming that n goes to infinity through even values $2, 4, 6, \dots$. Indeed if the above theorem holds for n going to infinity through even values, we can set $W = W^{(i_1)}$, for all $i_1 \in \{0, 1\}$, and then it follows that it also holds for n going to infinity through odd values. Therefore, from now on, we assume that $n = 2m$.

In (4.34)-(4.37), the upper bound on $Z(W^{(i_1)})^{[d]}$, for any $i_1 \in \{0, 1\}$ and $d \in \{1, 2\}$, is a function of $Z(W^{[d]})$, such that $\{d, d'\} = \{1, 2\}$. Therefore, applying the transform $W \rightarrow (W^{(0)}, W^{(1)})$ twice, we get an upper bound on $Z(W^{(i_1 i_2)})^{[d]}$, $\forall i_1 i_2 \in \{0, 1\}^2$, which is a function of $Z(W^{[d]})$. For this reason, it is convenient to consider even steps of polarization, i.e., $n = 2m$, and use $W \rightarrow (W^{(00)}, W^{(01)}, W^{(10)}, W^{(11)})$ as our basic transform for recursion. For any given sequence $i_1 \dots i_n \in \{0, 1\}^n$, we write $i_1 \dots i_n = \omega_1 \dots \omega_m$, such that $\omega_k = i_{2k-1} i_{2k} \in \{0, 1\}^2, \forall k > 0$.

To prove Theorem 103, we will express the limit therein as the probability of an event on a probability space. Therefore, suppose that $\{B_i : i = 0, 1, \dots, \infty\}$ is a sequence of random i.i.d variables defined on a probability space (Ω, \mathcal{F}, P) , where each B_i takes values in $\{0, 1\}^2$ with equal probability, meaning that $P(B_i = 00) = P(B_i = 01) = P(B_i = 10) = P(B_i = 11) = \frac{1}{4}$. Let $\mathcal{F}_0 = \{\phi, \Omega\}$ be the trivial σ -algebra and $\mathcal{F}_m, m \geq 1$ be the σ -field generated by (B_1, \dots, B_m) . Define a random sequence of channels $\{W_m : m \geq 0\}$ on the probability space, such that $W_0 = W$, and at any time $m \geq 1$, $W_m = W_{m-1}^{\omega_m}$, where $\omega_m \in \{0, 1\}^2$ is the value of B_m . Therefore, if $B_1 = \omega_1, B_2 = \omega_2, \dots, B_m = \omega_m$, we have that $W_m = W^{(\omega_1 \dots \omega_m)}$.

For a $0 < \delta < \frac{1}{2}$, we define the following events on probability space,

$$A = \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, W_m \text{ is } \delta\text{-noiseless}\}. \quad (4.49)$$

$$B = \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, W_m \text{ is } \delta\text{-half-noisy of type 1}\}. \quad (4.50)$$

$$C = \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, W_m \text{ is } \delta\text{-half-noisy of type 2}\}. \quad (4.51)$$

$$D = \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, W_m \text{ is } \delta\text{-noisy}\}. \quad (4.52)$$

The intersection of any two of the above sets is the null set. Note that the limit in Theorem 103 is equal to $P(A \cup B \cup C \cup D)$, hence, in other words, Theorem 103 states that one of the events from A, B, C, D occurs with probability 1, as n goes to infinity.

We first prove the following Lemmas 106, 107 and 108, and then use them to prove the above polarization theorem.

Lemma 106. Consider a stochastic process $\{T_m : m \geq 0\}$ defined on (Ω, \mathcal{F}, P) such that it satisfies the following properties:

1. T_m takes values in $[0, 1]$ and is measurable with respect to \mathcal{F}_m , that is, T_0 is a constant and T_m is a function of (B_1, \dots, B_m) .

2. Process $\{(T_m, \mathcal{F}_m) : m \geq 0\}$ is a super-martingale, i.e., $\mathbb{E}_{B_{m+1}}[T_{m+1} \mid T_m = t_m, \dots, T_1 = t_1, T_0] \leq t_m$, for any $m > 0$ and all possible t_1, \dots, t_m .
3. $T_{m+1} = T_m^2$ with probability $\frac{1}{2}$.

Then, the limit $T_\infty = \lim_{m \rightarrow \infty} T_m$ exists with probability 1, and T_∞ takes values in $\{0, 1\}$.

Proof. The proof is similar to [2, Proposition 9]. Since the process $\{(T_m, \mathcal{F}_m) : m \geq 0\}$ is a super-martingale, T_m converges with probability 1. This gives the proof of the first part, which implies that $\lim_{m \rightarrow \infty} |T_{m+1} - T_m| = 0$. As $T_{m+1} = T_m^2$ with probability $\frac{1}{2}$, it follows that T_m takes values in $\{0, 1\}$. \square

Lemma 107. For all $d = 1, 2$, the process $\{Z(W_m^{[d]}) : m \geq 0\}$ defined on (Ω, \mathcal{F}, P) , is a super-martingale and there exist $q_1 = q_1(d), q_2 = q_2(d) \in \{0, 1\}^2$, such that when $B_{m+1} \in \{q_1, q_2\}$, $Z(W_{m+1}^{[d]}) \leq Z(W_m^{[d]})^2$.

Proof. For $d = 1$, using (4.34)-(4.37) with $W = W_m$, we get

$$Z(W_m^{(00)[1]}) \leq 2Z(W_m^{(0)[2]}) - Z((W_m^{(0)})^{[2]})^2 = 2Z(W_m^{[1]}) - Z(W_m^{[1]})^2, \quad (4.53)$$

$$Z(W_m^{(01)[1]}) \leq Z_1(W_m^{(0)})Z(W_m^{(0)[2]}) \leq Z(W_m^{[1]})^2, \quad (4.54)$$

$$Z(W_m^{(10)[1]}) \leq 2Z(W_m^{(1)[2]}) - Z(W_m^{(1)[2]})^2 \leq 2Z(W_m^{[1]}) - Z(W_m^{[1]})^2, \quad (4.55)$$

$$Z(W_m^{(11)[1]}) = Z_1(W_m^{(1)})Z(W_m^{(1)[2]}) \leq Z_2(W_m)Z(W_m^{[1]}), \quad (4.56)$$

where the second inequality in (4.54) uses the inequality $Z_1(W_m^{(0)}) \leq Z(W_m^{(0)[2]})$ from [Lemma 98, point 2], and second inequality in (4.56) uses $Z_1(W_m^{[1]}) = Z_2(W_m)$ from (4.30). From (4.53)-(4.56) and $Z_2(W) \leq Z(W^{[1]})$ [Lemma 98, point (b)], it follows,

$$\sum_{i_1, i_2 \in \{0, 1\}} Z_1(W_m^{(i_1 i_2)[1]}) \leq 4Z_1(W_m^{[1]}). \quad (4.57)$$

Hence, the process $\{Z(W_m^{[1]}) : m \geq 0\}$ is a super-martingale and also when $B_{m+1} \in \{01, 11\}$, we have that $Z(W_{m+1}^{[1]}) \leq Z(W_m^{[1]})^2$.

For $d = 2$, from (4.34)-(4.37) with $W = W_m$, we have that

$$Z(W_m^{(00)[2]}) = Z(W_m^{(0)[1]}) \leq 2Z(W_m^{[2]}) - Z(W_m^{[2]})^2. \quad (4.58)$$

$$Z(W_m^{(01)[2]}) \leq Z(W_m^{(0)[1]}) \leq 2Z(W_m^{[2]}) - Z(W_m^{[2]})^2. \quad (4.59)$$

$$Z(W_m^{(10)[2]}) = Z(W_m^{(1)[1]}) = Z_1(W_m)Z(W_m^{[2]}). \quad (4.60)$$

$$Z(W_m^{(11)[2]}) \leq Z(W_m^{(1)[1]}) = Z_1(W_m)Z(W_m^{[2]}). \quad (4.61)$$

From (4.58)-(4.61) and using $Z_1(W) \leq Z(W^{[2]})$ [Lemma 98, point (b)], we have that

$$\sum_{i_1, i_2 \in \{0, 1\}} Z(W_m^{(i_1 i_2)[2]}) \leq 4Z(W_m^{[2]}). \quad (4.62)$$

Thus, process $\{Z(W_m^{[2]}) : m \geq 0\}$ is a super-martingale, and also when $B_{m+1} \in \{10, 11\}$, we have that $Z(W_{m+1}^{[2]}) \leq Z(W_m^{[2]})^2$. \square

Lemma 108. Define the following events for $d = 1, 2$,

$$S^{[d]}(\delta) := \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, Z(W_m^{[d]}) < \delta\}. \quad (4.63)$$

$$T^{[d]}(\delta) := \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, Z(W_m^{[d]}) > 1 - \delta\}. \quad (4.64)$$

$$S_d(\delta) := \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, Z_d(W_m) < \delta, Z_3(W_m) < \delta\}. \quad (4.65)$$

$$T_d(\delta) := \{\omega \in \Omega : \exists m_0, \forall m \geq m_0, Z_d(W_m) > 1 - \delta\}. \quad (4.66)$$

Then,

(a) $P(S^{[d]}(\delta) \cup T^{[d]}(\delta)) = 1, \forall d = 1, 2$.

(b) Given $\{d, d'\} = \{1, 2\}$, then

(i) $S^{[d]}(\delta) \subseteq S_{d'}(\delta)$.

(ii) $T^{[d]}(\delta) \subseteq T_{d'}(\delta)$ with probability 1.

Proof. **Point (a):** It follows directly from Lemmas 106 and 107. As a consequence, note that any $\omega \in \Omega$ belongs to one of the sets, $S^{[1]}(\delta) \cap S^{[2]}(\delta)$, $S^{[1]}(\delta) \cap T^{[2]}(\delta)$, $T^{[1]}(\delta) \cap S^{[2]}(\delta)$ and $T^{[1]}(\delta) \cap T^{[2]}(\delta)$ with probability 1. This will be used in the proof of Theorem 103.

Point (b).(i): From [lemma 98, point 2], we have that $Z_{d'}(W_m) \leq Z(W_m^{[d]})$ and $Z_3(W_m) \leq Z(W_m^{[d]})$, for $\{d, d'\} = \{1, 2\}$. Then, it immediately follows by definitions of $S^{[d]}(\delta)$ and $S_{d'}(\delta)$ that $S^{[d]}(\delta) \subseteq S_{d'}(\delta)$.

Point (b).(ii)²: We assume $T^{[d]}(\delta) \not\subseteq T_{d'}(\delta)$ with non-zero probability and disprove it by contradiction. The above assumption implies that the following event,

$$E = \{\omega \in \Omega : \omega \in T^{[d]}(\delta), \omega \notin T_{d'}(\delta)\}, \quad (4.67)$$

occurs with non-zero probability, that is, $P(E) > 0$.

Define an event \mathcal{E}_m such that $Z_{d'}(W_m) \leq 1 - \delta$, that is, given $B_1 = \omega_1, \dots, B_m = \omega_m$, we have $Z_{d'}(W^{(\omega_1 \dots \omega_m)}) \leq 1 - \delta$. Any $\omega \in E$ belongs to infinitely many \mathcal{E}_m because if there exists a m_0 such that $Z_{d'}(W_m) > 1 - \delta$, for all $m > m_0$, this would imply $\omega \in T_{d'}(\delta)$, which is not true by assumption. Given $\omega \in E$, consider $M = \{m_1, m_2, \dots\}$ as the set of instances such that for all $m_i \in M, \omega \in \mathcal{E}_{m_i}$. Further, take m such that $B_{m+1} = B_{m+2} = 11$ happens, probability of such an event is given by $P(B_{m+1} = B_{m+2} = 11) = \frac{1}{16} > 0$, for any $m \geq 1$, therefore, $\sum_{m_i \in M} P(B_{m_i+1} = B_{m_i+2} = 11) = \infty$. Since $\{B_m : m \geq 1\}$ are i.i.d. random variables, using Borel-Cantelli lemma, there are infinitely many $m_i \in M$ for which $B_{m_i+1} = B_{m_i+2} = 11$.

The condition $\omega \in T^{[d]}(\delta)$ implies that $Z(W_m^{[d]}) > 1 - \delta$, for all $m \geq m_0$. Take a $m \geq m_0$ such that $\omega \in \mathcal{E}_m$, and $B_{m+1} = B_{m+2} = 11$. Then, we have the following for all $d = 1, 2$,

$$\begin{aligned} Z(W_{m+2}^{[d]}) &\leq Z_{d'}(W_{m+1})Z(W_{m+1}^{[d]}) \\ &\leq Z_{d'}(W_m)^2 Z_{d'}(W_m) Z_d(W_m^{[d]}) \\ &\leq (1 - \delta)^3 < (1 - \delta), \end{aligned}$$

where both the first and second inequalities use (4.56) and (4.61), the second inequality also uses $Z_d(W_{m+1}) = Z_d(W_m)^2$ (from (4.30) and (4.31)), and the third inequality follows from the assumption that $Z_{d'}(W_m) \leq (1 - \delta)$. Hence, we have a contradiction with the statement that $Z(W_m^{[d]}) > 1 - \delta$ for all $m \geq m_0$. Therefore, $T^{[d]}(\delta) \subseteq T_{d'}(\delta)$ holds with probability 1. \square

²Note that $T_{d'}(\delta) \subseteq T^{[d]}(\delta)$, by the same reasoning as in the proof of previous point (b).(i). Hence, point (b).(ii) actually implies that $T^{[d]}(\delta) = T_{d'}(\delta)$ with probability 1.

Proof of Theorem 103: We have the following by definition,

$$\begin{aligned} S_1(\delta) \cap S_2(\delta) &= A. \\ T_1(\delta) \cap S_2(\delta) &= B. \\ S_1(\delta) \cap T_2(\delta) &= C. \\ T_1(\delta) \cap T_2(\delta) &= D. \end{aligned}$$

From point (ii).(a) of Lemma 108, we have that $S^{[1]}(\delta) \cap S^{[2]}(\delta) \subseteq S_1(\delta) \cap S_2(\delta)$, which means $\omega \in S^{[1]} \cap S^{[2]} \implies \omega \in S_1(\delta) \cap S_2(\delta)$. Similarly, from point (ii).(a) and point (ii).(b) of Lemma 108, we have that

$$\begin{aligned} S^{[1]}(\delta) \cap T^{[2]}(\delta) &\subset T_1(\delta) \cap S_2(\delta). \\ T^{[1]}(\delta) \cap S^{[2]}(\delta) &\subset S_1(\delta) \cap T_2(\delta). \\ T^{[1]}(\delta) \cap T^{[2]}(\delta) &\subset T_1(\delta) \cap T_2(\delta). \end{aligned}$$

From point (i) of Lemma 108, we know that one of the events from $S^{[1]}(\delta) \cap S^{[2]}(\delta)$, $S^{[1]}(\delta) \cap T^{[2]}(\delta)$, $T^{[1]}(\delta) \cap S^{[2]}(\delta)$ and $T^{[1]}(\delta) \cap T^{[2]}(\delta)$ happens with probability 1, therefore, we have that

$$S^{[1]}(\delta) \cap S^{[2]}(\delta) = S_1(\delta) \cap S_2(\delta) = A, \quad (4.68)$$

$$S^{[1]}(\delta) \cap T^{[2]}(\delta) = T_1(\delta) \cap S_2(\delta) = B, \quad (4.69)$$

$$T^{[1]}(\delta) \cap S^{[2]}(\delta) = S_1(\delta) \cap T_2(\delta) = C, \quad (4.70)$$

$$T^{[1]}(\delta) \cap T^{[2]}(\delta) = T_1(\delta) \cap T_2(\delta) = D, \quad (4.71)$$

Hence, $P(A \cup B \cup C \cup D) = 1$. □

4.3 Quantum Coding Scheme

Here, we take the same construction as in Section 2.5.1 on $N = 2^n$ copies of a Pauli channel \mathcal{W} , while using the same two-qubit Clifford L everywhere as channel combining operation. We know that the construction yields 2^n quantum virtual channels $\mathcal{W}^{(i_1 \dots i_n)}$, that are CMP channels. Further, we also consider the classical polar code construction on N copies of the classical counterpart channel $W := \mathcal{W}^\#$, using permutation $\Gamma(L)$ as channel combining operation, which synthesizes 2^n classical virtual channels $W^{i_1 \dots i_n}$. From Proposition 78 and Corollary 79, we know that the classical counterpart of the CMP channel $\mathcal{W}^{(i_1 \dots i_n)}$ is classically equivalent to $W^{(i_1 \dots i_n)}$, that is, $\mathcal{W}^{(i_1 \dots i_n)\#} \equiv W^{(i_1 \dots i_n)}$ in the sense of Definition 76. Moreover, from Theorem 103, we know that $W^{(i_1 \dots i_n)}$ tend to be completely noisy, half-noisy, or completely noiseless. We now give the encoding and decoding of the quantum polar code.

4.3.1 Encoding

Consider n steps of polarization with $n > 0$. The polar code construction synthesizes $N = 2^n$ virtual channels corresponding to each $i \in \{0, 1, \dots, N-1\}$. We shall denote, $\mathcal{W}^{(i)} := \mathcal{W}^{(i_1 \dots i_n)}$, where $i_1 \dots i_n$ is the binary representation of $i \in \{0, 1, \dots, N-1\}$.

Similar to Section 4.2, we define the following sets,

$$\mathcal{A} = \{i \in \{0, 1, \dots, N-1\} : W^{(i)} \text{ is } \delta\text{-noiseless}\}. \quad (4.72)$$

$$\mathcal{B} = \{i \in \{0, 1, \dots, N-1\} : W^{(i)} \text{ is } \delta\text{-half-noisy of type 1}\}. \quad (4.73)$$

$$\mathcal{C} = \{i \in \{0, 1, \dots, N-1\} : W^{(i)} \text{ is } \delta\text{-half-noisy of type 2}\}. \quad (4.74)$$

$$\mathcal{D} = \{i \in \{0, 1, \dots, N-1\} : W^{(i)} \text{ is } \delta\text{-noisy}\}. \quad (4.75)$$

From Theorem 103, it follows that for sufficiently large N , all but a vanishing fraction of elements from the set $\{0, 1, \dots, N-1\}$ belong to one of the above sets. Let $\bar{\mathcal{D}}$ denote the complement of $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$. The inputs to the virtual channels $\mathcal{W}^{(i)}$ are supplied as follows for i in $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and $\bar{\mathcal{D}}$,

- If $a \in \mathcal{A}$, the corresponding $\mathcal{W}^{(a)}$ is used for quantum communication.
- If $b \in \mathcal{B}$, the input of the corresponding $\mathcal{W}^{(b)}$ is set to $|+\rangle$, the eigenvector of the Pauli X operator with eigenvalue 1 (recall Z part of the input $x_1 x_2 \in \{0, 1\}^2$, that is x_2 , is randomized by $W^{(b)}$).
- If $c \in \mathcal{C}$, the input of the corresponding $\mathcal{W}^{(c)}$ is set to $|0\rangle$, the eigenvector of the Pauli Z operator with eigenvalue 1 (recall X part of the input $x_1 x_2 \in \{0, 1\}^2$, that is x_1 , is completely randomized by $W^{(c)}$).
- If $d \in \bar{\mathcal{D}}$, the input of the corresponding $\mathcal{W}^{(d)}$ is set to half of an EPR pair. The other half of the EPR pair is given to the decoder.

We shall use the following notation similar to Section 1.5.3 and Section 2.5.1.

- (i) With a slight abuse of notation, we shall denote $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and $\bar{\mathcal{D}}$ consisting of $|\mathcal{A}|, |\mathcal{B}|, |\mathcal{C}|$ and $|\bar{\mathcal{D}}|$ qubits, respectively.
- (ii) Let $\Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}$ be the maximally entangled state $\Phi_{\mathcal{D}\mathcal{D}'}$ as follows,

$$\Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'} = \otimes_{d \in \bar{\mathcal{D}}} \Phi_{dd'}, \quad (4.76)$$

where indices d and d' indicate the d -th qubits of systems $\bar{\mathcal{D}}$ and $\bar{\mathcal{D}}'$, respectively, and $\Phi_{dd'}$ is the density matrix corresponding to an EPR pair.

- (iii) Let $\rho_{\mathcal{B}}^+ := \otimes_{b \in \mathcal{B}} |+\rangle\langle +|_b$ and $\rho_{\mathcal{C}}^0 := \otimes_{c \in \mathcal{C}} |0\rangle\langle 0|_c$.
- (iv) Let also Q_N denote the quantum polar transform, that is, the N -qubit Clifford unitary obtained by applying the two-qubit Clifford unitary L for n levels of recursion (see also Section 2.5.1).

Let a quantum state $\rho_{\mathcal{A}}$ on the system \mathcal{A} is encoded by supplying it as input to the virtual channels corresponding to $a \in \mathcal{A}$. The encoded state, denoted by $\varphi_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'}$, is obtained by applying $G_q \otimes I_{\bar{\mathcal{D}}'}$ on the system $\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'$ as follows

$$\varphi_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'} := (G_q \otimes I_{\bar{\mathcal{D}}'}) (\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}^+ \otimes \rho_{\mathcal{C}}^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) (G_q^\dagger \otimes I_{\bar{\mathcal{D}}'}). \quad (4.77)$$

As no errors occur on the system $\bar{\mathcal{D}}'$, the following is the channel output,

$$\psi_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'} := (\mathcal{W}^{\otimes N} \otimes I_{\bar{\mathcal{D}}'}) (\varphi_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'}). \quad (4.78)$$

Since \mathcal{W} is a Pauli channel, we have that

$$\psi_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}\bar{\mathcal{D}}'} = (E_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}} Q_N \otimes I_{\bar{\mathcal{D}}'}) (\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}^+ \otimes \rho_{\mathcal{C}}^0 \otimes \Phi_{\bar{\mathcal{D}}\bar{\mathcal{D}}'}) (Q_N^\dagger E_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}}^\dagger \otimes I_{\bar{\mathcal{D}}'}), \quad (4.79)$$

for a random N -qubit Pauli error $E_{\mathcal{A}\mathcal{B}\mathcal{C}\bar{\mathcal{D}}} \in \bar{\mathcal{G}}_N$.

4.3.2 Decoding

The decoding is similar to Section 1.2.3, and which is performed in the three steps given below.

Step 1: Apply the inverse quantum polar transform on the channel output state. Applying Q_N^\dagger on the output state $\psi_{ABC\bar{D}\bar{D}'}$, we have that

$$\begin{aligned} Q_N^\dagger \psi_{ABC\bar{D}\bar{D}'} Q_N &= (Q_N^\dagger E_{ABC\bar{D}} Q_N \otimes I_{\bar{D}'}) (\rho_A \otimes \rho_B^+ \otimes \rho_C^0 \otimes \Phi_{\bar{D}\bar{D}'}) (Q_N^\dagger E_{ABC\bar{D}}^\dagger Q_N \otimes I_{\bar{D}'}) \\ &= (E'_{ABC\bar{D}} \otimes I_{\bar{D}'}) (\rho_A \otimes \rho_B^+ \otimes \rho_C^0 \otimes \Phi_{\bar{D}\bar{D}'}) (E'_{ABC\bar{D}} \otimes I_{\bar{D}'}), \end{aligned}$$

where $E'_{ABC\bar{D}} := Q_N^\dagger E_{ABC\bar{D}} Q_N$. Since Q_N is a N -qubit Clifford unitary, it follows that $E'_{ABC\bar{D}} \in \bar{\mathcal{G}}_N$ is also a Pauli error.

Step 2: Quantum measurement. Let $E'_{ABC\bar{D}} = \otimes_{a \in \mathcal{A}} E'_a \otimes_{b \in \mathcal{B}} E'_b \otimes_{c \in \mathcal{C}} E'_c \otimes_{d \in \bar{\mathcal{D}}} E'_d$, where $E'_a, E'_b, E'_c, E'_d \in \bar{\mathcal{G}}_1$. We know that any $E'_i \in \bar{\mathcal{G}}_1$ can be written as $X^{u_1} Z^{u_2}$, where $u_1 u_2 \in \{0, 1\}^2$. The decoder performs the Pauli X measurement on each $b \in \mathcal{B}$, which determines the Z part (u_2) corresponding to E'_b , and the Pauli Z measurement on each $c \in \mathcal{C}$, which determines the X part (u_1) corresponding to E'_c . Finally, the decoder performs the *Bell measurement*, that is, the measurement corresponding to the Pauli operators $X \otimes X$ and $Z \otimes Z$, on the two-qubit system dd' for each $d \in \bar{\mathcal{D}}$, which determines both X and Z parts ($u_1 u_2$) corresponding to E'_d .

Step 3: Decode the classical counterpart polar code. Note that when the *all-identity vector* $I^N \in \bar{\mathcal{G}}_1^N$ is input to the N instances of the classical counterpart $\mathcal{W}^\#$, denoted by $\mathcal{W}^{\#N}$, the error $E_{ABC\bar{D}} \in \bar{\mathcal{G}}_N$ can be considered as an output of $\mathcal{W}^{\#N}$. As $\mathcal{W}^\#$ is a symmetric channel, we have that $\mathcal{W}^{\#N}(E_{ABC\bar{D}} | I^N) = \mathcal{W}^{\#N}(I^N | E_{ABC\bar{D}})$, therefore, we can equivalently consider I^N as the observed channel output, and $E_{ABC\bar{D}}$ (unknown) the channel input. Hence, we have been given,

- u_2 corresponding to E'_b for any $b \in \mathcal{B}$.
- u_1 corresponding to E'_c for any $c \in \mathcal{C}$.
- $u_1 u_2$ corresponding to E'_d for any $d \in \bar{\mathcal{D}}$.
- A noisy observation (namely I^N) of the error $E_{ABC\bar{D}} = P_N E'_{ABC\bar{D}}$, where $P_N E'_{ABC\bar{D}} := Q_N E'_{ABC\bar{D}} Q_N^\dagger$, is the linear transformation corresponding to the quantum polar transform Q_N . Thus, E is classical polar code encoded version of E' .

Based on the above, we can use classical polar decoding, namely the *successive cancellation* decoding, to recover the value of $u_1 u_2$ corresponding to E'_a for all $a \in \mathcal{A}$, u_1 corresponding to E'_b for all $b \in \mathcal{B}$, and u_2 corresponding to E'_c for all $c \in \mathcal{C}$.

4.3.3 Number of Preshared EPR Pairs

In this section, we give an upper bound on $\frac{|\mathcal{D}|}{N}$, that is, the fraction of virtual channels requiring preshared EPR pairs, and also a lower bound on $\frac{|\mathcal{B}| + |\mathcal{C}|}{N}$, that is, the fraction of virtual channels frozen in either the Pauli X or Z basis.

Proposition 109. *Following inequalities hold for sufficiently large N ,*

$$(a) \quad \frac{|\mathcal{D}|}{N} \leq Z(W^{[1]})Z(W^{[2]}).$$

(b) $\frac{|\mathcal{B}|+|\mathcal{C}|}{N} \geq 2 - \mathbb{I}(W) - 2Z(W^{[1]})Z(W^{[2]})$, where $\mathbb{I}(W)$ is the symmetric mutual information of W .

Proof. Point (a): From (4.34)-(4.37), we have the following for any W ,

$$Z(W^{(0)[1]})Z(W^{(0)[2]}) \leq (2 - Z(W^{[2]}))Z(W^{[1]})Z(W^{[2]}).$$

$$Z(W^{(1)[1]})Z(W^{(1)[2]}) \leq Z_1(W)Z(W^{[1]})Z(W^{[2]}).$$

Using the above two equations, we have that

$$\sum_{i_1 \in \{0,1\}} Z(W^{(i_1)[1]})Z(W^{(i_1)[2]}) \leq 2Z(W^{[1]})Z(W^{[2]}) - (Z(W^{[2]}) - Z_1(W))Z(W^{[1]})Z(W^{[2]}), \quad (4.80)$$

$$\leq 2Z(W^{[1]})Z(W^{[2]}), \quad (4.81)$$

where the second inequality follows from $Z(W^{[2]}) \geq Z_1(W)$. Applying (4.81) recursively, for any $W^{(i)}$, with $i_1 \dots i_n \in \{0,1\}^n$ being the binary representation of $i \in \{0, \dots, N-1\}$, we have that

$$\sum_{i=0}^{N-1} Z(W^{(i)[1]})Z(W^{(i)[2]}) \leq 2^n Z(W^{[1]})Z(W^{[2]}). \quad (4.82)$$

We know from Theorem 103 that for sufficiently large $N = 2^n$, any $i \in \{0, \dots, N-1\}$ belongs to one of the sets \mathcal{A} , \mathcal{B} , \mathcal{C} and \mathcal{D} with probability 1. Further, we have that

$$Z(W^{(i)[1]})Z(W^{(i)[2]}) \rightarrow \begin{cases} 1, & \text{if } i \in \mathcal{D}. \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from (4.82), it follows that

$$|\mathcal{D}| \leq N Z(W^{[1]})Z(W^{[2]}).$$

Point (b): Recursively applying (3.26), we have that

$$\sum_{i=0}^{N-1} I(W^{(i)}) = N\mathbb{I}(W). \quad (4.83)$$

We know from Section 4.1 that $I(W^{(i)}) \rightarrow 2$ for $i \in \mathcal{A}$, $I(W^{(i)}) \rightarrow 1$ for $i \in \mathcal{B} \cup \mathcal{C}$, and $I(W^{(i)}) \rightarrow 0$ for $i \in \mathcal{D}$. Thus, we have that

$$2|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}| = N\mathbb{I}(W). \quad (4.84)$$

Any i belongs to one of the sets \mathcal{A} , \mathcal{B} , \mathcal{C} and \mathcal{D} with probability 1, therefore,

$$\frac{|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}| + |\mathcal{D}|}{N} \rightarrow 1. \quad (4.85)$$

From the above two equations, we have that

$$|\mathcal{B}| + |\mathcal{C}| + 2|\mathcal{D}| \approx N(2 - \mathbb{I}(W)). \quad (4.86)$$

Since $|\mathcal{D}| \leq N Z(W^{[1]})Z(W^{[2]})$ from part (a), we have that

$$|\mathcal{B}| + |\mathcal{C}| \geq N(2 - \mathbb{I}(W) - 2Z(W^{[1]})Z(W^{[2]})). \quad \square$$

The upper bounds in points (a) and (b) of the above lemma are not strict in general as one can get a stronger bound by recursively applying (4.80) instead of (4.81) to evaluate $\sum_{i=0}^{N-1} Z(W^{(i)[1]})Z_2(W^{(i)[2]})$ in (4.82). However, here it is not possible to apply (4.80) recursively as we only have upper bound for $Z(W^{(i_1)[2]})$ when $i_1 = 1$.

4.3.4 Fast Polarization

As mentioned in Section 1.2.4 and Section 3.4, reliability of the successive cancellation decoding depends on the speed of polarization, that is, if polarization happens fast enough, the block error probability of the successive cancellation decoding goes to zero. In this section, using the results from [25], we give a fast polarization property, which ensures reliable decoding of the quantum polar code constructed in the previous section.

Proposition 110. *Let $W := \mathcal{W}^\#$ be the classical counterpart of a CMP channel \mathcal{W} , and consider the quantum polar construction on \mathcal{W} for n polarization steps, using the two-qubit Clifford gate L as channel combining operation. If P_e^B is the block error probability of the successive cancellation decoding, then we have the following as $n \rightarrow \infty$,*

$$P_e^B = O(2^{-2^{\beta n}}), \quad (4.87)$$

for any $0 < \beta < \frac{1}{4}$.

Proof. From (4.53)-(4.56) and (4.58)-(4.61), for all $d = 1, 2$ and $\omega_i \in \{0, 1\}^2$, we have that

$$Z(W^{(\omega_i)^{[d]}}) \leq \begin{cases} 2Z(W^{[d]}), & \text{with probability } \frac{1}{2} \\ Z(W^{[d]})^2, & \text{with probability } \frac{1}{2} \end{cases}$$

Note that these are the same two main ingredients required for fast polarization mentioned in Section 3.4. Therefore, from [25], we have the following,

For any sequence $\omega = \omega_1 \cdots \omega_m$, with $n = 2m$, and $\omega_k \in \{0, 1\}^2, \forall k > 0$, such that $Z(W^{(\omega)^{[d]}}) \rightarrow 0$ as $m \rightarrow \infty$, we have the following

$$Z(W^{(\omega)^{[d]}}) \leq 2^{-2^{\alpha m}}, \text{ as } m \rightarrow \infty, \text{ for any } 0 < \alpha < \frac{1}{2}. \quad (4.88)$$

From (4.63), the condition $Z(W^{(\omega)^{[d]}}) \rightarrow 0$ as $m \rightarrow \infty$ implies that $\omega \in S^{[d]}(\delta)$ with $\delta \rightarrow 0$. From (4.68)-(4.71), we know that $S^{[1]}(\delta) = A \cup B$ and $S^{[2]}(\delta) = A \cup C$. Therefore, the above equation holds for $\omega \in A \cup B$, when $d = 1$, and $\omega \in A \cup C$, when $d = 2$.

From [24, Proposition 2], the symbol error probability of the maximum likelihood decoder, denoted by P_e , is upper bounded as $P_e(W) \leq 3Z(W)$, and $P_e(W^{[d]}) \leq Z(W^{[d]})$. Therefore, the block error probability of the successive cancellation decoding, P_e^B , can be upper bounded for sufficiently large codelength 2^n as follows

$$\begin{aligned} P_e^B &\leq \sum_{a \in \mathcal{A}} 3Z(W^{(a)}) + \sum_{b \in \mathcal{B}} Z(W^{(b)^{[1]}}) + \sum_{c \in \mathcal{C}} Z(W^{(c)^{[2]}}) \\ &\leq \sum_a 2 \left[Z(W^{(a)^{[1]}}) + Z(W^{(a)^{[2]}}) \right] + \sum_{b \in \mathcal{B}} Z(W^{(b)^{[1]}}) + \sum_{c \in \mathcal{C}} Z(W^{(c)^{[2]}}) \\ &\leq (4|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}|)2^{-2^{\alpha m}}, \text{ as } m \rightarrow \infty \\ &\leq 2^{n+2}2^{-2^{\alpha m}}, \end{aligned}$$

where the second inequality uses [Lemma 98, Point (b)] and the third inequality follows from (4.88). Therefore, $P_e^B = O(2^{-2^{\beta n}})$ for any $0 < \beta < \frac{1}{4}$. \square

The above proposition implies that $P_e^B \rightarrow 0$ as $N \rightarrow \infty$, hence, the decoding is reliable for sufficiently large N .

4.4 An Alternative Construction

In this section, we introduce an alternative construction, the goal of which is to improve the speed of multilevel polarization. For a quantum erasure channel, we show in the next section with the help of a computer program that the multilevel polarization occurs significantly faster for the alternative construction compared to the previous construction.

Firstly, we note the following circuit equivalence,

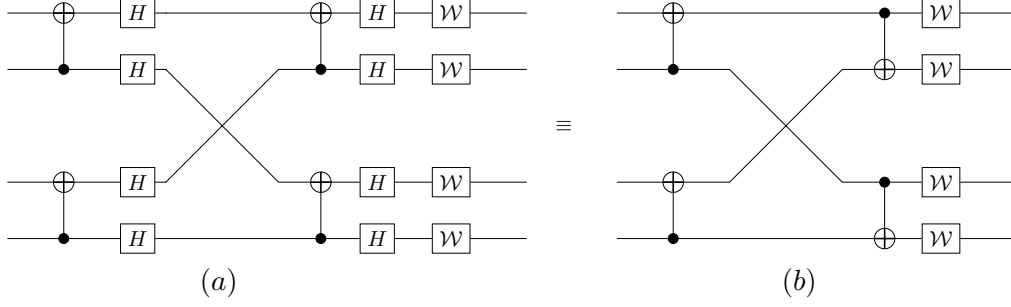


Figure 4.3: (a) and (b) are equivalent quantum circuits.

In circuit (b), the CNOT gate is used in both the first and second polarization step, however, the control and target are interchanged after the first step. To make this clear, we denote by L_1 and L_2 the CNOT gate in the first and second step, respectively. The quantum circuits (a) and (b) are equivalent in the sense that given any 4-qubit quantum state as input, the outputs of quantum circuits (a) and (b) are identical. Therefore, the virtual channels obtained after two steps of channel combining and splitting are equal for both circuits (a) and (b). Hence, the multilevel polarization theorem from the previous section also holds when CNOT gates L_1 and L_2 are used as channel combining operation alternatively for odd and even polarization steps, respectively. In other words, L_1 is used to combine two copies of \mathcal{W} , and then L_2 is used to combine two copies of \mathcal{W}^{i_1} , for all $i_1 \in \{0, 1\}$, again L_1 is used to combine two copies of $\mathcal{W}^{i_1 i_2}$, for all $i_1, i_2 \in \{0, 1\}$, and so on.

Here, we propose an alternative construction, where instead of using L_1 and L_2 for odd and even steps of polarization, an optimal choice is made at each polarization step, using the classical counterpart viewpoint as follows.

Let Γ_1 and Γ_2 be the permutations associated with L_1 and L_2 , respectively. We define $T(\Gamma, W) := Z((W \otimes_{\Gamma} W)^{[1]}) + Z((W \otimes_{\Gamma} W)^{[2]})$, where W is the classical counterpart of \mathcal{W} . For combining two copies of W , the permutation $\Gamma \in \{\Gamma_1, \Gamma_2\}$ is selected as channel combining operation if the following holds,

$$T(\Gamma, W) = \min\{T(\Gamma_1, W), T(\Gamma_2, W)\}. \quad (4.89)$$

A similar selection process takes place at each polarization step, so that two copies of a virtual channel $W^{(i_1 \dots i_k)}$ are combined using the permutation $\Gamma^{(i_1 \dots i_k)} \in \{\Gamma_1, \Gamma_2\}$ minimizing $T(\Gamma^{(i_1 \dots i_k)}, W^{(i_1 \dots i_k)})$.

We now give the following lemma for the Bhattacharya parameter of partial channels associated with virtual channels, $W \boxtimes W$ and $W \otimes W$, using permutations Γ_1 and Γ_2 .

Lemma 111. Let $W^{(0)} := W \boxtimes W$ and $W^{(1)} := W \otimes W$, and for $x \leq a$, $y \leq b$, we denote

$(x, y) \leq (a, b)$. When Γ_1 is used as channel combining operation, we have that

$$\left(Z(W^{(i_1)[1]}), Z(W^{(i_1)[2]}) \right) \leq \begin{cases} (Z(W^{[1]}), 2Z(W^{[2]} - Z(W^{[2]})^2), & \text{when } i_1 = 0, \\ (Z(W^{[1]}), Z_1(W)Z(W^{[2]})), & \text{when } i_1 = 1, \end{cases}$$

and when Γ_2 is used as channel combining operation, we have that

$$\left(Z(W^{(i_1)[1]}), Z(W^{(i_1)[2]}) \right) \leq \begin{cases} (2Z(W^{[1]} - Z(W^{[1]})^2), Z(W^{[2]})), & \text{when } i_1 = 0. \\ (Z_2(W)Z(W^{[1]}), Z(W^{[2]})), & \text{when } i_1 = 1. \end{cases}$$

Proof. We have omitted the proof of the lemma as it is basically the same proof as in Lemma 102. \square

It can be verified from the above inequalities that (4.81), i.e., $\sum_{i_1 \in \{0,1\}} Z(W^{(i_1)[1]})Z(W^{(i_1)[2]}) \leq 2Z(W^{[1]})Z(W^{[2]})$, holds for both Γ_1 and Γ_2 . This implies that the upper bound on the number of preshared EPR pairs from point (a) of Proposition 109 holds for the alternative construction. It is also easy to verify that point (b) of Proposition 109 holds as well.

4.5 Multilevel Polarization for the Quantum Erasure Channel

In this section, using both the first and second construction, we construct quantum polar codes for a quantum erasure channel with the help of a computer program, and compare the two constructions in terms of their speeds of polarization.

4.5.1 Bit-level Erasure Channel

Recall from Definition 47 that the quantum erasure channel with erasure probability $\epsilon > 0$ acts as follows,

$$\mathcal{W}_E(\rho_A) = (1 - \epsilon)|0\rangle\langle 0|_F \otimes \rho_A + \epsilon|1\rangle\langle 1|_F \otimes \frac{I_A}{2}. \quad (4.90)$$

The quantum erasure channel is a CMP channel as it can be written as,

$$\mathcal{W}_E(\rho_A) = (1 - \epsilon)|0\rangle\langle 0|_F \otimes \mathcal{W}_0(\rho_A) + \epsilon|1\rangle\langle 1|_F \otimes \mathcal{W}_1(\rho_A), \quad (4.91)$$

where $\mathcal{W}_0(\rho_A) = \rho_A$ and $\mathcal{W}_1(\rho_A) = \frac{1}{4}[\rho_A + X\rho_A X + Y\rho_A Y + Z\rho_A Z] = \frac{I_A}{2}$ for any ρ_A , are clearly Pauli channels. Therefore, the classical counterpart channel $\mathcal{W}_E^\#$ is the classical mixture of Pauli channels $\mathcal{W}_0^\#$ and $\mathcal{W}_1^\#$ with probabilities $1 - \epsilon$ and ϵ , respectively. Here, $\mathcal{W}_0^\#$ is the identity channel as $\mathcal{W}_0^\#(i | j) = \delta_{ij}, \forall i, j \in \{0, 1\}^2$, and $\mathcal{W}_1^\#$ completely randomizes the two-bit input as $\mathcal{W}_1^\#(i | j) = \frac{1}{4}, \forall i, j \in \{0, 1\}^2$. Thus, $\mathcal{W}_E^\#$ can be considered as a classical erasure channel with two-bits $x_1 x_2 \in \{0, 1\}^2$ as input and the erasure probability $\mathcal{W}_E^\#(? , ? | x_1, x_2) = \epsilon$. Here, symbol ? represents the erasure of a bit.

For the sake of clarity, we denote $W := \mathcal{W}_E^\#$ from now on. For W , the two bits of the input x_1, x_2 is either transmitted perfectly with the probability $1 - \epsilon$, or both bits are erased with the probability ϵ . However, polarizing W yields virtual channels that may erase only one bit either x_1 or x_2 (see also Lemma 114 below). For this reason, we define a more general erasure channel W' , referred to as the bit-level erasure channel, as follows.

Definition 112 (Bit-level erasure channel). *A bit-level erasure channel is defined by the following transition probabilities,*

$$W'(? , x_2 | x_1, x_2) = \epsilon_1, W'(x_1, ? | x_1, x_2) = \epsilon_2, W'(? , ? | x_1, x_2) = \epsilon_3, \forall x_1, x_2 \in \{0, 1\}.$$

The erasure channel W is a special case of the bit-level erasure channel with $\epsilon_1 = \epsilon_2 = 0$, and $\epsilon_3 = \epsilon$. In the next lemma, we give $Z(W'^{[1]})$, $Z(W'^{[2]})$, $Z_1(W')$ and $Z_2(W')$.

Lemma 113. *The following equalities hold for a bit-level erasure channel W' ,*

$$Z(W'^{[1]}) = Z_2(W') = \epsilon_1 + \epsilon_3.$$

$$Z(W'^{[2]}) = Z_1(W') = \epsilon_2 + \epsilon_3.$$

Proof. Given $x_1 x_2 \in \{0, 1\}^2$ as input to W' , the bits x_1 and x_2 are inputs to the partial channels $W'^{[1]}$ and $W'^{[2]}$, respectively. It is not very difficult to see that $W'^{[1]}$ and $W'^{[2]}$ are binary-input erasure channels with erasure probabilities $\epsilon_1 + \epsilon_3$ and $\epsilon_2 + \epsilon_3$, respectively. Since the Bhattacharyya parameter is equal to the erasure probability for a binary-input erasure channel, it follows that $Z(W'^{[1]}) = \epsilon_1 + \epsilon_3$ and $Z(W'^{[2]}) = \epsilon_2 + \epsilon_3$.

Moreover, $Z_1(W') = \epsilon_2 + \epsilon_3$ as for any x_1, x_2 , $\sqrt{W(y|x_1, x_2)W(y|x_1, x_2 \oplus 1)}$ is non-zero only when $y = x_1, ?$ or $y = ?, ?$. Similarly, $Z_2(W') = \epsilon_1 + \epsilon_3$. \square

Taking advantage of the above lemma, we will only use quantities $Z(W'^{[1]})$ and $Z(W'^{[2]})$ from now on. Also, from (3.2) and Lemma 113, the symmetric mutual information of W' is given by,

$$I(W') = 2 - Z(W'^{[1]}) - Z(W'^{[2]}). \quad (4.92)$$

4.5.2 Bhattacharyya Parameters for the First Construction

Here, we consider the quantum polar code construction given in Section 4.2. Firstly, we prove the following lemma for the partial channels.

Lemma 114. *Given W' is a bit-level erasure channel, let $W'^{(0)} := W' \boxtimes_{\Gamma} W'$ and $W'^{(1)} := W' \boxplus_{\Gamma} W'$ be the synthesized virtual channels for the channel combining operation $\Gamma = \Gamma(L)$ (Figure 4.2). Then, $W'^{(0)}$ and $W'^{(1)}$ are also bit-level erasure channels and the inequalities for partial channels in (4.34)-(4.37) are equalities, that is,*

$$Z(W'^{(0)[1]}) = 2Z(W'^{[2]}) - Z(W'^{[2]})^2. \quad (4.93)$$

$$Z(W'^{(0)[2]}) = Z(W'^{[1]}). \quad (4.94)$$

$$Z(W'^{(1)[1]}) = Z(W'^{[2]})^2. \quad (4.95)$$

$$Z(W'^{(1)[2]}) = Z(W'^{[1]}). \quad (4.96)$$

Proof. The erasure probabilities for $W'^{(0)}$,

$$\epsilon_1^0 := W'^{(0)}(? , x_2 | x_1, x_2) = \epsilon_2 + (1 - \epsilon_1 - \epsilon_2 - \epsilon_3) \times (\epsilon_2 + \epsilon_3).$$

$$\epsilon_2^0 := W'^{(0)}(x_1, ? | x_1, x_2) = \epsilon_1 \times (1 - \epsilon_2 - \epsilon_3).$$

$$\epsilon_3^0 := W'^{(0)}(? , ? | x_1, x_2) = \epsilon_3 + \epsilon_1 \times (\epsilon_2 + \epsilon_3).$$

The erasure probabilities for $W'^{(1)}$,

$$\epsilon_1^1 := W'^{(1)}(? , x_2 | x_1, x_2) = \epsilon_2 \times (\epsilon_2 + \epsilon_3).$$

$$\epsilon_2^1 := W'^{(1)}(x_1, ? | x_1, x_2) = \epsilon_1 + \epsilon_3 \times (1 - \epsilon_2 - \epsilon_3).$$

$$\epsilon_3^1 := W'^{(1)}(? , ? | x_1, x_2) = \epsilon_3 \times (\epsilon_2 + \epsilon_3).$$

Note that even when W' is an erasure channel, that is, $\epsilon_1 = \epsilon_2 = 0$, we have that $\epsilon_1^0 = \epsilon_2^1 = (1 - \epsilon_3)\epsilon_3$, which is non-zero except when $\epsilon_3 \in \{0, 1\}$. Therefore, the virtual channels $W'^{(0)}$ and $W'^{(1)}$ are bit-level erasure channels in general. From Lemma 113, we have that

$$\begin{aligned} Z(W'^{(0)[1]}) &= \epsilon_1^0 + \epsilon_3^0 = 2Z(W'^{[2]}) - Z(W'^{[2]})^2. \\ Z(W'^{(0)[2]}) &= \epsilon_2^0 + \epsilon_3^0 = Z(W'^{[1]}). \\ Z(W'^{(1)[1]}) &= \epsilon_1^1 + \epsilon_3^1 = Z(W'^{[2]})^2. \\ Z(W'^{(1)[2]}) &= \epsilon_2^1 + \epsilon_3^1 = Z(W'^{[1]}). \end{aligned}$$

□

Applying Lemma 114 recursively, we may compute $(Z(W^{(i)[1]}), Z(W^{(i)[2]}))$ for any virtual channel $W^{(i)}$ for $i \in \{0, \dots, N-1\}$.

4.5.3 Bhattacharyya Parameters for the Second construction

Here, we consider the alternative construction proposed in the Section 4.4. First of all, we give the following Lemma for Γ_1 and Γ_2 , the permutations associated with the CNOT gates L_1 and L_2 , respectively.

Lemma 115. *Given a bit-level erasure channel W' , let $W'^{(0)} := W' \boxtimes_{\Gamma} W'$ and $W'^{(1)} := W' \otimes_{\Gamma} W'$. Then, for $\Gamma = \Gamma_1$ as channel combining operation, we have that*

$$\left(Z(W'^{(i_1)[1]}), Z(W'^{(i_1)[2]}) \right) = \begin{cases} (Z(W'^{[1]}), 2Z(W'^{[2]}) - Z(W'^{[2]})^2), & \text{when } i_1 = 0, \\ (Z(W'^{[1]}), Z(W'^{[2]})^2), & \text{when } i_1 = 1, \end{cases}$$

and for $\Gamma = \Gamma_2$ as channel combining operation, we have that

$$\left(Z(W'^{(i_1)[1]}), Z(W'^{(i_1)[2]}) \right) = \begin{cases} (2Z(W'^{[1]}) - Z(W'^{[1]})^2, Z(W'^{[2]})), & \text{when } i_1 = 0. \\ (Z(W'^{[1]})^2, Z(W'^{[2]})), & \text{when } i_1 = 1. \end{cases}$$

Proof. The proof has been omitted as it is basically the same proof as in Lemma 114. □

Recall from Section 4.4 that $\Gamma \in \{\Gamma_1, \Gamma_2\}$ is chosen as channel combining operation if it satisfies (4.89). From Lemma 115, for a virtual channel $W'^{i_1 \dots i_k}$, we have that

$$T(\Gamma_1, W'^{(i_1 \dots i_k)}) = Z(W'^{(i_1 \dots i_k)[1]}) + Z(W'^{(i_1 \dots i_k)[2]})^2. \quad (4.97)$$

$$T(\Gamma_2, W'^{(i_1 \dots i_k)}) = Z(W'^{(i_1 \dots i_k)[1]})^2 + Z(W'^{(i_1 \dots i_k)[2]}). \quad (4.98)$$

Therefore, for a virtual channel $W'^{(i_1 \dots i_k)}$, we first determine the optimal permutation from $\{\Gamma_1, \Gamma_2\}$ using the above two equations, and subsequently compute $(Z(W'^{(i_1 \dots i_k i_{k+1})[1]}), Z(W'^{(i_1 \dots i_k i_{k+1})[2]}))$ using Lemma 115.

4.5.4 A Comparison of the Speed of Polarization between the Two Constructions

Here we will provide numerical results to compare the first and the second construction in terms of their speed of polarization. First, we note the following property.

It follows from Lemmas 114 and 115 that for a bit-level erasure channel W' , (4.81) is an equality for both the first and second construction, i.e., $\sum_{i_1 \in \{0,1\}} Z(W^{(i_1)[1]})Z(W^{(i_1)[2]}) =$

$2Z(W^{[1]})Z(W^{[2]})$. Therefore, the upper bound on $|\mathcal{D}|$ and the lower bound on $|\mathcal{B}| + |\mathcal{C}|$ from Proposition 109 are also equalities for both first and second constructions. Hence, as $N \rightarrow \infty$, we have that

$$\frac{|\mathcal{D}|}{N} \rightarrow Z(W'^{[1]})Z(W'^{[2]}), \quad (4.99)$$

$$\frac{|\mathcal{B}| + |\mathcal{C}|}{N} \rightarrow \left(Z(W'^{[1]}) + Z(W'^{[2]}) - 2Z(W'^{[1]})Z(W'^{[2]}) \right), \quad (4.100)$$

$$\frac{|\mathcal{A}|}{N} \rightarrow \left(1 - Z(W'^{[1]}) - Z(W'^{[2]}) + Z(W'^{[1]})Z(W'^{[2]}) \right), \quad (4.101)$$

where the first equation follows from [Proposition 109, point (a)], the second equation follows from [Proposition 109, point (b)] and (4.92), and the third equation is obtained by using $\frac{|\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}| + |\mathcal{D}|}{N} \rightarrow 1$.

We now consider a quantum erasure channel with erasure probability $W(?, ?|x_1, x_2) = 0.1$. From Lemma 113, $Z(W^{[1]}) = Z(W^{[2]}) = 0.1$. From above three equations, it follows that $\frac{|\mathcal{D}|}{N} \rightarrow 0.01$, $\frac{|\mathcal{B}| + |\mathcal{C}|}{N} \rightarrow 0.18$ and $\frac{|\mathcal{A}|}{N} \rightarrow 0.81$ as $N \rightarrow \infty$. Therefore, we have saved 9% of EPR pairs as compared to the construction from Chapter 3 (see also Section 3.6), and are left with only 1% of preshared EPR pairs. For this erasure channel, we perform a numerical simulation for $n = 20$ steps of polarization for both the first and second construction, and compare their speeds of polarization.

In Figure 4.4, the parameter $T^{(i)} := Z(W^{(i)[1]}) + Z(W^{(i)[2]})$ is plotted for both first and second constructions after $n = 20$ polarization steps.

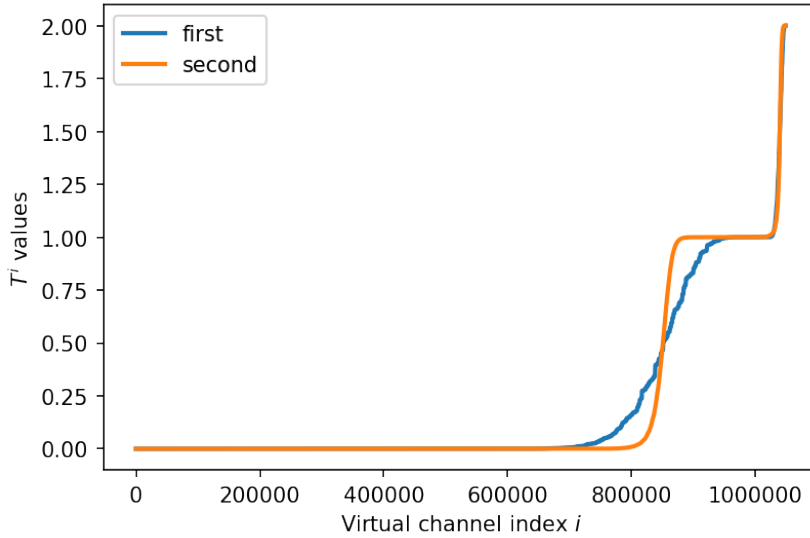


Figure 4.4: $T^{(i)}$ values for a quantum erasure channel with erasure probability $W(?, ?|x_1, x_2) = 0.1$ after $n = 20$ polarization steps. The virtual channel indices $i \in \{0, \dots, 2^n - 1\}$ are sorted according to increasing $T^{(i)}$ values.

The multilevel polarization is evident in the above figure. In particular, we have the following,

- When $T^{(i)} \rightarrow 0$, that is, the plateau in the beginning of the plot, $i \in \mathcal{A}$.
- When $T^{(i)} \rightarrow 1$, that is, the plateau in the middle of the plot, $i \in \mathcal{B} \cup \mathcal{C}$.
- When $T^{(i)} \rightarrow 2$, that is, the plateau in the end of the plot, $i \in \mathcal{D}$.

Further, it is clear from the slopes of the slopes of the two curves in Figure 4.4 that the polarization happens faster for the second construction.

Finally, for $\delta = 10^{-6}$, we compare the first and second construction in the following table,

	$\frac{ \mathcal{A} }{N}$	$\frac{ \mathcal{B} + \mathcal{C} }{N}$	$\frac{ \mathcal{D} }{N}$	$\frac{ \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} }{N}$
First construction	0.49438	0.03021	0.00046	0.52505
Second construction	0.64493	0.07359	0.00071	0.71923

Table 4.1: Fraction of polarized channels for $\delta = 10^{-6}$.

Thus, we have $\frac{|\mathcal{A}|+|\mathcal{B}|+|\mathcal{C}|+|\mathcal{D}|}{N}$ greater for the second construction meaning that more fraction of virtual channels are polarized for the second construction.

5

Towards Fault Tolerant Quantum Computing using Quantum Polar Codes

In principle, quantum computers can solve certain problems much faster than any classical computer [69, 70, 71]. However, in real life, qubits are subject to noise and the implementation of quantum gates are not perfect. This means that without error correction, noise will accumulate over time and render the computation useless. Therefore, we need to encode quantum information using a quantum error correction code to be able to perform any useful computation. However, having access to an efficient quantum code does not in itself implies the ability to do fault tolerant quantum computation. It must be complemented with several quantum procedures aimed at preparing encoded (logical) states, operating on encoded states, and extracting information about the error that has happened in the form of an error syndrome. Moreover, in order to be usable, these procedures must themselves be fault tolerant. Roughly speaking, a procedure is said to be fault tolerant if a component failure at any stage of the computation does not spread on many qubits, thus ensuring that it remains correctable by the code (see Definition 116 below). Hence, fault tolerant quantum computing incorporates both a quantum error correcting code and the above fault-tolerant procedures [72, 73, 74, 75, 76]. The quantum fault-tolerance theorem (or quantum threshold theorem) [73] states that if the physical error rate is below a threshold value, one can perform arbitrarily long quantum computation reliably.

We now discuss in more detail the quantum procedures needed for fault-tolerant quantum computing. First of all, we need to prepare encoded logical quantum states fault tolerantly. Further, we should be able to protect the encoded logical states from noise for arbitrarily long time. To do so, we need to perform error correction repeatedly after some time interval in order to prevent the accumulation of errors. For each round of error correction, an error syndrome needs to be extracted. To this end, ancilla qubits are first entangled with data qubits (*i.e.*, qubits of the encoded quantum state) through a *syndrome extraction circuit*, then quantum measurements are performed on ancilla qubits. This is done in a way that error syndrome is extracted without collapsing the encoded quantum information. These measurements also project the error to a Pauli error, that is decoded by a classical algorithm processing the error syndrome. Subsequently, the error may be corrected. Note that component failures that happen during syndrome extraction can also introduce errors on the encoded block of qubits. Hence, we would need a fault

tolerant procedure for the syndrome extraction [72, 77], so that we may hope to correct the errors introduced during syndrome extraction in the next round of correction.

Once we can protect encoded qubits from noise for arbitrarily long time, we would like to perform quantum algorithms on them. Here, we consider the circuit model of quantum computation, where quantum gates chosen from a universal set (see Definition 43) are applied sequentially. For fault-tolerant implementation of a quantum circuit, each qubit in the original circuit is replaced by many physical qubits, which are encoding a logical qubit. Each quantum gate in the circuit is replaced by a procedure involving many quantum gates on encoded logical qubits. We would require these procedures to be fault tolerant. To summarize, we will need the following fault tolerant procedures:

1. Fault tolerant preparation of encoded states.
2. Fault tolerant procedures for every gate in a universal set of quantum gates.
3. A fault tolerant procedure for syndrome extraction.

As mentioned before, for a round of error correction, first an error syndrome is extracted. This syndrome is a classical information that is processed by a classical decoding algorithm to produce an estimate of the error. In the remaining of the chapter, the term “decoder” or “decoding” refers to the classical decoding. A low complexity decoding is an essential ingredient for fault tolerance. Indeed, to keep up with the accumulating errors, one needs to perform error correction rapidly meaning that the time interval between two rounds of error correction needs to be small. Decoding must be faster than the syndrome generation rate, since otherwise the latency overhead becomes exponential in the number of non-Clifford gates, which would hinder any quantum advantage [78]. Hence, a low complexity decoder is needed.

Currently, quantum LDPC codes, and particularly their topological constructions, are one of the most promising candidates for fault tolerant quantum computation [79, 80, 81, 82, 83]. The main property of quantum LDPC codes is that they have low-weight stabilizer generators. Taking advantage of this property, logical state preparation and syndrome extraction is done fault tolerantly by performing measurements corresponding to the generators. However, high complexity decoding may be an obstacle in the implementation of quantum LDPC codes for fault tolerant quantum computing. For example, for topological quantum LDPC codes, the minimum-weight perfect matching (MWPM) [84, 85] is currently the standard decoding, whose complexity scales as the cube in the number of physical qubits. Although this means MWPM decoding is efficient, complexity needs to be further reduced for practical applications. To do so, many alternatives for decoding topological quantum codes have also been proposed [86, 87, 88, 78]. For the general class of LDPC codes, decoding algorithms have been proposed based on belief-propagation [89, 90, 91, 92, 93]. Despite these efforts, further optimizations in decoding are needed if quantum LDPC codes are to be used in practical devices for fault tolerance. This is an active field of research and a lot of efforts are currently being devoted to develop new decoding algorithms [94, 95, 96, 97, 98, 99, 100].

Motivated by the fact that quantum polar codes are equipped with a decoding of log-linear complexity (see Section 1.2.3), in this chapter, our goal is to investigate them in the context of fault tolerant quantum computing. However, an obvious shortcoming of quantum polar codes in the context of fault tolerant quantum computing is the entanglement assistance, which is required for both CSS and purely quantum constructions of the quantum polar code. However, we have that the entanglement assistance goes to zero for CSS quantum polar codes if the low noise condition in Lemma 53 is satisfied. For

this reason, we shall focus on CSS quantum polar codes instead of purely quantum polar codes. Throughout this chapter, we shall assume that the low noise condition is satisfied, so that the entanglement assistance is zero for CSS quantum polar codes.

In the following, for CSS quantum polar codes, we first provide a fault-tolerant procedure for syndrome extraction, which is based on Steane's technique [77]. This requires ancilla qubits prepared in an encoded logical state, hence the problem of syndrome extraction is basically reduced to logical state preparation. We then provide a procedure to prepare the required encoded logical state based on repetition. While encoding one qubit per polar block of length $N = 2^n$, we finally provide logical versions of Pauli, Hadamard and CNOT gates on the encoded state. Hence, the only thing that is missing for universal fault tolerant quantum computation is a fault-tolerant procedure for T gates (see Remark 44).

5.1 Quantum Circuits, Noise Model and Fault tolerant Procedures

Quantum circuit model is a universal language to represent quantum computations. A quantum circuit consists of qubits, unitary gates and quantum measurements. Generally, it is customary to start a quantum circuit on n qubits in an amplitude (computational) basis state $|x_1, \dots, x_n\rangle$, where $x_1, \dots, x_n \in \{0, 1\}$. Then, a sequence of quantum gates chosen from a universal set are applied. A subset of qubits then are measured to extract classical information and measured qubits are discarded. Here, we only consider single qubit measurements as any joint measurement on n qubits is equivalent to applying a unitary on n qubits and then doing single qubit measurements.

For noisy quantum circuits, we consider the following noise model for quantum gates and measurements [83].

1. **Noise model for Quantum Gates:** For any quantum gate U that entangles N -qubits, we assume that noise affects all N qubits upon which the gate acts as follows. While attempting to perform U , one instead performs U followed by a Pauli channel \mathcal{W}_N on N qubits. An N qubit Pauli channel is a channel that randomly applies a N qubit Pauli on N qubit input with some probability.
2. **Noise model for Single Qubit Measurements:** We consider that the outcome of a single qubit measurement is reported incorrectly with some probability. For example, for the measurement of Pauli- Z operator, the measurement wrongly reports 0 as the outcome, when actually the quantum state is projected into $|1\rangle$ state and similarly 1 as the outcome, when it is projected into $|0\rangle$ state.

We have considered the above noise model for the sake of simplicity. Other related noise models have also been investigated for fault tolerant quantum computation, *e.g.*, adversarial independent stochastic noise [101, 102, 103, 104], local stochastic noise [105], depolarizing noise [106, 107], and biased noise, where phase flips happens more often than bit flips [108, 109].

As discussed before, for fault tolerant implementation of a quantum circuit, each qubit in the original circuit is replaced by an encoded (logical) version containing many physical qubits. The qubit state initialization in the computational basis state is replaced by a procedure of logical state preparation, and each gate in the original circuit is replaced

by a procedure, which acts on encoded qubits. Based on the noise model defined above, we say that a quantum gate or measurement on physical qubits fails if a Pauli error apart from the identity occurs in the implementation. Note that a failure can cause a Pauli error that acts non-trivially on many qubits. The weight of a Pauli operator on N qubits is defined as the number of positions, on which it acts non-trivially. For example, the weight of the Pauli error $X^{u_0} \otimes \dots \otimes X^{u_{N-1}}$ is equal to the number of 1s in the vector $(u_0, \dots, u_{N-1}) \in \{0, 1\}^N$. In the following definition, we give the condition for a procedure to be fault tolerant.

Definition 116 (Fault tolerant procedures). *A procedure acting on encoded qubits is said to be fault tolerant if any $n \geq 1$ failures happened during the procedure (including both quantum gate and measurment failures) cause a Pauli error of weight at most n in any one encoded qubit.*

For example a procedure that has only single qubit gates is fault tolerant according to the above definition. Further, a gate, which acts on only one qubit in each encoded quantum state, generates only one error in each encoded quantum state if it fails. Hence, a procedure containing such quantum gates, which act in parallel, *i.e.*, they don't overlap, is also fault tolerant. An example of such a procedure is transversal application of a gate on qubits in encoded quantum states [72, 73] (see also Figures 5.1 and 5.2 below).

5.2 Fault Tolerant Syndrome Extraction

Recall from Section 1.5.1 that a CSS quantum polar code on a Pauli channel consists of two classical polar codes on induced amplitude and phase channels W_A and W_P . For a quantum polar code of length $N = 2^n$, we divide the set of N qubits denoted by indices $\mathcal{S} := \{0, 1, \dots, N-1\}$ into four subsets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$. The qubits in \mathcal{A} are used for quantum communication, each qubit \mathcal{B} is set to the phase basis state $|+\rangle$, each qubit in \mathcal{C} is set to the amplitude basis state $|0\rangle$, and each qubit in \mathcal{D} is set to half of an EPR pair that is preshared between the sender and receiver. Finally, the quantum information is encoded by applying the quantum polar transform Q_N . As discussed before, we assume the low noise condition, so that no preshared EPR pair is required, hence, the set \mathcal{D} is empty (see also the Appendix).

We now briefly recall the decoding from Section 1.5.4. Consider that a Pauli error E occurs on an encoded quantum state. We first apply the inverse of the quantum polar transform Q_N on the error corrupted encoded state. Then, we measure the qubits in \mathcal{B} in the phase basis and the qubits in \mathcal{C} in the amplitude basis. This gives the X and Z components of the error $E' = Q_N^\dagger E Q_N$ corresponding to the set \mathcal{C} and \mathcal{B} , respectively. This information is then given as input to the classical successive cancellation decoder, which outputs an estimate of the error E . The above decoding strategy is fine in a communication scenario, where the goal is to transmit qubits from one place to another using a noisy quantum channel. However, it does not work for computation due to the following reason. After a round of error correction, one needs to encode qubits again in order to perform the remaining computation. The errors that had occurred during the time when the quantum information was not encoded would not be corrected by the quantum polar code. Therefore, one needs to perform the error correction in such a way that quantum information remains encoded.

One way to do this is by performing X and Z type (stabilizer) generator measurements of the CSS quantum polar codes. From Proposition 117 below, it can be seen that generator

measurements also gives the X and Z components of the error $E' = Q_N^\dagger E Q_N$ corresponding to the set \mathcal{C} and \mathcal{B} , respectively. However, as we discuss below, this approach is problematic as an error in the outcome of a generator measurement may propagate to many qubits due to successive cancellation decoding.

Proposition 117. *Let $E' = X^{u'_0} Z^{v'_0} \otimes \dots \otimes X^{u'_{N-1}} Z^{v'_{N-1}}$. Then, the measurement of X type generators G_X gives the values of v'_b for all $b \in \mathcal{B}$ and the measurement of Z type generators G_Z gives the values of u'_c for all $c \in \mathcal{C}$.*

Proof. Recall from Section 1.5.3 that X type generating set G_X has elements of form $g_X^b = Q_N(X_b \otimes_{k \neq b} I_k) Q_N^\dagger$, where $b \in \mathcal{B}$ and Z type generating set G_Z has elements of form $g_Z^c = Q_N(Z_c \otimes_{k \neq c} I_k) Q_N^\dagger$, where $c \in \mathcal{C}$. We first observe the following

$$\begin{aligned} g_X^b E &= Q_N(X_b \otimes_{k \neq b} I_k) Q_N^\dagger E \\ &= Q_N(X_b \otimes_{k \neq b} I_k) E' Q_N^\dagger \\ &= (-1)^{v'_b} Q_N E' (X_b \otimes_{k \neq b} I_k) Q_N^\dagger \\ &= (-1)^{v'_b} E g_X^b, \end{aligned} \tag{5.1}$$

where third equality follows from $X_b Z_b = -Z_b X_b$. This means that g_X^b anticommutes with the error E if $v'_b = 1$, otherwise commutes. Therefore, the measurement of the stabilizer g_X^b gives the value of v'_b . Similarly for a Z type generator g_Z^c , we have that

$$g_Z^c E = (-1)^{u'_c} E g_Z^c. \tag{5.2}$$

Therefore, the measurement of g_Z^c gives the value of u'_c . \square

Now, the first question is whether measurement of generators can be done fault tolerantly. Indeed it can be accomplished using ancilla qubits prepared in a cat state [72]. The cat state consists of the same number of qubits as the weight of the generator that we want to measure, and it is a uniform superposition of computational basis states containing even number of 1. To extract the syndrome, first the (qubit-wise) transversal CNOT gate is applied between the qubits in the encoded quantum state (at positions where the generator does not have the identity) and the qubits in the cat state. Finally, the qubits of the cat state are measured. One can see that a n number of failures occurred during the generator measurement procedure (either while applying the transversal CNOT or while measuring the ancilla qubits) can produce at most n errors in the encoded block of quantum polar code. Hence, the procedure is fault tolerant.

The second question is if the decoding based on the error corrupted syndrome is fault tolerant. As shown in Proposition 117, measurement of a generator of the quantum polar code gives either X or Z component of $E' = Q_N^\dagger E Q_N$ corresponding to some index i in either \mathcal{C} or \mathcal{B} . Based on this information, the error E is estimated using successive cancellation decoding. Failures during the generator measurement procedure may cause error in the extracted syndrome. In other words, we get incorrect information about the error E' corresponding to the position i . The successive cancellation decoding based on incorrect information about E' even at one position may output an error \hat{E} , that is different from E at many positions. Hence, decoding is not fault tolerant as an error in the procedure may spread on many qubits after error correction based on the successive cancellation decoding.

Fault Tolerant Syndrome Extraction Using Steane's Technique: We now show that using Steane's syndrome extraction technique, a noisy version of some codeword for the

classical polar codes on W_A and W_P can be extracted in a fault tolerant way. Based on this information, we can further estimate the error E in a way that the error happened during the syndrome extraction does not propagate uncontrollably after decoding, unlike the syndrome decoding based on generator measurements as discussed before.

Consider an encoded quantum state $|\phi_L\rangle_S$ (encoded version of a quantum state $|\phi\rangle_A$) on which a Pauli error E occurs. Let $E := E_X E_Z$, such that the binary indicator vectors for E_X and E_Z are $\mathbf{e}_X \in \{0, 1\}^N$ and $\mathbf{e}_Z \in \{0, 1\}^N$, respectively. Let $|0\rangle_A := \otimes_{a \in A} |0\rangle_a$, i.e., all zero state on A , and $|+\rangle_B := \otimes_{b \in B} |+\rangle_b$, i.e., all plus state on B . Then, the error corrupted encoded state $E|\phi_L\rangle$ can be written as follows in the amplitude basis,

$$\begin{aligned} E|\phi_L\rangle_S &= EQ_N(|\phi\rangle_A |+\rangle_B |0\rangle_C) \\ &= \frac{1}{\sqrt{2^{|B|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{x} \in \{0,1\}^{|B|}} \phi_{\mathbf{u}} EQ_N(|\mathbf{u}, \mathbf{x}, 0\rangle)_S \\ &= \frac{1}{\sqrt{2^{|B|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{x} \in \{0,1\}^{|B|}} \phi_{\mathbf{u}} E|P_N(\mathbf{u}, \mathbf{x}, 0)\rangle_S \\ &= \frac{1}{\sqrt{2^{|B|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{x} \in \{0,1\}^{|B|}} \phi_{\mathbf{u}} E_Z |\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0)\rangle_S, \end{aligned} \quad (5.3)$$

where in the second equality, we have expanded $|\phi\rangle_A$ and $|+\rangle_B$ in the amplitude basis and the third equality follows from the fact that the quantum polar transform Q_N acts as the classical polar transform P_N in the amplitude basis.

We will now write $E|\phi_L\rangle_S$ in the phase basis. For this, we will use the notation $|\bar{0}\rangle := |+\rangle$ and $|\bar{1}\rangle := |-\rangle$ from Section 1.5.1.

$$\begin{aligned} E|\phi_L\rangle_S &= EQ_N(|\phi\rangle_A |\bar{0}\rangle_B |0\rangle_C) \\ &= \frac{1}{\sqrt{2^{|C|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{z} \in \{0,1\}^{|C|}} \phi_{\bar{\mathbf{u}}} EQ_N(|\mathbf{u}, 0, \mathbf{z}\rangle)_S \\ &= \frac{1}{\sqrt{2^{|C|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{z} \in \{0,1\}^{|C|}} \phi_{\bar{\mathbf{u}}} E|P_N^r(\mathbf{u}, 0, \mathbf{z})\rangle_S \\ &= \frac{1}{\sqrt{2^{|C|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{z} \in \{0,1\}^{|C|}} \phi_{\bar{\mathbf{u}}} E_X |\mathbf{e}_Z \oplus P_N^r(\mathbf{u}, 0, \mathbf{z})\rangle_S, \end{aligned} \quad (5.4)$$

where in the second equality, we have expanded $|\phi\rangle_A$ and $|0\rangle_C$ in the phase basis and the third equality follows from the fact that the quantum polar transform Q_N acts as the reverse classical polar transform P_N^r (see also (1.115)) in the phase basis.

Fault Tolerant Syndrome Extraction for E_X : For this, we need ancilla qubits prepared in the encoded logical state $|+_L\rangle$. We denote by $S' = A' \cup B' \cup C'$ the ancilla system. Then, we have the following in the amplitude basis,

$$\begin{aligned} |+_L\rangle_{S'} &= Q_N(|+\rangle_{A'} |+\rangle_{B'} |0\rangle_{C'}) \\ &= \frac{1}{\sqrt{2^{|A|+|B|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{x} \in \{0,1\}^{|B|}} Q_N(|\mathbf{u}, \mathbf{x}, 0\rangle)_{S'} \\ &= \frac{1}{\sqrt{2^{|A|+|B|}}} \sum_{\mathbf{u} \in \{0,1\}^{|A|}, \mathbf{x} \in \{0,1\}^{|B|}} |P_N(\mathbf{u}, \mathbf{x}, 0)\rangle_{S'}, \end{aligned} \quad (5.5)$$

where in the second equality, we have expanded $|+\rangle_{A'}$ and $|+\rangle_{B'}$ in the amplitude basis. We now apply transversal CNOT gate $C_{S \rightarrow S'}$ between S and S' such that qubits in S are control and qubits in S' are targets as illustrated in Figure 5.1.

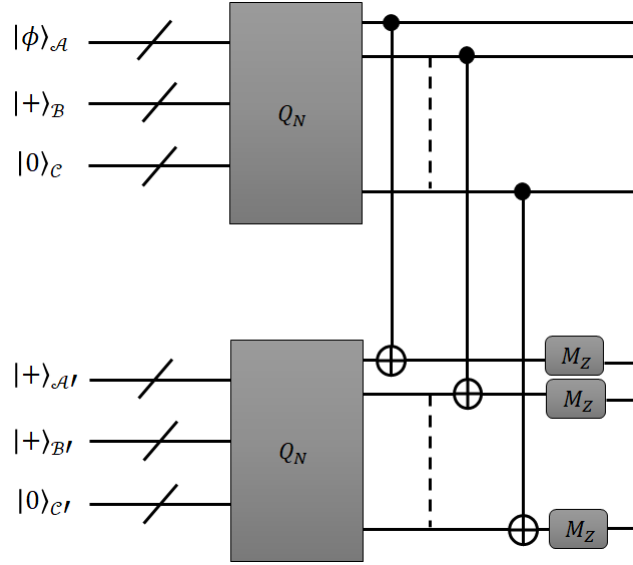


Figure 5.1: Syndrome extraction for E_X . Here, wires in the left of quantum polar transform Q_N represent qubits corresponding to quantum systems A , B and C , and each wire in the right of Q_N represents one qubit. M_Z corresponds to the Pauli Z measurement.

We have the following,

$$\begin{aligned}
 |\theta\rangle_{SS'} &= C_{S \rightarrow S'} (E|\phi_L\rangle_S |+_L\rangle_{S'}) \\
 &= \frac{1}{\sqrt{2^{|A|+2|B|}}} \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|A|}, \mathbf{x}, \mathbf{x}' \in \{0,1\}^{|B|}} \phi_{\mathbf{u}} C_{S \rightarrow S'} (E_Z |\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0)\rangle_S |P_N(\mathbf{u}', \mathbf{x}', 0)\rangle_{S'}) \\
 &= \frac{1}{\sqrt{2^{|A|+2|B|}}} \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|A|}, \mathbf{x}, \mathbf{x}' \in \{0,1\}^{|B|}} \phi_{\mathbf{u}} E_Z |\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0)\rangle_S |\mathbf{e}_X \oplus P_N(\mathbf{u} \oplus \mathbf{u}', \mathbf{x} \oplus \mathbf{x}', 0)\rangle_{S'} \\
 &= E|\phi_L\rangle_S E_X |+_L\rangle_{S'},
 \end{aligned} \tag{5.6}$$

where the second equality follows from (5.3) and (5.5). Therefore, the CNOT gate $C_{S \rightarrow S'}$ simply copies the X part E_X onto the ancilla qubits S' , while S and S' remain separated. We now measure the ancilla qubits S' in the amplitude basis, which yields $(\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0))$ for some $\mathbf{u} \in \{0,1\}^{|A|}$ and $\mathbf{x} \in \{0,1\}^{|B|}$. Clearly, the syndrome extraction is fault tolerant as it consists of the transversal CNOT gate and single qubit measurements.

Note that the measurement outcome here is not a syndrome in the traditional sense, *i.e.*, the value of parity checks of the error \mathbf{e}_X , but a noisy version of some random codeword. In this case, the input of the noisy virtual channels (channels in the set \mathcal{C}) are indeed frozen to zero. This enables the use of the successive cancellation decoder in a fault tolerant way, as explained later in more detail.

Fault Tolerant Syndrome Extraction for E_Z : For this, we need ancilla qubits prepared

in the encoded logical state $|0_L\rangle$. We have the following in the phase basis,

$$\begin{aligned}
 |0_L\rangle_{S'} &= Q_N(|0\rangle_{\mathcal{A}'}|\bar{0}\rangle_{B'}|0\rangle_{C'}) \\
 &= \frac{1}{\sqrt{2^{|\mathcal{A}|+|\mathcal{C}|}}} \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}, \mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} Q_N|\overline{(\mathbf{u}, 0, \mathbf{z})}\rangle_{S'} \\
 &= \frac{1}{\sqrt{2^{|\mathcal{A}|+|\mathcal{C}|}}} \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}, \mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} |\overline{P_N^r(\mathbf{u}, 0, \mathbf{z})}\rangle_{S'}, \tag{5.7}
 \end{aligned}$$

where in the second equality, we have expanded $|0\rangle_{\mathcal{A}}$ and $|0\rangle_{\mathcal{C}}$ in the phase basis. We now apply the transversal CNOT gate $C_{S' \rightarrow S}$ between qubits in S and S' , such that qubits in S' are control and qubits in S are targets as illustrated in Figure 5.2.

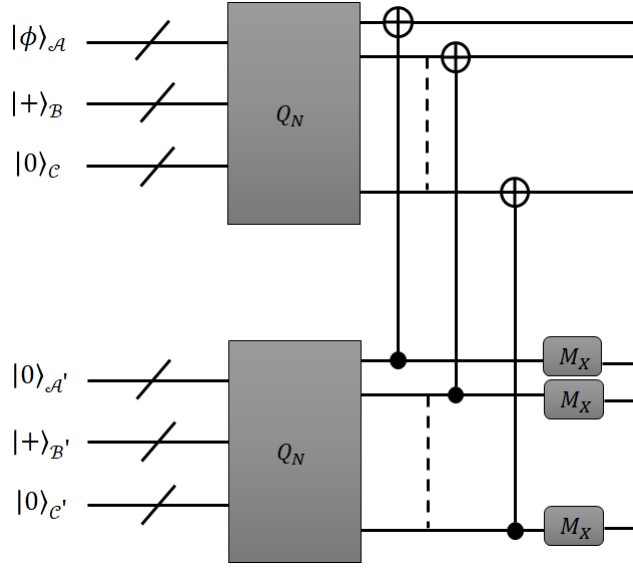


Figure 5.2: Syndrome extraction for E_Z . M_X corresponds to the Pauli X measurement.

We have the following

$$\begin{aligned}
 |\theta'\rangle_{SS'} &= C_{S' \rightarrow S}(E|\phi_L\rangle_S|0_L\rangle_{S'}) \\
 &= \frac{1}{\sqrt{2^{|\mathcal{A}|+2|\mathcal{C}|}}} \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|\mathcal{A}|}, \mathbf{z}, \mathbf{z}' \in \{0,1\}^{|\mathcal{C}|}} \phi_{\mathbf{u}} C_{S' \rightarrow S} \left(E_X |\overline{e\mathbf{z} \oplus P_N^r(\mathbf{u}, 0, \mathbf{z})}\rangle_S |\overline{P_N^r(\mathbf{u}', 0, \mathbf{z}')}\rangle_{S'} \right) \\
 &= \frac{1}{\sqrt{2^{|\mathcal{A}|+2|\mathcal{C}|}}} \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|\mathcal{A}|}, \mathbf{z}, \mathbf{z}' \in \{0,1\}^{|\mathcal{C}|}} \phi_{\mathbf{u}} E_X |\overline{e\mathbf{z} \oplus P_N^r(\mathbf{u}, 0, \mathbf{z})}\rangle_S |\overline{e\mathbf{z} \oplus P_N^r(\mathbf{u} \oplus \mathbf{u}', 0, \mathbf{z}' \oplus \mathbf{z})}\rangle_{S'} \\
 &= E|\phi_L\rangle_S E_Z|0_L\rangle_{S'}, \tag{5.8}
 \end{aligned}$$

where the second equality follows from (5.4) and (5.7). Therefore, the qubit-wise CNOT gate $C_{S' \rightarrow S}$ simply copies the Z part E_Z onto the ancilla qubits S' , while S and S' remain separated. We now measure the ancilla qubits S' in the phase basis, which yields $(e\mathbf{z} \oplus P_N^r(\mathbf{u}, 0, \mathbf{z}))$ for some $\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}$ and $\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}$, which is a noisy version of the codeword $P_N^r(\mathbf{u}, 0, \mathbf{z})$.

Decoding Based on the Extracted Error Syndrome: We first consider the decoding of X part E_X of the total error E . Our goal is to guess the indicator vector e_X corresponding to E_X with high probability using $(e_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0))$, which is a noisy version of the encoded vector $P_N(\mathbf{u}, \mathbf{x}, 0)$. Further, the components of the vector $(\mathbf{u}, \mathbf{x}, 0)$ corresponding

to \mathcal{C} are all zero and known to both the encoder and decoder. Recall from Section 1.5.3 that the set \mathcal{C} corresponds to the bad virtual channels for the classical polar code on the induced amplitude channel W_A . Therefore, using successive cancellation decoding, we can reliably estimate the vectors \mathbf{u} and \mathbf{x} , hence, also the vector \mathbf{e}_X . Similarly, for E_Z , the indicator vector \mathbf{e}_Z is decoded using (noisy version of a codeword) $(\mathbf{e}_Z \oplus P_N^r(\mathbf{u}, 0, \mathbf{z}))$.

We now see what happens if there are errors in the extracted syndrome. Suppose because of failures during the syndrome extraction (either in CNOT gates or single qubit measurements), one instead gets $(\mathbf{e}'_X \oplus \mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}, 0))$ and $(\mathbf{e}'_Z \oplus \mathbf{e}_Z \oplus P_N^r(\mathbf{u}, 0, \mathbf{z}))$ for some $\mathbf{e}'_X, \mathbf{e}'_Z \in \{0, 1\}^N$. Then, the successive cancellation decoding would output $\mathbf{e}'_X \oplus \mathbf{e}_X$ and $\mathbf{e}'_Z \oplus \mathbf{e}_Z$ instead of \mathbf{e}_X and \mathbf{e}_Z , respectively. Therefore, the error correction procedure only leaves errors \mathbf{e}'_X and \mathbf{e}'_Z (errors in the syndrome) on the encoded qubits. This means that successive cancellation decoding is fault tolerant. The remaining errors may be corrected in the next round of error correction.

5.3 Preparation of Encoded Logical States

We need logical polar code states corresponding to a computational basis state to initialize a fault tolerant quantum computation. Further, as discussed in Section 5.2, we need ancilla qubits prepared in logical states $|0_L\rangle$ and $|+_L\rangle$ to extract error syndromes corresponding to X and Z errors, that have happened on the encoded quantum state. First of all, note that to implement the circuit for quantum polar transform Q_N (see Section 1.5.1), we must be able to apply non-local CNOT gates. Further, this circuit is not fault tolerant as failure of a CNOT gate may cause an error of weight two in the prepared encoded state. Further, these errors may propagate to many more qubits through the CNOT gates that are applied afterwards. The propagation of errors may cause too many errors on the encoded state to be correctable by the polar code. Further, these errors are correlated, hence, polar codes may not be suitable as they have been designed to correct independent errors. Hence, we need to use some other method to suppress the errors that occurred during the execution of quantum polar transform circuit.

A construction to produce encoded states of CSS codes free of errors is based on verification of the encoded state, that may have been prepared in a non fault tolerant way [72, 110, 101]. The prepared state is checked for errors using a fault-tolerant procedure. If the errors are detected, it is discarded and the encoded state is prepared again. If no errors are detected, the prepared encoded state is further used for computation. Another construction that has been used for CSS codes is based on fault tolerant preparation of a graph state [111].

Here, we will provide fault tolerant procedures to make the polar code state $|0_L\rangle$ free of both X and Z errors. It is shown in Section 5.4 below that the Hadamard gate is transversal up to a renumbering of qubits, when only one qubit is encoded per quantum polar block. Hence, we can convert $|0_L\rangle$ to $|+_L\rangle$ fault tolerantly, when one qubit is encoded. However, in this section, we present things in general without imposing any constraint on how many qubits are encoded per quantum polar block. Our approach to prepare $|0_L\rangle$ is based on repetition. First, we prepare the encoded quantum state $|0_L\rangle$ free of X errors, using many independent copies of the corrupted logical state $|0_L\rangle$. Then, from many independent copies of $|0_L\rangle$, which are free of X errors, we prepare $|0_L\rangle$ free of both X and Z errors. Using a numerical simulation, we also estimate the logical error rate of producing the corrected polar code state for our method.

5.3.1 Removing X errors from the polar code state $|0_L\rangle$

Let $S = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ be a N -qubit system, on which we want to prepare the polar code state $|0_L\rangle$. We start by preparing the N qubits of S in state $|+\rangle$, denoted by $|+\rangle_S$, which is the equal superposition of all the computational basis states. We assume that this state can be prepared fault tolerantly, and to simplify the analysis, we shall actually assume that no errors occurred during the preparation of the $|+\rangle_S$ state. Hence, we have the following in the amplitude basis (throughout this section, we have omitted the normalization factor for the sake of clarity):

$$\begin{aligned} |+\rangle_S &= |+\rangle_{\mathcal{A}} |+\rangle_{\mathcal{B}} |+\rangle_{\mathcal{C}} \\ &= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} |\mathbf{u}, \mathbf{x}, \mathbf{z}\rangle_S \\ &= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} |P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_S, \end{aligned} \quad (5.9)$$

where the last equality follows by a change of variable.

Now, consider an ancilla system $S' = \mathcal{A}' \cup \mathcal{B}' \cup \mathcal{C}'$, on which we prepare the logical all-zero state, denoted $|0_L\rangle_{\mathcal{A}}$. Hence, if no error occurs during the state preparation, we have the following in the amplitude basis,

$$|0_L\rangle_{S'} = Q_N(|0\rangle_{\mathcal{A}'} |+\rangle_{\mathcal{B}'} |0\rangle_{\mathcal{C}'}) = \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} |P_N(0, \mathbf{x}, 0)\rangle_{S'}. \quad (5.10)$$

However, since the polar encoding is faulty, there may be some Pauli error corrupting the $|0_L\rangle_{S'}$ state. We first deal with X -type errors, and thus denote the error operator by E_X , with the corresponding indicator vector $\mathbf{e}_X = (e_1, \dots, e_N) \in \{0, 1\}^N$. Then, we have the following in the amplitude basis,

$$E_X |0_L\rangle_{S'} = \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} |\mathbf{e}_X \oplus P_N(0, \mathbf{x}, 0)\rangle_{S'}. \quad (5.11)$$

[Initial preparation] Applying the transversal CNOT gate $C_{S \rightarrow S'}$ on $|+\rangle_S E |0_L\rangle_{S'}$, we get:

$$\begin{aligned} C_{S \rightarrow S'} |+\rangle_S E |0_L\rangle_{S'} &= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x}, \mathbf{x}' \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} C_{S \rightarrow S'} (|P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_S |\mathbf{e}_X \oplus P_N(0, \mathbf{x}', 0)\rangle_{S'}) \\ &= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x}, \mathbf{x}' \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} |P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_S |\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x} \oplus \mathbf{x}', \mathbf{z})\rangle_{S'} \\ &= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x}, \mathbf{x}' \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} |P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_S |\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}', \mathbf{z})\rangle_{S'}. \end{aligned} \quad (5.12)$$

We now measure the S' system in the amplitude basis, giving the vector $\mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}', \mathbf{z})$ as measurement outcome, for some $\mathbf{u} \in \{0, 1\}^{|\mathcal{A}|}$, $\mathbf{x}' \in \{0, 1\}^{|\mathcal{B}|}$ and $\mathbf{z} \in \{0, 1\}^{|\mathcal{C}|}$. This leaves the system S in the following state,

$$|\Psi\rangle_S := \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} |P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_S = Q_N(|\mathbf{u}\rangle_{\mathcal{A}} |+\rangle_{\mathcal{B}} |z\rangle_{\mathcal{C}}). \quad (5.13)$$

Note that $|\Psi\rangle_S$ is the encoded logical state corresponding to $|\mathbf{u}\rangle_{\mathcal{A}}$, while assuming that the state of the \mathcal{C} subsystem is frozen to $|z\rangle_{\mathcal{C}}$ rather than all-zero $|0\rangle_{\mathcal{C}}$. Hence, we need to

determine the vectors $\mathbf{u} \in \{0, 1\}^{|\mathcal{B}|}$ and $\mathbf{z} \in \{0, 1\}^{|\mathcal{C}|}$. Indeed, the fact that the state of \mathcal{C} is frozen to $|z\rangle$ rather than all-zero is not problematic, as long as we are able to determine its value. Further, it's also important to determine \mathbf{u} , so that we may apply the logical X operator (Given in Section 5.4 below) to obtain the $|0_L\rangle_{\mathcal{A}}$.

To determine \mathbf{u} and \mathbf{z} , the only information we have is the measurement outcome, which is the binary vector

$$\mathbf{m} := \mathbf{e}_X \oplus P_N(\mathbf{u}, \mathbf{x}', \mathbf{z}). \quad (5.14)$$

For a vector $\mathbf{v} \in \{0, 1\}^{|\mathcal{S}|}$, we define its restriction to $\mathcal{B} \subseteq \mathcal{S}$, denoted by $\mathbf{v}|_{\mathcal{B}}$, as the vector obtained by keeping only the coordinates of \mathbf{v} corresponding to the set \mathcal{B} . Since E_X needs to be determined up to stabilizers, we may always assume that $P_N(\mathbf{e}_X)|_{\mathcal{B}} = 0$ ¹. Therefore, using $P_N(\mathbf{m}) = (\mathbf{u}, \mathbf{x}', \mathbf{z}) \oplus P_N(\mathbf{e}_X)$, it follows that $\mathbf{x}' = P_N(\mathbf{m})|_{\mathcal{B}}$. Since we can determine \mathbf{x}' , we may add $P_N(0, \mathbf{x}', 0)$ to the measurement result \mathbf{m} , therefore getting

$$\mathbf{m}' := \mathbf{m} \oplus P_N(0, \mathbf{x}', 0) = P_N(\mathbf{u}, 0, \mathbf{z}) \oplus \mathbf{e}_X, \quad (5.15)$$

with $P_N(\mathbf{e}_X)|_{\mathcal{B}} = 0$. To resume, we get a noisy version of $P_N(\mathbf{u}, 0, \mathbf{z})$, where the latter is corrupted by an error \mathbf{e}_X , satisfying $P_N(\mathbf{e}_X)|_{\mathcal{B}} = 0$.

[Repetition] Now, we prepare again the system \mathcal{S}' in the logical all-zero state (or we may use another ancilla system \mathcal{S}'' , that can be prepared in parallel with the first preparation of \mathcal{S}'). Since the preparation is again faulty, we get a state

$$E_X^1 |0_L\rangle_{\mathcal{S}'} = \sum_{\mathbf{x} \in \{0, 1\}^{|\mathcal{X}|}} |e_X^1 \oplus P_N(0, \mathbf{x}, 0)\rangle_{\mathcal{S}'}, \quad (5.16)$$

for some error E_X^1 with the indicator vector $e_X^1 \in \{0, 1\}^N$. Applying the transversal CNOT gate $C_{\mathcal{S} \rightarrow \mathcal{S}'}$ on $|\Psi\rangle_{\mathcal{S}} E_X^1 |0_L\rangle_{\mathcal{S}'}$, we get

$$C_{\mathcal{S} \rightarrow \mathcal{S}'} |\Psi\rangle_{\mathcal{S}} E_X^1 |0_L\rangle_{\mathcal{S}'} = \sum_{\mathbf{x}, \mathbf{x}' \in \{0, 1\}^{|\mathcal{B}|}} |P_N(\mathbf{u}, \mathbf{x}, \mathbf{z})\rangle_{\mathcal{S}} |e_X^1 \oplus P_N(\mathbf{u}, \mathbf{x} \oplus \mathbf{x}', \mathbf{z})\rangle_{\mathcal{S}'}. \quad (5.17)$$

Hence, \mathcal{S} and \mathcal{S}' systems are still separated, thus measuring the \mathcal{S}' system in the amplitude basis leaves the \mathcal{S} system in the same state $|\Psi\rangle_{\mathcal{S}}$, with measurement outcome $e_X^1 \oplus P_N(\mathbf{u}, \mathbf{x}', \mathbf{z})$ for some $\mathbf{x}' \in \{0, 1\}^{|\mathcal{B}|}$. Similarly to the above, we may determine a noisy version of $P_N(\mathbf{u}, 0, \mathbf{z})$,

$$\mathbf{m}'_1 = P_N(\mathbf{u}, 0, \mathbf{z}) \oplus e_X^1 \quad (5.18)$$

with $P(e_X^1)|_{\mathcal{B}} = 0$. We repeat the above procedure many times, and use a majority vote for each position $i \in \{0, \dots, N-1\}$ to determine the value of $P_N(\mathbf{u}, 0, \mathbf{z})$. For the majority voting decoding to be reliable, $\Pr(e_i = 1)$, i.e., the probability of the i th component e_i to be 1 for a random error vector \mathbf{e}_X , needs to be small for all $i \in \{0, \dots, N-1\}$. In Section 5.3.3 below, based on a numerical simulation, we estimate the logical error rate p_L , that is, probability that the majority voting decoding provides wrong prediction for $P_N(\mathbf{u}, 0, \mathbf{z})$.

¹Since we may replace \mathbf{e}_X by the equivalent error $\mathbf{e}'_X := \mathbf{e}_X \oplus P_N(0, P_N(\mathbf{e}_X)|_{\mathcal{B}}, 0)$, which satisfies $P_N(\mathbf{e}'_X) = P_N(\mathbf{e}_X) \oplus (0, P_N(\mathbf{e}_X)|_{\mathcal{B}}, 0)$, and therefore $P_N(\mathbf{e}'_X)|_{\mathcal{B}} = 0$.

Remark 118. So far, we have considered that errors occur when preparing the logical all-zero state on system S' , but we have neglected the errors that might be introduced when applying the transversal CNOT gate $C_{S \rightarrow S'}$. Errors introduced by the $C_{S \rightarrow S'}$ gates on the S' system do not propagate and do not accumulate during the repetition process, hence, they will have a negligible impact on the majority vote. Errors introduced by the $C_{S \rightarrow S'}$ gates on the S system will propagate to the S' system during the next repetitions. However, an error on the S system will propagate to only one error in S' , thus we expect their impact on the majority vote to be limited. Once we have determined the z and u values, it should be also possible to correct the S system.

5.3.2 Removing Z errors from the polar code state $|0_L\rangle$

Here, we will use many independent copies of the encoded state $|0_L\rangle$, which is corrupted only by Z errors and free of X errors, to produce the encoded state $|0_L\rangle$ free of both X and Z errors. The approach here is similar to the preparation of $|0_L\rangle$ free of X errors given above.

[Initial Preparation] We first start by preparing the N qubits of S in state $|0\rangle$, denoted by $|0\rangle_S$. We have the following in the phase basis

$$|0\rangle_S = |0\rangle_A |0\rangle_B |0\rangle_C \quad (5.19)$$

$$= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{(\mathbf{u}, \mathbf{x}, \mathbf{z})} \right\rangle_S \quad (5.20)$$

$$= \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{P_N^r(\mathbf{u}, \mathbf{x}, \mathbf{z})} \right\rangle_S, \quad (5.21)$$

where the last equality follows by a change of variable.

Now, consider an ancilla system $S' = \mathcal{A}' \cup \mathcal{B}' \cup \mathcal{C}'$, prepared in the logical all-zero state $|0_L\rangle_{S'}$, which is free of X errors. Let E_Z , with the corresponding indicator vector $\mathbf{e}_Z \in \{0,1\}^N$, be the Z type error that had occurred on $|0_L\rangle_{S'}$ during preparation. This error is the composition of Z errors occurred during the implementation of the polar transform circuit on S' , and during the procedure to make free of X errors. Hence, we have the following in the phase basis, using the notation $|\bar{0}\rangle := |+\rangle$ and $|\bar{1}\rangle := |-\rangle$

$$E_Z |0_L\rangle_{S'} = E_Z Q_N (|0\rangle_{\mathcal{A}'} |\bar{0}\rangle_{\mathcal{B}'} |0\rangle_{\mathcal{C}'}) = \sum_{\mathbf{u} \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{z} \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{\mathbf{e}_Z \oplus P_N^r(\mathbf{u}, 0, \mathbf{z})} \right\rangle_{S'}. \quad (5.22)$$

We now apply the transversal CNOT gate $C_{S' \rightarrow S}$ on $|0\rangle_S E_Z |0_L\rangle_{S'}$. Then, we have the following in the phase basis

$$\begin{aligned} & C_{S' \rightarrow S} |0\rangle_S E_Z |0_L\rangle_{S'} \\ &= \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z}, \mathbf{z}' \in \{0,1\}^{|\mathcal{C}|}} C_{S' \rightarrow S} \left(\left| \overline{P_N^r(\mathbf{u}, \mathbf{x}, \mathbf{z})} \right\rangle_S \left| \overline{\mathbf{e}_Z \oplus P_N^r(\mathbf{u}', 0, \mathbf{z}')} \right\rangle_{S'} \right) \\ &= \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z}, \mathbf{z}' \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{P_N^r(\mathbf{u}, \mathbf{x}, \mathbf{z})} \right\rangle_S \left| \overline{\mathbf{e}_Z \oplus P_N^r(\mathbf{u} \oplus \mathbf{u}', \mathbf{x}, \mathbf{z} \oplus \mathbf{z}')} \right\rangle_{S'} \\ &= \sum_{\mathbf{u}, \mathbf{u}' \in \{0,1\}^{|\mathcal{A}|}} \sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{B}|}} \sum_{\mathbf{z}, \mathbf{z}' \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{P_N^r(\mathbf{u}, \mathbf{x}, \mathbf{z})} \right\rangle_S \left| \overline{\mathbf{e}_Z \oplus P_N^r(\mathbf{u}', \mathbf{x}, \mathbf{z}')} \right\rangle_{S'}. \end{aligned} \quad (5.23)$$

We measure the S' system in the phase basis, giving the vector $e_Z \oplus P_N^r(u', x, z')$ as the measurement outcome, for some $u' \in \{0, 1\}^{|\mathcal{A}|}$, $x \in \{0, 1\}^{|\mathcal{B}|}$ and $z' \in \{0, 1\}^{|\mathcal{C}|}$. This leaves the system S in the following state,

$$|\Psi\rangle_S := \sum_{u \in \{0,1\}^{|\mathcal{A}|}} \sum_{z \in \{0,1\}^{|\mathcal{C}|}} \left| \overline{P_N^r(u, x, z)} \right\rangle_S = Q_N(|0\rangle_{\mathcal{A}} |\bar{x}\rangle_{\mathcal{B}} |0\rangle_{\mathcal{C}}). \quad (5.24)$$

Note that it is the encoded logical state corresponding to $|0\rangle_{\mathcal{A}}$, where the state of the \mathcal{B} subsystem is frozen to $|\bar{x}\rangle_{\mathcal{B}}$ rather than all-plus $|\bar{0}\rangle_{\mathcal{B}} = |+\rangle_{\mathcal{B}}$. We must determine the vector x .

To determine x , the only information we have is the measurement outcome, which is the binary vector

$$m := e_Z \oplus P_N^r(u', x, z'). \quad (5.25)$$

Since E_Z needs to be determined up to stabilizers, we may always assume that $P_N(e_Z)|_{\mathcal{A}} = P_N(e_Z)|_{\mathcal{C}} = 0$, similar to what we have done in the case of X error. Therefore, using $P_N^r(m) = (u', x, z') \oplus P_N^r(e_Z)$, it follows that $u' = P_N^r(m)|_{\mathcal{A}}$ and $z' = P_N^r(m)|_{\mathcal{C}}$. Since we can determine u' and z' , we may add $P_N^r(u', 0, z')$ to the measurement result m , therefore getting

$$m' := m \oplus P_N^r(u', 0, z') = P_N^r(0, x, 0) \oplus e_Z, \quad (5.26)$$

with $P_N^r(e_Z)|_{\mathcal{A}} = 0$ and $P_N^r(e_Z)|_{\mathcal{C}} = 0$. Observe that m' is a noisy version of $P_N^r(0, x, 0)$, corrupted by an error e_Z . Similarly to Section 5.3.1, we obtain many noisy versions of $P_N^r(0, x, 0)$ and, do a majority voting decoding to produce an estimate for $x \in \{0, 1\}^{|\mathcal{B}|}$.

5.3.3 Numerical Results

As the procedures presented above for removing X and Z errors from the polar code state $|0_L\rangle$ are almost identical, we consider only X errors in this section, and using a numerical simulation, we compute the logical error rate p_L for preparing $|0_L\rangle$ free of X errors. We consider a fault model for the CNOT gate, where each possible error $I \otimes X$, $X \otimes I$, and $X \otimes X$ occurs after a CNOT gate with probability $p/3$ (hence, the total error probability is p).

We consider that two qubits are encoded per polar block, *i.e.*, $|\mathcal{A}| = 2$, for the reasons explained below in Section 5.4. We first evaluate $\Pr(e_i = 1), \forall i \in \{1, \dots, N\}$, that is, the probability for i th component e_i of a random error vector e_X to be 1. Then, using this probability, we evaluate the error probability of the majority voting decoding $p_m^i, \forall i \in \{1, \dots, N\}$, depending on the number of repetitions R . Note that for an odd number of independent repetitions R , p_m^i is equal to the probability that at least $\frac{R+1}{2}$ times out of R repetitions the component e_i is equal to 1, that is,

$$p_m^i = \sum_{k=\frac{R+1}{2}}^R C(R, k) \Pr(e_i = 1)^k (1 - \Pr(e_i = 1))^{R-k}, \quad (5.27)$$

where $C(R, k) = \frac{R!}{k!(R-k)!}$. Finally, under the assumption that the majority voting decoding holds independently on each position, we compute the logical error rate p_L (probability of failing to recover $P_N(z, 0, u)$ value) as below,

$$p_L = 1 - \prod_{i=0}^{N-1} (1 - p_m^i). \quad (5.28)$$

To evaluate $\Pr(e_i = 1), \forall i \in \{0, \dots, N-1\}$, we proceed as follows:

1. We apply the polar transform on the all-zero vector, using the above fault model to implement CNOT gates, and generate a random error vector e_X .
2. We replace $e_X \leftarrow e_X \oplus P_N(0, P_N(e_X)|_B, 0)$, such that we get an equivalent error (still denoted by e_X here) for which $P_N(e_X)|_B = 0$.
3. We repeat the steps 1 and 2 several times and update $\Pr(e_i = 1)$ according to the positions of 1s in e_X .

Figure 5.3 below shows the logical error rate p_L , as a function of the CNOT fault probability p for different values of N , where each curve corresponds to a different number of repetitions R .

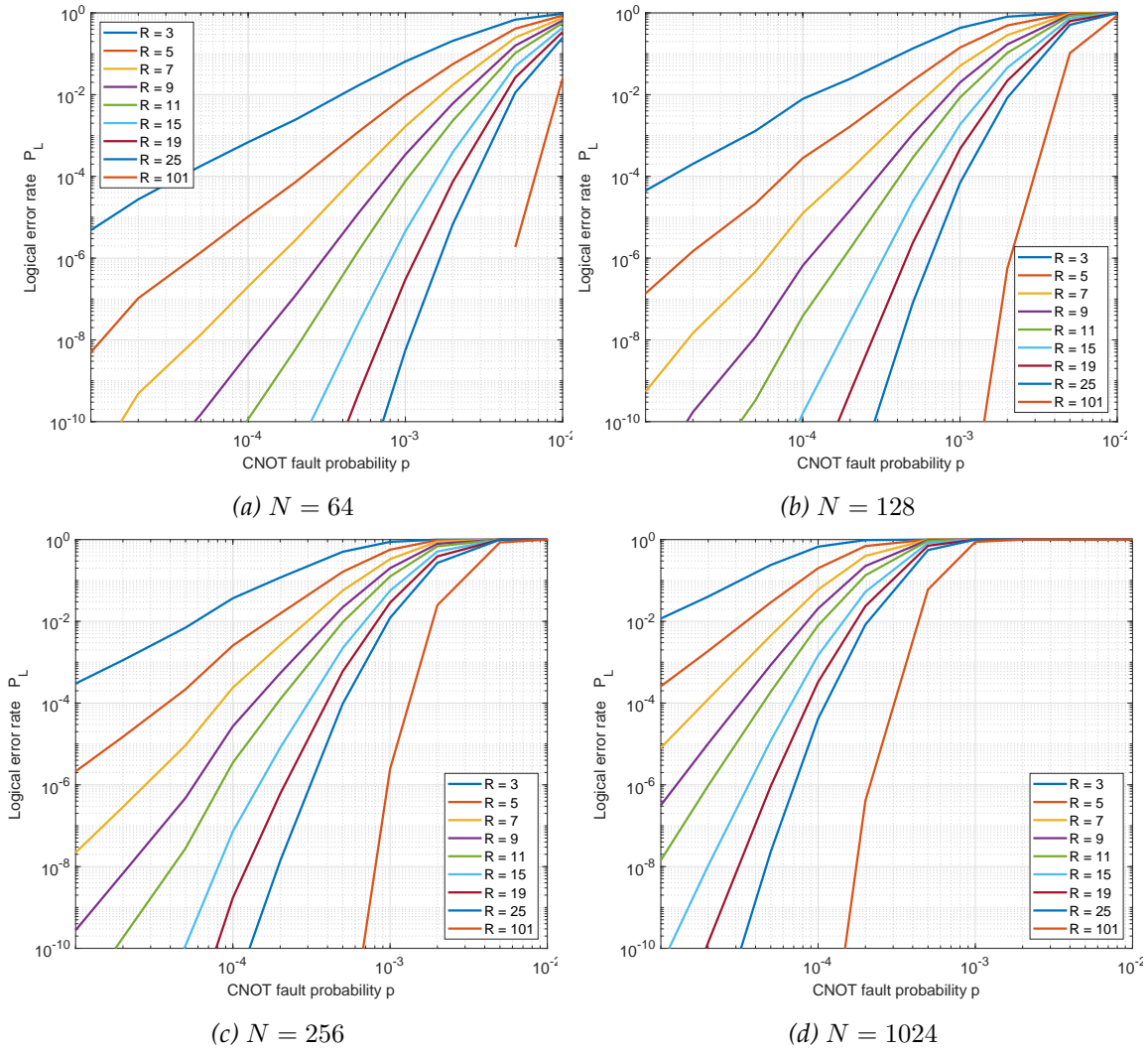


Figure 5.3: Logical error rate p_L as a function of the CNOT fault probability p . Number of physical qubits $N = 64, 128, 256, 1024$, encoding $|\mathcal{A}| = 2$ logical qubits.

It can be observed that the logical error rate p_L increases with the CNOT fault probability p . Further, it also increases with increasing number of physical qubits – this is not surprising, since our approach is based on repetition and does not use error correcting capability of the polar code, hence it is easier to prepare a shorter code state. This means that this approach can not be used to prepare a code state of arbitrary large N for a constant p . However, it may work for a given N if p is small enough and if use a sufficient

number of repetitions. For example, for $N = 64$ and $p = 10^{-4}$, the logical error rate is equal to 10^{-10} for $R = 11$ repetitions. For a given p , preparation of an encoded state of arbitrarily large code length N is still an open problem.

5.4 Fault Tolerant Logical Gates on Encoded Quantum States

In this section, we discuss the fault tolerant procedures for logical gates on encoded quantum states. While encoding one qubit per quantum polar block, we provide fault tolerant procedures to implement logical Paulis X and Z , CNOT, and Hadamard gates.

We consider a quantum polar code on a Pauli channel such that its induced amplitude and phase channels are equivalent, *i.e.*, $W_A \equiv W_P$. Then, from the definitions of \mathcal{B} and \mathcal{C} in Section 1.5.3, we have the following for sets \mathcal{B} and \mathcal{C} (see also the Appendix).

$$b \in \mathcal{B} \iff N - b - 1 \in \mathcal{C}. \quad (5.29)$$

Here, we take $|\mathcal{B}| = |\mathcal{C}| = \frac{N-2}{N}$. This implies that $|\mathcal{A}| = 2$, such that $\mathcal{A} = \{a, N - a - 1\}$, for some $a \in \{0, \dots, N - 1\}$. We will further freeze one qubit of \mathcal{A} in either $|0\rangle$ or $|+\rangle$ state, and use the remaining qubit to encode quantum information. Hence, we have two degrees of freedom associated with freezing a qubit in \mathcal{A} , the first is the choice of the qubit that is frozen and the second is the chosen quantum state in which it is frozen (either $|0\rangle$ or $|+\rangle$). We exploit these degrees of freedoms to implement the logical Hadamard gate transversally (up to a renumbering of the qubits). Finally, for logical X , Z , CNOT gates, it does not make any difference the specific way the qubit in \mathcal{A} is frozen.

To implement an arbitrary unitary gate, we need fault tolerant procedures for a universal set. A standard universal set is $\{C, H, T\}$, where C is the CNOT gate, H is the Hadamard gate, and T is phase gate $R_{\frac{\pi}{8}}$ (see Section 1.3.8 for the definitions of these quantum gates). Hence, the fault tolerant procedure for T gate is the only procedure that we need for universal fault-tolerant quantum computation with quantum polar codes. Finally, We would also like to point out that transversal procedures alone can not be used to implement a universal set according to the Eastin-Knill no go theorem [112]. However, given our implementation of the Hadamard gate, Eastin-Knill theorem does not directly imply the impossibility of having a transversal T gate. In any case, there are different ways to circumvent Eastin-Knill theorem such as quantum gate teleportation [113], magic state distillation [114], gauge fixing [115], concatenated schemes [116], code deformation [117], and lattice surgery [118].

We first give the definition of encoded logical gates.

Definition 119 (Encoded Logical Gates). *Let U be a unitary gate acting on n qubits, and $|\phi_L\rangle$ be the encoded version of a n qubit quantum state $|\phi\rangle$. Then, for any $|\phi\rangle$, the encoded version of U , denoted by $L(U)$, acts as follows on $|\phi_L\rangle$,*

$$L(U)|\phi_L\rangle = |(U\phi)_L\rangle, \quad (5.30)$$

where $|(U\phi)_L\rangle$ is the encoded version of $U|\phi\rangle$.

It can be seen that for any unitary gate U , there always exists an encoded version $L(U)$, however, we need a fault tolerant way to implement it. We now proceed with fault tolerant procedures for Pauli, Hadamard and CNOT gates.

5.4.1 Fault Tolerant Procedure for encoded Pauli X and Z Gates

Let $i \in \mathcal{A}$ be the position corresponding to the information qubit in the polar block ². Then, using Definition 119, for a single qubit unitary U_i acting on the uncoded information qubit, its logical version $L(U_i)$ is given by,

$$L(U) = Q_N U^{(i)} Q_N^\dagger, \quad (5.31)$$

where $U^{(i)} = U_i \otimes_{j \neq i} I_j$ and recall that Q_N is the quantum polar transform. Let $X^{\mathbf{u}} := X^{u_0} \otimes \dots \otimes X^{u_{N-1}}$ and $Z^{\mathbf{u}} := Z^{u_0} \otimes \dots \otimes Z^{u_{N-1}}$, where $\mathbf{u} = (u_0, \dots, u_{N-1}) \in \{0, 1\}^N$. Further, let $\mathbf{u}_i \in \{0, 1\}^N$ be a vector such that it has 1 at i th position and 0 everywhere else. Then, from (5.31), encoded gates corresponding to Pauli gates X_i and Z_i for the information qubit $i \in \mathcal{A}$ are given by,

$$L(X_i) = Q_N X^{\mathbf{u}_i} Q_N^\dagger = X^{P_N \mathbf{u}_i}, \quad (5.32)$$

$$L(Z_i) = Q_N Z^{\mathbf{u}_i} Q_N^\dagger = Z^{P_N^r \mathbf{u}_i}, \quad (5.33)$$

where the second equalities of (5.32) and (5.33) follow from (1.138)-(1.141). Therefore, encoded Pauli gates $L(X^{\mathbf{u}_i})$ corresponds to applying Pauli X on physical qubits at positions, where the vector $P_N \mathbf{u}_i \in \{0, 1\}^N$ has 1. Similarly, $L(Z^{\mathbf{u}_i})$ corresponds to applying Pauli Z on physical qubits at positions, where the vector $P_N^r \mathbf{u}_i \in \{0, 1\}^N$ has 1. As we are only applying single qubit gates on physical qubits, these procedures are fault tolerant.

5.4.2 Fault Tolerant Procedure for the encoded CNOT Gate

Let the first qubit encoded into \mathcal{S} and the second qubit encoded into \mathcal{S}' . Then, the transversal CNOT gate $C_{\mathcal{S} \rightarrow \mathcal{S}'}$ (see Figure 5.1) is equal to the encoded $C_{1 \rightarrow 2}$. This follows from the fact that setting $n = 2$ and $U = C_{1 \rightarrow 2}$, $C_{\mathcal{S} \rightarrow \mathcal{S}'}$ satisfies (5.30) for computational basis states, as shown below in (5.34). Note that it's sufficient that (5.30) holds for a basis, as then it extends to arbitrary quantum states by linearity.

For any $u, u' \in \{0, 1\}$, we have that,

$$\begin{aligned} C_{\mathcal{S} \rightarrow \mathcal{S}'} |u_L\rangle_{\mathcal{S}} |u_L\rangle_{\mathcal{S}'} &= \frac{1}{2^{|\mathcal{B}|}} \sum_{\mathbf{x}, \mathbf{x}' \in \{0, 1\}^{|\mathcal{B}|}} C_{\mathcal{S} \rightarrow \mathcal{S}'} (|P_N(u, \mathbf{x}, 0)\rangle_{\mathcal{S}} |P_N(u', \mathbf{x}', 0)\rangle_{\mathcal{S}'}) \\ &= \frac{1}{2^{|\mathcal{B}|}} \sum_{\mathbf{x}, \mathbf{x}' \in \{0, 1\}^{|\mathcal{B}|}} (|P_N(u, \mathbf{x}, 0)\rangle_{\mathcal{S}} |P_N(u \oplus u', \mathbf{x} \oplus \mathbf{x}', 0)\rangle_{\mathcal{S}'}) \\ &= |u_L\rangle_{\mathcal{S}} |(u \oplus u')_L\rangle_{\mathcal{S}'}, \end{aligned} \quad (5.34)$$

where in the first equality, we have expanded $|u_L\rangle_{\mathcal{S}}$ and $|u_L\rangle_{\mathcal{S}'}$ in the computational basis.

5.4.3 Fault Tolerant Procedure for the encoded Hadamard Gate

Here we take a different approach, using the stabilizer group of the quantum polar code rather than using Definition 119 directly. We first give the following proposition.

Proposition 120. *If a unitary U on physical qubits preserves the stabilizer group of the code via conjugation, i.e., it maps the stabilizer group onto itself via conjugation, then U corresponds to an encoded logical gate.*

²Recall that we are encoding only one qubit per quantum polar block.

Proof. First of all, we note that if a encoded quantum state $|\phi_L\rangle$ is stabilized by an operator A , then $U|\phi_L\rangle$ is stabilized by UAU^\dagger . It is easily seen as follows,

$$U|\phi_L\rangle = UA|\phi_L\rangle \quad (5.35)$$

$$= UAU^\dagger(U|\phi_L\rangle). \quad (5.36)$$

Therefore, the fact that U preserves the stabilizer group of the code via conjugation implies that for any encoded state $|\phi_L\rangle$, $U|\phi_L\rangle$ also belongs to the codespace. Hence, it can be written as,

$$U|\phi_L\rangle = |(U'\phi)_L\rangle, \quad (5.37)$$

for some U' acting on the uncoded state $|\phi\rangle$. This completes the proof. \square

We now prove the following lemma.

Lemma 121. *For a CSS quantum polar code, such that frozen sets \mathcal{B} and \mathcal{C} satisfy the condition $b \in \mathcal{B} \iff N - b - 1 \in \mathcal{C}$ given in (5.29). Then, the gate $R_q\bar{H}$ is an encoded logical gate, where \bar{H} is the transversal Hadamard gate on physical qubits and R_q is the quantum gate that permutes qubits according to the permutation R defined in (1.114), that is, it acts as follows in the computational basis*

$$R_q|u_0, \dots, u_i, \dots, u_{N-1}\rangle := |u_{N-1}, \dots, u_{N-1-i}, \dots, u_0\rangle, \quad (5.38)$$

where $(u_0, \dots, u_i, \dots, u_{N-1}) \in \{0, 1\}^N$.

Proof. To prove the lemma, we show that given the condition in (5.29), the gate $R_q\bar{H}$ preserves the stabilizer group under conjugation.

Similarly to (5.32) and (5.33), the set of X and Z type stabilizer generators of the quantum polar code (Section 1.5.3) can be written as follows.

$$G_X = \{X^{P_N \mathbf{u}_b} \mid b \in \mathcal{B}\}, \quad (5.39)$$

$$G_Z = \{Z^{P_N^r \mathbf{u}_c} \mid c \in \mathcal{C}\}, \quad (5.40)$$

where $\mathbf{u}_b \in \{0, 1\}^N$ is a vector such that it has 1 at position $b \in \mathcal{B}$ and 0 everywhere else and similarly $\mathbf{u}_c \in \{0, 1\}^N$ has 1 at position $c \in \mathcal{C}$ and 0 everywhere else. Further, using $HXH^\dagger = Z$ and $HZH^\dagger = X$ and the definition of R_q , we have the following,

$$R_q\bar{H}X^{P_N \mathbf{u}_b}\bar{H}^\dagger R_q^\dagger = Z^{RP_N \mathbf{u}_b}. \quad (5.41)$$

$$R_q\bar{H}Z^{P_N^r \mathbf{u}_c}\bar{H}^\dagger R_q^\dagger = X^{RP_N^r \mathbf{u}_c}. \quad (5.42)$$

We now show that $Z^{RP_N \mathbf{u}_b} \in G_Z$ and $X^{RP_N^r \mathbf{u}_c} \in G_X$. From $b \in \mathcal{B} \iff N - b - 1 \in \mathcal{C}$, it follows that

$$R\mathbf{u}_b = \mathbf{u}_{c'} \text{ for } c' = N - b - 1 \in \mathcal{C}. \quad (5.43)$$

$$R\mathbf{u}_c = \mathbf{u}_{b'} \text{ for } b' = N - c - 1 \in \mathcal{B}. \quad (5.44)$$

Using (5.43)-(5.44), we have that

$$Z^{RP_N \mathbf{u}_b} = Z^{RP_N R R \mathbf{u}_b} = Z^{P_N^r \mathbf{u}_{c'}} \in G_Z. \quad (5.45)$$

$$X^{RP_N^r \mathbf{u}_c} = X^{RP_N^r R R \mathbf{u}_c} = X^{P_N \mathbf{u}_{b'}} \in G_X. \quad (5.46)$$

Hence, $R_q\bar{H}$ fixes the stabilizer group via conjugation. Finally, from Proposition 120, it follows that $R_q\bar{H}$ is an encoded logical gate. \square

Turning $R_q\bar{H}$ into the Logical Hadamard Gate: For the encoded Hadamard gate $L(H)$, we need a procedure on physical qubits that satisfies the following,

- It is an encoded logical gate (see Proposition 120).
- It exchanges the logical $L(X)$ and $L(Z)$ via conjugation.

As shown in Lemma 121, the first property from the above is satisfied for $R_q\bar{H}$. For the second property, we first consider that two information qubits are encoded corresponding to positions a and $N - a - 1$ in \mathcal{A} . From (5.32)-(5.33), (5.41)-(5.42) and (5.45)-(5.46), logical Pauli X and Z gates corresponding to a and $N - a - 1$ are mapped as follows via conjugation with $R_q\bar{H}$.

$$\begin{aligned} L(X_a) &\rightarrow L(Z_{N-a-1}). \\ L(Z_a) &\rightarrow L(X_{N-a-1}). \\ L(X_{N-a-1}) &\rightarrow L(Z_a). \\ L(Z_{N-a-1}) &\rightarrow L(X_a). \end{aligned}$$

This means $R_q\bar{H}$ on physical qubits is equivalent to applying $L(S(H \otimes H))$, *i.e.*, the encoded version of $S(H \otimes H)$, where S is the swap gate. Finally, we freeze one of the qubits in \mathcal{A} by setting its quantum state to either $|0\rangle$ or $|+\rangle$, and encode only one qubit per quantum polar block. Then, $L(S(H \otimes H))$ basically corresponds to applying encoded Hadamard on the information qubit, and changing the basis of the ancilla qubit and then swapping the frozen and information qubits with each other. Hence, we successfully managed to apply the encoded Hadamard gate. However, the information qubit changed its position in the polar block from a to $N - a - 1$ if a th qubit is used to encode information, or from $N - a - 1$ to a if $N - a - 1$ th qubit is used and the frozen qubit in \mathcal{A} changes its value from $|0\rangle$ to $|+\rangle$, or from $|+\rangle$ to $|0\rangle$.

Conclusion and Perspectives

The goal of this thesis has been to investigate polar coding in the context of quantum communication and computation. We have introduced purely quantum polar codes, which achieve the symmetric channel capacity with entanglement assistance. Purely quantum polar codes make effective use of a quantum polarization phenomenon, where polarization happens at the quantum level, so that the synthesized virtual channels tend to be completely noisy or noiseless as quantum channels. This quantum polarization relies on a channel combining and splitting procedure, using randomized Clifford unitary as the channel combining operation. We have introduced the necessary definitions and worked out the proof of quantum polarization for general quantum channels with qudit input. Further, for qubit Pauli channels, we provide an alternative proof of quantum polarization, and also a decoding algorithm. Moreover, for Pauli channels, we show another polarization phenomenon, namely, multilevel polarization, which uses a fixed channel combining operation and allows to reduce the number of preshared EPR pair compared to the first construction. Finally, we have investigated CSS quantum polar codes for fault tolerant quantum computing. For CSS quantum polar codes, we have provided fault tolerant procedures to (i) prepare a polar code state, (ii) perform logical Pauli, Hadamard, and CNOT gates, and (iii) extract error syndrome.

We now discuss some important open problems arising from this thesis. First of all, we believe that two-level purely quantum polarization may also happen while using one single specific Clifford gate throughout the polarization process. For this, we need guaranteed improvement or degradation of a virtual channel after two steps of polarization as discussed in Section 3.3.2, wherein we have provided numerical evidence in this regard. However, a rigorous proof of guaranteed improvement or degradation is still missing.

For Pauli channels, we have an efficient decoding algorithm. For non-Pauli quantum channels, there are two approaches to deal with the problem. The first approach is based on syndrome measurement, *i.e.*, measuring a generating set of the stabilizer group, which projects the error to a Pauli error. This is the standard way to deal with general quantum noise, and it has been discussed in Section 2.5.2. However, this approach does not achieve symmetric mutual information in general due to the loss of information incurred during the syndrome measurement. The second approach is the twirling of a general quantum channel, with respect to the group of Pauli operators (see Section 1.3.5, for a definition of channel twirling). This operation projects the original quantum channel into a Pauli channel [61](see also Lemma 63), and one may reduce the decoding problem to decoding the latter. This approach is also suboptimal, since channel twirling induces a loss of information. Hence, a decoding algorithm for non-Pauli channels, allowing reliable communication at rates close to the symmetric coherent information remains an open problem.

For fault tolerant quantum computation with CSS quantum polar codes, our procedure to prepare a polar code state (see Section 5.3) works, if the code length is not too large for a constant CNOT gate fault probability p . Larger code lengths imply smaller logical error rate for the quantum polar code, hence one may want to prepare a larger code state. It may be possible to prepare a larger polar code state by better exploiting the structure of the quantum polar code or using concatenation [73]. The other procedures for preparing encoded quantum state for CSS codes such as fast filtering [110] and preparation based on a graph state [111] may also be investigated in this regard.

Finally, we have provided fault tolerant procedures to implement the logical CNOT and Hadamard gates, and we are missing a fault tolerant procedure for logical T gate to have a universal fault tolerant quantum computation with CSS quantum polar codes. Given that our procedure to implement the Hadamard gate is not exactly transversal, Eastin-Knill no go theorem does not directly apply. This means that a transversal (or local) procedure for logical T gate could be possible. This intuition is further strengthened by the fact that some Reed-Muller codes allows transversal implementation of the T gate [119, 114, 111], since polar codes are very close to Reed-Muller codes. Moreover, there are several ways to get round Eastin-Knill theorem (see Section 5.4), out of which a widely used fault tolerant construction is based on magic state distillation. In this context, it is also worth noticing that polar codes can be used for magic state distillation as shown in [120].

Appendix

Here we consider a Pauli channel \mathcal{W} , such that both its amplitude and phase channels W_A and W_P are equivalent, and are given by a BSC W . Alternatively, we may consider that X and Z errors are corrected independently, without taking into account the correlation between them (in this case, if we take for instance \mathcal{W} to be the depolarizing channel, W_A and W_P are equivalent BSCs). Hence, a CSS quantum polar code on \mathcal{W} relies on a classical polar code on W to correct both X and Z errors. Using a numerical simulation, we show that one can construct entanglement free polar codes, without sacrificing the error correction performance. We will use the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and \mathcal{D} defined in Section 1.5.3. We first note the following property.

Proposition 122. *Let $\mathcal{J} \subset \{0, \dots, N-1\}$ be the frozen set for a classical polar code on W , i.e., the set of indices corresponding to the bad virtual channels. Then, the entanglement free condition (the frozen set \mathcal{D} is empty) holds for the CSS quantum polar code on \mathcal{W} , constructed using this classical polar code on W , if and only if the following holds*

$$\mathcal{J} \cap \pi(\mathcal{J}) = \emptyset, \quad (5.47)$$

where $\pi(\mathcal{J}) := \{j \in \{0, \dots, N-1\} \mid N-j-1 \in \mathcal{J}\}$.

Proof. As the polar transform on $W_A = W$ is P_N (not reversed), it directly follows that $\mathcal{J} = \mathcal{C} \cup \mathcal{D}$. From the definitions of \mathcal{B} and \mathcal{C} , it can be seen that if W_A and W_P are equivalent, we have the following

$$j \in \mathcal{C} \iff N-j-1 \in \mathcal{B}. \quad (5.48)$$

Further, using the fact that if $j \in \mathcal{D}$, both $W^{(j)}$ and $W^{(N-j-1)}$ are bad, we have that

$$j \in \mathcal{D} \iff N-j-1 \in \mathcal{D}. \quad (5.49)$$

From (5.48) and (5.49), we have that $\pi(\mathcal{J}) = \mathcal{B} \cup \mathcal{D}$. Since, the sets \mathcal{B}, \mathcal{C} and \mathcal{D} are disjoint, it follows that

$$\mathcal{J} \cap \pi(\mathcal{J}) = \mathcal{D}. \quad (5.50)$$

Hence, the entanglement free condition holds if and only if $\mathcal{J} \cap \pi(\mathcal{J}) = \emptyset$. \square

Corollary 123. *For the entanglement free condition, the rate R of the classical polar code on W must satisfy,*

$$R = 1 - \frac{|\mathcal{J}|}{N} \geq \frac{1}{2}. \quad (5.51)$$

Note that (5.51) is not sufficient for the entanglement free condition.

Numerical Results

To construct a classical polar code on W , with rate $R \geq 1/2$ and code length N , we estimate the error probability of each virtual channel, based on the density evolution technique [121]. Then we consider two different constructions, as follows.

1. The first construction is simply the conventional construction of a Polar code, in which the frozen set \mathcal{J} consists of the $(1 - R)N$ virtual channels with highest error probabilities. We note that $\mathcal{J} \cap \pi(\mathcal{J})$ may be non empty, thus the CSS code based on this classical polar code may need entanglement assistance.
2. For the second construction, we impose the condition that $\mathcal{J} \cap \pi(\mathcal{J}) = \emptyset$ to make it “entanglement free”. Precisely, we freeze virtual channels, starting again from the worst (highest probability value) to the best one (lowest probability value). However, each time a virtual channel, say of index j , is frozen (that is, we add j to \mathcal{J}), we declare the virtual channel of index $N - 1 - j$ as being “unfreezable”. Unfreezable channels are skipped (not added to \mathcal{J}) as the freezing process continues, and the latter stops when $|\mathcal{J}| = (1 - R)N$. This way, we make sure that the set of frozen virtual channels satisfies $\mathcal{J} \cap \pi(\mathcal{J}) = \emptyset$, thus the CSS code based on this construction does not need entanglement assistance.

Once we have fixed the frozen set, we may get an upper-bound of the word error rate (that is, the rate at which the successive cancellation decoder fails), based on the estimated error probabilities of the unfrozen virtual channels [121]. Moreover, this upper-bound is known to be tight, especially in the low word error rate region. In Figures 5.4, 5.5, and 5.6 below, we plot this upper-bound for the two constructions, for various rate R and code length N values. Blue curves correspond to the conventional construction (first construction), while the red markers correspond to the entanglement-free construction (second construction). One can observe that both constructions exhibit the same word error rate performance, which, based on a careful analysis of our numerical results, can be explained as follows. First, the entanglement assistance requirement for the conventional construction (that is, $|\mathcal{J} \cap \pi(\mathcal{J})|$) is low. Second, in the entanglement free construction, virtual channels declared as “unfreezable” will not be frozen, but instead some other virtual channels will be added to \mathcal{J} . Practically, this amounts to replacing a small amount of the channels frozen by the conventional construction, by some others which were not frozen before. It turns out that this replacement operates an exchange between virtual channels with almost the same error probability.

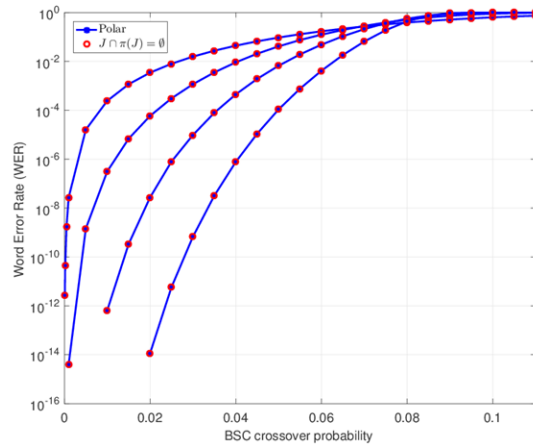


Figure 5.4: Word error rate performance of the conventional and entanglement-free constructions, for $R = 0.5$.

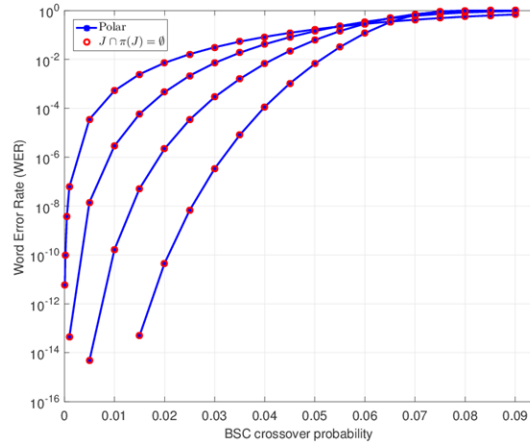


Figure 5.5: Word error rate performance of the conventional and entanglement-free constructions, for $R = 0.55$.

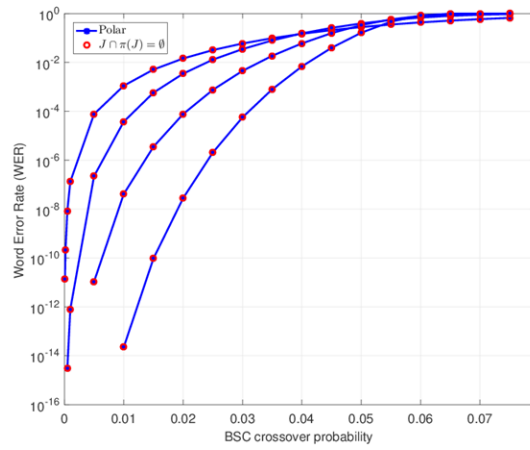


Figure 5.6: Word error rate performance of the conventional and entanglement-free constructions, for $R = 0.6$.

Bibliography

- [1] Claude Elwood Shannon. “A mathematical theory of communication”. In: *Bell System Technical Journal* 27 (1948), pp. 379–423.
- [2] Erdal Arkan. “Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels”. In: *IEEE Transactions on Information Theory* 55.7 (July 2009), pp. 3051–3073. DOI: [10.1109/TIT.2009.2021379](#). arXiv: [0807.3917](#).
- [3] Charles H. Bennett and Gilles Brassard. “Quantum cryptography: Public key distribution and coin tossing”. In: *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*. Vol. 175. 1984, p. 8.
- [4] Richard P Feynman. *Simulating physics with computers*. 1982.
- [5] Alexander S. Holevo. “Information theoretical aspects of quantum measurements”. In: *Problems in Information Transmission (URSS)* 9 (2 1973), pp. 31–42.
- [6] Stephen Wiesner. “Conjugate coding”. In: *ACM Sigact News* 15.1 (1983), pp. 78–88.
- [7] William K Wootters and Wojciech H Zurek. “A single quantum cannot be cloned”. In: *Nature* 299.5886 (1982), pp. 802–803.
- [8] Mark M. Wilde. *Quantum Information Theory*. Cambridge Books Online. Cambridge University Press, 2013. DOI: [10.1017/CBO9781139525343](#). arXiv: [1106.1445](#).
- [9] Seth Lloyd. “Capacity of the noisy quantum channel”. In: *Physical Review A* 55.3 (1997), p. 1613. DOI: [10.1103/PhysRevA.55.1613](#). arXiv: [quant-ph/9604015](#).
- [10] Peter W. Shor. “The quantum channel capacity and coherent information”. In: *Lecture notes, MSRI Workshop on Quantum Computation* (2002).
- [11] Igor Devetak. “The private classical capacity and quantum capacity of a quantum channel”. In: *IEEE Transactions on Information Theory* 51.1 (2005), pp. 44–55. DOI: [10.1109/TIT.2004.839515](#). arXiv: [quant-ph/0304127](#).
- [12] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. “Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem”. In: *IEEE Transactions on Information Theory* 48.10 (2002), pp. 2637–2655. DOI: [10.1109/TIT.2002.802612](#). arXiv: [quant-ph/0106052](#).
- [13] Joseph M. Renes, Frédéric Dupuis, and Renato Renner. “Efficient Polar Coding of Quantum Information”. In: *Physical Review Letters* 109 (5 Aug. 2012), p. 050504. DOI: [10.1103/PhysRevLett.109.050504](#). arXiv: [1109.3195](#).
- [14] Fazlollah M. Reza. *An introduction to information theory*. Courier Corporation, 1994.
- [15] David E. Muller. “Application of Boolean algebra to switching circuit design and to error detection”. In: *Transactions of the IRE Professional Group on Electronic Computers* EC-3.3 (1954), pp. 6–12. DOI: [10.1109/IREPGELC.1954.6499441](#).
- [16] Irving S. Reed. “A class of multiple-error-correcting codes and the decoding scheme”. In: *Transactions of the IRE Professional Group on Information Theory* 4.4 (1954), pp. 38–49. DOI: [10.1109/TIT.1954.1057465](#).

- [17] Irving S. Reed and Gustave Solomon. "Polynomial Codes Over Certain Finite Fields". In: *Journal of the Society for Industrial and Applied Mathematics* 8.2 (1960), pp. 300–304. DOI: [10.1137/0108018](#).
- [18] Alexis Hocquenghem. "Codes correcteurs d'erreurs". In: *Chiffers* 2 (1959), pp. 147–156.
- [19] Raj Chandra Bose and Dwijendra K. Ray-Chaudhuri. "On a class of error correcting binary group codes". In: *Information and Control* 3.1 (1960), pp. 68–79. DOI: [10.1016/S0019-9958\(60\)90287-4](#).
- [20] Valerii Denisovich Goppa. "Codes on algebraic curves". In: *Dokl. Akad. Nauk SSSR* 259.1 (1981), pp. 1289–1290.
- [21] Valerii Denisovich Goppa. "Algebraico-geometric codes". In: *Izvestiya Rossiiskoi Akademii Nauk* 46.4 (1982), pp. 762–789.
- [22] Robert G. Gallager. "Low-density parity-check codes". In: *IRE Transactions on information theory* 1.1962 (8), pp. 21–28. DOI: [10.1109/TIT.1962.1057683](#).
- [23] Shrinivas Kudekar, Tom Richardson, and Rüdiger L. Urbanke. "Spatially coupled ensembles universally achieve capacity under belief propagation". In: *IEEE Transactions on Information Theory* 59.12 (2013), pp. 7761–7813. DOI: [10.1109/TIT.2013.2280915](#). arXiv: [1201.2999](#).
- [24] Eren Şaşoğlu, Emre Telatar, and Erdal Arikan. "Polarization for arbitrary discrete memoryless channels". In: *IEEE Information Theory Workshop (ITW)*. 2009, pp. 144–148. arXiv: [0908.0302 \[cs.IT\]](#).
- [25] Erdal Arikan and Emre Telatar. "On the Rate of Channel Polarization". In: *IEEE International Symposium on Information Theory*. 2009, pp. 1493–1495. DOI: [10.1109/ISIT.2009.5205856](#). arXiv: [0807.3806](#).
- [26] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. New York, NY, USA: Cambridge University Press, 2000. DOI: [10.1017/CBO9780511976667](#).
- [27] Christopher A. Fuchs and Jeroen van de Graaf. "Cryptographic distinguishability measures for quantum-mechanical states". In: *IEEE Transactions on Information Theory* 45.4 (1999), pp. 1216–1227. DOI: [10.1109/18.761271](#). arXiv: [quant-ph/9712042](#).
- [28] Benjamin Schumacher. "Quantum coding". In: *Physical Review A* 51 (4 Apr. 1995), pp. 2738–2747. DOI: [10.1103/PhysRevA.51.2738](#).
- [29] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. "Quantum state merging and negative information". In: *Communications in Mathematical Physics* 269.1 (2007), pp. 107–136. DOI: [10.1007/s00220-006-0118-x](#). arXiv: [quant-ph/0512247](#).
- [30] Peter W. Shor and John A. Smolin. *Quantum error-correcting codes need not completely reveal the error syndrome*. (1996). arXiv: [quant-ph/9604006](#).
- [31] David P. DiVincenzo, Peter W. Shor, and John A. Smolin. "Quantum-channel capacity of very noisy channels". In: *Physical Review A* 57.2 (1998), p. 830. DOI: [10.1103/PhysRevA.57.830](#). arXiv: [quant-ph/9706061](#).
- [32] Graeme Smith and John A. Smolin. "Degenerate quantum codes for Pauli channels." In: *Physical Review Letters* 98.3 (2007), p. 030501. DOI: [10.1103/PhysRevLett.98.030501](#). arXiv: [quant-ph/0604107](#).

-
- [33] Alfréd Rényi. “On measures of entropy and information”. In: *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.* 1961, pp. 547–561. URL: <http://projecteuclid.org/euclid.bsmsp/1200512181>.
- [34] S. Kullback. *Information theory and statistics*. Courier Corporation, 1997.
- [35] Tim Van Erven and Peter Harremoës. “Rényi divergence and Kullback-Leibler divergence”. In: *IEEE Transactions on Information Theory* 60.7 (2014), pp. 3797–3820. DOI: [10.1109/TIT.2014.2320500](https://doi.org/10.1109/TIT.2014.2320500). arXiv: [1206.2459](https://arxiv.org/abs/1206.2459).
- [36] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. “On quantum Rényi entropies: a new generalization and some properties”. In: *Journal of Mathematical Physics* 54.12, 122203 (2013). DOI: [10.1063/1.4838856](https://doi.org/10.1063/1.4838856). arXiv: [1306.3142](https://arxiv.org/abs/1306.3142).
- [37] Dénes Petz. “Quasi-entropies for finite quantum systems”. In: *Reports on Mathematical Physics* 23.1 (1986), pp. 57–65. DOI: [10.1016/0034-4877\(86\)90067-4](https://doi.org/10.1016/0034-4877(86)90067-4).
- [38] Milán Mosonyi and Fumio Hiai. “On the quantum Rényi relative entropies and related capacity formulas”. In: *IEEE Transactions on Information Theory* 57.4 (2011). DOI: [10.1109/TIT.2011.2110050](https://doi.org/10.1109/TIT.2011.2110050). arXiv: [0912.1286](https://arxiv.org/abs/0912.1286).
- [39] Mark M. Wilde, Andreas Winter, and Dong Yang. “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy”. In: *Communications in Mathematical Physics* 331.2 (2014), pp. 593–622. DOI: [10.1007/s00220-014-2122-x](https://doi.org/10.1007/s00220-014-2122-x). arXiv: [1306.1586](https://arxiv.org/abs/1306.1586).
- [40] Huzihiro Araki. “On an inequality of Lieb and Thirring”. In: *Letters in Mathematical Physics* 19.2 (1990), pp. 167–170. DOI: [10.1007/BF01045887](https://doi.org/10.1007/BF01045887).
- [41] Elliott H. Lieb and Walter Thirring. “Inequalities for the moments of the eigenvalues of the Schrödinger hamiltonian and their relation to Sobolev inequalities”. In: *Studies in Mathematical Physics*. Princeton University Press, 1976, pp. 269–303.
- [42] Salman Beigi. “Sandwiched Rényi divergence satisfies data processing inequality”. In: *Journal of Mathematical Physics* 54.12, 122202 (2013). DOI: [10.1063/1.4838855](https://doi.org/10.1063/1.4838855). arXiv: [1306.5920](https://arxiv.org/abs/1306.5920).
- [43] Hamza Fawzi and Omar Fawzi. “Defining quantum divergences via convex optimization”. In: *Quantum* 5 (2021), p. 387. DOI: [10.22331/q-2021-01-26-387](https://doi.org/10.22331/q-2021-01-26-387). arXiv: [2007.12576v2](https://arxiv.org/abs/2007.12576v2).
- [44] Marco Tomamichel, Mario Berta, and Masahito Hayashi. “Relating different quantum generalizations of the conditional Rényi entropy”. In: *Journal of Mathematical Physics* 55.8, 082206 (2014). DOI: [10.1063/1.4892761](https://doi.org/10.1063/1.4892761). arXiv: [1311.3887](https://arxiv.org/abs/1311.3887).
- [45] Renato Renner. “Security of quantum key distribution”. PhD thesis. ETH Zürich, 2005. DOI: [10.3929/ethz-a-005115027](https://doi.org/10.3929/ethz-a-005115027). arXiv: [quant-ph/0512258](https://arxiv.org/abs/quant-ph/0512258).
- [46] Igor Devetak and Peter W. Shor. “The capacity of a quantum channel for simultaneous transmission of classical and quantum information”. In: *Communications in Mathematical Physics* 256.2 (2005), pp. 287–303. DOI: [10.1007/s00220-005-1317-6](https://doi.org/10.1007/s00220-005-1317-6). arXiv: [quant-ph/0311131](https://arxiv.org/abs/quant-ph/0311131).
- [47] Daniel Gottesman. “Stabilizer codes and quantum error correction”. PhD thesis. California Institute of Technology, 1997. arXiv: [quant-ph/9705052](https://arxiv.org/abs/quant-ph/9705052).
- [48] A. R. Calderbank and Peter W. Shor. “Good quantum error-correcting codes exist”. In: *Physical Review A* 54.2 (1996). DOI: [10.1103/PhysRevA.54.1098](https://doi.org/10.1103/PhysRevA.54.1098). arXiv: [quant-ph/9512032](https://arxiv.org/abs/quant-ph/9512032).
-

- [49] Andrew Steane. “Multiple-particle interference and quantum error correction”. In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 452.1954 (1996). DOI: [10.1098/rspa.1996.0136](#). arXiv: [quant-ph/9601029](#).
- [50] Todd Brun, Igor Devetak, and Min-Hsiu Hsieh. “Correcting quantum errors with entanglement”. In: *science* 314.5718 (2006), pp. 436–439. DOI: [10.1126/science.1131563](#). arXiv: [quant-ph/0610092](#).
- [51] Todd Brun, Igor Devetak, and Min-Hsiu Hsieh. “General entanglement-assisted quantum error-correcting codes”. In: *Physical Review A* 76.6 (2007), p. 062313. DOI: [10.1103/PhysRevA.76.062313](#). arXiv: [0708.2142](#).
- [52] Mark M. Wilde and Todd Brun. “Optimal entanglement formulas for entanglement-assisted quantum coding”. In: *Physical Review A* 77.6 (2008), p. 064302. DOI: [10.1103/PhysRevA.77.064302](#). arXiv: [0804.1404](#).
- [53] Mark M. Wilde and Saikat Guha. “Polar Codes for Degradable Quantum Channels”. In: *IEEE Transactions on Information Theory* 59.7 (2013), pp. 4718–4729. DOI: [10.1109/TIT.2013.2250575](#). arXiv: [1109.5346](#).
- [54] Mark M. Wilde, Olivier Landon-Cardinal, and Patrick Hayden. “Towards efficient decoding of classical-quantum polar codes”. In: *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)*. Vol. 22. 2013, pp. 157–177. DOI: [10.4230/LIPIcs.TQC.2013.157](#). arXiv: [1302.0398](#).
- [55] Joseph M. Renes, David Sutter, Frédéric Dupuis, and Renato Renner. “Efficient quantum polar codes requiring no preshared entanglement”. In: *IEEE Transactions on Information Theory* 61.11 (Nov. 2015), pp. 6395–6414. DOI: [10.1109/TIT.2015.2468084](#). arXiv: [1307.1136](#).
- [56] Joseph M. Renes and Mark M. Wilde. “Polar Codes for Private and Quantum Communication Over Arbitrary Channels”. In: *IEEE Transactions on Information Theory* 60.6 (2014), pp. 3090–3103. DOI: [10.1109/TIT.2014.2314463](#). arXiv: [1212.2537](#).
- [57] Zak Webb. “The Clifford group forms a unitary 3-design”. In: *Quantum Information and Computation* 16 (2016), pp. 1379–1400. DOI: [10.26421/QIC16.15-16-8](#). arXiv: [1510.02769](#).
- [58] Alexander S. Holevo. “Complementary channels and the additivity problem”. In: *Theory of Probability and Its Applications* 51.1 (2007), pp. 92–100. DOI: [10.1137/S0040585X97982244](#). arXiv: [quant-ph/0509101](#).
- [59] Andreas Winter. “Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints”. In: *Communications in Mathematical Physics* 347.1 (2016), pp. 291–313. DOI: [10.1007/s00220-016-2609-8](#). arXiv: [1507.07775](#).
- [60] Frédéric Dupuis. “The decoupling approach to quantum information theory”. PhD thesis. Université de Montréal, 2009. arXiv: [1004.1641](#).
- [61] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. “Exact and approximate unitary 2-designs and their application to fidelity estimation”. In: *Physical review A* 80.1 (2009), p. 012304. DOI: [10.1103/PhysRevA.80.012304](#). arXiv: [quant-ph/0606161](#).
- [62] Olivia Di Matteo. *A short introduction to unitary 2-designs*. notes: [Unitary2Designs.pdf](#).

-
- [63] Joseph Emerson, Robert Alicki, and Karol Życzkowski. “Scalable noise estimation with random unitary operators”. In: *Journal of Optics B: Quantum and Semiclassical Optics* 7.10 (2005). DOI: [10.1088/1464-4266/7/10/021](https://doi.org/10.1088/1464-4266/7/10/021). arXiv: [quant-ph/0503243](https://arxiv.org/abs/quant-ph/0503243).
 - [64] S Hamed Hassani and Rüdiger Urbanke. “Universal polar codes”. In: *IEEE International Symposium on Information Theory*. 2014, pp. 1451–1455. DOI: [10.1109/ISIT.2014.6875073](https://doi.org/10.1109/ISIT.2014.6875073). arXiv: [1307.7223](https://arxiv.org/abs/1307.7223).
 - [65] Edwin Hewitt and Karl Stromberg. *Real and abstract analysis: a modern treatment of the theory of functions of a real variable*. Springer-Verlag, 2013.
 - [66] Wassily Hoeffding. “Probability inequalities for sums of bounded random variables”. In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30. DOI: [10.1080/01621459.1963.10500830](https://doi.org/10.1080/01621459.1963.10500830).
 - [67] Woomyoung Park and Alexander Barg. “Polar Codes for Q-Ary Channels, $q = 2^r$ ”. In: *IEEE Transactions on Information Theory* 59.2 (2012), pp. 955–969. DOI: [10.1109/TIT.2012.2219035](https://doi.org/10.1109/TIT.2012.2219035). arXiv: [1107.4965](https://arxiv.org/abs/1107.4965).
 - [68] Aria G Sahebi and S Sandeep Pradhan. “Multilevel Polarization of Polar Codes over Arbitrary Discrete Memoryless Channels”. In: *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2011, pp. 1718–1725. DOI: [10.1109/Allerton.2011.6120375](https://doi.org/10.1109/Allerton.2011.6120375). arXiv: [1107.1535](https://arxiv.org/abs/1107.1535).
 - [69] Peter Shor. “Algorithms for quantum computation: discrete logarithms and factoring”. In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 124–134. DOI: [10.1109/SFCS.1994.365700](https://doi.org/10.1109/SFCS.1994.365700). arXiv: [quant-ph/9508027](https://arxiv.org/abs/quant-ph/9508027).
 - [70] Lov Grover. “A fast quantum mechanical algorithm for database search”. In: *STOC '96: Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*. 1996, pp. 212–219. DOI: [10.1145/237814.237866](https://doi.org/10.1145/237814.237866). arXiv: [quant-ph/9605043](https://arxiv.org/abs/quant-ph/9605043).
 - [71] Ashley Montanaro. “Quantum algorithms: an overview”. In: *npj Quantum Information* 2.15023 (2016), pp. 1–8. DOI: [10.1038/npjqi.2015.23](https://doi.org/10.1038/npjqi.2015.23).
 - [72] Peter Shor. “Fault-tolerant Quantum Computation”. In: *Proceedings of 37th Conference on Foundations of Computer Science* (1996), pp. 56–65. DOI: [10.1109/SFCS.1996.548464](https://doi.org/10.1109/SFCS.1996.548464). arXiv: [quant-ph/9605011](https://arxiv.org/abs/quant-ph/9605011).
 - [73] Dorit Aharonov and Michael Ben-Or. “Fault-tolerant quantum computation with constant error rate”. In: *SIAM Journal on Computing* 38.4 (2008), pp. 1207–1282. DOI: [10.1137/S0097539799359385](https://doi.org/10.1137/S0097539799359385). arXiv: [quant-ph/9906129](https://arxiv.org/abs/quant-ph/9906129).
 - [74] Andrew M. Steane. “Efficient fault-tolerant quantum computing”. In: *Nature* 399 (1999), pp. 124–126. DOI: <https://doi.org/10.1038/20127>. arXiv: [quant-ph/9809054](https://arxiv.org/abs/quant-ph/9809054).
 - [75] John Preskill. “Fault-tolerant Quantum Computation”. In: *Introduction to quantum computation and information* (1998), pp. 213–269. DOI: [10.1142/9789812385253_0008](https://doi.org/10.1142/9789812385253_0008). arXiv: [quant-ph/9712048](https://arxiv.org/abs/quant-ph/9712048).
 - [76] Daniel Gottesman. *An introduction to quantum error correction and fault-tolerant quantum computation*. (2009). arXiv: [0904.2557](https://arxiv.org/abs/0904.2557).
 - [77] Andrew M. Steane. “Active Stabilization, Quantum Computation, and Quantum State Synthesis”. In: *Physical review letters* 78 (1997), p. 2252. DOI: <https://doi.org/10.1103/PhysRevLett.78.2252>. arXiv: [quant-ph/9611027](https://arxiv.org/abs/quant-ph/9611027).
-

- [78] Adam Holmes, Mohammad Reza Jokar, Ghasem Pasandi, Yongshan Ding, Mas-soud Pedram, and Frederic T Chong. “NISQ+: Boosting quantum computing power by approximating quantum error correction”. In: *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*. IEEE. 2020, pp. 556–569. DOI: [10.1109/ISCA45697.2020.00053](https://doi.org/10.1109/ISCA45697.2020.00053). arXiv: [2004.04794](https://arxiv.org/abs/2004.04794).
- [79] David J. C. MacKay, Graeme Mitchison, and Paul L. McFadden. “Sparse-graph codes for quantum error correction”. In: *IEEE Transactions on Information Theory* 50.10 (2004), pp. 2315–2330. DOI: [10.1109/TIT.2004.834737](https://doi.org/10.1109/TIT.2004.834737). arXiv: [quant-ph/0304161](https://arxiv.org/abs/quant-ph/0304161).
- [80] Alexei Yu. Kitaev. “Fault-tolerant quantum computation by anyons”. In: *Annals of Physics* 303.1 (2003), pp. 2–30. DOI: [10.1016/S0003-4916\(02\)00018-0](https://doi.org/10.1016/S0003-4916(02)00018-0). arXiv: [quant-ph/9707021.pdf](https://arxiv.org/abs/quant-ph/9707021.pdf).
- [81] Sergey B. Bravyi and Alexei Yu. Kitaev. *Quantum codes on a lattice with boundary*. (1998). arXiv: [quant-ph/9811052](https://arxiv.org/abs/quant-ph/9811052).
- [82] Héctor Bombin and Miguel A. Martin-Delgado. “Topological Quantum Distillation”. In: *Phys. Rev. Lett.* 97 (18 2006), p. 180501. DOI: [10.1103/PhysRevLett.97.180501](https://doi.org/10.1103/PhysRevLett.97.180501). arXiv: [quant-ph/0605138](https://arxiv.org/abs/quant-ph/0605138).
- [83] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. “Surface codes: Towards practical large-scale quantum computation”. In: *Phys. Rev. A* 86 (3 2012), p. 032324. DOI: [10.1103/PhysRevA.86.032324](https://doi.org/10.1103/PhysRevA.86.032324). arXiv: [1208.0928](https://arxiv.org/abs/1208.0928).
- [84] Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. “Topological quantum memory”. In: *Journal of Mathematical Physics* 43.9 (2002), pp. 4452–4505. DOI: [10.1063/1.1499754](https://doi.org/10.1063/1.1499754). arXiv: [quant-ph/0110143](https://arxiv.org/abs/quant-ph/0110143).
- [85] Austin G Fowler, Adam C Whiteside, Angus L McInnes, and Alimohammad Rab-bani. “Topological code autotune”. In: *Physical Review X* 2.4 (2012), p. 041003. DOI: [10.1103/PhysRevX.2.041003](https://doi.org/10.1103/PhysRevX.2.041003). arXiv: [1202.6111](https://arxiv.org/abs/1202.6111).
- [86] James Wootton. “A simple decoder for topological codes”. In: *Entropy* 17.4 (2015), pp. 1946–1957. DOI: [10.3390/e17041946](https://doi.org/10.3390/e17041946). arXiv: [1310.2393](https://arxiv.org/abs/1310.2393).
- [87] Nicolas Delfosse and Naomi H Nickerson. “Almost-linear time decoding algo-rithm for topological codes”. In: *Quantum Information and Computation* 8.10 (2019), pp. 300–311. arXiv: [1709.06218](https://arxiv.org/abs/1709.06218).
- [88] Savvas Varsamopoulos, Koen Bertels, and Carmen Garcia Almudever. “Compar-ing neural network based decoders for the surface code”. In: *IEEE Transactions on Computers* 69.2 (2019), pp. 300–311. DOI: [10.1109/TC.2019.2948612](https://doi.org/10.1109/TC.2019.2948612). arXiv: [1811.12456](https://arxiv.org/abs/1811.12456).
- [89] David Poulin and Yeojin Chung. “On the iterative decoding of sparse quantum codes”. In: (2008). arXiv: [0801.1241](https://arxiv.org/abs/0801.1241).
- [90] Zunaira Babar, Panagiotis Botsinis, Dimitrios Alanis, Soon Xin Ng, and Lajos Hanzo. “Fifteen years of quantum LDPC coding and improved decoding strate-gies”. In: *IEEE Access* 3 (2015), pp. 2492–2519. DOI: [10.1109/ACCESS.2015.2503267](https://doi.org/10.1109/ACCESS.2015.2503267).
- [91] Ye-Hua Liu and David Poulin. “Neural belief-propagation decoders for quantum error-correcting codes”. In: *Physical review letters* 122.20 (2019), p. 200501. DOI: [10.1103/PhysRevLett.122.200501](https://doi.org/10.1103/PhysRevLett.122.200501). arXiv: [1811.07835](https://arxiv.org/abs/1811.07835).

-
- [92] Anthony Leverrier, Jean-Pierre Tillich, and Gilles Zémor. “Quantum expander codes”. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE. 2015, pp. 810–824. DOI: [10.1109/FOCS.2015.55](#). arXiv: [1504.00822](#).
 - [93] Omar Fawzi, Antoine Grospellier, and Anthony Leverrier. “Efficient decoding of random errors for quantum expander codes”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018, pp. 521–534. DOI: [10.1145/3188745.3188886](#). arXiv: [1711.08351](#).
 - [94] Pavel Panteleev and Gleb Kalachev. *Degenerate quantum LDPC codes with good finite length performance*. (2019). arXiv: [1904.02703](#).
 - [95] Shai Evra, Tali Kaufman, and Gilles Zémor. *Decodable quantum LDPC codes beyond the \sqrt{n} distance barrier using high dimensional expanders*. (2020). arXiv: [2004.07935](#).
 - [96] Joschka Roffe, David R White, Simon Burton, and Earl T Campbell. “Decoding across the quantum LDPC code landscape”. In: (2020). arXiv: [2005.07016](#).
 - [97] Nikolas P Breuckmann and Vivien Londe. *Single-shot decoding of linear rate LDPC quantum codes with high performance*. (2020). arXiv: [2001.03568](#).
 - [98] Nicolas Delfosse and Matthew B Hastings. “Union-find decoders for homological product codes”. In: *Quantum* 5 (2021), p. 406. DOI: [10.22331/q-2021-03-10-406](#). arXiv: [2009.14226](#).
 - [99] Nicolas Delfosse, Vivien Londe, and Michael Beverland. “Toward a Union-Find decoder for quantum LDPC codes”. In: (2021). arXiv: [2103.08049](#).
 - [100] Omar Fawzi, Lucien Grouès, and Anthony Leverrier. “Linear programming decoder for hypergraph product quantum codes”. In: *IEEE Information theory workshop*. 2020. DOI: [10.1109/ITW46852.2021.9457611](#).
 - [101] Panos Aliferis, Daniel Gottesman, and John Preskill. “Quantum accuracy threshold for concatenated distance-3 codes”. In: *Quantum Information and Computation* 6 (2006), pp. 97–165. arXiv: [quant-ph/0504218](#).
 - [102] Panos Aliferis and Andrew W Cross. “Subsystem fault tolerance with the Bacon-Shor code”. In: *Physical review letters* 98.22 (2007), p. 220502. DOI: [10.1103/PhysRevLett.98.220502](#). arXiv: [quant-ph/0610063](#).
 - [103] Panos Aliferis, Daniel Gottesman, and John Preskill. “Accuracy threshold for post-selected quantum computation”. In: *Quantum Information and Computation* 8.3 (2008), pp. 181–244. arXiv: [quant-ph/0703264](#).
 - [104] Panos Aliferis and John Preskill. “Fibonacci scheme for fault-tolerant quantum computation”. In: *Physical Review A* 79.1 (2009), p. 012332. DOI: [10.1103/PhysRevA.79.012332](#). arXiv: [0809.5063](#).
 - [105] Omar Fawzi, Antoine Grospellier, and Anthony Leverrier. “Constant overhead quantum fault-tolerance with quantum expander codes”. In: *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2018, pp. 743–754. arXiv: [1808.03821](#).
 - [106] Andrew W Cross, David P DiVincenzo, and Barbara M Terhal. “A comparative code study for quantum fault-tolerance”. In: *Quantum Information and Computation* 9.7/8 (2009), pp. 0541–0572. arXiv: [0711.1556](#).
 - [107] Andrew J. Landahl, Jonas T. Anderson, and Patrick R. Rice. *Fault-tolerant quantum computing with color codes*. (2011). arXiv: [1108.5738](#).
 - [108] Panos Aliferis and John Preskill. “Fault-tolerant quantum computation against biased noise”. In: *Physical Review A* 78.5 (2008), p. 052331. DOI: [10.1103/PhysRevA.78.052331](#). arXiv: [0710.1301](#).
-

- [109] David K Tuckett, Stephen D Bartlett, Steven T Flammia, and Benjamin J Brown. “Fault-tolerant thresholds for the surface code in excess of 5% under biased noise”. In: *Physical review letters* 124.13 (2020), p. 130501. arXiv: [1907.02554](#).
- [110] Andrew M. Steane. *Fast fault-tolerant filtering of quantum codewords*. (2002). arXiv: [quant-ph/0202036](#).
- [111] Robert Raussendorf, Jim Harrington, and Kovid Goyal. “A fault-tolerant one-way quantum computer”. In: *Annals of physics* 321.9 (2006), pp. 2242–2270. DOI: [10.1016/j.aop.2006.01.012](#). arXiv: [quant-ph/0510135](#).
- [112] Bryan Eastin and Emanuel Knill. “Restrictions on transversal encoded quantum gate sets”. In: *Physical review letters* 102.11 (2009), p. 110502. DOI: [10.1103/PhysRevLett.102.110502](#). arXiv: [0811.4262](#).
- [113] Daniel Gottesman and Isaac L Chuang. “Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations”. In: *Nature* 402.6760 (1999), pp. 390–393. DOI: [10.1038/46503](#).
- [114] Sergey Bravyi and Alexei Kitaev. “Universal quantum computation with ideal Clifford gates and noisy ancillas”. In: *Physical Review A* 71.2 (2005), p. 022316. DOI: [10.1103/PhysRevA.71.022316](#). arXiv: [quant-ph/0403025](#).
- [115] Adam Paetznick and Ben W Reichardt. “Universal fault-tolerant quantum computation with only transversal gates and error correction”. In: *Physical review letters* 111.9 (2013), p. 090505. DOI: [10.1103/PhysRevLett.111.090505](#). arXiv: [1304.3709](#).
- [116] Tomas Jochym-O’Connor and Raymond Laflamme. “Using concatenated quantum codes for universal fault-tolerant quantum gates”. In: *Physical review letters* 112.1 (2014), p. 010505. DOI: [10.1103/PhysRevLett.112.010505](#). arXiv: [1309.3310](#).
- [117] Héctor Bombín and Miguel A. Martin-Delgado. “Quantum measurements and gates by code deformation”. In: *Journal of Physics A: Mathematical and Theoretical* 42.9 (2009), p. 095302. DOI: [10.1088/1751-8113/42/9/095302](#). arXiv: [0704.2540](#).
- [118] Clare Horsman, Austin G Fowler, Simon Devitt, and Rodney Van Meter. “Surface code quantum computing by lattice surgery”. In: *New Journal of Physics* 14.12 (2012), p. 123011. DOI: [10.1088/1367-2630/14/12/123011](#). arXiv: [1111.4022](#).
- [119] Emanuel Knill, Raymond Laflamme, and Wojciech Zurek. Threshold accuracy for quantum computation. (1996). arXiv: [quant-ph/9610011](#).
- [120] Anirudh Krishna and Jean-Pierre Tillich. Magic state distillation with punctured polar codes. (2018). arXiv: [1811.03112](#).
- [121] Ryuhei Mori and Toshiyuki Tanaka. “Performance and construction of polar codes on symmetric binary-input memoryless channels”. In: *2009 IEEE International symposium on information theory*. IEEE. 2009, pp. 1496–1500. DOI: [10.1109/ISIT.2009.5205857](#). arXiv: [0901.2207](#).

