Universidade Federal de Campina Grande

TELECOM ParisTech

Thesis

In partial fulfillment of the requirements for the degree of Doctor of Science from Universidade Federal de Campina Grande and TELECOM ParisTech

Specializing: Telecommunications

Zero-Error Capacity of Quantum Channels

Rex Antonio da Costa Medeiros

May 9th, 2008

Raimundo C. Silvério Freire, Dr., UFCG	President
Braulio Maia Junior, D.Sc., UFCG	Readers
Romain Alléaume, Dr., TELECOM ParisTech	
Carlile Campos Lavor, D.Sc., Unicamp	Examiners
Valdemar Cardoso da Rocha, Ph.D., UFPE	
Francisco Marcos de Assis, D.Sc., UFCG	Thesis advisors
Gérard Cohen, Dr., TELECOM ParisTech	

Capacidade Erro-Zero de Canais Quânticos

Rex Antonio da Costa Medeiros

Tese submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, Brasil, e de TELECOM ParisTech, França, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no domínio da Engenharia Elétrica.

Área de Concentração: Processamento de Informação – Comunicações

Francisco Marcos de Assis, D.Sc., UFCG
Gérard Cohen, Dr., TELECOM ParisTech
Orientadores

Campina Grande, Paraíba, Brasil

“Ainda que eu andasse pelo vale da sombra da morte,
não temeria mal algum,
porque tu estás comigo;
a tua vara e o teu cajado me consolam.”

Salmos 23:4

Dedicatória

Aos meus pais, irmã, esposa e ao meu filho Mateus que está a caminho, fonte da minha inspiração, do meu desejo de vencer.

Agradecimentos

A Deus, pela minha existência, sempre ...

À minha esposa, por se fazer presente em todos os momentos desta minha trajetória, acreditando que eu seria capaz, apoiando-me nas dificuldades, bem como torcendo e vibrando com cada obstáculo superado. Agradeço por ter aberto mão dos seus projetos em favor dos meus.

Tenho uma felicidade imensa de agradecer aos meus orientadores, prof. Francisco M. de Assis e prof. Gérard Cohen, que sempre acreditaram no meu potencial. Particularmente, agradeço ao prof. Assis pelo companheirismo e cumplicidade desde o meu mestrado, e ao prof. Cohen pela recepção impecável que tive na TELECOM ParisTech.

Gostaria de expressar minha profunda gratidão a dois professores que colaboraram decisivamente para o sucesso deste trabalho: Prof. Romain Alléaume e Prof. Huges Randriam, da TELECOM ParisTech. Espero que nossa parceria continue dando bons frutos por longos e longos tempos. Agradeço em especial ao Prof. Raimundo Freire pela sua ajuda indispensável na elaboração da proposta Alban, bem como pela sua disponibilidade e atenção para comigo.

Aos meus pais, irmã, a quem eu dedico este trabalho, pois eles são partes de mim, acreditam e me apoiam incondicionalmente. Aos sogros, amigos e familiares, pelas palavras de incentivo e credibilidade no meu profissionalismo.

A todo o pessoal administrativo da UFCG/UAEE/COPELE, de TELECOM ParisTech e do Alban Office, especialmente à Ângela, Florence Besnard, Sophie Beranger e Salete Figueiredo, pela eficiência no trato das questões burocráticas.

Por fim aos colegas que tanto partilharam minhas desesperanças como também as alegrias ao longo da tese: Edmar, Alfranque, Felipe, Protásio, Bruno, Júnior, Myriam, Xiaoyun, Sandrine, Rodrigo, Judson, Gilson ...

Abstract

In this thesis, we generalise Shannon's zero-error capacity of discrete memoryless channels to quantum channels. We propose a new kind of capacity for transmitting classical information through a quantum channel. The *quantum zero-error capacity* (QZEC) is defined as being the maximum amount of classical information per channel use that can be sent over a noisy quantum channel, with the restriction that the probability of error *must be* equal to zero. The communication protocol restricts codewords to tensor products of input quantum states, whereas collective measurements can be performed between several channel outputs. Hence, our communication protocol is similar to the Holevo-Schumacher-Westmoreland protocol. We reformulate the problem of finding the QZEC in terms of graph theory. This equivalent definition allows us to demonstrate some properties of ensembles of quantum states and measurements attaining the QZEC. We show that the capacity of a d-dimensional quantum channel can always be achieved by using an ensemble of at most d pure quantum states, and collective von Neumann measurements are necessary and sufficient to attain the channel capacity. We discuss whether the QZEC is a non-trivial generalisation of the classical zero-error capacity. By non-trivial we mean that there exist quantum channels requiring two or more channel uses in order to reach the capacity, and the capacity can only be attained by using ensembles of non-orthogonal quantum states at the channel input. We also calculate the QZEC of some quantum channels. We show that finding the QZEC of classical-quantum channels is a purely classical problem. In particular, we exhibit a quantum channel for which we claim the QZEC can only be reached by a set of non-orthogonal states. If the conjecture holds, it is possible to give an exact solution for the capacity, and construct an error-free quantum block code reaching the capacity. Finally, we demonstrate that the QZEC is upper bounded by the Holevo-Schumacher-Westmoreland capacity.

Resumo

Nesta tese, a capacidade erro-zero de canais discretos sem memória é generalizada para canais quânticos. Uma nova capacidade para a transmissão de informação clássica através de canais quânticos é proposta. A capacidade erro-zero de canais quânticos (CEZQ) é definida como sendo a máxima quantidade de informação por uso do canal que pode ser enviada através de um canal quântico ruidoso, considerando uma probabilidade de erro igual a zero. O protocolo de comunicação restringe palavras-código a produtos tensoriais de estados quânticos de entrada, enquanto que medições coletivas entre várias saídas do canal são permitidas. Portanto, o protocolo empregado é similar ao protocolo de Holevo-Schumacher-Westmoreland. O problema de encontrar a CEZQ é reformulado usando elementos da teoria de grafos. Esta definição equivalente é usada para demonstrar propriedades de famílias de estados quânticos e medições que atingem a CEZQ. É mostrado que a capacidade de um canal quântico num espaço de Hilbert de dimensão d pode sempre ser alcançada usando famílias compostas de, no máximo, d estados puros. Com relação às medições, demonstra-se que medições coletivas de von Neumann são necessárias e suficientes para alcançar a capacidade. É discutido se a CEZQ é uma generalização não trivial da capacidade erro-zero clássica. O termo não trivial refere-se a existência de canais quânticos para os quais a CEZQ só pode ser alcançada através de famílias de estados quânticos não-ortogonais e usando códigos de comprimento maior ou igual a dois. É investigada a CEZQ de alguns canais quânticos. É mostrado que o problema de calcular a CEZQ de canais clássicos-quânticos é puramente clássico. Em particular, é exibido um canal quântico para o qual conjectura-se que a CEZQ só pode ser alcançada usando uma família de estados quânticos não-ortogonais. Se a conjectura é verdadeira, é possível calcular o valor exato da capacidade e construir um código de bloco quântico que alcança a capacidade. Finalmente, é demonstrado que a CEZQ é limitada superiormente pela capacidade de Holevo-Schumacher-Westmoreland.

Contents

Acronyms **xix**

Notation **xxi**

1 Resumo detalhado em Português **1**

1.1 Introdução . . . 1

1.1.1 Transmissão de informação clássica através de canais quânticos . . . 1

1.1.2 Capacidade erro-zero de canais clássicos . . . 2

1.1.3 Organização da tese . . . 3

1.2 Fundamentos da mecânica quântica . . . 3

1.2.1 Postulados da mecânica quântica . . . 3

1.2.2 O operador de densidade . . . 6

1.3 Capacidades de canais quânticos . . . 6

1.3.1 Entropia de von Neumann . . . 6

1.3.2 Canais quânticos . . . 7

1.3.3 Capacidades clássicas de canais quânticos . . . 8

1.4 Teoria da informação de erro-zero . . . 11

1.4.1 Capacidade ordinária de canais clássicos . . . 11

1.4.2 A capacidade erro-zero . . . 12

1.4.3 Função teta de Lovász . . . 14

1.5 Capacidade erro-zero de canais quânticos . . . 15

1.5.1 Capacidade erro-zero quântica . . . 15

1.5.2 Estados quânticos que atingem a CEZQ . . . 17

1.5.3 Medições que alcançam a CEZQ . . . 18

1.5.4 Exemplos . . . 18

1.5.5 Capacidade erro-zero quântica e a capacidade HSW . . . 21

1.6 Conclusões e perspectivas . . . 21

2 Introduction **23**

2.1 Classical information over quantum channels 23

2.2 Zero-error capacity of classical channels 24

2.3 Thesis outline 26

3 Fundamentals of Quantum Mechanics **29**

3.1 Introduction 29

3.2 Linear algebra and Hilbert spaces 29

3.2.1 Inner product 30

3.2.2 Linear operators 33

3.2.3 Pauli operators 34

3.2.4 Eigenvectors and eigenvalues 34

3.2.5 Hermitians and unitary operators 34

3.2.6 Tensor products 35

3.3 Quantum mechanics postulates 37

3.3.1 State space 37

3.3.2 Evolution 37

3.3.3 Measurements 38

3.3.4 Composite quantum systems 40

3.4 The density operator 42

3.4.1 Quantum mechanics postulates and density operators 43

3.5 Conclusions 44

4 Quantum channel capacities **45**

4.1 Introduction 45

4.2 Von Neumann entropy and quantum channels 45

4.2.1 The von Neumann entropy 45

4.2.2 Quantum channels 47

4.3 Accessible information and the Holevo bound 49

4.4 The Holevo-Schumacher-Westmoreland theorem 50

4.5 The adaptive capacity 52

4.6 Entanglement-assisted capacity 52

4.7 Conclusions 54

5 Zero-error information theory **55**

5.1 Ordinary capacity of classical channels 55

5.2 The zero-error capacity 58

5.2.1 The adjacency-reducing mapping 61

5.2.2 Relation with graph theory 62
5.3 Lovász theta function 66
5.4 Channels with complete feedback and list codes 69
5.5 The sum and product of channels 71
5.6 Conclusions 72

6 Zero-error capacity of quantum channels 73
6.1 Introduction 73
6.2 Quantum zero-error capacity 75
6.2.1 A graph-theoretic approach 79
6.3 Quantum states achieving the QZEC 80
6.3.1 The cardinality of the set $\mathcal{S}$ achieving the QZEC 83
6.4 Measurements reaching the capacity 87
6.5 Examples 88
6.5.1 Bit flip channel 88
6.5.2 Depolarizing channel 89
6.5.3 Zero-error capacity of classical-quantum channels 89
6.5.4 A particular classical-quantum channel 91
6.5.5 Non-orthogonal states attaining the QZEC 93
6.6 Zero-error capacity and HSW capacity 95
6.7 Conclusions 97
6.A Matlab m-file 98

7 Conclusions and Perspectives 103
7.1 Conclusions 103
7.2 Perspectives 104
7.2.1 A generalisation of the Lovász's theta function 104
7.2.2 Variations in the communication protocol 104
7.2.3 Decoherence-free subspaces and noiseless subsystems 105
7.2.4 Graph states 105

Acknowledgements 107

List of publications 109

Bibliography 111

Index 117

List of Figures

3.1 A POVM apparatus. 41

5.1 A classical communication system. 56

5.2 A binary erasure channel (BEC) with erasure probability p. 57

5.3 The G_5 channel. 57

5.4 Two distinguishable sequences $\mathbf{x}'$ and $\mathbf{x}''$. 59

5.5 Some discrete memoryless channels. 60

5.6 Characteristic graphs G of discrete memoryless channels in Figure 5.5. . . . 63

5.7 Adjacency graphs of DMC of Figure 5.5. 65

5.8 Graphs that can be covered by a number of cliques. 66

5.9 A spherical triangle delimited by the vectors $\mathbf{v}_1$, $\mathbf{v}_3$ and the handle $\mathbf{c}$. . . . 68

5.10 A discrete memoryless channel with feedback. 70

6.1 General representation of a quantum zero-error communication system. . . 75

6.2 Two distinguishable tensor product sequences of length n. 78

6.3 An extended-by-clonning graph. 84

6.4 Characteristic graph of $\mathcal{S}$ 92

6.5 Two characteristic graphs corresponding to ensembles $\mathcal{S}$ and $\mathcal{S}'$. 94

Acronyms

BEC	Binary Erasure Channel. 56
c-q	classical-quantum. 89
DFS	Decoherence-Free Subspaces. 105
DMC	Discrete Memoryless Channel. 25
EbC	Extended-by-cloning graph. 83
EPR	Einstein, Podolsky and Rosen. 41
HSW	Holevo-Schumacher-Westmoreland. 24
NS	Noiseless Subsystems. 105
POVM	Positive Operator-Valued Measurement. 74, 75
q-c	quantum-classical. 89
QZEC	Quantum zero-error capacity. 76
w.l.o.g	Without loss of generality. 81

Notation

$\|v\rangle\|w\rangle$	Tensor product between $\|v\rangle$ and $\|w\rangle$. 36
$(\mathcal{X}, p(y\|x), \mathcal{Y})$	DMC with input alphabet $\mathcal{S}$, output alphabet $\mathcal{Y}$ and transition probabilities $p(y\|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. 57
A_{ij}	Elements of an adjacency matrix. 64
$C^{(0)}(\mathcal{E})$	Zero-error capacity of the quantum channel $\mathcal{E}$. 76
C_0	Zero-erro capacity of DMC. 60
$C_E(\mathcal{E})$	Entanglement-assisted capacity of the quantum channel $\mathcal{E}$. 26, 53
$C_{1,1}(\mathcal{E})$	One-shot capacity of the quantum channel $\mathcal{E}$. 26, 50
$C_{1,A}(\mathcal{E})$	Adaptive capacity of the quantum channel $\mathcal{E}$. 26
$C_{1,\infty}(\mathcal{E})$	Holevo-Schumacher-Westmoreland capacity of the quantum channel $\mathcal{E}$. 24, 26
C	Shannon ordinary capacity of a DMC. 58
H_p	Binary Shannon entropy. 49, 56
$I(X;Y)$	Mutual information between X and Y. 56
I_{acc}	Accessible information. 49
$N(i)$	The vertex set of neighbors of the vertex i. 83
$R_{\mathcal{S}}$	maximum information transmission rate using zero-error quantum codes with alphabet $\mathcal{S}$. 90
R	Rate of a block code. 58
X	Random variable X. 56
$\mathcal{E}(\rho)$	Quantum channel output for an input ρ. 47
$\chi(G)$	Chromatic number of the graph G. 63
$\langle v\|w\rangle$	Inner product between $\|v\rangle$ and $\|w\rangle$. 30
$\mathbf{x}$	A sequence of input symbols for a DMC. 59
$\mathcal{P}$	A set of POVM elements. 87
$\mathcal{S}$	A subset of quantum input states. 74
$\mathcal{X}$	Input alphabet of a DMC. 55
$\mathcal{Y}$	Output alphabet of a DMC. 55

$\omega(G)$	Clique number of the graph G. 62
$\mathbb{1}_d$	Identity operator of the d-dimensional Hilbert space. 33, 89
$\|w\rangle\langle v\|$	Outer product between $\|w\rangle$ and $\|v\rangle$. 33
$\|\psi\rangle^{\otimes n}$	n- tensor product of $\|\psi\rangle$. 37
$\|v\rangle$	A pure quantum state. 29
$\rho_i \perp_{\mathcal{E}} \rho_j$	ρ_i is non-adjacent to ρ_j. 77
ρ_i	Density matrix. 23
supp ρ	Support of the state ρ: the Hilbert space spanned by eigenvectors of ρ with nonzero eigenvalues. 91
$\theta(G)$	Lovász theta function of the graph G. 67
tr $[\rho]$	Trace of the density operator ρ. 42
$\{p_i, \rho_i\}$	Ensemble of quantum states. 23
$p(y\|x)$	Conditional probability. 55

Chapter 1

Resumo detalhado em Português

1.1 Introdução

1.1.1 Transmissão de informação clássica através de canais quânticos

Uma das problemáticas mais estudadas em teoria da informação quântica é o conceito de capacidade de canais quânticos [1, 2]. De forma geral, a capacidade de um canal é definida como sendo o supremo das taxas alcançáveis, i.e., o supremo das taxas em que a informação pode ser transmitida confiavelmente através do canal.

A mecânica quântica provê diversos recursos que permitem definir capacidade de canais quânticos de várias maneiras [1, 2]. Para um canal quântico dado, a capacidade pode assumir diferentes valores dependendo: (a) do tipo de informação a ser transmitida – clássica ou quântica; (b) recursos externos, como entrelaçamento ou realimentação; e (c) do protocolo de comunicação. O protocolo de comunicação determina os procedimentos de codificação, medição e decodificação dos estados quânticos na saída do canal.

Nesta tese, serão consideradas capacidades de canais quânticos sem memória para a transmissão de informação clássica. De acordo com o protocolo, as capacidades podem ser agrupadas em três categorias:

1. palavras-código são restritas a produtos tensoriais e medições são feitas individualmente na saída do canal [3, 4, 5, 6];

2. palavras-código são restritas a produtos tensoriais, enquanto que medições entrelaçadas entre várias saídas do canal são permitidas [7, 8, 9, 10];

3. são permitidas palavras-código entrelaçadas, como também medições coletivas na saída do canal [11].

Exemplos de capacidades que empregam o protocolo 1 são a capacidade *one-shot* [3, 4, 5] e a capacidade adaptativa de Shor [6]. A principal capacidade que emprega o protocolo 2 é a capacidade de Holevo-Schumacher-Westmoreland (HSW) [7, 8], que é considerada uma generalização da capacidade ordinária de Shannon.

As capacidades que empregam o protocolo 3 estão diretamente conectadas a um dos problemas em aberto mais importantes da teoria da informação quântica: a conjectura de Holevo [7]. Esta conjectura afirma que a utilização de estados entrelaçados entre vários usos do canal não aumenta a capacidade de canais quânticos sem memória. Entretanto, é sabido que palavras-código entrelaçadas podem aumentar a capacidade HSW de canais quânticos com memória [11].

1.1.2 Capacidade erro-zero de canais clássicos

Em 1956, oito anos após seu primeiro trabalho introduzindo a teoria da informação e a capacidade de canais, Shannon [12] demonstrou que era possível transmitir informação *sem erro* através de um canal discreto sem memória (DSM), ao invés de permitir uma probabilidade de erro assintoticamente pequena [13]. A *capacidade erro-zero* foi definida como sendo o supremo das taxas em que informação pode ser transmitida através de um canal DSM com probabilidade de erro igual a zero.

No artigo original, Shannon sugeriu que a capacidade erro-zero poderia ser descrita usando elementos da teoria de grafos. Ao associar um grafo com um canal DSM, ele introduziu uma nova quantidade, a capacidade de Shannon de um grafo [14, 15, 16]. Diferentemente da capacidade ordinária, calcular a capacidade erro-zero é um problema combinatorial. Devido a sua natureza restritiva − uma probabilidade de erro igual a zero é imposta, a teoria da informação de erro-zero é freqüentemente desconhecida dos pesquisadores em teoria da informação. Entretanto, seus métodos possuem importantes aplicações em combinatória e teoria de grafos.

Esta tese propõe uma generalização da capacidade erro-zero para canais quânticos. Inicialmente, é definido um código quântico de erro-zero, como também os procedimentos de codificação e decodificação. Então, a capacidade erro-zero quântica é definida como sendo o supremo das taxas em que informação clássica pode ser transmitida *sem erro* através de um canal quântico sem memória. O problema de encontrar a capacidade erro-zero quântica é reformulado em termos da teoria de grafos. São investigadas propriedades de estados quânticos e medições que atingem a capacidade erro-zero quântica. Através de um exemplo, é conjecturado que a capacidade erro-zero quântica é uma generalização não-trivial da capacidade erro-zero clássica. Por último, é mostrado que a capacidade HSW é um limitante superior da capacidade erro-zero quântica.

1.1.3 Organização da tese

As contribuições são apresentadas no Capítulo 6. Leitores familiarizados com a teoria da informação quântica e a teoria da informação de erro-zero clássica podem ler diretamente o Capítulo 6. Esta tese está organizada como segue:

Os Capítulos 3 e 4 contêm conceitos de informação quântica relacionados à tese. A Seção 3.2 objetiva introduzir a notação de Dirac, ao mesmo tempo que discute conceitos importantes em informação quântica, como operadores unitários e produtos tensoriais. Os quatro postulados da mecânica quântica são apresentados na Seção 3.3, seguidos de uma discussão sobre o formalismo dos operadores de densidade. Uma breve revisão das capacidades clássicas de canais quânticos é dada no Capítulo 4. O Capítulo 5 traz um resumo das principais definições e resultados da teoria da informação de erro-zero clássica. A Seção 5.2 introduz a capacidade erro-zero. Uma abordagem baseada na teoria de grafos é discutida na Seção 5.2.2. Na Seção 5.3 é definida a função teta de Lovász, que é usada para calcular a capacidade erro-zero do pentágono. As Seções 5.4 e 5.5 ilustram o quão diferente é o comportamento da capacidade erro-zero face à capacidade ordinária.

A capacidade erro-zero quântica (CEZQ) é introduzida no Capítulo 6. Na Seção 6.2 é definido um código de erro-zero quântico, bem como a CEZQ. Uma definição equivalente para a CEZQ em termos da teoria de grafos é apresentada na Seção 6.2.1. A Seção 6.3 é dedicada ao estudo de estados quânticos e medições que atingem a capacidade. A CEZQ de alguns canais quânticos é calculada na Seção 6.5. É mostrado um exemplo de um canal quântico em que conjectura-se que a capacidade erro-zero só possa ser alcançada usando uma família de estados quânticos não-ortogonais. Finalmente, a Seção 6.6 apresenta um limitante superior para a CEZQ: a capacidade de Holevo-Schumacher-Westmoreland [7, 8].

No Capítulo 7 é feito um resumo das contribuições e são dadas algumas direções para trabalhos futuros.

1.2 Fundamentos da mecânica quântica

Esta seção introduz a mecânica quântica de forma breve e objetiva. Uma abordagem mais detalhada pode ser encontrada em livros específicos [17, 2].

1.2.1 Postulados da mecânica quântica

Os postulados da mecânica quântica são discutidos brevemente nas seções seguintes.

Espaço de estados

Postulado 1 *Associado a todo sistema quântico está um espaço vetorial complexo com produto interno, i.e., um espaço de Hilbert, chamado de espaço de estado do sistema quântico. O estado do sistema quântico é completamente descrito pelo vetor de estado, que é um vetor unitário pertencente ao espaço de estado do sistema.*

O sistema quântico mais simples é o *qubit*, que é uma referência a *bit quântico*. O qubit pertence ao espaço de estado de dimensão dois. Portanto, qualquer qubit pode ser escrito como

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{1.1}$$

em que a, b são números complexos. Uma das propriedades mais interessantes dos sistemas quânticos é que o estado $|0\rangle$ pode coexistir com o estado $|1\rangle$ num estado de superposição: $|\psi\rangle = a|0\rangle + b|1\rangle$.

Evolução

Postulado 2 *A evolução de um sistema quântico isolado é descrita por transformações unitárias. O estado do sistema $|\psi_1\rangle$ no tempo t_1 está relacionado com $|\psi_2\rangle$, que é o estado do sistema no tempo t_2, por meio de um operador unitário U, que depende somente dos tempos t_1 e t_2,*

$$|\psi_2\rangle = U|\psi_1\rangle. \tag{1.2}$$

Na maioria dos textos sobre mecânica quântica, a evolução é descrita por uma equação diferencial

$$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle, \tag{1.3}$$

em que $\hbar$ é chamada de constante de Planck e H é um operador Hermitiano do sistema quântico fechado, conhecido como Hamiltoniano do sistema. A equação acima é creditada ao físico austríaco Erwin Schrödinger.

Medições

Quando sistemas quânticos são expostos a um ambiente externo, sua evolução pode não mais ser unitária. O postulado seguinte descreve o comportamento de sistemas quânticos quando sujeitos à medições.

Postulado 3 *As medições em sistemas quânticos são descritas por um conjunto de operadores de medição* $\{M_m\}$*, os quais atuam no espaço de estado do sistema medido. Se o estado do sistema quântico antes da medição é* $|\psi\rangle$*, então a probabilidade de se obter uma saída* m *é dada por*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle. \tag{1.4}$$

O estado do sistema após a medição será

$$|\psi'\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{1.5}$$

Como a soma das probabilidades deve ser igual a um, os operadores de medição devem satisfazer a equação de completude

$$\sum_m M_m^\dagger M_m = \mathbb{1}. \tag{1.6}$$

O postulado acima descreve medições quânticas de forma mais geral. No entanto, existem dois casos particulares que são de interesse para esta tese, as medições projetivas e as medições POVM (*Positive Operator-Valued Measurements*).

As medições projetivas, ou medições de von Neumann, são descritas por um conjunto de projetores $\{P_m\}$, satisfazendo $\sum_m P_m = \mathbb{1}$ e $P_i P_j = \delta_{ij} P_i$. Ao se medir o estado $|\psi\rangle$, a probabilidade de se obter a saída m é dada por $p(m) = \langle\psi|P_m|\psi\rangle$. Dado que m ocorre, o estado do sistema após a medição será $|\psi'\rangle = \frac{P_m|\psi\rangle}{\sqrt{p(m)}}$.

As medições POVM são descritas por operadores de medição tais que $E_m \equiv M_m^\dagger M_m$ (geralmente não se tem acesso aos operadores M_m). A probabilidade de obter a saída m dado que o estado $|\psi\rangle$ é medido é dada por $p(m) = \langle\psi|E_m|\psi\rangle$. O conjunto $\{E_m\}$ é comumente chamado de POVM. Note que, no caso das medições POVM, não é possível escrever o estado de saída em função do estado original. Entretanto, na maioria das aplicações em teoria da informação quântica, o estado do sistema resultante não é importante, e sim as probabilidades associadas a cada um deles.

Sistemas quânticos compostos

Diversos sistemas quânticos podem interagir para formar sistemas compostos. O postulado seguinte descreve o espaço de estado de sistemas compostos.

Postulado 4 *O espaço de estado de um sistema quântico composto é o produto tensorial dos espaços de estado dos sistemas físicos individuais. Adicionalmente, se* n *sistemas são preparados cada um no estado* $|\psi_i\rangle$*, então o estado do sistema global é* $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.

As notações seguintes são usadas para representar sistemas compostos: $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle \equiv |\psi_1\rangle|\psi_2\rangle \ldots |\psi_n\rangle \equiv |\psi_1\psi_2 \ldots \psi_n\rangle$.

1.2.2 O operador de densidade

O estado de um sistema quântico é dito ser *puro* quando pode ser representado por um vetor unitário num espaço de Hilbert. No entanto, existem situações em que o sistema quântico em questão pode estar em qualquer um dos estados puros $|\psi_1\rangle, |\psi_2\rangle, \ldots,$ com probabilidades $p_1, p_2, \ldots$. O formalismo usado para lidar com esta situação é o operador de densidade.

Definição 1 (Operador de densidade [2]) *Considere que um sistema quântico está num estado $|\psi_i\rangle$ com probabilidade p_i. O operador de densidade que descreve o sistema é definido como sendo*

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1.7}$$

Neste caso, o sistema é dito estar num estado *misto*. O operador de densidade é também chamado de matriz de densidade do sistema. Operadores de densidade são matrizes bem caracterizadas: possuem traço igual a um, $\mathrm{tr}\,[\rho] = 1$, e também são operadores positivos. Claramente, a matriz de densidade de um sistema puro $|\psi\rangle$ é dada por $\rho = |\psi\rangle\langle\psi|$. Ainda, dada uma matriz de densidade ρ, o sistema está num estado puro se e somente se $\mathrm{tr}\,[\rho^2] = 1$. Caso contrário, se $\mathrm{tr}\,[\rho^2] < 1$, o sistema está num estado misto.

O formalismo de vetores e de matrizes de densidade são equivalentes. Portanto, os postulados da mecânica quântica podem ser enunciados em termos de operadores de densidade.

1.3 Capacidades de canais quânticos

Será feito nesta seção um resumo das principais capacidades canais quânticos possuem para a transmissão de informação clássica. Antes, porém, é dada uma definição da entropia de von Neumann e de canais quânticos. É importante salientar que todas as capacidades discutidas nesta seção permitem uma probabilidade de erro de decodificação assintoticamente nula, ou seja, embora pequena ela é diferente de zero.

1.3.1 Entropia de von Neumann

A entropia de von Neumann entropy [2, pp. 510] é uma generalização da entropia de Shannon para estados quânticos. A entropia de von Neumann de um estado ρ é definida como sendo

$$S(\rho) \equiv -\mathrm{tr}\left[\rho \log \rho\right], \tag{1.8}$$

em que o logaritmo é tomado na base 2. Num espaço de Hilbert de dimensão d, o máximo valor da entropia é $\log d$, correspondente ao estado $\rho = \mathbb{1}_d/d$, que é chamado de completamente despolarizado. A entropia relativa é definida de maneira análoga à entropia de Shannon,

$$S(\rho||\sigma) \equiv \mathrm{tr}\left[\rho \log \rho\right] - \mathrm{tr}\left[\rho \log \sigma\right]. \tag{1.9}$$

Como no caso clássico, a entropia relativa é não negativa, $S(\rho||\sigma) \geq 0$.

A entropia de von Neumann possui algumas propriedades interessantes, dentre elas: (1) a entropia é não negativa e zero se e somente se ρ é um estado puro; (2) se um sistema composto AB está num estado puro, então $S(A) = S(B)$; e (3) suponha que p_i são probabilidades e ρ_i possuam seus suportes em subespaços ortogonais. Então,

$$S\left(\sum_i p_i \rho_i\right) = H(p) + \sum_i p_i S(\rho_i). \tag{1.10}$$

Por analogia à entropia de Shannon, define-se as entropias de von Neumann conjunta e condicional, como também a informação mútua relacionada a sistemas compostos. A entropia conjunta $S(A,B)$ de um sistema composto AB é definida por $S(A,B) = -\mathrm{tr}\left[\rho^{AB} \log \rho^{AB}\right]$, em que ρ^{AB} é o operador de densidade do sistema AB. A entropia condicional e a informação mútua são definidas respectivamente como

$$S(A|B) \equiv S(A,B) - S(B), \tag{1.11}$$

$$S(A:B) \equiv S(A) + S(B) - S(A,B) \tag{1.12}$$

$$= S(A) - S(A|B) = S(B) - S(B|A). \tag{1.13}$$

Um resultado bastante útil é que a entropia de von Neumann é subaditiva [2, pp.515]: $S(A,B) \leq S(A) + S(B)$, com igualdade se e somente se $\rho_{AB} = \rho_A \otimes \rho_B$. Outras propriedades da entropia de von Neumann podem ser encontradas em Nielsen e Chuang [2].

1.3.2 Canais quânticos

Suponha que um sistema quântico ρ inicialmente fechado interaja com um sistema aberto, chamado de *ambiente*. Suponha ainda que, após a interação, o sistema volte ao seu estado fechado. Em geral, o estado final do sistema, denotado por $\mathcal{E}(\rho)$, não pode ser relacionado com o estado ρ por meio de uma transformação unitária. O formalismo usado para lidar com esta situação é conhecido como operação quântica, que é um mapeamento do conjunto de operadores do espaço de estado de entrada para operadores do espaço de estado de saída com as propriedades seguintes [2, pp. 367]:

1. tr $[\mathcal{E}(\rho)]$ é a probabilidade que o processo representado por $\mathcal{E}$ ocorra, dado que ρ é o estado inicial. Assim, $0 \leq \text{tr}\,[\mathcal{E}(\rho)] \leq 1$ para qualquer estado ρ.

2. $\mathcal{E}$ é um mapeamento linear e convexo no conjunto dos operadores de densidade, i.e., para probabilidades p_i,
$$\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i). \tag{1.14}$$

3. $\mathcal{E}$ é um mapeamento completamente positivo, de forma que $\mathcal{E}(\rho)$ seja positivo para qualquer operador positivo ρ.

A prova do teorema abaixo pode ser encontrada em Nielsen e Chuang [2, pp. 368].

Teorema 1 *Um mapeamento $\mathcal{E}$ satisfaz as propriedades 1, 2 e 3 se e somente se*
$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger \tag{1.15}$$
para algum conjunto de operadores $\{E_i\}$ tal que $\sum_i E_i^\dagger E_i \leq \mathbb{1}$.

Canais quânticos são modelados por operações quânticas que preservam o traço dos operadores de densidade. Ou seja, canais quânticos são operações quânticas lineares, completamente positivas e que preservam o traço. Neste caso, a restrição imposta aos operadores $\{E_i\}$ é $\sum_i E_i^\dagger E_i = \mathbb{1}$. Canais quânticos são definidos para um estado de entrada ρ que é uma matriz de densidade. No caso em que o estado de entrada é puro $|\psi\rangle$, basta representá-lo usando o formalismo dos operadores de densidade, $\rho = |\psi\rangle\langle\psi|$.

1.3.3 Capacidades clássicas de canais quânticos

A capacidade *one-shot* $C_{1,1}(\mathcal{E})$

Considere uma fonte quântica que emite estados ρ_i com probabilidades p_i. Suponha que após cada emissão os estados são medidos, e que X e Y são variáveis aleatórias associadas aos índices dos estados e às saídas das medições, respectivamente. A informação acessível [3, 4, 5] é definida como sendo o máximo da informação mútua $I(X;Y)$, em que o máximo é tomado sobre todas as medições POVMs:
$$I_{acc} = \max_{\{M_m\}} I(X;Y). \tag{1.16}$$

O limitante de Holevo é um resultado bastante interessante e útil em teoria da informação quântica. Ele é um limitante superior para a informação acessível. Defina a *quantidade de Holevo* como sendo
$$\chi = S(\rho) - \sum_i p_i S(\rho_i), \tag{1.17}$$

em que $\rho = \sum_i p_i \rho_i$. O limitante de Holevo afirma que $I_{acc} \leq \chi$. A igualdade se observa desde que todos os estados quânticos comutem entre si [2, pp. 77].

A capacidade $C_{1,1}(\mathcal{E})$ é definida como sendo a informação acessível de uma família de estados quânticos na saída do canal quântico.

Definição 2 (Capacidade $C_{1,1}(\mathcal{E})$ [18, 19]) *Seja $\mathcal{E}(\cdot)$ um canal quântico como definido na Seção 4.2.2. A capacidade $C_{1,1}(\mathcal{E})$ é definida com sendo o máximo da informação acessível na saída de um canal quântico, em que o máximo é tomado sobre todas as famílias na entrada do canal.*

$$C_{1,1}(\mathcal{E}) = \max_{\{\rho_x, p_x\}} I_{acc_{out}}, \tag{1.18}$$

em que $I_{acc_{out}}$ é a informação acessível da família $\{\mathcal{E}(\rho_x), p_x\}$.

A capacidade de Holevo-Schumacher-Westmoreland

Considere o problema de enviar uma mensagem clássica escolhida aleatoriamente de um conjunto $\{1, \ldots, 2^{nR}\}$ por meio de um canal quântico. No protocolo é permitido que Alice prepare palavras-código como sendo produtos tensoriais e que Bob possa realizar medições coletivas na saída do canal. A capacidade $C_{1,\infty}(\mathcal{E})$ é a análoga quântica da capacidade ordinária de Shannon.

Teorema 2 (Holevo-Schumacher-Westmoreland [7, 8]) *A capacidade HSW de um canal quântico $\mathcal{E}$ é*

$$C_{1,\infty}(\mathcal{E}) \equiv \max_{\{p_i, \rho_i\}} \left[S\left(\mathcal{E}\left(\sum_i p_i \rho_i\right)\right) - \sum_i p_i S(\mathcal{E}(\rho_i)) \right]. \tag{1.19}$$

O máximo é tomado sobre todas as famílias $\{p_i, \rho_i\}$ de estados quânticos de entrada.

A capacidade adaptativa

A capacidade adaptativa de um canal quântico, definida por Shor [6], é derivada da capacidade $C_{1,1}$ pela mudança no protocolo de comunicação. Com relação às medições, é permitido que Bob realize medições adaptativas nos estados recebidos: ele faz medições num estado de saída que somente reduz parcialmente o estado. Em seguida, ele faz uso da saída da medição para definir medições em outros estados. Bob pode retornar e realizar outras medições no estado parcialmente reduzido, em que esta última medição pode depender de todas as outras.

A *taxa de informação* para uma dada codificação e uma estratégia de medição é a informação mútua entre as palavras-código preparadas por Alice e as saídas das medições, dividido pelo número de estados usados na palavra-código (usos do canal).

Definição 3 *A capacidade adaptativa $C_{1,A}$ é definida como sendo o supremo das taxas de informação sobre todas as codificações e estratégias de medição que usam operações quânticas locais com relação aos estados separados, bem como computação clássica para coordená-los.*

No seu trabalho, Shor mostrou que a capacidade adaptativa é um limitante superior para a capacidade $C_{1,1}$ e que ela própria é limitada pela capacidade HSW.

Capacidade auxiliada por entrelaçamento

O fenômeno do entrelaçamento é um dos recursos mais impressivos da mecânica quântica. Suas aplicações incluem, por exemplo, o teletransporte de estados quânticos e a codificação superdensa. O teletransporte pode ser visto como uma forma de elevar de zero a meio qubit por uso a capacidade quântica de um canal clássico. Por outro lado, a codificação superdensa dobra a capacidade clássica de um canal quântico perfeito [2, pp. 26]. Em ambos os casos, um par EPR, que é um estado quântico maximamente entrelaçado, deve ter sido compartilhado previamente entre o transmissor e o receptor. Bennett *et. al.* [9, 10] mostraram que entrelaçamento compartilhado entre transmissor e receptor pode aumentar a capacidade HSW de canais quânticos. A chamada capacidade auxiliada por entrelaçamento é a máxima taxa de transmissão de informação clássica num cenário em que uma quantidade arbitrária de estados entrelaçados é compartilhada entre o transmissor e o receptor.

Definição 4 (Capacidade auxiliada por entrelaçamento [9]) *A capacidade auxiliada por entrelaçamento de um canal quântico $\mathcal{E}$ é*

$$C_E(\mathcal{E}) = \max_{\rho \in \mathcal{H}_{in}} S(\rho) + S(\mathcal{E}(\rho)) - S((\mathcal{E} \otimes \mathcal{I})(\Phi_\rho)), \tag{1.20}$$

em que $\rho \in \mathcal{H}_{in}$ é a matriz de densidade sobre os estados de entrada, Φ_ρ é um estado puro sobre o produto tensorial dos espaços de estado $\mathcal{H}_{in} \otimes \mathcal{H}_R$ tal que $tr_R[\Phi_\rho] = \rho$. $\mathcal{H}_{in}$ é o espaço de estado de entrada e $\mathcal{H}_R$ é o espaço de referência. O terceiro termo do lado direito da equação, $S((\mathcal{E} \otimes \mathcal{I})(\Phi_\rho))$, denota a entropia de von Neumann da purificação [2, pp. 109] Φ_ρ de ρ sobre o sistema de referência $\mathcal{H}_R$, metade do qual ($\mathcal{H}_{in}$) foi enviado através do canal quântico $\mathcal{E}$, enquanto que a outra metade ($\mathcal{H}_R$) foi enviado através do canal identidade (esta parte corresponde à porção do estado entrelaçado que Bob possuía no início do protocolo).

Para transmitir informação usando o protocolo acima, Alice e Bob "consomem" entrelaçamento. Em geral, $S(\rho)$ qubits de entrelaçamento (i.e., pares EPR) por uso do canal são necessários para atingir a capacidade auxiliada por entrelaçamento.

1.4 Teoria da informação de erro-zero

1.4.1 Capacidade ordinária de canais clássicos

Considere que um sistema A (Alice) deseja se comunicar com um sistema B (Bob). Fundamentalmente, a comunicação entre Alice e Bob é bem sucedida quando uma sinalização por parte de Alice induz um estado físico desejado em Bob. A análise quantitativa de um sistema de sinalização para prover comunicação é feita usando um arcabouço matemático introduzido por Claude E. Shannon em 1948 [13]. A ferramenta matemática usada para descrever o meio em que a informação é transmitida é o canal de comunicação.

Definição 5 (Canal discreto sem memória [20]) *Considere um alfabeto de entrada $\mathcal{X}$ e um alfabeto de saída $\mathcal{Y}$. Um canal clássico discreto sem memória (DSM) $C : \mathcal{X} \rightarrow \mathcal{Y}$, denotado por $(\mathcal{X}, p(y|x), \mathcal{Y})$, é definido por uma matriz estocástica cujas linhas são indexadas por elementos do conjunto finito $\mathcal{X}$, enquanto que as colunas são indexadas por índices de $\mathcal{Y}$. O elemento (x, y) da matriz estocástica é a probabilidade $p(y|x)$ que $y \in \mathcal{Y}$ seja recebido quando $x \in \mathcal{X}$ é transmitido. O canal é dito ser sem memória se a distribuição de probabilidade da saída depende somente da entrada naquele tempo, e que ela é condicionalmente independente de entradas ou saídas prévias.*

Definição 6 (Capacidade de canais DSM) *A capacidade informacional de canais discretos sem memória é dada por*

$$C = \max_{p(x)} I(X, Y), \tag{1.21}$$

em que o máximo é tomado sobre todas as distribuições de entrada $p(x)$. $I(X, Y)$ é a informação mútua entre as variáveis aleatórias X e Y que representam a entrada e a saída do canal DSM, respectivamente.

Para enunciar o teorema da codificação de Shannon, é necessário definir um código (M, n) para um canal DSM:

Definição 7 *Um código de blocos (M, n) para um canal DSM $(\mathcal{X}, p(y|x), \mathcal{Y})$ é composto como segue:*

1. *Um conjunto de índices $\{1, \ldots, M\}$, em que cada índice está associado a uma mensagem clássica.*

2. *Uma função de codificação*

$$X^n : \{1, \ldots, M\} \rightarrow \mathcal{X}^n,$$

originando palavras-código $\mathbf{x}^1 = X^n(1), \ldots, \mathbf{x}^M = X^n(M)$. *Um livro de códigos é o conjunto de todas as palavras-código.*

3. *Uma função de decodificação*

$$g : \mathcal{Y}^n \to \{1, \ldots, M\},$$

que mapeia cada palavra-código recebida numa mensagem do conjunto $\{1, \ldots, M\}$.

A probabilidade de erro do código é dada por $P_e = \Pr\left(g(Y^n) \neq i | X^n = X^n(i)\right)$, e a taxa de transmissão da informação é $R = \frac{1}{n} \log M$ bits por símbolo. O teorema da codificação de canal garante a existência de códigos que alcançam a capacidade do canal com uma probabilidade de erro arbitrariamente pequena.

Teorema 3 (Codificação de canal [20]) *Todas as taxas abaixo da capacidade* C *são alcançáveis, ou seja, existe uma seqüência de códigos tal que a probabilidade média de erro tende a zero quando o comprimento do código tende para infinito. Equivalentemente, qualquer seqüência de códigos com uma probabilidade de erro assintoticamente baixa possui taxa* $R \leq C$.

1.4.2 A capacidade erro-zero

O teorema da codificação de canal afirma que existe uma probabilidade de erro positiva mesmo para as melhores famílias de códigos. Shannon mostrou que era possível transmitir informação sem erro através de canais DSM. Shannon [12] definiu um código (M, n) de erro-zero da mesma forma que um código de blocos (M, n), mas com a restrição seguinte à probabilidade de erro:

$$\Pr\left(g(Y^n) \neq i | X^n = X^n(i)\right) = 0 \; \forall \, i \in \{1, \ldots, M\}, \tag{1.22}$$

que garante a inexistência de erros de decodificação. Dois símbolos de entrada $x_i, x_j \in \mathcal{S}$ são ditos ser não-adjacentes (ou distinguíveis) se existe pelo menos um símbolo $y \in \mathcal{Y}$ tal que ambas $p(y|x_i)$ e $p(y|x_j)$ são diferentes de zero. Caso contrário, os símbolos são adjacentes. Dado que uma seqüência de n símbolos $\mathbf{x} = x_1 x_2 \ldots x_n$ é transmitida por um canal DSM, a seqüência $\mathbf{y} = y_1 y_2 \ldots y_n$ é recebida com probabilidade

$$p^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i). \tag{1.23}$$

Se duas seqüências $\mathbf{x}'$ e $\mathbf{x}''$ podem ambas resultar numa seqüência $\mathbf{y}$ com probabilidade positiva, então as seqüências são ditas ser *indistinguíveis* ou adjacentes, já que o decodificador não consegue distinguí-las na saída do canal. Caso contrário, as seqüências são não-adjacentes. As seqüências $\mathbf{x}'$ e $\mathbf{x}''$ são distinguíveis se e somente se existir pelo menos um índice $1 \leq i \leq n$ tal que x_i' e x_i'' são não-adjacentes. Pode-se pensar nas distribuições $p(y|x)$ e $p^n(\cdot|\mathbf{x})$ como vetores de dimensão $|\mathcal{X}|$ e $|\mathcal{X}|^n$, respectivamente. Assim, duas seqüências $\mathbf{x}', \mathbf{x}'' \in \mathcal{X}^n$ são distinguíveis se os vetores correspondentes são ortogonais.

Definição 8 (Capacidade erro-zero) *Defina $N(n)$ como a cardinalidade máxima de um conjunto de vetores mutuamente ortogonais entre $p^n(\cdot|\mathbf{x})$, $\mathbf{x} \in \mathcal{X}^n$. A capacidade erro-zero de um canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ é dada por*

$$C_0 = \limsup_{n\to\infty} \frac{1}{n} \log N(n). \tag{1.24}$$

Intuitivamente, C_0 é a taxa máxima de transmissão de informação sem erro do canal.

Devido ao fato de que $N(n)$ ser supermultiplicativo, o limite superior coincide com o supremo (em n) dos números $\frac{1}{n}\log N(n)$.

O mapa de redução de adjacência

O cálculo da capacidade erro-zero de alguns canais simples pode ser feito usando o conceito de *mapeamento de redução de adjacência*, que é uma função $f : \mathcal{X} \to \mathcal{X}$ com a propriedade de que se x_i e x_j são não-adjacentes no canal, então $f(x_i)$ e $f(x_j)$ são não-adjacentes.

Teorema 4 *Seja $(\mathcal{X}, p(y|x), \mathcal{Y})$ um canal DSM. Se todos os símbolos $\mathcal{X}$ podem ser mapeados usando um mapa de redução de adjacência f num subconjunto $\mathcal{X}' \subset \mathcal{X}$ de símbolos não-adjacentes, então $C_0 = \log|\mathcal{X}'|$.*

Relação com a teoria de grafos

O problema de calcular a capacidade erro-zero de canais DSM pode ser reformulado usando elementos da teoria de grafos. Dado um canal $(\mathcal{X}, p(y|x), \mathcal{Y})$ é possível construir um grafo característico G como segue. Tome tantos vértices quanto for o número de símbolos em $\mathcal{X}$ e conecte dois vértices se os símbolos correspondentes são não-adjacentes. Defina o n-ésimo produto de Shannon de G como sendo um grafo para o qual $V(G^n) = \mathcal{X}^{\times n}$ e $\{\mathbf{x}', \mathbf{x}''\} \in E(G^n)$ se para pelo menos um $1 \leq i \leq n$ as i-ésimas coordenadas de $\mathbf{x}'$ e $\mathbf{x}''$ satisfazem $\{x_i', x_i''\} \in E(G)$. É fácil verificar que o número máximo de seqüências distinguíveis de comprimento n é o número de clique de G^n, i.e, $N(n) = \omega(G^n)$. Portanto, a capacidade erro-zero é dada por

$$C_0 = \sup_n \frac{1}{n} \log \omega(G^n). \tag{1.25}$$

Seja $\chi(G)$ o número cromático do grafo característico G. Shannon [12] demonstrou o seguinte resultado:

Teorema 4' *Seja $(\mathcal{X}, p(y|x), \mathcal{Y})$ um canal DSM e G o grafo característico correspondente. Se $\omega(G) = \chi(G)$, então $C_0 = \chi(G)$.*

Originalmente Shannon usou uma abordagem diferente, mas equivalente, para relacionar a capacidade erro-zero e grafos. Para um dado canal DSM $(\mathcal{X}, p(y|x), \mathcal{Y})$, é possível associar uma matriz de adjacência como segue:

$$A_{ij} = \begin{cases} 1 & \text{se } x_i \text{ é adjacente a } x_j \text{ ou se } i = j \\ 0 & \text{caso contrário,} \end{cases} \tag{1.26}$$

em que $x_i, x_j \in \mathcal{X}$. Shannon definiu um grafo de adjacência em que os vértices são símbolos do conjunto $\mathcal{X}$ e dois vértices são conectados se os símbolos correspondentes são adjacentes. Este grafo é complementar ao grafo característico. Então foi mostrado o resultado seguinte [21]:

Teorema 4" *Seja G o grafo de adjacência de um canal DSM $(\mathcal{X}, p(y|x), \mathcal{Y})$. Se G pode ser coberto por $N(1)$ cliques, então $C_0 = \log N(1)$.*

1.4.3 Função teta de Lovász

A conexão entre a capacidade erro-zero e a teoria de grafos motivou a definição de estruturas interessantes na teoria de grafos. Uma delas é a função teta de Lovász [21], que é um funcional que pode ser calculado em tempo polinomial e seu valor se encontra entre duas grandezas com complexidades NP-completas: o número de clique e o número cromático de um grafo [22].

Dado um canal DSM $(\mathcal{X}, p(y|x), \mathcal{Y})$ e um grafo de adjacência G correspondente com vértices $\mathcal{X}$, uma representação ortonormal de G é um conjunto de $|\mathcal{X}|$ vetores $\mathbf{v}_{x_i}$ num espaço Euclidiano tal que se $x_i, x_j \in \mathcal{X}$ são não-adjacentes, então $\mathbf{v}_{x_i}$ e $\mathbf{v}_{x_j}$ são ortogonais. O *valor* de uma representação é definido como sendo

$$\min_{\mathbf{c}} \max_{x_i \in \mathcal{X}} \frac{1}{(\mathbf{c}^T \mathbf{v}_{x_i})^2},$$

em que o mínimo é tomado sobre todos os vetores unitários $\mathbf{c}$. O vetor $\mathbf{c}$ que alcança o mínimo é chamado de *handle* da representação. A função $\theta(G)$ de Lovász é definida como sendo o mínimo valor sobre todas as representações de G. A representação é dita ser ótima se ela alcança o valor mínimo. Lovász provou o resultado seguinte:

Teorema 5 ([21]) *A capacidade erro-zero de um canal DSM $(\mathcal{X}, p(y|x), \mathcal{Y})$ é limitada superiormente pelo logaritmo da função θ do seu grafo de adjacência G:*

$$C_0 \leq \log \theta(G). \tag{1.27}$$

A definição da função θ abriu caminho para a resolução de um problema apontado por Shannon mas que só viria a ser resolvido treze anos depois: o cálculo da capacidade erro-zero do canal DSM que origina o pentágono como grafo característico. No seu artigo, Shannon [12] mostrou que a capacidade do pentágono G_5 era tal que $\frac{1}{2}\log 5 \leq C_0(G_5) \leq \log\frac{5}{2}$. Lovász então construiu uma representação ortonormal para o pentágono e mostrou que o valor daquela representação coincidia com o limite inferior encontrado por Shannon, provando assim que a capacidade erro-zero do pentágono era $C_0(G_5) = \frac{1}{2}\log 5$.

1.5 Capacidade erro-zero de canais quânticos

1.5.1 Capacidade erro-zero quântica

Dado um canal quântico, investiga-se qual a máxima quantidade de informação por uso do canal que Alice pode transmitir para Bob com uma probabilidade de erro igual a zero. A comunicação é feita considerando o seguinte protocolo: o alfabeto da fonte é um conjunto $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ de estados de dimensão d, em que d é a dimensão do canal quântico; Alice prepara palavras-código que são produtos tensoriais de estados do alfabeto da fonte e medições coletivas são permitidas na saída do canal. Essencialmente, este protocolo é similar ao da capacidade HSW [7, 8]. Um código de bloco quântico de erro-zero é definido como segue.

Definição 9 *Um código de bloco de erro-zero quântico (K_n, n) é composto de:*

1. *um conjunto de índices $\{1, \ldots, K_n\}$, em que cada índice está associado a uma mensagem clássica;*

2. *uma função de codificação*

$$X^n : \{1, \ldots, K_n\} \to \mathcal{S}^{\otimes n}, \tag{1.28}$$

levando à palavras-código quânticas $\overline{\rho}_1 = X^n(1), \ldots, \overline{\rho}_{K_n} = X^n(K_n)$;

3. *uma função de decodificação*

$$g : \{1, \ldots, m\} \to \{1, \ldots, K_n\}, \tag{1.29}$$

que associa deterministicamente uma saída $y \in \{1, \ldots, m\}$ de uma medição POVM a uma mensagem clássica com a seguinte propriedade

$$Pr\left(g(Y = y) \neq i | X^n = X^n(i)\right) = 0 \; \forall\, i \in \{1, \ldots, K_n\}. \tag{1.30}$$

A taxa desse código é dada por $R_n = \frac{1}{n}\log K_n$ *(bits por uso do canal).*

Definição 10 *A capacidade erro-zero de um canal quântico $\mathcal{E}(\cdot)$, denotada por $C^{(0)}(\mathcal{E})$, é o supremo das taxas alcançáveis com probabilidade de erro igual a zero,*

$$C^{(0)}(\mathcal{E}) = \sup_{\mathcal{S}} \sup_{n} \frac{1}{n} \log K_n, \tag{1.31}$$

em que K_n é o número máximo de mensagens clássicas que o sistema pode transmitir sem erro quando um código de bloco quântico de erro-zero (K_n, n) e alfabeto $\mathcal{S}$ é usado.

Por definição, dois estados quânticos $\rho_i, \rho_j \in \mathcal{S}$ são ditos ser não-adjacentes em $\mathcal{E}$ se $\mathcal{E}(\rho_i)$ e $\mathcal{E}(\rho_j)$ são distinguíveis. Caso contrário eles são adjacentes. Usa-se a notação $\rho_i \perp_{\mathcal{E}} \rho_j$ para denotar que ρ_i é não-adjacente a ρ_j. Da mesma forma, duas seqüências de produtos tensoriais $\hat{\rho}_i, \hat{\rho}_j \in \mathcal{S}^{\otimes n}$ são não-adjacentes se elas são distinguíveis na saída do canal. Caso contrário elas são adjacentes.

Proposição 1 *Para um dado canal quântico $\mathcal{E}$ e um código com alfabeto $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$, $\hat{\rho}_i, \hat{\rho}_j \in \mathcal{S}^{\otimes n}$ são não-adjacentes se e somente se para ao menos um $1 \leq k \leq n$, ρ_{i_k} é não-adjacente a ρ_{j_k}.*

Proposição 2 *A capacidade erro-zero quântica de um canal $\mathcal{E}$ é maior que zero se e somente se existe pelo menos dois estados $\rho_i, \rho_j \in \mathcal{S}$ tais que $\rho_i \perp_{\mathcal{E}} \rho_j$.*

Relação com a teoria de grafos

A capacidade erro-zero quântica (CEZQ) é redefinida usando elementos da teoria de grafos. Dado um canal quântico $\mathcal{E}$ e um conjunto $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ de estados de entrada, é possível construir um grafo característico $\mathcal{G}$ como segue:

$$V(\mathcal{G}) = \{1, \ldots, l\}, \tag{1.32}$$

$$E(\mathcal{G}) = \{(i, j); \rho_i \perp_{\mathcal{E}} \rho_j;\ \rho_i, \rho_j \in \mathcal{S}; i \neq j\}. \tag{1.33}$$

Defina o n-ésimo produto de Shannon de $\mathcal{G}$, $\mathcal{G}^n$, como sendo o grafo

$$V(\mathcal{G}^n) = \{1, \ldots, l\}^n, \tag{1.34}$$

$$E(\mathcal{G}^n) = \{(i_1 \ldots i_n, j_1 \ldots j_n); \rho_{i_k} \perp_{\mathcal{E}} \rho_{j_k} \text{ para ao menos um } 1 \leq k \leq n; \rho_{i_k}, \rho_{j_k} \in \mathcal{S}\}. \tag{1.35}$$

Sendo assim, o número máximo de mensagens que o sistema pode transmitir sem erro usando um código quântico de erro-zero com alfabeto $\mathcal{S}$ é dado pelo número de clique de $\mathcal{G}^n$, $\omega(\mathcal{G}^n)$.

Definição 11 *A capacidade erro-zero de um canal quântico $\mathcal{E}$ é dada por*

$$C^{(0)}(\mathcal{E}) = \sup_{\mathcal{S}} \sup_{n} \frac{1}{n} \log \omega\left(\mathcal{G}^n\right), \tag{1.36}$$

em que o supremo é tomado sobre todos os conjuntos $\mathcal{S}$ de estados de entrada e $\omega(\mathcal{G}^n)$ é o número de clique do grafo n-ésimo produto de Shannon de $\mathcal{G}$, que é o grafo característico associado ao conjunto $\mathcal{S}$.

1.5.2 Estados quânticos que atingem a CEZQ

É sabido que a capacidade de HSW [7, 8] pode ser sempre alcançada usando no máximo d^2 estados puros [2, pp. 555]. Um resultado análogo é mostrado para a capacidade erro-zero quântica.

Proposição 3 *A capacidade erro-zero de um canal quântico $\mathcal{E}$ num espaço de Hilbert de dimensão d pode sempre ser alcançada por um conjunto $\mathcal{S}$ composto de, no máximo, d estados quânticos puros, i.e., $\mathcal{S} = \{\rho_i = |v_i\rangle\langle v_i|\}_{i=1}^{d}$.*

É importante ressaltar que, na demonstração do resultado acima, foi necessário definir o conceito de grafo k-clonado. Seja $G = (V, E)$ um grafo não-direcionado tal que $V = \{0, \ldots, l-1\}$ e $E \subset \{(i,j); i,j \in V; i \neq j\}$. Para cada vértice $i \in V(G)$, denote por $N(i)$ o conjunto de vizinhos de i,

$$N(i) = \{j \in V(G); (i,j) \in E(G)\}. \tag{1.37}$$

Definição 12 *O grafo k-clonado de G, denotado por G', é um grafo com $l+1$ vértices obtido de G "clonando" o vértice k de G:*

1. *$V(G') = \{0, \ldots, l\}$, em que l é o rótulo do vértice clonado;*
2. *$E(G') = E(G) \cup \{(l,j); j \in N(k)\}$, i.e., ambos os vértices l e k possuem os mesmos vizinhos.*

Teorema 6 *Para todo n, $\omega(G^n) = \omega(G'^n)$.*

O teorema implica que a capacidade erro-zero (clássica ou quântica) de um canal associado com o grafo G é igual a capacidade erro-zero do canal associado com G. Uma versão menos restritiva do teorema também foi mostrada.

Corolário 1 *Suponha que ao invés de clonar um vértice de G, todo um subgrafo induzido de G seja clonado, dando origem a um novo grafo G'. Então, $\omega(G'^n) = \omega(G^n)$ para todo n.*

Corolário 2 *Na definição de grafo k-clonado, suponha que o conjunto de arestas do grafo k-clonado seja tal que $E(G') = E(G) \cup \{(l,j); j \in N(l)\}$, em que $N(l) \subseteq N(k)$. i.e., o vértice l de G' possui os mesmos vizinhos do vértice original k em G, mas o último pode ter outros vizinhos. Então, $\omega(G'^n) = \omega(G^n)$ ainda se verifica para todo n.*

1.5.3 Medições que alcançam a CEZQ

Suponha que a CEZQ seja alcançada para um conjunto $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ e um dado n. Então, o conjunto $\mathcal{S}^{\otimes n}$ possui exatamente $K_n = \omega(\mathcal{G}^n)$ palavras-código não-adjacentes entre si, i.e., todos os estado quânticos

$$\begin{aligned}
\mathcal{E}(\overline{\rho}_1) &= \underbrace{\mathcal{E}(\rho_{1_1}) \otimes \mathcal{E}(\rho_{1_2}) \otimes \cdots \otimes \mathcal{E}(\rho_{1_n})}_{P_1} \\
\mathcal{E}(\overline{\rho}_2) &= \underbrace{\mathcal{E}(\rho_{2_1}) \otimes \mathcal{E}(\rho_{2_2}) \otimes \cdots \otimes \mathcal{E}(\rho_{2_n})}_{P_2} \\
\vdots \quad & \quad \vdots \\
\mathcal{E}(\overline{\rho}_{K_n}) &= \underbrace{\mathcal{E}(\rho_{K_{n_1}}) \otimes \mathcal{E}(\rho_{K_{n_2}}) \otimes \cdots \otimes \mathcal{E}(\rho_{K_{n_n}})}_{P_{K_n}}
\end{aligned} \tag{1.38}$$

são dois a dois ortogonais no espaço de Hilbert de saída de dimensão d^n. Defina P_i como sendo o projetor sobre o subespaço de Hilbert gerado pelos estados no suporte de $\mathcal{E}(\overline{\rho}_i)$. O conjunto

$$\mathcal{P} = \{P_1, \ldots, P_{K_n}, P_{K_n+1}\}, \tag{1.39}$$

$P_{K_n+1} = \mathbb{1} - \sum_{i=1}^{K_n} P_i$, é uma medição de von Neumann (projetiva) que permite distinguir as K_n seqüências de estados quânticos. Portanto, medições projetivas coletivas são suficientes para decodificar qualquer código quântico de erro-zero. Ainda, foi mostrado que tais medições são necessárias para alcançar a CEZQ.

1.5.4 Exemplos

Canal de troca de bit

O canal de troca de bit num espaço de Hilbert de dimensão dois,

$$\mathcal{E}(\rho) = p\rho + (1-p)X\rho X, \tag{1.40}$$

possui capacidade erro-zero igual a $C^0(\mathcal{E}) = \frac{1}{1}\log(2) = 1$ bit por uso. A capacidade é alcançada por um conjunto de estados $\mathcal{S} = \{|v_1\rangle, |v_2\rangle\}$ em que

$$\begin{aligned}
|v_1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\
|v_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).
\end{aligned}$$

O canal de despolarização

O canal de despolarização num espaço de Hilbert de dimensão d,

$$\mathcal{E}(\rho) = p\frac{1}{d}\mathbb{1}_d + (1-p)\rho, \tag{1.41}$$

possui capacidade erro-zero quântica igual a zero desde que $0 < p < 1$, visto que quaisquer dois estados quânticos de entrada, ρ_i, ρ_j, não podem ser distinguidos na saída do canal, ou seja, são adjacentes.

Capacidade erro-zero de canais clássicos-quânticos

Na literatura, um canal quântico $\mathcal{E}$ para qual o estado $(\mathbb{1} \otimes \mathcal{E})(\Gamma)$ é sempre separável (mesmo para um Γ entrelaçado) é chamado de canal de quebra de entrelaçamento [23, 24], e pode ser sempre escrito na forma

$$\mathcal{E}(\rho) = \sum_i \sigma_i \mathrm{tr}\left[\rho X_i\right], \tag{1.42}$$

em que $\{\sigma_i\}$ é uma família fixa de estados quânticos e $\{X_i\}$ define uma medição POVM. O canal é chamado de *clássico-quântico* (c-q) se $X_i = |\psi_i\rangle\langle\psi_i|$, em que $\{|\psi_i\rangle\}$ é uma base ortonormal, i.e., os elementos de POVM são projetores de dimensão um.

Proposição 4 *Seja $\mathcal{E}_{c-q}$ um canal quântico c-q num espaço de Hilbert de dimensão d definido por $\{\sigma_i\}$ e $\{X_i = |\psi_i\rangle\langle\psi_i|\}_{i=1}^d$, em que $\{|\psi_i\rangle\}$ é uma base ortonormal. Então a capacidade erro-zero quântica pode sempre ser alcançada pelo conjunto*

$$\mathcal{S} = \{|\psi_1\rangle, \ldots, |\psi_d\rangle\}. \tag{1.43}$$

O resultado afirma que calcular a CEZQ de um canal c-q é um problema completamente clássico, sendo necessário somente explicitar as relações de adjacência entre os estados do conjunto $\mathcal{S}$ que define as medições POVM. Explicitadas as relações de adjacência, o grafo característico $\mathcal{G}$ é encontrado e a CEZQ será dada por $C^0(\mathcal{E}_{c-q}) = \sup_n \frac{1}{n}\log\omega(\mathcal{G}^n)$.

Um canal clássico-quântico particular

Considere o canal c-q de dimensão 5 definido por

$$|\sigma_i\rangle = \frac{|i\rangle + |i+1 \mod 5\rangle}{\sqrt{2}}, \sigma_i = |\sigma_i\rangle\langle\sigma_i| \qquad \text{e} \qquad X_i = |i\rangle\langle i|, \quad 0 \leq i \leq 4, \tag{1.44}$$

em que $\{|0\rangle, \ldots, |4\rangle\}$ é a base computacional do espaço de Hilbert de dimensão 5.

O conjunto $\mathcal{S}$ que alcança a capacidade é dado por $\mathcal{S} = \{|0\rangle, \ldots, |4\rangle\}$. Os estados correspondentes na saída do canal são $\mathcal{E}(|i\rangle) = \sigma_i$. As relações de adjacência são dadas por

$$|0\rangle \perp_{\mathcal{E}} |2\rangle \qquad |0\rangle \perp_{\mathcal{E}} |3\rangle \qquad |1\rangle \perp_{\mathcal{E}} |3\rangle \qquad |1\rangle \perp_{\mathcal{E}} |4\rangle \qquad |2\rangle \perp_{\mathcal{E}} |4\rangle.$$

Neste caso, o conjunto $\mathcal{S}$ que atinge a CEZQ dá origem ao pentágono como grafo característico. Portanto, a CEZQ do canal em questão é dada por $C^{(0)}(\mathcal{E}) = C_0(G_5) = \frac{1}{2}\log 5$ bits/uso. Apesar da capacidade do canal ser alcançada usando um conjunto de estados quânticos dois a dois ortogonais, são necessários dois usos do canal para atingir a CEZQ. Um código quântico de erro-zero que alcança a capacidade é dado por

$$\begin{aligned} \overline{\rho}_1 = |0\rangle|0\rangle, \quad \overline{\rho}_2 = |1\rangle|2\rangle, \quad \overline{\rho}_3 = |2\rangle|4\rangle \\ \overline{\rho}_4 = |3\rangle|1\rangle, \qquad \overline{\rho}_5 = |4\rangle|3\rangle. \end{aligned} \tag{1.45}$$

Estados quânticos não-ortogonais atingindo a CEZQ

Esta seção discute um exemplo de um canal quântico ilustrando que a CEZQ pode ser uma generalização não-trivial da capacidade erro-zero de Shannon para canais quânticos. Por generalização não-trivial entende-se que existem canais quânticos para os quais a capacidade é alcançada para dois ou mais usos do canal ($n > 1$) e que a capacidade só pode ser alcançada por um conjunto $\mathcal{S}$ contendo estados não-ortogonais.

Seja $\mathcal{E}(\cdot)$ um canal quântico com operadores de Kraus $\{E_1, E_2, E_3\}$ dados por

$$E_1 = \begin{bmatrix} 0.5 & 0 & 0 & 0 & \frac{\sqrt{49902}}{620} \\ 0.5 & -0.5 & 0 & 0 & 0 \\ 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -\frac{\sqrt{457}}{50} & \frac{\sqrt{457}}{50} \\ 0 & 0 & 0 & -0.62 & -\frac{289}{1550} \end{bmatrix} \quad E2 = \begin{bmatrix} 0.5 & 0 & 0 & 0 & -\frac{\sqrt{49902}}{620} \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & \frac{\sqrt{457}}{50} & -\frac{\sqrt{457}}{50} \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

$$E3 = 0.3|4\rangle\langle 4|,$$

em que $\beta = \{|0\rangle, \ldots, |4\rangle\}$ é a base computacional do espaço de Hilbert de dimensão 5. Considere o conjunto $\mathcal{S}$ de estados de entrada para $\mathcal{E}$:

$$\mathcal{S} = \left\{ |v_1\rangle = |0\rangle, |v_2\rangle = |1\rangle, |v_3\rangle = |2\rangle, |v_4\rangle = |3\rangle, |v_5\rangle = \frac{|3\rangle + |4\rangle}{\sqrt{2}} \right\}. \tag{1.46}$$

O grafo característico associado ao conjunto $\mathcal{S}$ é construído explicitando as relações de adjacência em $\mathcal{S}$, as quais são dadas por

$$|v_1\rangle \perp_{\mathcal{E}} |v_3\rangle, \qquad |v_1\rangle \perp_{\mathcal{E}} |v_4\rangle, \qquad |v_2\rangle \perp_{\mathcal{E}} |v_4\rangle, \qquad |v_2\rangle \perp_{\mathcal{E}} |v_5\rangle \qquad \text{e} \qquad |v_3\rangle \perp_{\mathcal{E}} |v_5\rangle.$$

Estas relações também dão origem ao pentágono como grafo característico. É interessante notar que se o estado $|v_5\rangle$ em $\mathcal{S}$ é substituído pelo estado $|4\rangle$, a fim de construir um conjunto $\mathcal{S}' = \beta$ de estados dois a dois ortogonais, o grafo característico resultante possui capacidade de Shannon de 1 bit/uso, i.e., menor do que a capacidade do pentágono. Isto significa que é possível transmitir mais informação usando estados não-ortogonais na entrada do canal.

Devido a forma como o canal quântico foi construído (código-fonte transcrito no Appendix 6.A, é conjecturado que a CEQZ do canal em questão é alcançada para o conjunto $\mathcal{S}$, o que implica que a CEQZ é uma generalização não-trivial da capacidade erro-zero de Shannon.

1.5.5 Capacidade erro-zero quântica e a capacidade HSW

Nesta seção é demonstrado que a capacidade erro-zero de canais quânticos, $C^{(0)}(\mathcal{E})$, é limitada superiormente pela capacidade de Holevo-Schumacher-Westmoreland, $C_{1,\infty}(\mathcal{E})$ [7, 8].

Teorema 7 *Seja $\mathcal{E}$ um canal quântico num espaço de Hilbert de dimensão d. Então*

$$C^{(0)}(\mathcal{E}) \leq C_{1,\infty}(\mathcal{E}). \tag{1.47}$$

1.6 Conclusões e perspectivas

Nesta tese foi proposta uma nova capacidade para a transmissão de informação clássica através de canais quânticos. A capacidade erro-zero quântica foi definida como sendo o supremo das taxas em que informação clássica é transmitida através de um canal quântico ruidoso com probabilidade de erro igual a zero. A CEZQ é uma generalização da capacidade erro-zero de canais clássicos discretos sem memória proposta por Shannon [12].

As principais contribuições desta tese foram:

1. foi proposta uma nova capacidade para canais quânticos;
    - a capacidade erro-zero foi generalizada para canais quânticos;
    - um código de erro-zero quântico foi formalmente definido;
2. a capacidade erro-zero quântica foi definida usando elementos da teoria de grafos;
3. foi definido o conceito de grafo k-clonado; os resultado obtidos a partir da definição são úteis tanto no cálculo da capacidade erro-zero de canais quânticos quanto de canais clássicos;
4. com relação aos estados quânticos que atingem a CEZQ

- foi mostrado que a capacidade erro-zero quântica pode sempre ser alcançada por uma família de d estados puros, em que d é a dimensão do canal quântico;

5. com relação às medições que atingem a CEZQ

    - foi mostrado que medições de von Neumann (projetivas) coletivas são necessárias e suficientes para atingir a capacidade erro-zero quântica;

6. a capacidade erro-zero quântica de canais clássicos-quânticos foi estudada, mostrando-se que ela pode sempre ser alcançada usando a base ortonormal como alfabeto do código quântico, ou seja, o cálculo da CEZQ de canais c-q é puramente clássico;

7. alguns exemplos do cálculo da capacidade erro-zero quântica foram exibidos:

    - foi apresentado um canal c-q cujo conjunto $\mathcal{S}$ que atinge a capacidade dá origem ao pentágono como grafo característico. Desta forma, a CEZQ deste canal pôde ser calculada e um código quântico de erro-zero que atinge a capacidade foi explicitado;
    - um exemplo de um canal quântico que dá origem ao pentágono como grafo característico para um conjunto $\mathcal{S}$ de estados não-ortogonais foi mostrado;

8. com base no exemplo acima, foi conjecturado que a capacidade erro-zero quântica é uma generalização não-trivial da capacidade erro-zero de Shannon;

9. por último, foi mostrado que a capacidade erro-zero quântica é limitada superiormente pela capacidade HSW.

Algumas propostas para trabalhos futuros são (lista não-exaustiva):

1. generalização da função teta de Lovász para o caso quântico;

2. variações do protocolo de comunicação − presença de um canal de realimentação clássico entre o receptor e o transmissor, disponibilidade de uma quantidade arbitrária de entrelaçamento compartilhado entre o transmissor e o receptor e o caso em que há múltiplos transmissores e receptores;

3. pode-se investigar ligações entre a CEZQ e a teoria de subespaços livres de decoerência e subsistemas sem ruído;

4. como a capacidade erro-zero clássica possui aplicações na teoria de complexidade computacional, é provável que a CEZQ possa estar relacionada à complexidade computacional quântica.

Chapter 2

Introduction

2.1 Classical information over quantum channels

One of the main issues in quantum information theory is the concept of quantum channel capacity [1, 2]. In a more fundamental way, the capacity of a channel is defined as the least upper bound of rates at which information can be transmitted through the channel with arbitrarily high reliability.

Quantum mechanics provides many features allowing of several ways to define quantum channel capacity [1, 2]. For a given quantum channel, the capacity may assume different values depending on: (a) the kind of information to be carried − although channel signalling is always performed using quantum states, one may wish to use a quantum channel to transmit classical messages or quantum systems, e.g., quantum states generated by a quantum source; (b) external resources, like entanglement of a feedback classical channel from the receiver (Bob) to the sender (Alice); and (c) the communication protocol. The communication protocol determines how information should be encoded at the transmitter and decoded at the receiver end.

In this work we focus on the capacity of memoryless quantum channels to carry classical information. Several such capacities have already been defined. According to the communication protocol, quantum channel capacities can be grouped into three categories:

1. codewords are restricted to tensor products of input quantum states and measurements are performed individually at the channel output [3, 4, 5, 6];

2. codewords are restricted to tensor products of input quantum states, whereas entangled measurements between several channel outputs are allowed [7, 8, 9, 10];

3. entanglement between several channel inputs is allowed, as well as collective measurements at the channel output [11].

Examples of capacities employing the protocol 1 are the one-shot capacity [3, 4, 5] and the Shor's adaptive capacity [6]. Suppose that Alice prepares states ρ_i with probability p_i and gives a prepared state to Bob. Accessible information is the maximum amount of information about the prepared state that Bob can extract from the received states by performing only individual measurements. The one-shot capacity is defined as the maximum over all input ensembles $\{p_i, \rho_i\}$ of the accessible information of the corresponding output ensemble. Shor's protocol is similar to the above, except that Bob can perform partial measurements on one signal which only partially reduces the quantum state, use the outcome of this measurement to determine which measurements to make on different signals, return to redefine the measurement on the first state, and so forth. It was showed that the adaptive capacity is always greater than or equal to the one-shot capacity.

The main example of quantum channel capacity making use of protocol 2 is the Holevo-Schumacher-Westmoreland (HSW) capacity [7, 8]. The HSW capacity, denoted by $C_{1,\infty}(\mathcal{E})$, is also known as *the classical capacity of quantum channels*. The HSW capacity is the generalisation of the Shannon's ordinary capacity [13], in the sense that the Shannon coding theorem can be derived from the HSW coding theorem [25, 23]. The quantum channel coding theorem asserts that for each rate $R \leq C_{1,\infty}$ there exists a sequence of codes for which the error probability goes asymptotically to zero as the code length goes to infinity. Conversely, every achievable rate R must be less than or equal to the capacity $C_{1,\infty}$.

Capacities employing protocol 3 are directly connected with one of the most important open issues in quantum information theory, the additivity conjecture of the Holevo information [7]. The conjecture asserts that entanglement between several channel inputs does not increase the HSW capacity of memoryless quantum channels. However, it is known that entangled codewords may increase the HSW capacity of quantum channels with memory [11].

2.2 Zero-error capacity of classical channels

Information theory was introduced by Claude E. Shannon in 1948 [13]. In his paper, Shannon defined a number C representing the capacity of a communication channel for transmitting information reliably. He proved the existence of codes that allow reliable transmission, provided that the communication rate is less than the channel capacity. A randomly generated code with large block size has a high probability to be a good code. By reliable transmission we mean that the error probability can be made as close to zero as possible, but not actually zero. Most of information theory issues, including channel capacity, are based on probability theory and statistics. This asymptotic capacity is hereafter denoted the *ordinary capacity*.

In 1956, eight years after his first paper introducing information theory, Shannon demonstrated how Discrete Memoryless Channels (DMCs) could be used to transmit information in a scenario where *no errors* are permitted, instead of allowing an asymptotically small probability of error. The so-called *zero-error capacity* was defined as the least upper bound of rates at which information can be transmitted through a DMC with a probability of error *equal to zero* [12]. Körner and Orlitsky [26] pointed out some situations in which it would be interesting to consider a scenario where no transmission errors are allowed and ask for the maximum rate at which information can be transmitted:

- Applications where no errors can be tolerated.
- In some models, only a small number of channel uses or a few source instances are available. Therefore, we cannot appeal to results ensuring that the error probability decreases as the number of uses or instances increases.
- The zero-error information theory can be used to study the communication complexity of error-free protocols and functions.
- Functionals and methods originally used in zero-error information theory are often applied in mathematics and computer science.

In the original paper, Shannon gave a graph theoretic approach to the zero-error capacity. By associating a DMC with a graph, Shannon introduced a new quantity in graph theory, the Shannon capacity of a graph [14, 15, 16]. Differently from the ordinary capacity, finding the zero-error capacity of a DMC (or a graph) is a combinatorial problem. Because of its restrictive nature − a vanishing probability of error is required − the zero-error information theory is frequently unknown to many information theorists. Nevertheless, its methods play an important role in areas like combinatorics and graph theory.

In this work we generalise the zero-error capacity to quantum channels. Initially, we formally define an error-free quantum code as well as the encoding and decoding procedures. Then, we define the quantum zero-error capacity as the least upper bound of rates at which classical information can be transmitted without error through a noisy memoryless quantum channel. The problem of finding the zero-error capacity is reformulated in the language of graph theory and an equivalent definition is given. We also investigate some properties of quantum states and measurements reaching the quantum zero-error capacity. A mathematically motivated example is used to claim that the quantum zero-error capacity is a non-trivial generalisation of the Shannon zero-error capacity, in the sense that there exist quantum channels for which the capacity can only be reached by using an ensemble of non-orthogonal quantum input states, and two or more channel uses are necessary in order to attain the capacity, i.e., the capacity can only be reached by using a quantum code of length two or more. We formally relate the quantum zero-error capacity to the HSW capacity, given a proof that the former is upper bounded by the latter.

2.3 Thesis outline

Contributions are entirely presented in Chapter 6. Readers familiarized with quantum information and classical zero-error information theory can skip Chapters 2 to 5 and go directly to Chapter 6. This thesis is organized as follows:

Chapters 3 and 4 give an overview of quantum information concepts related to the thesis. Section 3.2 aims to introduce the Dirac's notation to the reader, whereas discusses important tools in quantum information, as unitary operators and tensor products. The four quantum mechanics postulates are further presented in Section 3.3, followed by a discussion about the density operator formalism. A brief survey about classical capacities of quantum channels is given in Chapter 4. Initially, we introduce the von Neumann entropy and we give a mathematical definition of quantum channels. Sections 4.3 to 4.6 review the one-short capacity $C_{1,1}(\mathcal{E})$, the Holevo-Schumacher-Westmoreland capacity $C_{1,\infty}(\mathcal{E})$, the adaptive capacity $C_{1,A}(\mathcal{E})$, and the entanglement-assisted capacity $C_E(\mathcal{E})$, respectively.

Chapter 5 introduces some definitions and results in classical zero-error information theory. Section 5.1 presents the ordinary capacity of DMC and some examples are given. Section 5.2 introduces the zero-error capacity, and a method for calculating the capacity of simple channels is discussed in Section 5.2.1. Section 5.2.2 presents a graph-theoretic approach for the zero-error capacity. The representation of a DMC using either an adjacency graph or its complementary graph gives two different but equivalent ways of calculating the zero-error capacity. In Section 5.3 we present the Lovász theta function [21], a polynomially computable functional which is sandwiched in between the clique and the chromatic numbers of a graph. This functional was used by Lovász to calculate de zero-error capacity of the pentagon graph, a five vertices graph for which Shannon was not able to give an exact value for the capacity. Sections 5.4 and 5.5 illustrate how different is the behaviour of the zero-error capacity and the ordinary capacity.

The quantum zero-error capacity is introduced in Chapter 6. In Section 6.2 we define a zero-error quantum block code and we give a formal definition of the quantum zero-error capacity. In Section 6.2.1 we present a graph-theoretic approach for the quantum zero-error capacity and we demonstrate that the two definitions are equivalent. We study in Section 6.3 some properties of quantum states attaining the quantum zero-error capacity. We show that the capacity can always be reached using an ensemble of at most d pure states, where d is the dimension of the quantum channel. We also investigate in Section 6.4 quantum measurements archiving the capacity. We have shown that collective von Neumann measurements are necessary and sufficient in order to reach the channel capacity. Section 6.5 gives some examples of the quantum zero-error capacity calculation. We explicit a mathematically motivated quantum channel and we conjecture that its zero-error capacity cannot be achieved using an ensemble of pairwise orthogonal quantum states. Moreover, this channel requires two or more channel uses in order to transmit a given message at higher rates. Finally, we demonstrate in Section 6.6 that the quantum zero-error capacity is upper bounded by the HSW capacity [7, 8].

In Chapter 7 we summarize our contributions and we give some directions for further research and perspectives.

Chapter 3

Fundamentals of Quantum Mechanics

3.1 Introduction

This chapter introduces quantum mechanics in a brief and objective way. However, special attention was given to ensure that almost all concepts and definitions amongst subsequent chapters are discussed here. A more detailed approach can be found in specific textbooks [17, 2].

3.2 Linear algebra and Hilbert spaces

Although linear algebra is a well known topic in engineering, the notation used by physicists to describe quantum mechanics is different to that used in most courses of linear algebra. As we will see among this chapter, the Dirac's notation is more convenient to describe quantum systems and their evolutions. Such notation is widely used by physicists and it is standard in textbooks of quantum information and computation [2]. Dirac's notation will be gradually introduced in this chapter,

Definition 1 (Vector space [17]) *Let F be a field. A vector space V over F, with elements (vectors) represented by $|v\rangle$, is a structure composed by a set and two binary operations, $(+) : V \times V \longrightarrow V$ and $(\cdot) : F \times V \longrightarrow V$, such that*

1. $(|v\rangle + |w\rangle) + |1\rangle = |v\rangle + (|w\rangle + |1\rangle)$ *for all* $|v\rangle, |w\rangle, |1\rangle \in V$;
2. $|v\rangle + |w\rangle = |w\rangle + |v\rangle$ *for all* $|v\rangle, |w\rangle \in V$;
3. $\exists \mathbf{0} \in V$ *such that* $|v\rangle + \mathbf{0} = |v\rangle$ *for all* $|v\rangle \in V$;
4. *for any* $|v\rangle \in V$, *there exists an element* $|w\rangle \in V$ *such that* $|v\rangle + |w\rangle = \mathbf{0}$;
5. $k_1 \cdot (k_2 \cdot |v\rangle) = (k_1 k_2) \cdot |v\rangle$ *for all* $k_1, k_2 \in F$ *and* $|v\rangle \in V$;

6. $1 \cdot |v\rangle = |v\rangle$ *for all* $|v\rangle \in V$;

7. $k \cdot (|v\rangle + |w\rangle) = (k \cdot |v\rangle) + (k \cdot |w\rangle)$ *for all* $k \in F$, *and* $|v\rangle, |w\rangle \in V$;

8. $(k_1 + k_2) \cdot |v\rangle = (k_1 \cdot |v\rangle) + (k_2 \cdot |v\rangle)$ *for all* $k_1, k_2 \in F$ *and* $|v\rangle \in V$.

Elements of V *are referred as vectors, and* $\mathbf{0} \in V$ *is the zero vector of* V.

In Definition 1, the symbol $|v\rangle$ denotes an arbitrary vector in V, where v is its label. In the Dirac's notation, the structure $|\cdot\rangle$ is called a *ket*. Note that for the zero vector the *ket* is not used.

A vector subspace of a space V is a subset W of V such that W is also a vector space, i.e., W should satisfy all the conditions of Definition 1.

A set of nonzero vectors $|v_1\rangle, \dots, |v_n\rangle$, belonging to a vector space V over a field F, is said to be linearly independent if for any scalars $a_1, a_2, \dots, a_n \in F$,

$$a_1|v_1\rangle + a_2|v_2\rangle + \cdots + a_n|v_n\rangle = 0$$

implies $a_1 = a_2 = \cdots = a_n = 0$. Otherwise, the set is called linearly depedent.

A set of vectors $\beta = \{|v_1\rangle, \dots, |v_n\rangle\}$ generates the vector space V if any vector $|v\rangle$ of V can be written as a linear combination $|v\rangle = \sum_i a_i|v_i\rangle$, where $a_i \in F$. A linearly independent set β that generates V is called a *basis* of V. The dimension of V, $\dim(V)$, is defined as being the cardinality of a basis β.

3.2.1 Inner product

Let V be a vector space over the field C of complex numbers. This space is particularly important in quantum mechanics. For such space, an inner product is defined as follows.

Definition 2 (Inner product [17]) *An inner product in a vector space* V *over the field* C *of complex numbers is a function* $(\ ,\ ) : V \times V \longrightarrow C$ *such that, for all* $k_1, k_2 \in C$ *and* $|v_1\rangle, |v_2\rangle, |v\rangle, |w\rangle \in V$, *the properties below are verified:*

1. $(|w\rangle, k_1|v_1\rangle + k_2|v_2\rangle) = k_1(|w\rangle, |v_1\rangle) + k_2(|w\rangle, |v_2\rangle)$[1];

2. $(|v\rangle, |w\rangle) = (|w\rangle, |v\rangle)^*$, *where* $(^*)$ *denotes complex conjugation;*

3. $(|v\rangle, |v\rangle) \geq 0$, *e* $(|v\rangle, |v\rangle) = 0$ *if and only if* $|v\rangle = \mathbf{0}$.

[1] Some authors impose the linearity condition to the first argument instead of the second.

The above notation for inner product is not standard in quantum mechanics. Instead, it is widely used $\langle v|w\rangle$ to denote the inner product between $|v\rangle$ and $|w\rangle$. $\langle v|$ stands for the dual vector of $|v\rangle$, which will be formally defined later in this section.

The vectors $|v\rangle$ and $|w\rangle$ are said to be *orthogonal* if the inner product $\langle v|w\rangle$ is zero. The norm of vector $|v\rangle$ is defined as

$$|||v\rangle|| \equiv \sqrt{\langle v|v\rangle}. \tag{3.1}$$

A unitary vector $|v\rangle$ is a vector such that $|||v\rangle|| = 1$. A unitary vector $|v'\rangle = |v\rangle/|||v\rangle||$ is referred as the normalization of $|v\rangle$. The set of vectors $\{|i\rangle\}$, with indexes i, is said to be an orthonormal set if all vector are unitary, and vectors are pairwise orthogonal, i.e., $\langle i|j\rangle = \delta_{ij}$, where i, j are chosen from the index set. An orthogonal basis for the vector space of dimension d is a set of d pairwise orthogonal vectors. An orthogonal basis is orthonormal if all vectors are unitary.

The following definitions are necessary to introduce Hilbert spaces.

Definition 3 (Metric [27]) *A metric in a set X is a function $d : X \times X \to \Re$, which associates each pair of elements $x, y \in X$ with a real number $d(x, y)$ satisfying the following conditions for any $x, y, z \in X$:*

1. $d(x, x) = 0$;
2. *If* $x \neq y$ *then* $d(x, y) > 0$;
3. $d(y, x) = d(x, y)$;
4. $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 4 (Metric spaces [27]) *A metric space , denoted by (X, d), is composed by two parts: a set X and a metric $d(x, y)$.*

Definition 5 (Cauchy sequences [17]) *A sequence $\{x_m\}$ in a metric space (X, d) is a Cauchy sequence if for each $\epsilon > 0$ exists a N such that $d(x_n, x_m) \leq \epsilon$ for any $n, m \geq N$.*

As an example, consider the metric space consisting of all points in the interval $[0, 1]$, $X = \{x \in R : 0 \leq x \leq 1\}$, and the usual metric, $d(x, y) = |x - y|$. The sequence $\{1/n\} = \{1, 1/2, 1/4, \ldots\}$ is a Cauchy sequence. Given $\epsilon > 0$, choose $N \geq 2/\epsilon$. If $n, m \geq N$, then $1/n \leq \epsilon/2$ and $1/m \leq \epsilon/2$. Consequently, $|1/n - 1/m| \leq 1/n + 1/m \leq \epsilon$ for all $n, m \geq N$. Moreover, the sequence is convergent, since $\lim_{n\to\infty} 1/n = 0 \in X$ [27, pp. 116]. There exist Cauchy sequences that are divergent. Consider, for example, the same sequence in a metric space consisting of points $(0, 1]$, $X = \{x \in R : 0 < x \leq 1\}$, and the usual metric. Clearly, such sequence is a Cauchy sequence. However, the sequence is divergent, since the point 0 does not belong to the metric space.

Definition 6 ([27]) *A metric space (X, d) is complete if each Cauchy sequence in (X, d) is convergent.*

By definition, all vector space with inner product have an associated metric, and therefore they are metric spaces.

Definition 7 (Hilbert space [27]) *A Hilbert space is a vector space, together with a inner product, which are complete with relation to the norm defined by the inner product.*

As early mentioned, we are interest here in the vector space of n-tuples of complex numbers $(z_1, z_2, \ldots, z_n)$, denoted by C^n. The notation of column matrix will be used to refer to such vectors,

$$|z\rangle \equiv \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}. \tag{3.2}$$

The usual inner product in C^n is defined by

$$\langle y|z\rangle \equiv \begin{bmatrix} y_1^* & \ldots & y_n^* \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \tag{3.3}$$

where $(y_1, \ldots, y_n)$ and $(z_1, \ldots, z_n)$ are, respectively, the vector components of $|y\rangle$ and $|z\rangle$ with relation to the same orthonormal basis.

One can verify that the vector space C^n, together the inner product defined in Equation (3.3), is a Hilbert space of dimension n [17]. As we will see later, the state of a give quantum system can be represented by a unitary vector $|v\rangle$ belongs to a Hilbert space of dimension n. According with this notation, let $|v\rangle = \sum_i v_i|i\rangle$ and $|w\rangle = \sum_j w_j|j\rangle$ be representations of the vectors $|v\rangle$ and $|w\rangle$ with relation to an orthonormal basis $\{|i\rangle\}$, respectively. Since $\langle i|j\rangle = \delta_{ij}$,

$$\begin{aligned} \langle v|w\rangle &= \left(\sum_i v_i|i\rangle, \sum_j w_j|j\rangle\right) = \sum_{ij} v_i^* w_j \langle i|j\rangle = \sum_i v_i^* w_i \\ &= \begin{bmatrix} v_1^* & \ldots & v_n^* \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \end{aligned} \tag{3.4}$$

Equation 3.4 shows that the inner product between two vectors is equal to the inner product between the corresponding matrix representation with relation to a same orthonormal basis. Note that the dual vector $\langle v|$ is as a line vector whose components are complex conjugates of components of $|v\rangle$.

According to definitions above, the vectors

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{3.5}$$

form an orthonormal basis for the Hilbert space of dimension 2, i.e., any vector

$$|v\rangle = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \tag{3.6}$$

can be written as a linear combination $|v\rangle = a_0|0\rangle + a_1|1\rangle$ of $|0\rangle$ and $|1\rangle$.

In quantum mechanics, the basis $\{|0\rangle, |1\rangle\}$ is called the *computational basis* of the 2-dimensional Hilbert space. The computational basis for the n-dimensional Hilbert space is $\{|0\rangle, |1\rangle, \ldots, |n-1\rangle\}$, where $|k\rangle \equiv \begin{bmatrix} a_0 = 0 & a_1 = 0 & \ldots & a_k = 1 & \ldots & a_{n-1} = 0 \end{bmatrix}^T$.

3.2.2 Linear operators

A linear operator between two vector spaces V and W is defined as a function $A : V \longrightarrow W$ which is linear with relation to their inputs:

$$A\left(\sum_i a_i|v_i\rangle\right) = \sum_i a_i A|v_i\rangle. \tag{3.7}$$

It is usual to use the notation $A|v\rangle$ instead of $A(|v\rangle)$. Two important linear operators are the identity operator $\mathbb{1}$ and the operator 0, where $\mathbb{1}|v\rangle \equiv |v\rangle$ and $0|v\rangle \equiv \mathbf{0}$, respectively. The notation $\mathbb{1}_d$ is referred to the identity operator of the d-dimensional vector space.

An interesting representation of a linear operator, known as outer product, is obtained via inner product. Let $|v\rangle$ and $|w\rangle$ be two vectors belonging to vector spaces V and W with inner product, respectively. Define $|w\rangle\langle v|$ as being a linear operator from V to W in the following way:

$$(|w\rangle\langle v|)(|v'\rangle) \equiv |w\rangle\langle v|v'\rangle = \langle v|v'\rangle|w\rangle. \tag{3.8}$$

Dirac's powerful notation suggests two interpretations to Equation (3.8). The first is the application of the operator $|w\rangle\langle v|$ to the vector $|v'\rangle$, and the second is the product of the complex number $\langle v|v'\rangle$ by the vector $|w\rangle$.

Theorem 5 (Completeness relation [2]) *Let $\{|\psi_i\rangle\}$ be an orthonormal basis for a d-dimensional vector space V with inner product. Then*

$$\sum_i |\psi_i\rangle\langle\psi_i| = \mathbb{1}_d. \tag{3.9}$$

3.2.3 Pauli operators

We introduce Pauli operators, which are four 2 by 2 matrices that play a fundamental hole in quantum mechanics and quantum information [2].

$$\sigma_0 \equiv I \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \sigma_1 \equiv X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{3.10}$$

$$\sigma_2 \equiv Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \text{ and } \qquad \sigma_3 \equiv Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.11}$$

3.2.4 Eigenvectors and eigenvalues

An eigenvector of a linear operator A in a vector space V is a nonzero vector $|v\rangle$ such that $A|v\rangle = \lambda|v\rangle$. The number λ is the eigenvalue associated with the eigenvector $|v\rangle$. The eigenspace of λ is the union of the zero vector $\mathbf{0}$ together the set of all eigenvectors corresponding to λ.

The diagonal representation of an operator A in a vector space V is defined as being $A = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, where $\{|\psi_i\rangle\}$ is a set of orthonormal eigenvectors of A with corresponding eigenvalues λ_i. An operator is said to be diagonalizable if it has a diagonal representation.

3.2.5 Hermitians and unitary operators

We define in this section two important classes of operators in a Hilbert space. Let A be an operator in V and $|v\rangle, |w\rangle \in V$ two vectors that belong to V.

Definition 8 (Hermitian operator.) *The unique operator $A^\dagger \in V$ such that for all vectors $|v\rangle, |w\rangle \in V$,*

$$(|v\rangle, A|w\rangle) = (A^\dagger|v\rangle, |w\rangle), \tag{3.12}$$

is called the adjoint or Hermitian conjugate of A. An operator is said to be Hermitian *or self-adjoint if $A^\dagger = A$.*

From the definition, $(AB)^\dagger = B^\dagger A^\dagger$. By convention $|v\rangle^\dagger \equiv \langle v|$, and hence $(A|v\rangle)^\dagger = \langle v|A^\dagger$. The Hermitian conjugate of a matrix representation of an operator is the conjugate-transpose matrix of A, $A^\dagger \equiv (A^*)^T$, where (*) indicates complex conjugation and T indicates transposition.

Let $|1\rangle, \ldots, |d\rangle$ be an orthonormal basis for a d-dimension Hilbert subspace W of a n-dimensional Hilbert space V.

Definition 9 (Projector over a Hilbert subspace) *The Hermitian operator*

$$P \equiv \sum_{i=1}^{k} |i\rangle\langle i| \tag{3.13}$$

is a projector *onto the subspace W spanned by the vectors $|1\rangle, \ldots, |k\rangle$.*

It is easy to see that if $|w\rangle \in W$ then $P|w\rangle = |w\rangle$. The orthogonal complement of P, $Q = \mathbb{1} - P$, is a projector over the orthogonal subspace spanned by $|k+1\rangle, \ldots, |n\rangle$.

An operator A is said to be *normal* if $AA^\dagger = A^\dagger A$. Clearly, every Hermitian operator is also a normal operator. An important result in linear algebra stands that every normal operator M in a Hilbert space V has a *spectral decomposition* [2, pp. 72]

$$M = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle\langle\psi_i|, \tag{3.14}$$

where $|\psi_i\rangle$ are eigenvectors of M with eigenvalues λ_i, d is the dimension of V, and the set of vectors $\{|\psi_i\rangle\}$ forms an orthonormal basis for V.

Unitary operators defined below play an important hole in quantum mechanics, since they describe the evolution of a closed quantum system.

Definition 10 (Unitary operators [2]) *An operator U is* unitary *if $U^\dagger U = UU^\dagger = \mathbb{1}$.*

Geometrically, unitary operators have the property that they preserve the inner product between vectors, i.e., if $|v\rangle, |w\rangle \in V$ then

$$(U|v\rangle, U|w\rangle) = \langle v|U^\dagger U|w\rangle = \langle v|\mathbb{1}|w\rangle = \langle v|w\rangle. \tag{3.15}$$

Definition 11 (Positive operators [2]) *A Hermitian operator A in a Hilbert space V is* positive *if, for every $|v\rangle \in V$, the number $\langle v|A|v\rangle$ is positive. If $\langle v|A|v\rangle$ is a real greater than zero for every $|v\rangle \neq \mathbf{0}$, then the operator A is said to be positive definite.*

Positive operators have a spectral decomposition $\sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ with nonnegative eigenvalues λ_i.

3.2.6 Tensor products

As we will see later in this chapter, the Hilbert space of composite quantum systems is the tensor product of individual Hilbert spaces. Thus, tensor product is a way of putting together two or more Hilbert spaces to produce a larger space.

Definition 12 (Tensor product [2]) *Let V and W be Hilbert spaces of dimension m and n, respectively. Then $V \otimes W$ is a Hilbert space of dimension mn. Elements of $V \otimes W$ are linear combinations of tensor products $|v_i\rangle \otimes |w_i\rangle$ of elements $|v_i\rangle \in V$ and $|w_i\rangle \in W$, satisfying the following properties:*

P1. $z(|v\rangle \otimes |w\rangle) = (z|v\rangle) \otimes |w\rangle = |v\rangle \otimes (z|w\rangle)$, $z \in C$, $|v\rangle \in V$ *and* $|w\rangle \in W$;

P2. $(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle$, $|v_1\rangle, |v_2\rangle \in V$ *and* $|w\rangle \in W$;

P3. $|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle$, $|v\rangle \in V$, $|w_1\rangle, |w_2\rangle \in W$.

If A and B are linear operators in V and W, respectively, then

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A|v\rangle \otimes B|w\rangle, \tag{3.16}$$

where $|v\rangle \in V$ and $|w\rangle \in W$. Naturally,

$$(A \otimes B)\left(\sum_i a_i|v_i\rangle \otimes |w_i\rangle\right) \equiv \sum_i a_i A|v_i\rangle \otimes B|w_i\rangle, \tag{3.17}$$

for $a_i \in C$, $|v_i\rangle \in V$ and $|w_i\rangle \in W$.

Depending on the context, notations to tensor product of operators and vectors can vary. The following notations will be used in this thesis. If A and B are linear operators in V and W, respectively, then

$$A \otimes B \equiv A_V B_W. \tag{3.18}$$

We often use the abbreviated form $|v\rangle \otimes |w\rangle \equiv |v\rangle|w\rangle \equiv |v, w\rangle \equiv |vw\rangle$. Therefore, if A is an operator acting in V and B acting in W, the following equations are equivalent:

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A_V B_W|v\rangle|w\rangle \equiv A_V B_W|vw\rangle. \tag{3.19}$$

The inner product in $V \otimes W$ is defined in a natural way, i.e., in terms of inner products in V and W, respectively.

$$\left(\sum_i a_i|v_i\rangle \otimes |w_i\rangle, \sum_j b_j|v'_j\rangle \otimes |w'_j\rangle\right) \equiv \sum_{ij} a_i^* b_j \langle v_i|v'_j\rangle\langle w_i|w'_j\rangle, \tag{3.20}$$

where $a_i, b_j \in C$, $|v_i\rangle, |v'_j\rangle \in V$ and $|w_i\rangle, |w'_j\rangle \in W$. From the definition of inner product, one can verify that if $|v_i\rangle$ and $|w_i\rangle$ are two orthonormal basis for V and W, respectively, then the product $|v_i\rangle \otimes |w_i\rangle$ is an orthonormal basis for $V \otimes W$.

In terms of matrix representation, the tensor product between operators A and B is equivalent to the Kronecker product between such matrices. Therefore, if the orders of matrices A and B are $m \times n$ and $p \times q$, respectively, then

$$A \otimes B \equiv \overbrace{\begin{bmatrix} A_{11}B & A_{12}B & \dots & A_{1n}B \\ A_{21}B & A_{22}B & \dots & A_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1}B & A_{m2}B & \dots & A_{mn}B \end{bmatrix}}^{nq} \left.\vphantom{\begin{bmatrix} A \\ A \\ \vdots \\ A \end{bmatrix}}\right\} mp. \tag{3.21}$$

It is easy to verify that transposition and complex conjugation are distributive with relation to the tensor product. Moreover, the tensor product: (a) of two unitary matrices is a unitary matrix; (b) of two Hermitian matrices is a Hermitian matrix; (c) of two positive operators is a positive operator; (d) of two projectors is a projector.

Finally, the notation $|\psi\rangle^{\otimes n}$ is often used to denote the n-tensor product of $|\psi\rangle$, e.g., $|\psi\rangle^{\otimes 3} = |\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle = |\psi\psi\psi\rangle$.

3.3 Quantum mechanics postulates

In this section we briefly review the four postulates of quantum mechanics. A more detailed approach can be found in Nielsen and Chuang [2].

3.3.1 State space

The first postulate establishes the mathematical environment where quantum systems are defined. Such framework is the already mentioned Hilbert space.

Postulate 1 *Associated to any isolated quantum system is a complex vector space with inner product, i.e., a Hilbert space, called state space of the quantum system. The state of a quantum system is completely described by their state vector, which is a unitary vector belonging to the state space of the system.*

The simplest quantum system is the *qubit*, which is a reference to *quantum bit*. The qubit belongs to a state space of dimension two. Therefore, any qubit can be written as

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{3.22}$$

where a, b are complex numbers and $|0\rangle$, $|1\rangle$ are defined in Equation (3.5). The postulate imposes unitary norm to $|\psi\rangle$, $\langle\psi|\psi\rangle = 1$, which means $|a|^2 + |b|^2 = 1$.

In quantum information and computation, the states $|0\rangle$ and $|1\rangle$ are, intuitively, analogous to classical bits 0 and 1, respectively. The main, fundamental difference is that the states $|0\rangle$ and $|1\rangle$ can coexist in a same system $|\psi\rangle$. This property is known as superposition: $|\psi\rangle = a|0\rangle + b|1\rangle$. The linear combination $\sum_i \alpha_i|\psi_i\rangle$ is referred to a superposition of states $|\psi_i\rangle$ with amplitudes α_i.

3.3.2 Evolution

The time evolution of a closed quantum system is the subject of the next postulate.

Postulate 2 *The evolution of an isolated quantum system is described by unitary transformations. The system state $|\psi_1\rangle$ at the time t_1 is related to $|\psi_2\rangle$, the system state at the time t_2, by means of a unitary operator U, which only depends on times t_1 and t_2,*

$$|\psi_2\rangle = U|\psi_1\rangle. \tag{3.23}$$

The Austrian physicist Ervin Schrödinger, in his formulation of quantum mechanics, described the continuos time evolution of a closed quantum system by a differential equation. The continuos time version of Postulate 2 is presented below.

Postulate 2' *The time evolution of an isolated quantum system $|\psi\rangle$ is described by the Schrödinger equation,*

$$i\hbar\frac{d|\psi\rangle}{dt} = H|\psi\rangle, \tag{3.24}$$

where $\hbar$ is the Planck constant and H is an Hermitian operator known as the Hamiltonian of the closed quantum system.

One can readily verify [2] that the two enunciates of Postulate 2 are equivalent.

3.3.3 Measurements

When a quantum system does not interact with the *external world*, its evolution is completely described by unitary operations. In order to obtain some *information* about the system, the experimentalist should introduce an external device which makes the system no longer closed, and thus not necessarily subjected to unitary evolution. Postulate 3 explains the behaviour of quantum systems when they are submitted to measurements.

Postulate 3 *Measurements in quantum systems are described by a set of measurement operators $\{M_m\}$ acting on the state space of the system being measured. If the state of the quantum system before the measurement is $|\psi\rangle$, then the probability that outcome m occurs is given by*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle. \tag{3.25}$$

The state of the system after the measurement will be

$$|\psi'\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{3.26}$$

Because probabilities sum to one, measurement operators must satisfy the completeness equation

$$\sum_m M_m^\dagger M_m = \mathbb{1}. \tag{3.27}$$

Postulate 3 is the most general description of a quantum measurement. Many physicists are unfamiliar with it, specially the experimentalists. The main reason is because they do not know how to implement such measurements using physical devices. There are two special cases of general measurements which are important to our work: projective and Positive Operator-Valued Measurements (POVM).

Projective measurements. *A projective measurement, also called a von Neumann measurement, is described by an* observable M*, which is a Hermitian operator on the state space of the system being measured. The observable has a spectral decomposition*

$$M = \sum_m \lambda_m P_m, \tag{3.28}$$

where P_m *is a projector onto the eigenspace of* M *with eigenvalue* λ_m*. Measurement outcomes correspond to eigenvalue indices* m*. When a system in a state* $|\psi\rangle$ *is observed, the probability of get outcome* m *is given by*

$$p(m) = \langle\psi|P_m|\psi\rangle. \tag{3.29}$$

Given that the outcome m *occurred, the state of the system immediately after the measurement will be*

$$|\psi'\rangle = \frac{P_m|\psi\rangle}{\sqrt{p(m)}}. \tag{3.30}$$

Instead of given an observable to describe a von Neumann measurement, one can simply construct a list of projectors P_m satisfying $\sum_m P_m = \mathbb{1}$ and $P_i P_j = \delta_{ij} P_i$, i.e., projectors must be pairwise orthogonal. The corresponding observable is then $M = \sum_m m P_m$. We say that a quantum system is "measured in a basis $|m\rangle$", where $|m\rangle$ is an orthonormal basis, when a projective measurement with projectors $P_m = |m\rangle\langle m|$ is performed.

As an example, let P_{+1} and P_{-1} be two projectors such that

$$P_{+1} = |+\rangle\langle+| = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad P_{-1} = |-\rangle\langle-| = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{3.31}$$

Note that because $P_{+1} + P_{-1} = \mathbb{1}_2$, the set $\mathcal{P} = \{P_{+1}, P_{-1}\}$ defines a quantum projective measurement. Suppose we are measuring the state $|0\rangle$ using $\mathcal{P}$. The probability of getting outcomes $+1$ and -1 are, respectively,

$$p(+1) = \langle\psi|+\rangle\langle+|\psi\rangle = 1/2 \quad \text{and} \quad p(-1) = \langle\psi|-\rangle\langle-|\psi\rangle = 1/2. \tag{3.32}$$

Given that outcome +1 occurs, the post measurement state will be

$$\frac{P_{+1}|0\rangle}{\sqrt{p(+1)}} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{3.33}$$

Instead, if the experimentalist gets outcome −1, the post measurement state will be

$$\frac{P_{-1}|0\rangle}{\sqrt{p(-1)}} = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{3.34}$$

Alternatively, this is equivalent to perform a measurement of the observable (Pauli operator) X on the state $|0\rangle$, since $X = (+1)P_{+1} + (-1)P_{-1}$.

POVM Measurements. *Consider a quantum measurement as described in postulate 3, with measurement elements* $\{M_m\}$. *Define*

$$E_m \equiv M_m^\dagger M_m. \tag{3.35}$$

POVM measurements are defined by a set of POVM operators $\{E_m\}$, *where* E_m *are positive operators satisfying* $\sum_m E_m = \mathbb{1}$. *The probability of get outcome* m *given that the state* $|\psi\rangle$ *is measured is*

$$p(m) = \langle\psi|E_m|\psi\rangle. \tag{3.36}$$

The set $\{E_m\}$ *is often called a POVM.*

Differently than general and projective measurements, we are not able to predict the state of the post measurement quantum system once a POVM measurement is performed. Fortunately, most of the applications in quantum computation and information theory does not care about post measurement states. Instead, we are often interested in measurement outcomes and the corresponding associated probabilities. For example, in quantum error correction theory, the received codeword is subjected to projective measurements - which are a special case of POVM measurements; outcomes correspond to error syndromes, which are used in order to choose unitary operations, whose application on the received state can recovery the transmitted state. Figure 3.1 illustrates a POVM measurement apparatus. When an unknown quantum state ρ is measured, a led turns on to indicate the outcome.

3.3.4 Composite quantum systems

Individual quantum systems can interact to produce composite quantum systems. The following postulate determines the state space of the composite system as a tensor product of individual state spaces.

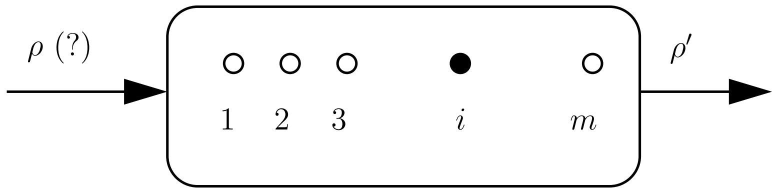

Figure 3.1: A POVM measurement apparatus. When a quantum state is measured using a set $\{E_1, \ldots, E_m\}$, a led is turned on indicating the outcome.

Postulate 4 *The state space of a composite quantum system is the tensor product of the state spaces of the indivudual physical systems. Moreover, if n systems are prepared in the state $|\psi_i\rangle$, then the joint system state is $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.*

We should use any of the following equivalent notations for representing composite systems: $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle \equiv |\psi_1\rangle|\psi_2\rangle \ldots |\psi_n\rangle \equiv |\psi_1\psi_2 \ldots \psi_n\rangle$.

Postulate 4 allows the definition of one of the most interesting concepts in quantum mechanics - *entanglement*. By definition, a composite systems is said to be entangled if we can not write the state of the whole system as a tensor product of states in each of the individual systems. For example, consider the two qubit state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \tag{3.37}$$

This state is entangled, since there are no single qubit states $|a\rangle$ and $|b\rangle$ for which $|\psi\rangle = |a\rangle|b\rangle$.

The Bell basis is a set of four entangled states that forms a basis for the 4-dimensional Hilbert space:

$$|\beta_{00}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{3.38}$$

$$|\beta_{01}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \tag{3.39}$$

$$|\beta_{10}\rangle = \frac{|00\rangle - |01\rangle}{\sqrt{2}}, \tag{3.40}$$

$$|\beta_{11}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{3.41}$$

The Bell basis plays an important hole in quantum computation and information applications. States in the Bell basis are also known as Einstein, Podolsky and Rosen (EPR) pairs.

3.4 The density operator

Until now the state of a quantum system has been represented by a unitary vector in an appropriated Hilbert space. Such systems are said to be in a *pure* state . They suggest a situation of minimum ignorance, where there is nothing more to be determined but the system state itself. However, there are situations where such formalism does not apply. In particular:

- with ensembles $\mathcal{F}$, where the system can be in any of the pure states $|\psi_1\rangle, |\psi_2\rangle, \ldots$, with probabilities $p_1, p_2, \ldots$;
- in a situation where the system (called A) is part of a larger system AB which is in a pure, entangled state Ψ.

Quantum systems in any of the states above are said to be in a *mixed* state. The mathematical formalist to deal with these situations is the *density operator*:

Definition 13 (Density operator [2]) *Assume that a quantum system is in some state $|\psi_i\rangle$ with probability p_i. The density operator describing the state of the system is defined as being*

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{3.42}$$

The density operator if often called the density matrix of the system. Density operators are well characterized matrices.

Theorem 6 (Characterization of density operators [2]) *An operator ρ is a density matrix associated with an ensemble $\{p_i, |\psi_i\rangle\}$ if and only if the following are true:*

1. (Trace condition) *$tr[\rho] = 1$, where $tr[\rho]$ stands for the trace of the operator ρ;*

2. (Positivity) *ρ is a positive operator.*

It is straightforward to see that the density matrix of a pure system is $\rho = |\psi\rangle\langle\psi|$, which is cleary a trace one matrix. Given a density operator ρ of an unknown quantum system, how can we infer whether the system is in a pure or mixed state? It turns out that the system is in a pure state if and only if $\mathrm{tr}[\rho^2] = 1$. In fact, $\mathrm{tr}[\rho^2] \leq 1$ with equality if and only if ρ is pure. Since this result plays an important role in this work, we demonstrate it below.

If ρ is a density operator, then $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$. Moreover

$$\rho^2 = \sum_i \lambda_i^2 |\psi_i\rangle\langle\psi_i|.$$

By the trace condition, $\sum_i \lambda_i = 1$. Since $0 \leq \lambda_i \leq 0$ and $\lambda_i^2 \leq \lambda_i$, we have

$$\begin{aligned} \mathrm{tr}\left[\rho^2\right] &= \mathrm{tr}\left[\sum_i \lambda_i^2 |\psi_i\rangle\langle\psi_i|\right] \\ &= \sum_i \lambda_i^2 \mathrm{tr}\left[|\psi_i\rangle\langle\psi_i|\right] \\ &= \sum_i \lambda_i^2 \\ &\leq \sum_i \lambda_i \\ &= 1. \end{aligned} \quad (3.43)$$

If ρ is a pure state, then $\rho = |\psi\rangle\langle\psi|$ and

$$\begin{aligned} \mathrm{tr}\left[\rho^2\right] &= \mathrm{tr}\left[|\psi\rangle\langle\psi|\right] \\ &= \langle\psi|\psi\rangle \\ &= 1. \end{aligned} \quad (3.44)$$

Conversely, if $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ is a state such that $\mathrm{tr}\left[\rho^2\right] = 1$, then

$$\sum_i \lambda_i^2 = 1.$$

Such condition is verified if and only if $\lambda_k = 1$ and $\lambda_{i \neq k} = 0$. Therefore,

$$\rho = |\psi_k\rangle\langle\psi_k|$$

is a pure state.

Vector and density matrix formalisms are equivalent. Hence, one can enunciate the four postulates of quantum mechanics in terms of density operators.

3.4.1 Quantum mechanics postulates and density operators

We revisit the four postulates of Section 3.3.

Postulate 1: Associated with any quantum system is a complex vector space with inner product (i.e., a Hilbert space), called state space of the system. The system state is completely described by its density operator, which is a trace one positive operator ρ acting on the state space of the system. If the quantum system is in the state ρ_i with probability p_i, then the density operator of the system is $\rho = \sum_i p_i \rho_i$.

Postulate 2: The evolution of a closed quantum system is described by unitary transformations. The state ρ_1 of the system at t_1 is related to the system state ρ_2 at time t_2 by means of a unitary operation U, which depends only on times t_1 and t_2,

$$\rho_2 = U \rho_1 U^\dagger. \quad (3.45)$$

Postulate 3: Measurements in quantum systems are defined by a set $\{M_m\}$ of measurement operators. Operators M_m act on the state space of the system being measured. Indices m are the measurement outcomes. If the system state rather before the measurement is ρ, then the probability that the outcome m occurs is

$$p(m) = \mathrm{tr}\left[M_m^\dagger M_m \rho\right]. \tag{3.46}$$

The state of the system immediately after the measurement is

$$\rho' = \frac{M_m \rho M_m^\dagger}{\mathrm{tr}\left[M_m^\dagger M_m \rho\right]}. \tag{3.47}$$

Measurement operators satisfy the completeness relation

$$\sum_m M_m^\dagger M_m = \mathbb{1}. \tag{3.48}$$

Postulate 4: The state space of a composite quantum system is the tensor product of the individual state spaces. Moreover, if we have n quantum systems, namely 1 to n, and the system i is prepared in the state ρ_i, then the whole state of the composite system is $\rho_i \otimes \rho_2 \otimes \cdots \otimes \rho_n$.

As we already pointed out, the formulation in terms of density operators is equivalent to the formulation in terms of state vectors. For example, assume that the evolution of a closed quantum system is described by the unitary operator U. If the system i is initially in the state $|\psi_i\rangle$ with probability p_i, then the post evolution state of the system will be $U|\psi_i\rangle$ with probability p_i. Therefore, the evolution of the density operator will be

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \xrightarrow{U} \sum_i p_i U|\psi_i\rangle\langle\psi_i| U^\dagger = U\rho U^\dagger. \tag{3.49}$$

3.5 Conclusions

We have given in this chapter an overview of the main aspects of quantum mechanics, which are important to the best understanding of this thesis. We have started by defining a Hilbert space and linear operators. Then we defined Hermitian and unitary operators, as well as tensor products. In the second part of the chapter, we presented the four postulates of quantum mechanics based on the Heisenberg formulation and using the Dirac's notation.

In the next chapter, we introduce some capacities of quantum channels for transmitting classical information, which is the main subject of this thesis.

Chapter 4

Quantum channel capacities

4.1 Introduction

Given a noisy quantum channel, the maximum amount of classical information per channel use Alice can transmit to Bob is called the classical capacity of the quantum channel. As we already discussed in Section 2.1, the capacity depends on the communication protocol and on available physical resources, such as entanglement.

In this chapter we begin by introducing the von Neumann entropy and the mathematical framework to describe quantum channels. Then, we present an overview of quantum channel capacities for transmitting classical information. We emphasize that all capacities discussed here allow for an asymptotically small probability of error whenever code rates approach the channel capacity, even if the best coding scheme is used.

4.2 Von Neumann entropy and quantum channels

4.2.1 The von Neumann entropy

The Shannon entropy measures the uncertainty associated with a probability distribution. Quantum states are described in a similar way, where density operators replace the distributions. In this section we introduce the von Neumann entropy [2, pp. 510], which is a generalisation of the Shannon entropy for quantum states.

The von Neumann entropy of a quantum state ρ is defined as

$$S(\rho) \equiv -\mathrm{tr}\left[\rho \log \rho\right]. \tag{4.1}$$

In Equation (4.1), the logarithm is taken to base 2. The logarithm of the operator ρ is calculated by taking its spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, where $\log \rho =$

$\sum_i \log(\lambda_i)|\psi_i\rangle\langle\psi_i|$. Because λ_i are eigenvalues of ρ and $\{|\psi_i\rangle\}$ forms an orthonormal set, the von Neumann entropy can be written as

$$S(\rho) = -\text{tr}\left[\sum_i \lambda_i|\psi_i\rangle\langle\psi_i| \sum_j \log\lambda_j|\psi_j\rangle\langle\psi_j|\right] \tag{4.2}$$

$$= -\text{tr}\left[\lambda_i \log\lambda_i|\psi_i\rangle\langle\psi_i|\right] \tag{4.3}$$

$$= -\sum_i \lambda_i \log\lambda_i, \tag{4.4}$$

where $0\log 0 \equiv 0$. In a Hilbert space of dimension d, the maximum of the von Neumann entropy is $\log d$, corresponding to the quantum state $\rho = \mathbb{1}_d/d$. In this case, we have a maximum ignorance about the state of the system, which we call of *completely depolarized* system.

The relative entropy between two quantum states ρ and σ is defined in a similar fashion to the relative entropy between two probability distributions,

$$S(\rho||\sigma) \equiv \text{tr}\left[\rho\log\rho\right] - \text{tr}\left[\rho\log\sigma\right]. \tag{4.5}$$

As in the classical case, the relative entropy can be infinite. In particular, the relative entropy is $+\infty$ if the kernel of σ (the vector space spanned by eigenvectors of σ with eigenvalues 0) has a non-trivial intersection with the support of ρ, the vector space spanned by eigenvectors of ρ with nonzero eigenvalues. Otherwise, the relative entropy is finite. Moreover, the relative entropy is non-negative, $S(\rho||\sigma) \geq 0$.

The von Neumann entropy has some interesting properties [2]:

(1) The entropy is non-negative. $S(\rho)$ is zero if and only if ρ is a pure state.

(2) In a Hilbert space of dimension d, the entropy takes its maximum value $\log d$. The state for which $S(\rho) = \log d$ is $\rho = \mathbb{1}_d/d$, and corresponds to the completely depolarized state.

(3) Assume that the composite system AB is in a pure state. Then $S(A) = S(B)$.

(4) Suppose that p_i are probabilities and ρ_i have their support on orthogonal subspaces. Then

$$S\left(\sum_i p_i\rho_i\right) = H(p) + \sum_i p_i S(\rho_i). \tag{4.6}$$

By analogy with the Shannon entropy, it is possible to define the joint and conditional von Neumann entropies, as well as mutual information for composite systems. The joint entropy $S(A,B)$ of a composite quantum system AB is defined as

$$S(A,B) = -\text{tr}\left[\rho^{AB}\log\rho^{AB}\right], \tag{4.7}$$

where ρ^{AB} is the density operator of the system AB. The conditional entropy and the mutual information are defined respectively as

$$S(A|B) \equiv S(A,B) - S(B), \quad (4.8)$$

$$S(A:B) \equiv S(A) + S(B) - S(A,B) \quad (4.9)$$

$$= S(A) - S(A|B) = S(B) - S(B|A). \quad (4.10)$$

An useful result is the subadditivity of the entropy [2, pp.515],

$$S(A,B) \leq S(A) + S(B), \quad (4.11)$$

with equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$. Besides properties already mentioned, the von Neumann entropy has many others that can be found in texbooks [2].

4.2.2 Quantum channels

The time evolution of a closed quantum system ρ is completely described by unitary operators. If the system remains closed, it is always possible to return to the initial system state. Suppose that a closed quantum system interacts in some way with an open system, here called *environment* . Additionally, suppose that after the interaction the system becomes closed again. We denote by $\mathcal{E}(\rho)$ the state of the system after interaction. In general, the final state $\mathcal{E}(\rho)$ can not be related by a unitary transformation to the initial state ρ. The formalism used to deal with such situation is known as quantum operations. A quantum operation is a map $\mathcal{E}$ from the set of operators of the input space $\mathcal{H}_1$ to the set of operators of the output state space $\mathcal{H}_2$ with the following properties: (for simplicity we consider $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$.) [2, pp. 367]

1. $\mathrm{tr}\,[\mathcal{E}(\rho)]$ is the probability that the process represented by $\mathcal{E}$ occurs, when ρ is the initial state. Thus, $0 \leq \mathrm{tr}\,[\mathcal{E}(\rho)] \leq 1$ for any state ρ.

2. $\mathcal{E}$ is a convex-linear map on the set of density operators, that is, for probabilities p_i,

$$\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i). \quad (4.12)$$

3. $\mathcal{E}$ is a completely positive map. That is, if $\mathcal{E}$ maps density operators of system $\mathcal{H}_1$ to density operators of system $\mathcal{H}_2$, then $\mathcal{E}(A)$ must be positive for any positive operator A. Furthermore, $(\mathcal{I} \otimes \mathcal{E})(B)$ must be positive for any positive operator B on a composite system $R\mathcal{H}_1$, where $\mathcal{I}$ denotes the identity map on R.

The proof of the next theorem can be found in Nielsen and Chuang [2, pp. 368].

Theorem 7 *A map $\mathcal{E}$ satisfies properties 1, 2 and 3 if*

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger, \tag{4.13}$$

for some set of operators $\{E_i\}$ from the input Hilbert space to the output Hilbert space, and $\sum_i E_i^\dagger E_i \leq \mathbb{1}$.

Quantum operations for which $\sum_i E_i^\dagger E_i$ is strictly less than the identity are non-trace-preserving maps. This means a map that takes trace one density matrices into operators such that $\mathrm{tr}\,[\mathcal{E}(\rho)] < 1$. The class of non-trace-preserving maps are particularly useful to describe process in which extra information about what occurred in the interaction is obtained by measurement.

To model a quantum channel, it is required that the map $\mathcal{E}$ takes a valid density operator ρ into another valid one $\mathcal{E}(\rho)$. Hence, quantum channels form a class of maps called *completely positive trace-preserving quantum operations*, which are completely positive maps that preserve the trace of operators,

$$\begin{aligned} 1 &= \mathrm{tr}\,[\rho] && (4.14)\\ &= \mathrm{tr}\,[\mathcal{E}(\rho)] && (4.15)\\ &= \mathrm{tr}\left[\sum_i E_i \rho E_i^\dagger\right] && (4.16)\\ &= \mathrm{tr}\left[\sum_i E_i^\dagger E_i \rho\right]. && (4.17) \end{aligned}$$

Since this relationship is true for all ρ, we must have

$$\sum_i E_i^\dagger E_i = \mathbb{1}. \tag{4.18}$$

Equation (4.13) is known as the operator-sum representation of the quantum channel $\mathcal{E}$. Operators in $\{E_i\}$ are called operation elements.

As an example, consider the *depolarizing* channel. In a 2-dimensional Hilbert space, this channel leaves a qubit intact with probability p and replaces the input state by a completely depolarized state $\frac{1}{2}\mathbb{1}_2$ with probability $1-p$:

$$\mathcal{E}(\rho) = p\rho + (1-p)\frac{1}{2}\mathbb{1}_2. \tag{4.19}$$

Cleary, the map above is not in the operator-sum representation. However, for any qubit ρ,

$$\frac{\mathbb{1}_2}{2} = \frac{\rho + X\rho X + Y\rho Y + Z\rho Z}{4}, \tag{4.20}$$

where X, Y and Z are Pauli operators. Therefore, the operator-sum representation of the depolarizing channel is

$$\mathcal{E}(\rho) = \left(\frac{1}{4} + \frac{3}{4}p)\right)\rho + \frac{1-p}{4}(X\rho X + Y\rho Y + Z\rho Z), \tag{4.21}$$

with operation elements given by

$$\left\{\sqrt{\frac{1}{4} + \frac{3p}{4}}\mathbb{1}_2, \frac{\sqrt{1-p}}{2}X, \frac{\sqrt{1-p}}{2}Y, \frac{\sqrt{1-p}}{2}Z\right\}. \tag{4.22}$$

As we have seen, a quantum channel is defined for an input mixed state ρ. However, we can always represent a pure state $|\psi\rangle$ using the density operator formalism, $\rho = |\psi\rangle\langle\psi|$. Therefore, the output of the channel for an input pure state will be $\mathcal{E}(|\psi\rangle\langle\psi|) = \sum_i E_i|\psi\rangle\langle\psi|E_i^\dagger$. For the sake of brevity, we should write $\mathcal{E}(|\psi\rangle)$ instead of $\mathcal{E}(|\psi\rangle\langle\psi|)$.

4.3 Accessible information and the Holevo bound

Consider a classical source emitting symbols $\mathcal{X} = 1, \ldots, n$ with probabilities $p_1, \ldots, p_n$. Suppose that symbols emitted by the source are used by Alice to prepare quantum states $\rho_1, \ldots, \rho_n$. After preparation, Alice gives the quantum state to Bob, which is allowed to perform individual measurements aiming to infer the symbol emitted by the source. Define X and Y as being random variables representing the classical source and measurement outcomes, respectively. The *accessible information* [3, 4, 5] is defined as the maximum of the mutual information $I(X;Y)$, where the maximum is taken over all measurement schemes:

$$I_{acc} = \max_{\{M_m\}} I(X;Y). \tag{4.23}$$

In classical information theory, the accessible information is not interesting, since in principle it is always possible to distinguish between classical states (e.g. two voltage levels). In contrast, quantum mechanics does not allow for perfectly distinguishing arbitrary quantum states. For example, if Alice prepares two non-orthogonal states $|\psi\rangle$ and $|\varphi\rangle$ with probabilities p and $1-p$, respectively, then the accessible information is strictly less than H_p, where $H_p = p\log p - (1-p)\log(1-p)$ stands for the binary Shannon entropy.

A very useful result in quantum information theory is the Holevo bound.

Theorem 8 (Holevo bound [18]) *Consider a quantum memoryless source and an ensemble $\{p_i, \rho_i\}$ of quantum states. Suppose that the source emits ρ_i with probabilities p_i. Define*

$$\chi = S(\rho) - \sum_i p_i S(\rho_i), \tag{4.24}$$

where $\rho = \sum_i p_i \rho_i$. *Then,*

$$I_{acc} \leq \chi. \tag{4.25}$$

The real number χ is known as Holevo quantity, and it is an upper bound on the accessible information. In terms of POVM measurements, the Holevo bound can be enunciated in the following way:

Theorem 8 (Holevo bound [18]) *Suppose that Alice prepares states* ρ_x*, where* $\mathcal{X} = 1, \ldots, n$*, with probabilities* $p_1, \ldots, p_n$*. Alice gives Bob a particular state to be measured according to a POVM* $\{E_{\mathcal{Y}}\} = \{E_1, \ldots, E_m\}$*. Measurement outcomes are represented by the random variable* Y*. The Holevo bound asserts that, for any measurement chosen by Bob:*

$$I(X;Y) \leq S(\rho) - \sum_x p_x S(\rho_x), \tag{4.26}$$

where $\rho = \sum_x p_x \rho_x$. The equality holds since all quantum states ρ_x commutes [2, pp. 77].

The $C_{1,1}(\mathcal{E})$ capacity of a quantum channel, often called one-shot capacity, is defined below.

Definition 14 ($C_{1,1}(\mathcal{E})$ capacity [18, 19]) *Let* $\mathcal{E}$ *be a quantum channel as stated in Section 4.2.2. The* $C_{1,1}$ *capacity of* $\mathcal{E}$ *is defined as the maximum over all input ensembles of the accessible information of the corresponding output ensemble,*

$$C_{1,1}(\mathcal{E}) = \max_{\{\rho_x, p_x\}} I_{acc_{out}}, \tag{4.27}$$

where $I_{acc_{out}}$ *is the accessible information of the ensemble* $\{\mathcal{E}(\rho_x), p_x\}$*.*

The information transmission protocol of the $C_{1,1}$ capacity has three constrains: entangled states are not allowed between two or more uses of the channel; measurements at the channel output must be individual; and adaptive measurements are denied, i.e., Bob is not allowed to perform a "partial" measurement over the state, use such result to choose the next measurement and return to complete the first measurement. Adaptive measurements are proved to improve the $C_{1,1}$ capacity, as described in Section 4.5. The first "1" in the index of $C_{1,1}$ refer to the first restriction on the communication protocol, whereas the second "1" is due to the second restriction.

4.4 The Holevo-Schumacher-Westmoreland theorem

Consider the problem of sending classical messages randomly chosen from a set $\{1, \ldots, 2^{nR}\}$ by means of a quantum channel. Differently from the first assumption of the communication protocol of the $C_{1,1}$ capacity, Alice is allowed to prepare codewords as tensor products of quantum states $\rho_1 \otimes \rho_2 \otimes \ldots$, where each of the states $\rho_1, \rho_2, \ldots$ is chosen from an ensemble $\{p_i, \rho_i\}$. The notation $C_{1,\infty}(\mathcal{E})$ stands for the classical capacity of a quantum channel in a scenario where Alice can not use entangled states between two or more uses of the channel but Bob is allowed to perform collective measurements at the channel output. This means that Bob can wait for a number of states and measure all the states together (the "∞" in the second index of $C_{1,\infty}(\mathcal{E})$). The $C_{1,\infty}(\mathcal{E})$ capacity is the quantum analogous of the Shannon ordinary capacity.

The problem of finding the $C_{1,\infty}(\mathcal{E})$ capacity was studied simultaneously and independently by Holevo [7] and by Schumacher and Westmoreland [8]. The following result is known as the HSW theorem.

Theorem 9 (Holevo-Schumacher-Westmoreland) *The $C_{1,\infty}(\mathcal{E})$ capacity of a quantum channel $\mathcal{E}$ is*

$$C_{1,\infty}(\mathcal{E}) \equiv \max_{\{p_i,\rho_i\}} \left[S\left(\mathcal{E}\left(\sum_i p_i \rho_i\right)\right) - \sum_i p_i S(\mathcal{E}(\rho_i)) \right]. \tag{4.28}$$

The maximum is taken over all ensembles $\{p_i, \rho_i\}$ of input quantum states.

The proof of the theorem makes use of random coding and typical subspaces. A detailed demonstration can be found in Nielsen e Chuang [2, pp. 555].

As an example, consider the 2-dimensional depolarizing channel already discussed in Section 4.2.2. Consider an ensemble $\{p_j, |\psi_j\rangle\}$. Then

$$\mathcal{E}(|\psi_j\rangle\langle\psi_j|) = p|\psi_j\rangle\langle\psi_j| + (1-p)\frac{\mathbb{1}}{2}. \tag{4.29}$$

The quantum state $\mathcal{E}(|\psi_j\rangle\langle\psi_j|)$ has eigenvalues $(1+p)/2$. Therefore,

$$S(\mathcal{E}(|\psi_j\rangle\langle\psi_j|)) = H_2\left(\frac{1+p}{2}\right), \tag{4.30}$$

which does not depend on $|\psi_j\rangle$ at all. Hence, maximization of Equation (4.28) can be done by maximizing the entropy $S\left(\sum_j \mathcal{E}(|\psi_j\rangle\langle\psi_j|)\right)$. Note that if $\{|\psi_i\rangle\}$ is a set of orthonormal states, then $\sum_j \mathcal{E}(|\psi_j\rangle\langle\psi_j|) = p_i(\sum_j |\psi_i\rangle\langle\psi_i|) + (1-p)\mathbb{1}_2 = \mathbb{1}_2$, which maximizes $S\left(\sum_j \mathcal{E}(|\psi_j\rangle\langle\psi_j|)\right)$. Therefore, the HSW capacity of the qubit depolarizing channel is given by

$$C_{1,\infty}(\mathcal{E}) = 1 - H_2\left(\frac{1+p}{2}\right). \tag{4.31}$$

4.5 The adaptive capacity

The adaptive capacity of a quantum channel, defined by Shor [6], is derived from the $C_{1,1}$ capacity by varying the communication protocol. In his paper, Shor illustrated the adaptive capacity using the lifted trine states

$$T_0(\alpha) = \sqrt{1-\alpha}|000\rangle + \sqrt{\alpha}|001\rangle, \quad (4.32)$$

$$T_1(\alpha) = -\frac{1}{2}\sqrt{1-\alpha}|000\rangle + \frac{\sqrt{3}}{2}\sqrt{1-\alpha}|010\rangle + \sqrt{\alpha}|001\rangle, \quad (4.33)$$

$$T_2(\alpha) = -\frac{1}{2}\sqrt{1-\alpha}|000\rangle - \frac{\sqrt{3}}{2}\sqrt{1-\alpha}|010\rangle + \sqrt{\alpha}|001\rangle. \quad (4.34)$$

The communication protocol is similar to the $C_{1,1}$, except that Bob can perform adaptive measurements on the received states: he makes a measurement on one state which only partially reduces the quantum state, uses the outcome of this measurement to make intervening measurements on other states, and returns to make a further measurement on the reduced state of the original signal (the last measurement may depend on the outcomes of intervening measurements).

The *information rate* for a given encoding and measurement strategies is the mutual information between Alice's prepared codewords and Bob's measurement outcomes at the channel output, divided by the number of states used (channel uses) in the codeword.

Definition 15 *The adaptive capacity $C_{1,A}$ is defined to be the supremum of the information rate over all encodings and all measurement strategies that use quantum operations local to the separate states and classical computation to coordinate them.*

In his paper, Shor demonstrated that the adaptive capacity considering the lifted trine states is strictly greater that the $C_{1,1}$ capacity and less than the $C_{1,\infty}$ capacity for $\alpha > 0$. Moreover, it was shown that for any ensemble of two pure states at the channel input, the adaptive capacity is equal to the $C_{1,1}$ capacity.

4.6 Entanglement-assisted capacity

Entanglement is an amazing feature of quantum mechanics. Several protocols and applications in quantum information and computation use entanglement as a physical resource. Maybe the most interesting of such applications are teleportation and superdense coding [2, pp. 26]. In both cases, an maximally bipartite entangled state (EPR pair) is produced, possibly by a third part, and shared along two participants, Alice and Bob. Suppose that Alice has an unknown and arbitrary qubit state $|\psi\rangle$ she aims to delivery to Bob. Although Alice owns her part of an EPR pair, Alice and Bob do not have a quantum channel in order to communicate the state $|\psi\rangle$. The *teleporting protocol* makes use of local measurements and a noiseless classical channel among the two participants to teleport the Alice's state $|\psi\rangle$ to Bob. The only thing Alice should do is perform a collective measurement in the Bell basis on the state $|\psi\rangle$ and her part of the EPR pair. Then, Alice sends Bob the classical two bits corresponding to measurement outcomes. In order to get the qubit $|\psi\rangle$, Bob only needs to apply one of the four Pauli operators on his part of the EPR pair depending on the received bits. The counterpart of teleporting is superdense coding. Given that Alice and Bob have a shared EPR pair, it is shown that Alice can send Bob two classical bits coded into one qubit state. Straightforward we conclude that: (a) shared entanglement can increase the quantum capacity of a noiseless classical channel from zero to half qubit per channel use; and (b) it can duplicate the classical capacity of a noiseless quantum channel.

Bennett and his collaborators [9, 10] have demonstrated that shared entanglement can increase the classical capacity (HSW capacity) of noisy quantum channels. The entanglement-assisted capacity $C_E(\mathcal{E})$ of a noisy quantum channel is defined to the asymptotical classical information transmission rate in a scenario where an arbitrary amount of entanglement is shared between the sender and the receiver.

Definition 16 (Entanglement-assistided capacity [9]) *The entanglement-assisted capacity of a quantum channel $\mathcal{E}$ is*

$$C_E(\mathcal{E}) = \max_{\rho \in \mathcal{H}_{in}} S(\rho) + S(\mathcal{E}(\rho)) - S((\mathcal{E} \otimes \mathcal{I})(\Phi_\rho)), \tag{4.35}$$

where $\rho \in \mathcal{H}_{in}$ is a density matrix over the input states. In Equation (4.35), Φ_ρ is a pure state over the tensor product of state spaces $\mathcal{H}_{in} \otimes \mathcal{H}_R$ such that $tr_R[\Phi_\rho] = \rho$. $\mathcal{H}_{in}$ is the input state space and $\mathcal{H}_R$ is a space of reference. The third term on the right site of Equation (4.35), $S((\mathcal{E} \otimes \mathcal{I})(\Phi_\rho))$, denotes the von Neumann entropy of the purification [2, pp. 109] Φ_ρ of ρ over the reference system $\mathcal{H}_R$, half of which ($\mathcal{H}_{in}$) has been sent through the channel $\mathcal{E}$ while the other half ($\mathcal{H}_R$) has been sent through the identity channel (this corresponds to the portion of the entangled state that Bob holds at the start of the protocol).

The quantity being maximized in Equation (4.35) is denoted quantum mutual information, which is a generalisation of the Shannon mutual information to quantum systems [20]. In order to transmit information using the protocol described above, Alice and Bob "consume" entanglement. In general, $S(\rho)$ qubits of entanglement (i.e., EPR pairs) per channel use are necessary to reach the entanglement-assisted capacity.

4.7 Conclusions

We have presented in this chapter a brief overview of classical capacities of quantum channels. We have first explained the one-shot capacity $C_{1,1}$. After that, we have discussed the Holevo-Schumacher-Westmoreland capacity, which is a generalisation of the ordinary Shannon capacity. Finally, we presented the adaptive and entanglement-assisted capacities. At the beginning of the chapter, we shortly introduced the von Neumann entropy and quantum operations, which is a formalism to model interactions of closed quantum system with the environment. The next chapter is devoted to the zero-errro capacity of classical channels.

Chapter 5

Zero-error information theory

5.1 Ordinary capacity of classical channels

Consider a system A hereafter referred to as Alice, and a system B hereafter referred to as Bob. We say that Alice communicates with Bob when the physical acts of Alice have induced a desired physical state in Bob. As this transfer of information is a physical process, it is subject to the uncontrollable ambient noise and imperfections of the physical signalling process itself. The communication is successful if the receiver Bob and the transmitter Alice agree on what was sent.

The quantitative analysis of the above physical signaling system is made using a mathematical framework introduced by Claude E. Shannon in 1948 [13]. This framework includes a mathematical analog of the signaling systems shown in Figure 5.1. The encoder maps source symbols from some finite alphabet into some sequence of channel symbols, afterwards called codeword, which is sent through the channel. The channel produces an output sequence which is random but has a probability distribution that depends on the input sequence. From the output sequence, we attempt to recover the transmitted message. Two input sequences are said to be confusable when these sequences induce the same output sequence in the output. Shannon showed that we can choose a "non-confusable" subset of input sequences in a manner that with high probability, there is only one highly likely input that could have caused the particular output. Essentially, this means that we can reconstruct input sequences at the output with negligible probability of error. The maximum rate at which this can be done is called the ordinary capacity of the channel. It is convenient to define formally a discrete memoryless channel.

Definition 17 (Discrete memoryless channel [20]) *Consider an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$. A classical discrete memoryless channel (DMC) $C : \mathcal{X} \to \mathcal{Y}$,*

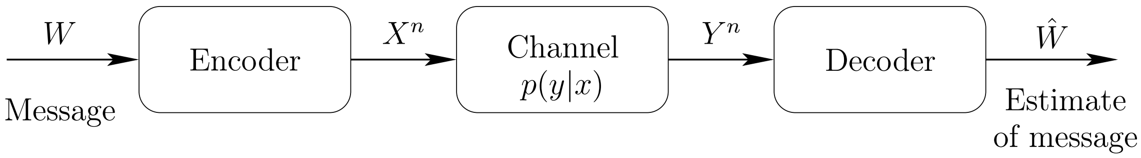

Figure 5.1: A classical communication system.

denoted by $(\mathcal{X}, p(y|x), \mathcal{Y})$*, is defined by a stochastic matrix whose rows are indexed by the elements of the finite set* $\mathcal{X}$*, while the columns are indexed by those of another finite set* $\mathcal{Y}$*. The* (x, y)*th element of the stochastic matrix is the probability* $p(y|x)$ *that* $y \in \mathcal{Y}$ *is received when* $x \in \mathcal{X}$ *is transmitted. The channel is said to be memoryless if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous inputs or outputs.*

Definition 18 (Information capacity) *The information capacity of a classical discrete channel is given by*

$$C = \max_{p(x)} I(X;Y), \tag{5.1}$$

where the maximum is taken over all input distributions $p(x)$*.* $I(X;Y)$ *stands for the mutual information between random variables* X *and* Y *representing the input and output of the DMC, respectively.*

Example 1 (Binary erasure channel) *The Binary Erasure Channel (BEC) is illustrated in Figure 5.2. When a bit is transmitted through this channel, it is received unchanged with probability* $1-p$ *or it is lost (erased) with probability* p*. The BEC has two inputs* $\mathcal{X} = \{0, 1\}$ *and three outputs* $\mathcal{Y} = \{0, \Delta, 1\}$*, where the symbol* Δ *represents an erasure. The capacity of the binary erasure channel is calculated as follows:*

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} [H(Y) - H(Y|X)] \\ &= \max_{p(x)} H(Y) - H_p, \end{aligned} \tag{5.2}$$

where H_p *stands for the binary entropy . The output distribution* $p(y)$ *depends on the input distribution* $p(x)$ *for* X *in the following way: Let* $Pr\ (X = 0) = \delta$*. Then* $Pr\ (Y = 0) = (1-p)\delta$*,* $Pr\ (Y = \Delta) = p$ *and* $Pr\ (Y = 1) = (1-p)(1-\delta)$*. Therefore,*

$$\begin{aligned} C &= \max_{\delta} H((1-p)\delta, p, (1-p)(1-\delta)) - H_p \\ &= \max_{\delta} H_p + (1-p)H_\delta - H_p \\ &= \max_{\delta} (1-p)H_\delta \\ &= 1 - p, \end{aligned} \tag{5.3}$$

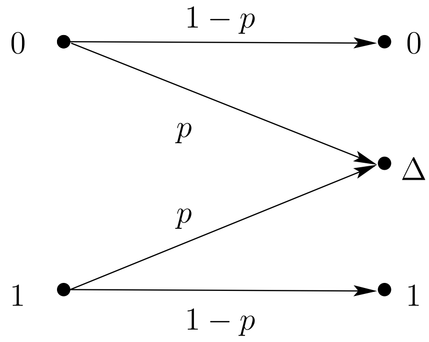

Figure 5.2: A binary erasure channel (BEC) with erasure probability p.

where the capacity is reached by $\delta = 1/2$.

Example 2 (The G_5 channel) *The discrete memoryless channel of Figure 5.3, denoted by* G_5*, will play an important role in the study of the zero-error capacity of DMC in Section 5.2. This channel models a situation in which an input symbol* $i \in \{0, \ldots, 4\}$ *is either received unchanged at the output with probability* $\frac{1}{2}$ *or it is transformed into the next symbol* $i + 1 \mod 5$ *with probability* $\frac{1}{2}$*. The ordinary capacity of the* G_5 *DMC is given by*

$$\begin{aligned} C(G_5) &= \max_{p(x)}[H(X) - H(X|Y)] \\ &= \log 5 - \log 2 \\ &= \log 5/2, \end{aligned} \tag{5.4}$$

where the maximum is achieved by a uniform probability distribution over the input.

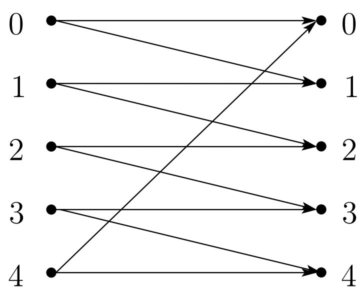

Figure 5.3: The G_5 channel.

In order to enunciate Shannon's coding theorem, we need to define an (M, n) code for the a DMC:

Definition 19 *An* (M, n) *block code for a DMC* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *is composed of the following:*

1. *A set of indexes* $\{1, \ldots, M\}$*, where each index is associated with a classical message.*

2. *An encoding function*

$$X^n : \{1, \ldots, M\} \to \mathcal{X}^n,$$

yielding codewords $\mathbf{x}^1 = X^n(1), \ldots, \mathbf{x}^M = X^n(M)$*. A codebook is the set of all codewords.*

3. *A decoding function*

$$g : \mathcal{Y}^n \to \{1, \ldots, M\},$$

which maps each received codeword on a message in the set $\{1, \ldots, M\}$*.*

The error probability of this code is $P_e = \Pr\left(g(Y^n) \neq i | X^n = X^n(i)\right)$, and its information transmission rate is $R = \frac{1}{n} \log M$ bits per symbol. The channel coding theorem guarantees the existence of codes attaining the channel capacity with an arbitrary small probability of error.

Theorem 10 (Channel coding theorem [20]) *All rates below capacity* C *are achievable, namely, there exists a sequence of codes such that the error probability goes asymptotically to zero as the code length tends to infinity. Conversely, any sequence of codes with an asymptotically small probability of error must have a rate* $R \leq C$*.*

5.2 The zero-error capacity

The channel coding theorem, presented in Section 5.1, asserts that even the best coding scheme attaining the ordinary capacity C allows for an asymptotically small but non-vanishing probability of error. From now, we will be interested in the case where no transmission errors are permitted.

Consider a classical discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. Symbols in the input and output alphabets will be hereafter called input and output symbols, respectively. Shannon [12] defined an error-free code as follows:

Definition 20 *An* (M, n) *error-free code for the DMC* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *in Figure 5.1 is composed of the following:*

1. *A set of indexes* $\{1, \ldots, M\}$*, where each index is associated with a classical message.*

2. *An encoding function*

$$X^n : \{1, \ldots, M\} \to \mathcal{X}^n,$$

yielding codewords $\mathbf{x}^1 = X^n(1), \ldots, \mathbf{x}^M = X^n(M)$*. The set of all codewords is called a codebook.*

3. A decoding function

$$g : \mathcal{Y}^n \rightarrow \{1, \ldots, M\},$$

which deterministically assigns a guess to each possible received codeword with the following property:

$$Pr\ (g(Y^n) \neq i | X^n = X^n(i)) = 0\ \forall\ i \in \{1, \ldots, M\}. \tag{5.5}$$

The only difference between Definitions 20 and 19 is the Equation (5.5) in Definition 20, which guarantees the nonexistence of decoding errors. In the zero-error context, we are particulary interested in symbols that can be fully distinguished at the channel output. They are called non-adjacent symbols.

Definition 21 (Adjacency) *Consider a DMC* $(\mathcal{X}, p(y|x), \mathcal{Y})$. *Two input symbols* $x_i, x_j \in \mathcal{X}$ *are said to be adjacent (or indistinguishable) if there exists an output symbol in* $\mathcal{Y}$ *which can be caused by either of these two, i.e., there is an* $y \in \mathcal{Y}$ *such that both* $p(y|x_i)$ *and* $p(y|x_j)$ *do not vanish. Otherwise, they are said to be non-adjacent (or distinguishable).*

Consider the sequence $\mathbf{x} = x_1 x_2 \ldots x_n$ being transmitted through a DMC. The output sequence $\mathbf{y} = y_1 y_2 \ldots y_n$ is received with probability

$$p^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i). \tag{5.6}$$

If two sequences $\mathbf{x}'$ and $\mathbf{x}''$ can both result in the sequence $\mathbf{y}$ with positive probability, then no decoder can decide with zero probability of error which of the two sequences has been transmitted by the sender. Such sequences will be called *indistinguishable* or adjacent at the receiving end of the DMC. In fact, if all input symbols in $\mathcal{X}$ are adjacent to each other, any code with more than one codeword has a probability of error great than zero. This is equivalent to say that $\mathbf{x}'$ and $\mathbf{x}''$ are distinguishable if and only if there exists at least one i, $1 \leq i \leq n$, such that x'_i and x''_i are non-adjacent, as illustrated in Figure 5.4.

$$\begin{aligned} \mathbf{x}' &= x'_1 x'_2 \ldots \ x'_i \ \ldots x'_{n-1} x'_n \\ \mathbf{x}'' &= x''_1 x''_2 \ldots \ x''_i \ \ldots x''_{n-1} x''_n \end{aligned}$$

Figure 5.4: Two distinguishable sequences $\mathbf{x}'$ and $\mathbf{x}''$: for at least one i, $1 \leq i \leq n$, the input symbols x'_i and x''_i are non-adjacent.

It is useful to think of probability distributions $p(y|x)$ and $p^n(\cdot|\mathbf{x})$ as vectors of dimension $|\mathcal{X}|$ and $|\mathcal{X}|^n$, respectively. Using this approach, we can restate the statement given earlier by saying that two sequences $\mathbf{x}', \mathbf{x}'' \in \mathcal{X}^n$ are distinguishable at the receiving end of the DMC channel if and only if the corresponding vectors $p^n(\cdot|\mathbf{x}')$ and $p^n(\cdot|\mathbf{x}'')$ are orthogonal.

Definition 22 (Zero-error capacity.) *Define $N(n)$ as the maximum cardinality of a set of mutually orthogonal vectors among the $p^n(\cdot|\mathbf{x})$, $\mathbf{x} \in \mathcal{X}^n$. The zero-error capacity of the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ is given by*

$$C_0 = \limsup_{n\to\infty} \frac{1}{n} \log N(n). \tag{5.7}$$

Intuitively, C_0 is the bit-per-symbol error-free information transmission rate capability of the channel.

The number $N(n)$ in Equation (5.7) is super multiplicative, i.e.,

$$N(n+m) \geq N(n) \cdot N(m). \tag{5.8}$$

To verify this, let $\mathbf{x}'$ and $\mathbf{x}''$ be sequences of length n and m, respectively. Then, there exist at least $N(n) \cdot N(m)$ non-adjacent sequences of length $n+m$, obtained by concatenating sequences of length n with sequences of length m. Hence, we can use the Fekete's lemma (see [28, pp. 85]) to demonstrate that the limit superior in Equation (5.7) is a true limit and actually coincides with the supremum of numbers $\frac{1}{n} \log N(n)$.

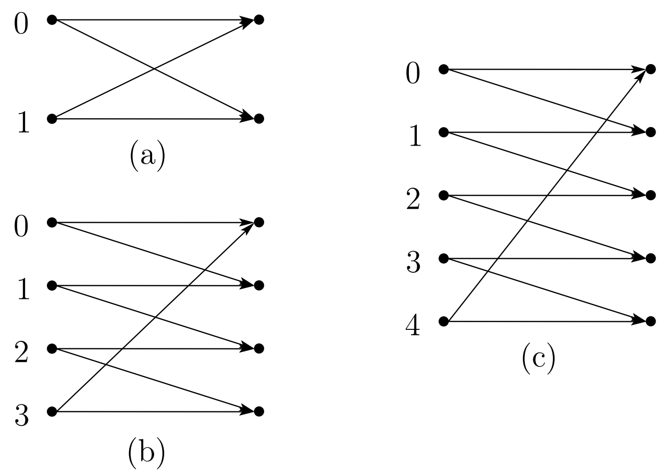

Figure 5.5: Some discrete memoryless channels. Since we are interested on adjacency relations, we omit the transition probabilities.

Shannon pointed out that the zero-error capacity of a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ depends only on which symbols in $\mathcal{X}$ are adjacent to each other. This is a major difference between the error-free capacity and the ordinary capacity of Definition 18, since in the latter the capacity depends on the choice of the probability distribution of the input symbols $\mathcal{X}$. It is easy to demonstrate that a DMC $(\mathcal{X}, p(x|y), \mathcal{Y})$ has a non-vanishing error-free capacity if and only if there exist at least two non-adjacent symbols in $\mathcal{X}$. Figure 5.5 shows some discrete memoryless channels. For the binary symmetric channel with $0 < p < 1$, the two input symbols are adjacent yielding $C_0 = 0$. Both channels in Figures 5.5(b) and 5.5(c) have at most two pairs of non-adjacent symbols. For example, if we consider codewords of length one, we can perform error-free communication by choosing to send only symbols $\{0, 2\}$ or $\{1, 3\}$ of the channel in Figure 5.5(b). In this case, the rate of the code is $\log 2 = 1$ bit per channel use.

One might ask whether we can increase the transmission rate by varying the code length or whether $C_0 = \log N(1)$. It turns out that we can. Consider the sequences $\{00, 12, 24, 31, 43\}$ of length 2 for the G_5 DMC of Figure 5.5(c). Clearly, these sequences are pairwise distinguishable at the channel output and hence are codewords of an error-free code of length two. The ordinary capacity of G_5 was calculated in the Example 2. Therefore, the zero-error capacity of G_5 is lower and upper bounded by

$$\frac{1}{2}\log 5 \leq C_0(G_5) \leq \log 5/2. \tag{5.9}$$

These bounds were given by Shannon in 1956, and the problem of finding the capacity $C_0(G_5)$ remained open during twenty years until Lovász [21] gave a brilliant solution. He showed that the Shannon's lower bound was tight

$$C_0(G_5) = \frac{1}{2}\log 5.$$

We demonstrate such result in Section 5.3, where we introduce the Lovász θ function.

As we can see, the calculation of the zero-error capacity is a very difficult problem even for simple channels. Although some methods we discuss in the next sections enable the computation of the zero-error capacity of particular classes of discrete memoryless channels, the general problem remains wide open.

5.2.1 The adjacency-reducing mapping

The calculation of the zero-error capacity of simple channels can be done using the notion of *adjacency-reducing mapping*. This means a mapping $f : \mathcal{X} \to \mathcal{X}$ with the property that if x_i and x_j are not adjacent in the channel, then $f(x_i)$ and $f(x_j)$ are not adjacent. Given any error-free code for a channel, we can always apply such a mapping symbol by symbol to the code in order to obtain another error-free code, since f never produces new adjacencies. Suppose that for a given DMC the mapping f takes all symbols in $\mathcal{X}$ onto a subset $\mathcal{X}' \subset \mathcal{X}$ of the symbols no two of which are adjacent. Clearly, there are at least $|\mathcal{X}'|^n$ n-length distinguishable sequences for this channel. However, any error-free code of length n has at most $|\mathcal{X}'|^n$ sequences, given that the application of f on this code leads to a new error-free code whose alphabet contains only $|\mathcal{X}'|$ symbols. These observations imply the following theorem enunciated by Shannon:

Theorem 11 *Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ be a discrete memoryless channel. If all symbols in $\mathcal{X}$ can be mapped by an adjacency-reducing mapping f into a subset $\mathcal{X}' \subset \mathcal{X}$ of non-adjacent symbols, then $C_0 = \log|\mathcal{X}'|$.*

As an example, consider the DMC illustrated in Figure 5.5(b). Let f be a mapping with $f(0) = 0$, $f(1) = 0$, $f(2) = 2$ and $f(3) = 2$. It is easy to see that f is an adjacency-reducing mapping satisfying the condition of Theorem 11, where $\mathcal{X}' = \{0, 2\}$. Therefore, the zero-error capacity of the channel is $C_0 = \log|\mathcal{X}'| = 1$ bit per channel use. It is easy to see that we cannot construct an adjacency-reducing mapping f for the G_5. In his paper, Shannon used this theorem to find the zero-error capacity of all discrete memoryless channels up to five input symbols, except for the G_5 channel. All DMCs with six input symbols were analyzed and their zero-error capacity computed, except for four channels whose capacity can be given in terms of $C_0(G_5)$.

In the next section, we show how a graph (and its complement) can be associated with a discrete memoryless channel. Theorem 11 is restated in a graph-based language.

5.2.2 Relation with graph theory

The problem of computing the zero-error capacity of discrete memoryless channels can be reformulated in terms of graph theory. Given a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ we can construct a characteristic graph G as follows. Take as many vertices as the number of input symbols in $\mathcal{X}$ and connect two vertices with an edge if the corresponding input symbols in $\mathcal{X}$ are distinguishable. Shortly, we can say that the vertex set of G is $V(G) = \mathcal{X}$ and its set of edges $E(G)$ is composed of pairs of orthogonal rows in $[p(y|x)]$. The characteristic graph of channels in Figure 5.5 are shown in Figure 5.6.

In graph theory, the order of a graph is the cardinality of its vertex set. A *clique* is defined as any complete subgraph of G, and the *clique number* [29] of a graph G, denoted by $\omega(G)$, stands for the maximal order of a clique in G. It is easy to see that the maximum number of non-adjacent symbols in G is $\omega(G)$, Namely $N(1) = \omega(G)$. For example, the pentagon graph G_5 of Figure 5.6(c) has the clique number $\omega(G_5) = 2$. Note that the vertex set of any clique corresponds to a set of distinguishable symbols in the channel.

Define the n-product G^n of the graph G as a graph for which $V(G^n) = \mathcal{X}^n$ and $\{\mathbf{x}', \mathbf{x}''\} \in E(G^n)$ if for at least one i, $1 \leq i \leq n$, the ith coordinates of $\mathbf{x}'$ and $\mathbf{x}''$ satisfy $\{x_i', x_i''\} \in E(G)$. Such product of graphs, often called *Shannon's product*, has

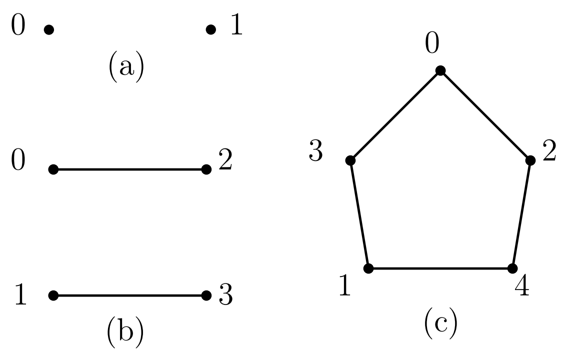

Figure 5.6: Characteristic graphs G of discrete memoryless channels in Figure 5.5. The vertex set of G is the set of input symbols $\mathcal{X}$ and its set of edges corresponds to all pairs of distinguishable symbols in $\mathcal{X}$.

the following meaning: the vertex set of G^n is composed of all n-length sequences, and we connect the vertices $\mathbf{x}'$ and $\mathbf{x}''$ if the corresponding sequences are distinguishable, as illustrated in Figure 5.4.

It is clear that the number of distinguishable sequences of length n is the clique number of G^n, i.e, $N(n) = \omega(G^n)$. Moreover, the sequences in the vertex set of the corresponding complete subgraph define a n-length error-free code for the channel. Therefore, the zero-error capacity of the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ can be rewritten as

$$C_0 = \sup_n \frac{1}{n} \log \omega(G^n). \tag{5.10}$$

In graph theory, the value of C_0 in Equation (5.10) refers to the Shannon capacity of the Graph G, and is denoted by $C_0(G)$.

The chromatic number of a graph G, denoted by $\chi(G)$, is the smallest number of colours necessary to colour the vertices of a graph so that no two adjacent vertices have the same colour. More formally, $\chi(G)$ is the smallest cardinality of a set K for which there exists a function $f : V(G) \to K$ with the property that adjacent vertices are mapped into different elements of K. Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ be a DMC for which the clique and the chromatic numbers of the characteristic graph G are the same, $\omega(G) = \chi(G)$. For any colouration of G, if we define the set $\mathcal{X}'$ in Theorem 11 as being the vertex set of the maximal clique in G, then we can always construct an adjacency-reducing mapping f fulfilling the requirement of the theorem: all symbols whose vertices share a given colour are mapped into the corresponding symbol in $\mathcal{X}'$ that own such colour. Because different colours are associated with non-adjacent symbols, such mapping ensures that any two non-adjacent symbols in $\mathcal{X}$ will be mapped into non-adjacent ones in $\mathcal{X}'$. Moreover, because symbols in $\mathcal{X}'$ correspond to the vertex set of the maximal clique, they are mutually distinguishable. Therefore, Theorem 11 can be entirely reformulated.

Theorem 11' *Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ be a discrete memoryless channel and G the corresponding characteristic graph. If $\omega(G) = \chi(G)$ then $C_0 = \chi(G)$.*

The best known graphs for which $\omega(G) = \chi(G)$ are the so-called *perfect graphs* [29]. A perfect graph is a graph G such that for every induced subgraph of G, the chromatic number equals the clique number. The class of perfect graphs includes bipartite graphs, interval graphs and wheel graphs with an odd number of vertices. The smallest vertex set on which a graph exists with $\omega(G) \neq \chi(G)$ has five vertices, and corresponds to the pentagon graph G_5 already discussed in the previous section.

Although $\omega(G) = \chi(G)$ is a sufficient condition for $\omega(G^n) = [\omega(G)]^n$, Lovász [21] showed that it is not a necessary condition. An example is the complement of the Petersen graph, which is isomorphic with the Kneser graph $KG_{5,2}$. However, it is unknown whether the equality $\log \omega(G') = C_0(G')$ for every induced subgraph $G' \subseteq G$ implies that G is perfect.

Originally, Shannon used a different but equivalent approach for relating the zero-error capacity with elements of graph theory. For a given DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$, we can associate an adjacency matrix $[A_{ij}]$ as follows:

$$A_{ij} = \begin{cases} 1 & \text{if } x_i \text{ is adjacent to } x_j \text{ or if } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{5.11}$$

where $x_i, x_j \in \mathcal{X}$. If two channels give rise to the same adjacency matrix, then it is obvious that an error-free code for one will be an error-free code for the other and, hence, the zero-error capacity C_0 for one will also apply to the other [12]. Such approach considers the adjacency structure of the adjacency matrix to construct a linear graph, called adjacency graph, which is the complementary of the characteristic graph. Therefore, both graphs have the same vertex set $\mathcal{X}$ and two vertices in the adjacency graph are connected by an edge if and only if they are not connected in the characteristic graph. Equivalently, an edge connects two vertices in the adjacency graph if and only if the corresponding input symbols in $\mathcal{X}$ are adjacent. In this case, we say that two vertices in the adjacent graph are independent if the corresponding symbols are non-adjacent in the channel. Clearly, there are $N(1)$ independent vertices in G. Figure 5.7 shows the adjacency graphs of the discrete memoryless channels of Figure 5.5.

Shannon [12] proved the following bounds on the zero-error capacity:

Theorem 12 *Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ be a DMC. The error-free capacity is bounded by the*

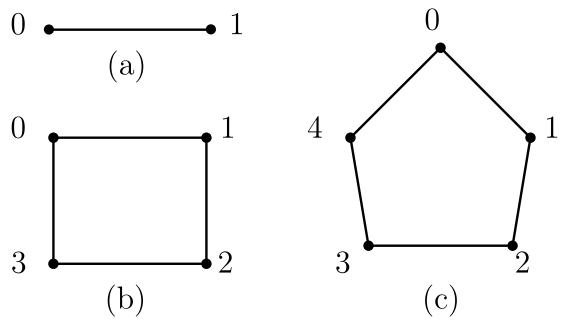

Figure 5.7: Adjacency graphs of discrete memoryless channels corresponding to the channels of Figure 5.5. These graphs are construct by taking as many vertices as the number of symbols in $\mathcal{X}$, and connecting two vertices if the corresponding symbols are adjacent in the channel.

inequalities:

$$-\log \min_{p(x_i)} \sum_{ij} A_{ij} p(x_i) p(x_j) \leq C_0 \leq \min_{p(y|x)} C, \tag{5.12}$$

where C is the ordinary capacity of any discrete memoryless channel with stochastic matrix $p(y|x)$ giving rise to the adjacency matrix A_{ij}; $p(x_i)$ stands for the input probability distribution.

The proof of the theorem can be found in [12]. Although the upper bound is fairly obvious, it has an interesting formulation in graph theory [30] according to which

$$C_0 \leq \log \chi^*(G), \tag{5.13}$$

where $\chi^*(G)$ is the fractional chromatic number of the adjacency graph G, a well-studied concept in polyhedral combinatorics [31] defined as follows. We assign nonnegative weights $p(x_i)$ to the vertices $\mathcal{X}$ of G such that

$$\sum_{x_i \in C} p(x_i) \leq 1$$

for every complete subgraph C in G. This assignment is called a fractional coloring. The fractional chromatic number is the maximum of $\sum_{x_i \in \mathcal{X}} p(x_i)$, where the maximum is taken over all fractional colorings of G. Actually, the fractional chromatic number is the solution of the real-valued relaxation of the integer programming problem that defines the chromatic number of G [26].

Suppose that a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ gives rise to an adjacency graph G such that G can be covered by $N(1)$ cliques. By this we mean that there are $N(1)$ cliques in G, namely $C_1, \ldots, C_{N(1)}$, in a way that their vertex sets, $V(C_1), \ldots, V(C_{N(1)})$, form a partition of $V(G)$. Theorem 11 can be rewritten as [21].

Theorem 11" *Let G be the adjacency graph of a discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. If G can be covered by $N(1)$ cliques, then $C_0 = \log N(1)$.*

Figure 5.8 illustrates Theorem 11". The maximum number of independent vertices in the adjacency graph of Figure 5.8(a) is $N(1) = 2$, e.g., 0 and 3. An adjacency-reducing mapping f for the corresponding DMC takes $f(0) = f(1) = f(2) = 0$ and $f(3) = f(4) = 3$. This mapping can be readily obtained by associating 0 and 3 with vertices of the order-2 and order-3 cliques, respectively. The cube graph has $N(1) = 4$, and can be covered by four clique of order 2 as illustrated in Figure 5.8(b). Therefore, the zero-error capacity of the equivalent DMC is $C_0 = \log 4 = 2$ bits per channel use.

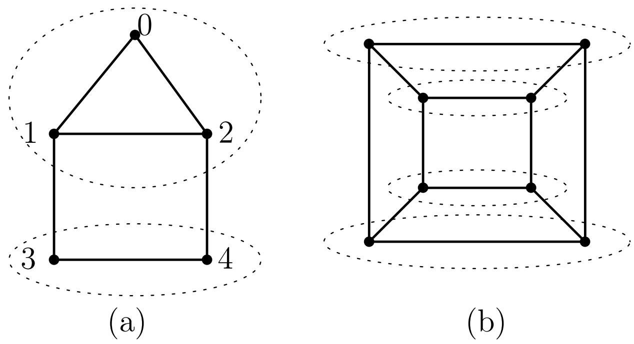

Figure 5.8: Graphs that can be covered by a number of cliques. (a) An adjacency graph with two independent vertices. This graph can be covered by two cliques and therefore there is an adjacency reducing map satisfying the requirement of Theorem 11. (b) The cube graph can be covered by four cliques of order two.

5.3 Lovász theta function

The redefinition of the zero-error capacity in terms of graph has yielded interesting constructions in combinatorics and graph theory. An example of such constructions is the Lovász theta function θ. The functional θ has many application in computer science and combinatorics. Particulary, the θ function is a polynomially computable functional sandwiched in between two NP-complete problems in graph theory: the clique and the chromatic numbers of a graph [22].

The very nice formulation we present in this section was used to compute the zero-error capacity of the pentagon graph. Such graph plays a crucial role in our study of the zero-error capacity of quantum channels. More precisely, we studied a quantum channel for which its zero-error capacity is given by the capacity of the pentagon graph G. Most of the following development can be found in [21].

Given a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ and the corresponding adjacency graph G with vertex set $\mathcal{X}$, an orthonormal representation of G is a set of $|\mathcal{X}|$ vectors $\mathbf{v}_{x_i}$ in an Euclidian space, such that if $x_i, x_j \in \mathcal{X}$ are non-adjacent, then $\mathbf{v}_{x_i}$ and $\mathbf{v}_{x_j}$ are orthogonal. The *value* of an orthonormal representation is defined as

$$\min_{\mathbf{c}} \max_{x_i \in \mathcal{X}} \frac{1}{(\mathbf{c}^T \mathbf{v}_{x_i})^2},$$

where the minimum is taken over all unitary vectors $\mathbf{c}$. The vector $\mathbf{c}$ yielding the minimum is called the handle of the representation. The Lovász $\theta(G)$ function of a graph is defined as the minimum value over all representations of G, and a representation is called optimal if it attains this minimum value. Lovász proved the following result:

Theorem 13 ([21]) *The zero-error capacity of a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is upperbounded by the logarithm of the θ function of its adjacency graph, G:*

$$C_0 \leq \log \theta(G). \tag{5.14}$$

Proof. First, we should note that if G and H are two graphs, and GH is their product as defined in Section 5.2.2, then $\theta(GH) \leq \theta(G)\theta(H)$. Let $\{\mathbf{v}_{x'_i}\}$ and $\{\mathbf{u}_{x''_j}\}$ be optimal orthonormal representations of G and H with handles $\mathbf{c}$ and $\mathbf{d}$, respectively. It is easy to see that $\{\mathbf{v}_{x'_i} \otimes \mathbf{u}_{x''_j}\}$ is an orthonormal representation of GH and $\mathbf{c} \otimes \mathbf{d}$ is a unitary vector. Therefore,

$$\begin{aligned} \theta(GH) &\leq \max_{x'_i, x''_j} \frac{1}{\left((\mathbf{c} \otimes \mathbf{d})^T (\mathbf{v}_{x'_i} \otimes \mathbf{u}_{x''_j})\right)^2} \\ &= \max_{x'_i} \frac{1}{(\mathbf{c}^T \mathbf{v}_{x_i})^2} \max_{x''_j} \frac{1}{(\mathbf{d}^T \mathbf{u}_{x''_j})^2} \\ &= \theta(G)\theta(H). \end{aligned}$$

By definition, if G is an adjacency graph and $\{\mathbf{v}_{x_i}\}$ is an optimum representation with handle $\mathbf{c}$, then there are $N(1)$ vectors $\{\mathbf{v}_{x_1}, \ldots, \mathbf{v}_{x_{N(1)}}\}$ pairwise orthogonal in this representation, where $N(1)$ is the maximum number of independent vertices in G. Hence,

$$1 = ||\mathbf{c}||^2 \geq \sum_{i=1}^{N(1)} (\mathbf{c}^T \mathbf{v}_{x_i})^2 \geq \frac{N(1)}{\theta(G)}. \tag{5.15}$$

Equation (5.8) implies $N(1)^n \leq N(n)$. Finally,

$$C_0 = \sup_n \frac{1}{n} \log N(n) \leq \sup_n \frac{1}{n} \log \theta(G^n) \leq \sup_n \frac{1}{n} \log \theta(G)^n = \log \theta(G).$$

■

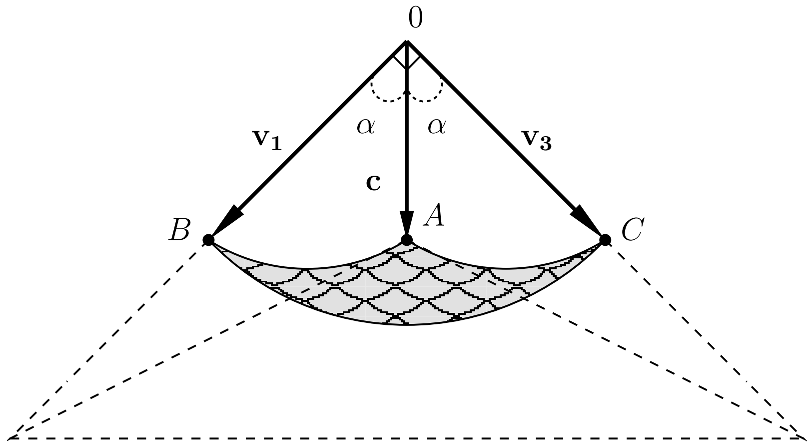

Figure 5.9: A spherical triangle delimited by the vectors $\mathbf{v}_1$, $\mathbf{v}_3$ and the handle $\mathbf{c}$. In a plane normal to the handle, the angle between two consecutive projections $\mathbf{v'}_i, \mathbf{v'}_{i+1 \mod 5}$ of the vectors $\mathbf{v}_i$ is $2\pi/5$. The spherical angle $\angle A$ is the angle between the vectors $\mathbf{v'}_1$ and $\mathbf{v'}_3$, i.e., $\angle A = 4\pi/5$.

Theorem 13 allows of the calculation of the zero-error of the pentagon graph. Remember that Shannon was only able to give lower and upper bounds for the capacity, $\frac{1}{2}\log 5 \le C_0(G_5) \le \log\frac{5}{2}$.

Construct an orthonormal representation for the pentagon G_5 of Figure (5.7)(c) as follows. Consider an umbrella whose handle and five ribs have unitary length. Let $\mathbf{v_0}, \ldots, \mathbf{v_4}$ be the ribs and $\mathbf{c}$ the handle, as vectors oriented away from their common point. Open the umbrella to the point where the maximum angle between the ribs is $\pi/2$. Note that the angle between two consecutive ribs must be the same, and that the angle between alternate ribs must be $\pi/2$. It is clear that $\{\mathbf{v}_0, \ldots, \mathbf{v}_4\}$ forms an orthonormal representation of G_5. Figure 5.9 illustrates this scenario, at which we plot the handle $\mathbf{c}$ and the two orthogonal vectors $\mathbf{v}_1$ and $\mathbf{v}_3$. The extremities of the six vectors are points on a unitary three-dimensional sphere centered in 0, and the points defined by the handle and any two alternated vectors delimit a spherical triangle identical to the triangle ABC of Figure 5.9. We are interested in the value of the representation, i.e.,

$$\min_{\mathbf{c}} \max_{0\le i\le 4} \frac{1}{(\mathbf{c}^T\mathbf{v}_i)^2}.$$

Note that $\mathbf{c}^T\mathbf{v}_i$ stands for the cosine of the angle between the handle and the rib $\mathbf{v}_i$, namely $\mathbf{c}^T\mathbf{v}_i = \cos(\alpha)$. Let $\beta = \pi/2$ be the angle between $\mathbf{v}_1$ and $\mathbf{v}_3$. The first spherical cosine theorem states that

$$\cos(\beta) = \cos^2(\alpha) + \sin^2(\alpha)\cos(\angle A).$$

Because angles α between the ribs and the handle are the same, the spherical angle $\angle A$ is the angle between the projection of the vectors $\mathbf{v}_1$ and $\mathbf{v}_3$ on the plane normal to the handle $\mathbf{c}$, i.e., $\angle A = 4\pi/5$. Finally, we can write

$$0 = \cos^2(\alpha) + \sin^2(\alpha)\cos(4\pi/5),$$

which gives $\cos^2(\alpha) = (\mathbf{c}^T\mathbf{v}_i)^2 = 1/\sqrt{5}$. Hence,

$$C_0(G_5) \leq \log\theta(G_5) \leq \log\left(\frac{1}{\cos^2(\alpha)}\right) = \log\sqrt{5} = \frac{1}{2}\log 5.$$

The opposite inequality is known and the Shannon's lower bound is tight.

The definition of $\theta(G)$ is not unique. In his paper [21], Lovász pointed out four equivalent definitions for $\theta(G)$. For example, he showed that $\theta(G)$ is the minimum of the largest eigenvalue of any symmetric matrix $(a_{ij})_{i,j=1}^{|\mathcal{X}|}$ such that $a_{ij} = 1$ if $i = j$ or if x_i and x_j are non-adjacent. Although the Lovász θ function behaves very beautifully, the value of $\log\theta(G)$ is generally different from the capacity. A new bound on the zero-error capacity was derived by Haemers [15], and it is sometimes better but quite often much worse than $\theta(G)$. A quadratic matrix of order $|\mathcal{X}|$ is said to *fit* the graph G if its diagonal entries are all nonzero and the element $a_{i,j}$ is zero if and only if the symbols x_i and x_j are adjacent in the channel. Haemers proved that the logarithm of the ranking of any these matrix upper-bounds the zero-error capacity of G. This result was illustrated with some examples for which his bound is better that $\theta(G)$. However, this is not true for the pentagon graph G_5.

In the next two sections we present two variants of the original problem: the zero-error capacity of DMC with feedback and the zero-error capacity of sum and product of discrete memoryless channels.

5.4 Channels with complete feedback and list codes

A complete feedback is characterized by a noiseless channel from the receiver to the sender, as illustrated in Figure 5.10. It is assumed that the actual received symbol is sent back immediately and noiselessly to the transmitter, which can use the feedback information in order to choose which symbol to transmit next. Although feedback can help in simplifying encoding and decoding processes, it was proved that this additional resource cannot increase the ordinary capacity of a discrete memoryless channel [20, pp.212]. Surprisedly, Shannon and Elias [12] showed that feedback may increase the zero-error capacity of such channels.

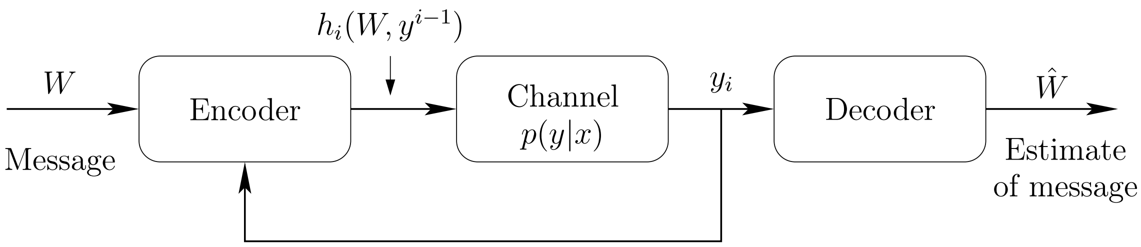

Figure 5.10: A discrete memoryless channel with feedback.

We define an error-free block code as a sequence of mappings $h_i(W, y^{i-1})$, where each h_i is a function only of the message W and the previous received symbols $y_1, y_2, \ldots, y_{i-1}$, and a sequence of decoding functions $g : \mathcal{Y}^n \to \{1, \ldots, M\}$. We define the probability of error as $P_e^{(n)} = \Pr\{\hat{W} \neq W\}$ and we require $P_e^{(n)} = 0$. Although the following result appeared in a Shannon's paper [12], it is due to Shannon and P. Elias.

Theorem 14 *Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ be a discrete memoryless channel and define $S_{y_j} = \{x_i \in \mathcal{X} | p(y_j|x_i) > 0\}$, the set of input symbols which cause output y_j with positive probability. Let Π be the set of probability functions P defined on subsets of $\mathcal{X}$. Then, the zero-error capacity of the DMC with feedback C_{0F} is zero if all input symbols in $\mathcal{X}$ are pairwise adjacent. Otherwise*

$$C_0 \leq C_{0F} = -\min_{P \in \Pi} \max_{y_j \in \mathcal{Y}} \log \sum_{x_i \in S_{y_j}} P(x_i). \tag{5.16}$$

As an example, consider the DMC of Figure 5.5(c). The zero-error capacity of this channel is $C_0 = \log \sqrt{5} \simeq 1.161$ bits per symbol. By symmetry, the minmax distribution P in Theorem 14 is the uniform with $p(x_i) = 1/5$, $i = 0, \ldots, 4$. Then, the zero-error capacity of the pentagon with feedback is

$$C_{0F} = -\log 2/5 \simeq 1.322.$$

The zero-error capacity of discrete memoryless channels with feedback is related to list decoding, a well-studied topic in information theory [32]. In the zero-error context, an error-free *list code* of size L and blocklength n for the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is a set $\mathcal{C} \subseteq \mathcal{X}^n$ such that for every $\mathbf{y} \in \mathcal{Y}^n$

$$|\{\mathbf{x} \in \mathcal{C} : p^n(\mathbf{y}|\mathbf{x}) > 0\}| \leq L.$$

Intuitively, for every transmitted codeword $\mathbf{x}$, the decoder should decide on a list of at most L transmitted codewords. For a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$, let $N(n, L)$ be the maximum cardinality of a list code $\mathcal{C} \subseteq \mathcal{X}^n$ with list size L and blocklength n. The list code capacity $C_{0,L}$ of list size L of the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is

$$C_{0,L} = \limsup_{n \to \infty} \frac{1}{n} \log N(n, L).$$

Then, the list code zero-error capacity of the channel is defined as

$$C_{0,\infty} = \sup_L C_{0,L}. \tag{5.17}$$

Note that the problem of finding the zero-error capacity of a DMC is a special case of the list code zero-error capacity with $L = 1$. Elias [33] demonstrated that Equations (5.16) and (5.17) are equivalent. Namely, the zero-error capacity of a DMC with feedback is equal to the list code zero-error capacity. Essentially, a feedback code can be viewed as a sequence of list codes with successively reduced list sizes.

5.5 The sum and product of channels

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with zero-error capacities C_{0_1} and C_{0_2}, respectively. We are interested in transmitting information using the two channels and we ask for the zero-error capacity of the joint system [12]. Basically, there are two natural ways of assembling two channels to form a single channel, which we call the *sum* and the *product* of two channels.

The sum of two channels is a new channel $(\mathcal{X}_1 \amalg \mathcal{X}_2, p(y_1|x_1) \oplus p(y_2|x_2), \mathcal{Y}_1 \amalg \mathcal{Y}_2)$ where the stochastic matrix of the sum channel is the direct sum of the two stochastic matrices, and the input (output) set is the disjoint union of $\mathcal{X}_1$ $(\mathcal{Y}_1)$ and $\mathcal{X}_2$ $(\mathcal{Y}_2)$, respectively. Intuitively, the sum channel behaves as $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ if an input symbol $x_{1_i} \in \mathcal{X}_1$ is used, otherwise, it behaves as $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$. This corresponds physically to a situation where either of two channels may be used but not both. Analogously, the product channel is a new DMC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \otimes p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ where the stochastic matrix is the direct product of the two matrices, and the input (output) set is the cartesian product of $\mathcal{X}_1$ $(\mathcal{Y}_1)$ and $\mathcal{X}_2$ $(\mathcal{Y}_2)$, respectively. In this case, we can think of the product DMC as of a nonstationary memoryless channel over which transmission is governed in strick alternation by the stochastic matrices $p(y_1|x_1)$ and $p(y_2|x_2)$:

$$p(y_{1_i}, y_{2_i}|x_{1_i}, x_{2_i}) = p(y_{1_i}|x_{1_i})p(y_{2_i}|x_{2_i}).$$

Consider two DMCs, $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$, $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$, and let C_1, C_2 be their corresponding ordinary capacities. It is well known [13] that the ordinary capacity of the sum channel is $C_+ = \log(\exp C_1 + \exp C_2)$. For the product channel, the ordinary capacity is proved to be $C_\times = C_1 + C_2$.

The error-free communication capacity of the sum and product channels was studied by Shannon [12]. If C_{0_+} and $C_{0_\times}$ denote the zero-error capacity of the sum and product channels, respectively, then Shannon demonstrated that

$$C_{0_+} \geq \log(\exp C_{0_1} + \exp C_{0_2}) \tag{5.18}$$

and

$$C_{0_\times} \geq C_{0_1} + C_{0_2}, \tag{5.19}$$

with equality if and only if the adjacent graph G of either of the two channels can be coloured using $\alpha(G)$ colours. In an analogy with the ordinary capacity, Shannon conjectured that, in fact, equalities always holds for zero-error capacities. The product channel conjecture was implicitly disproved in a example of Haemers [15]. More recently, Alon [16] proved the existence of channels for which $C_{0_\times} > C_{0_1} + C_{0_2}$. Such results, together with those of Section 5.4, suggest that the zero-error capacity behaves quite different from the ordinary capacity.

5.6 Conclusions

We have presented in this Chapter a survey of fundamental concepts in zero-error information theory. We have started by presenting the ordinary capacity of discrete memoryless channels, for which a small probability of error is allowed even if we make use of the best coding scheme to encode information. Next, the zero-error capacity of a DMC was introduced and a method to calculate the capacity of simple channel has been derived.

The problem of finding the zero-error capacity was reformulated in terms of graph theory. It was shown how several results in zero-error theory can be restated in a graph language. The most famous upper bound on the zero-error capacity, the Lovász θ function, was presented and used to calculate the zero-error capacity of the pentagon graph, a problem that remained open during more than twenty years. This example is particulary interesting because we have found a quantum channel for which its zero-error capacity equals the capacity of the pentagon. Finally, we presented two variations of the original problem: the zero-error capacity of a DMC with feedback and the zero-error capacity of sum and products of discrete memoryless channels.

Chapter 6

Zero-error capacity of quantum channels

6.1 Introduction

As we have already mentioned in Section 2.1, quantum channel capacities to carry classical information allow for an asymptotically small probability of error, even when the best quantum coding scheme is used. Such capacities include the one-shot capacity [3, 4, 5], the HSW capacity [7, 8], the adaptive capacity [6] and the entanglement-assisted capacity [9, 10]. The main reason of having a non-vanishing error probability is the decoding process, which is based on the concept of typical sequences and typical Hilbert subspaces [2]. More specifically, a received quantum codeword of a sufficiently long random code always has a high probability of belonging to a given Hilbert subspace, called typical subspace. An error is detected when the respective output codeword belongs to the orthogonal subspace, also called non-typical subspace. Although the probability of a received quantum codeword does not belong to a typical subspace is small, it is always different from zero. Hence, ordinary quantum error-correction schemes [34, 35] consist of embedding, in a controlled way, a given quantum state into another state that belongs to a higher dimensional Hilbert space. Depending on the encoding strategy, errors due to decoherence in the encoded stated might be detected and corrected in order to recovery the original quantum state.

Quantum perfect transmission, computing and storage are not a recent subject in quantum information and computation. In 1997, Zanardi *et al* [36, 37] pointed out that the symmetry between some quantum states and the environment might provide a new strategy for protecting quantum states from decoherence. Instead of use an *active* error detection/correction scheme, the authors showed that in the presence of a “coherent” environmental noise, where the original state and the environment share some kind of symmetry, one can design states that are immune to the noise rather than states that can be easily corrected. Hence, their approach consists in a *passive*, i.e., a intrinsic stabilization of quantum information. More recently, Kribs *et. al.* [38, 39] described a mathematical framework, called operator quantum-error correction, that incorporates the two techniques of error prevention/correction under a single approach.

An algebraic study of symmetries in the Zanardi model motivated the definition of the so called decoherence-free subspaces (DFS) [40], which are subspaces of the whole system's Hilbert space that are not affected by the noise under certain assumptions about the symmetry of the noise processes. Bacon *et. al.* [41] developed a general formalism, called noiseless subsystems, to find Hamiltonians involving one- and two-qubits interactions, which can be used to implement universal quantum gates without leaving a given decoherence-free (noiseless) subspace. Therefore, when computation is performed in this manner, the system is never exposed to errors. Such approach leads to a naturally fault tolerant quantum computation [42, 43, 44].

Although concepts of noiseless quantum codes and fault tolerant quantum computation are well developed, a number to quantify the maximum amount of classical information per channel use that can be sent without error through a noisy quantum channel was not defined until now. In this thesis, we generalise the concept of classical zero-error capacity to include quantum channels, in a scenario where they are used to transmit classical information. We define the *quantum zero-error capacity* as the supremum of rates at which classical information can be transmitted through a noisy quantum channel with a probability of error *equal* to zero.

Since our first paper in 2005 [45], some developments have been made by other researches. In a recent work, Beigi and Shor [46] demonstrated that finding the quantum zero-error capacity is a QMA-complete problem [47]. An interesting feature of quantum channels concerning the quantum zero-error capacity was pointed out by Duan and Shi [48]. The authors found a quantum channel allowing of perfect classical information transmission (i.e., quantum zero-error capacity greater than zero) once the channel is used two times, whereas no information could be sent in a single use of the channel. In their work, the communication protocol involves two senders and two receivers, where senders, as well as receivers, are able to exchange classical information between them.

In the next section, we first describe the zero-error communication protocol, which is similar to the HSW protocol. Then, a quantum error-free block code is formally defined. Once we define a protocol and a quantum code, we are able to quantity the maximum amount of error-free classical information per channel use that a quantum channel can transmit, i.e, the quantum zero-error capacity.

6.2 Quantum zero-error capacity

Given a quantum channel, we ask for the maximum amount of classical information per channel use Alice can transmit to Bob with a zero probability of error. Consider a d-dimensional quantum channel $\mathcal{E} \equiv \{E_a\}$ modelled by a linear, completely positive trace-preserving quantum operation. Hereafter, we denote $\mathcal{S}$ a subset of input quantum states of dimension d for $\mathcal{E}$. States $\rho_i \in \mathcal{S}$ are referred as input states. Figure 6.1 is a block diagram of a quantum communication system enabling Alice to transmit classical messages to Bob with a zero probability of error. Initially, Alice chooses a message from a set $\{1, \ldots, K_n\}$ of K_n classical messages. Then, the encoder maps such message onto a n-tensor product of quantum states in $\mathcal{S}$. The d^n-dimensional encoded state is called a quantum codeword. The quantum codeword is transmitted through a noisy quantum channel $\mathcal{E}$. At the receiver end, Bob performs a Positive Operator-Valued Measurement (POVM) on the whole received state. Measurement outcomes are arguments of a decoding function. The decoder should decide which message was sent by Alice with the property that no errors are allowed.

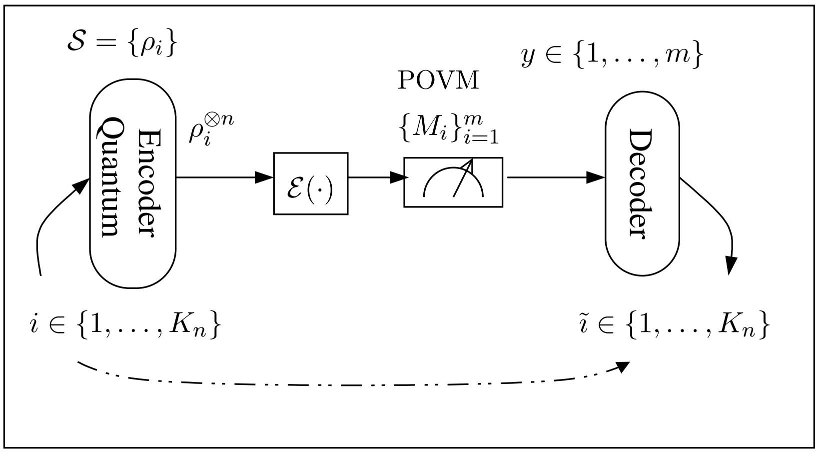

Figure 6.1: General representation of a quantum zero-error communication system.

The error-free communication protocol can be summarized as follows:

- The source alphabet is a set $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ of d-dimensional input quantum states;
- in order to be transmitted through a quantum channel, classical messages are mapped onto quantum codewords, which are tensor products of quantum states in $\mathcal{S}$;
- although input states are not allowed to be entangled between two or more channel uses, collective POVM measurements are authorized at the quantum channel output.

As we will see, such measurements are necessary and sufficient in order to reach the quantum zero-error capacity.

Essentially, the proposed protocol is similar to the protocol employed by the Holevo-Schumacher-Westmoreland [7, 8] capacity. In order to generalise the zero-error capacity for quantum channels, we should define a quantum error-free block code.

Definition 23 ((K_n, n) error-free quantum block code) *An (K_n, n) error-free quantum block code for a quantum channel $\mathcal{E}$ is composed of the following:*

1. *A set of indexes $\{1, \ldots, K_n\}$, where each index is associated with a classical message.*
2. *An encoding function*
$$X^n : \{1, \ldots, K_n\} \rightarrow \mathcal{S}^{\otimes n}, \tag{6.1}$$
yielding quantum codewords $\overline{\rho}_1 = X^n(1), \ldots, \overline{\rho}_{K_n} = X^n(K_n)$. The set of all quantum codewords is called a quantum codebook.
3. *A decoding function*
$$g : \{1, \ldots, m\} \rightarrow \{1, \ldots, K_n\}, \tag{6.2}$$
which deterministically assigns a guess to each possible measurement outcome $y \in \{1, \ldots, m\}$ performed by a POVM $\mathcal{P} = \{M_1, \ldots, M_m\}$. The decoding function has the following property:
$$Pr\left(g(Y = y) \neq i | X^n = X^n(i)\right) = 0 \; \forall \; i \in \{1, \ldots, K_n\}. \tag{6.3}$$

The reason why we put an index n in K_n is to remember that a given error-free quantum code of length n has exactly K_n codewords. It is easy to see that the transmission rate of a (K_n, n) error-free quantum block code is

$$R_n = \frac{1}{n} \log K_n \;\; \textit{(bits per channel use)}.$$

Definition 24 is a generalisation of the zero-error capacity for quantum channels.

Definition 24 (Quantum zero-error capacity (QZEC)) *Let $\mathcal{E}$ be a linear, completely positive trace-preserving quantum operation representing a noisy quantum channel. The zero-error capacity of $\mathcal{E}(\cdot)$, denoted by $C^{(0)}(\mathcal{E})$, is the least upper bound of achievable rates with probability of error equal to zero. That is,*

$$C^{(0)}(\mathcal{E}) = \sup_{\mathcal{S}} \sup_{n} \frac{1}{n} \log K_n, \tag{6.4}$$

where K_n stands for the maximum number of classical messages that the system can transmit without error, when a (K_n, n) error-free quantum block code with input alphabet $\mathcal{S}$ is used.

A fundamental property of quantum systems concerns the distinguishability of two quantum states [2]. In a given Hilbert space of dimension d, two quantum states ρ_1 and ρ_2 are perfectly distinguishable if and only if the Hilbert subspaces spanned by the supports of ρ_1 and ρ_2 are orthogonal. Equivalently, if ρ_1 is non-orthogonal to ρ_2 then such states are indistinguishable. It is clear that in a d-dimensional Hilbert space there are at most d pairwise distinguishable quantum states. Given a quantum channel $\mathcal{E}$, we are particularly interested in input quantum states ρ_i and ρ_j which are distinguishable at the channel output.

Definition 25 (Non-adjacent quantum states) *Consider a quantum channel $\mathcal{E}$ and a set $\mathcal{S}$ of input states. Two quantum states $\rho_i, \rho_j \in \mathcal{S}$ are said to be* non-adjacent *with relation to $\mathcal{E}$ if $\mathcal{E}(\rho_i)$ and $\mathcal{E}(\rho_j)$ are distinguishable. Otherwise, they are said to be* adjacent. *For short, we should use $\rho_i \perp_{\mathcal{E}} \rho_j$ to denote that ρ_i is non-adjacent to ρ_j.*

For the classical case, Shannon showed that the zero-error capacity of a discrete memoryless channel depends only on the adjacency relations between input symbols. Moreover, it was demonstrated that the classical zero-error capacity is greater that zero if and only if there exist at least two non-adjacent input symbols in $\mathcal{X}$. In order to demonstrate an analogous result for the quantum zero-error capacity, we need to investigate adjacency between two tensor product sequences of input states.

Consider a set $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ of input quantum states for a quantum channel $\mathcal{E}$. The set of all n-tensor products is denoted by $\mathcal{S}^{\otimes n}$. Let $\hat{\rho}_i = \rho_{i_1} \otimes \cdots \otimes \rho_{i_n}$ and $\hat{\rho}_j = \rho_{j_1} \otimes \cdots \otimes \rho_{j_n}$ be two n-tensor products of quantum states in $\mathcal{S}$. We say that $\hat{\rho}_i$ is *non-adjacent* to $\hat{\rho}_j$ if $\mathcal{E}(\hat{\rho}_i)$ and $\mathcal{E}(\hat{\rho}_j)$ are distinguishable, i.e, if $\mathcal{E}(\hat{\rho}_i)$ and $\mathcal{E}(\hat{\rho}_j)$ have orthogonal supports. Otherwise, they are said to be adjacent in $\mathcal{E}$.

Proposition 15 *For a given quantum channel $\mathcal{E}$ and a set $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ of input quantum states, let $\hat{\rho}_i, \hat{\rho}_j \in \mathcal{S}^{\otimes n}$ be two tensor product sequences of n states. Sequences $\hat{\rho}_i$ and $\hat{\rho}_j$ are non-adjacent in $\mathcal{E}$ if and only if for at least one k, $1 \leq k \leq n$, ρ_{i_k} is non-adjacent to ρ_{j_k}.*

Proof. Because the quantum channel is memoryless, we can write the channel output

$$\begin{aligned}\mathcal{E}(\hat{\rho}_i) &= \mathcal{E}(\rho_{i_1}) \otimes \cdots \otimes \mathcal{E}(\rho_{i_k}) \otimes \cdots \otimes \mathcal{E}(\rho_{i_n})\\ \mathcal{E}(\hat{\rho}_j) &= \mathcal{E}(\rho_{j_1}) \otimes \cdots \otimes \mathcal{E}(\rho_{j_k}) \otimes \cdots \otimes \mathcal{E}(\rho_{j_n})\end{aligned}$$

Figure 6.2: Two distinguishable tensor product sequences $\mathcal{E}(\hat{\rho}_i)$ and $\mathcal{E}(\hat{\rho}_j)$. The distinguishibility of the sequences depends only on the distinguisibility of states $\mathcal{E}(\rho_{i_j})$. Essentially, this means that a quantum channel has a nonzero error-free capacity if and only if there exists a set $\mathcal{S}$ of input states containing at least two non-adjacent states, $\rho_i \perp_{\mathcal{E}} \rho_j$; $\rho_i, \rho_j \in \mathcal{S}$.

as illustrated in Figure 6.2. If $\hat{\rho}_i \perp_{\mathcal{E}} \hat{\rho}_j$ then

$$\begin{aligned}\mathrm{tr}\left[\mathcal{E}(\hat{\rho}_i)\mathcal{E}(\hat{\rho}_j)\right] &= \mathrm{tr}\left[\left(\bigotimes_{k=1}^{n}\mathcal{E}(\rho_{i_k})\right)\left(\bigotimes_{k=1}^{n}\mathcal{E}(\rho_{j_k})\right)\right]\\ &= \prod_{k=1}^{n}\mathrm{tr}\left[\mathcal{E}(\rho_{i_k})\mathcal{E}(\rho_{j_k})\right]\\ &= 0,\end{aligned}$$

which means that $\rho_{i_k} \perp_{\mathcal{E}} \rho_{j_k}$ for at least one k, $1 \leq k \leq n$. The proof of the converse is trivial. ■

Proposition 15 guarantees that the distinguishibility of any two n-tensor product sequences depends only on adjacency relations of states $\rho_i \in \mathcal{S}$.

Proposition 16 *A quantum channel $\mathcal{E}$ has a non-vanishing zero-error capacity if and only if there exists a set $\mathcal{S}$ containing at least two non-adjacent states, $\rho_i \perp_{\mathcal{E}} \rho_j$, $\rho_i, \rho_j \in \mathcal{S}$.*

Proof. Suppose that $C^{(0)}(\mathcal{E}) > 0$. In this case, it should exist at least two codewords, $\overline{\rho}_i$ and $\overline{\rho}_j$, of a (K_n, n) quantum error-free code with alphabet $\mathcal{S}$ such that $\overline{\rho}_i \perp_{\mathcal{E}} \overline{\rho}_j$. By Proposition 15, $\rho_{i_k} \perp_{\mathcal{E}} \rho_{j_k}$ for at least one k, $1 \leq k \leq n$, $\rho_{i_k}, \rho_{j_k} \in \mathcal{S}$. The converse is trivial. ■

The previous analysis allows for a comprehensive understanding of the quantum zero-error capacity. Let $\mathcal{E}$ be a d-dimensional quantum channel. Fix a set of input quantum states $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ for $\mathcal{E}$. By Definitions 23 and 25, the maximum number of classical messages Alice can transmit to Bob without error using an $(K_1, 1)$ error-free quantum code with alphabet $\mathcal{S}$ is K_1, the maximum number of pairwise non-adjacent quantum states in $\mathcal{S}$. More specifically, if we consider subsets $\mathcal{S}' \subseteq \mathcal{S}$ such that $\forall \rho_i, \rho_j \in \mathcal{S}'; i \neq j; \rho_i \perp_{\mathcal{E}} \rho_j$, then

$$K_1 = \max_{\mathcal{S}' \subseteq \mathcal{S}} |\mathcal{S}'| \leq d. \tag{6.5}$$

Analogously, if n-tensor products of states in $\mathcal{S}$ are considered, then we have l^n possible sequences, namely, $\hat{\rho}_1, \ldots, \hat{\rho}_{l^n}$. Clearly, the maximum number of classical messages Alice can communicate to Bob using a (K_n, n) error-free quantum code with alphabet $\mathcal{S}$ will be the maximum number of pairwise non-adjacent sequences, denoted by K_n. The zero-error capacity of the quantum channel will be the supremum of the information transmission rate over all sets $\mathcal{S}$ of input states and code length n.

6.2.1 A graph-theoretic approach

Developments in the previous section allow of a nice interpretation of the zero-error capacity in terms of graph theory. Given a quantum channel $\mathcal{E}$ and a set of input states $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$, we can construct a characteristic graph $\mathcal{G}$ as follows: The vertex set of $\mathcal{G}$ is the index set of $\mathcal{S}$, and two vertices are connected if the corresponding input states in $\mathcal{S}$ are non-adjacent. Mathematically,

$$V(\mathcal{G}) = \{1, \ldots, l\}, \tag{6.6}$$

$$E(\mathcal{G}) = \{(i, j); \rho_i \perp_{\mathcal{E}} \rho_j;\ \rho_i, \rho_j \in \mathcal{S}; i \neq j\}. \tag{6.7}$$

It is easy to see that quantum states corresponding to vertices in any complete subgraph of $\mathcal{G}$ are mutually non-adjacent. Therefore, the maximum number of pairwise non-adjacent states in $\mathcal{S}$ is the clique number of $\mathcal{G}$, $\omega(\mathcal{G})$, which is the maximum cardinality of any complete subgraph of $\mathcal{G}$. Define a n-product $\mathcal{G}^n$ of $\mathcal{G}$ as a graph whose vertex set and the set of edges are given by

$$V(\mathcal{G}^n) = \{1, \ldots, l\}^n, \tag{6.8}$$

$$\begin{aligned} E(\mathcal{G}^n) = \{&(i_1 \ldots i_n, j_1 \ldots j_n); \rho_{i_k} \perp_{\mathcal{E}} \rho_{j_k} \text{ for at least one } k, 1 \leq k \leq n; \\ &\rho_{i_k}, \rho_{j_k} \in \mathcal{S}\}. \end{aligned} \tag{6.9}$$

If we denote $\mathcal{S}^{\otimes n}$ the set of all n-tensor product sequences of states in $\mathcal{S}$, then the vertex set of $\mathcal{G}^n$ is the index set of $\mathcal{S}^{\otimes n}$, whereas the set of edges is composed of pairs of such indexes whose corresponding sequences are non-adjacent in the channel $\mathcal{E}$. It turns out that the maximum number of messages we can transmit without error with a (K_n, n) error-free quantum code with alphabet $\mathcal{S}$ is the clique number of $\mathcal{G}^n$, $\omega(\mathcal{G}^n)$. Moreover, an error-free codebook is given by sequences of the corresponding vertices in the maximal clique of $\mathcal{G}^n$. If we consider the supremum over all possible sets of input states $\mathcal{S}$, we get an alternative and equivalent definition of the zero-error capacity in terms of graph theory.

Definition 26 (Equivalent definition of the QZEC) *The zero-error capacity of a quantum channel $\mathcal{E}$ is given by*

$$C^{(0)}(\mathcal{E}) = \sup_{\mathcal{S}} \sup_{n} \frac{1}{n} \log \omega\left(\mathcal{G}^n\right), \tag{6.10}$$

where the supremum is taken over all sets $\mathcal{S}$ of input states, and $\omega(\mathcal{G}^n)$ is the clique number of the $n-$product of the characteristic graph $\mathcal{G}$ associated with $\mathcal{S}$.

The quantum error-free capacity may also be interpreted as the supremum over zero-error capacities of classical discrete memoryless channels. For each set $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ of input states for a given quantum channel $\mathcal{E}$, we can associate an adjacency matrix $A(\mathcal{S})$ (see Section 5.2.2), which is a $l \times l$ matrix defined as follows:

$$A(\mathcal{S})_{ij} = \begin{cases} 1 & \text{if } \rho_i \text{ is adjacent to } \rho_j \text{ or if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{6.11}$$

A given adjacency matrix may correspond to an infinity number of classical DMCs. Shannon [12] has showed a procedure to find a DMC $(\mathcal{X}, p(y|x), \mathcal{S})$ that gives rise to a particular adjacency matrix A. Moreover, he demonstrated that discrete memoryless channels giving rise to a given adjacency matrix have the same zero-error capacity. If we denote $C_0(A(\mathcal{S}))$ the zero-error capacity of any equivalent DMC obtained from the $A(\mathcal{S})$, then a straightforward consequence of the Equation (6.10) is that

$$C_0(\mathcal{E}) = \sup_{\mathcal{S}} C_0(A(\mathcal{S})). \tag{6.12}$$

These equivalent definitions of the quantum zero-error capacity are used to prove most of our results in the next sections.

The next section investigates quantum states and measurements attaining the quantum error-free capacity. It is showed that we only need to consider pure quantum states at the channel input in order to reach the supremum in Equation (6.10). Moreover, we demonstrate that the capacity can always be reached by using a set $\mathcal{S}$ of at most d pure states. Concerning the measurements, we prove that collective measurements are necessary to attain the quantum zero-error capacity in Definition 24.

6.3 Quantum states achieving the QZEC

In this section we discuss some properties of quantum states reaching the quantum zero-error capacity, namely, quantum states in the set $\mathcal{S}$ achieving the supremum in Equation (6.4). It is well-known that the Holevo-Schumacher-Westmoreland (HSW) capacity [7, 8] can be reached using an ensemble $\{p_i, \rho_i\}$ of at most d^2 pure quantum states [2, pp. 555]. We use the equivalent definition of the quantum zero-error capacity to obtain an analogous result for the quantum case.

Proposition 17 *The zero-error capacity of quantum channels $\mathcal{E}$ can be achieved by a set $\mathcal{S}$ composed only of pure quantum states, i.e., $\mathcal{S} = \{\rho_i = |v_i\rangle\langle v_i|\}$.*

Proof. Consider a quantum channel $\mathcal{E}$ with operation elements $\{E_a\}$, as defined in Section 4.2.2. Suppose that the set $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$ achieving the supremum in Equation (6.4) may contain mixed states. We call $\mathcal{G}$ the characteristic graph associated with $\mathcal{S}$. To demonstrate the proposition, we show that it is always possible to obtain a set $\mathcal{S}'$ from $\mathcal{S}$, such that $\mathcal{S}'$ contains only pure states and $\mathcal{S}'$ also achieves the supremum in Equation (6.4).

Let $\rho_i \in \mathcal{S}$, $\rho_i = \sum_r \lambda_{i_r} |v_{i_r}\rangle\langle v_{i_r}|$, be an input quantum state. Then, the output of the channel when ρ_i is transmitted is given by

$$\begin{aligned}
\mathcal{E}(\rho_i) &= \sum_a E_a \rho_i E_a^\dagger \\
&= \sum_a E_a \left[\sum_r \lambda_{i_r} |v_{i_r}\rangle\langle v_{i_r}| \right] E_a^\dagger \\
&= \sum_a \sum_r \lambda_{i_r} E_a |v_{i_r}\rangle\langle v_{i_r}| E_a^\dagger. \qquad (6.13)
\end{aligned}$$

As we already explained in Section 6.2, the trace $\mathrm{tr}\,[\mathcal{E}(\rho_i)\mathcal{E}(\rho_j)]$ gives the adjacency relation between ρ_i and ρ_j. if $\rho_j = \sum_s \lambda_{j_s} |v_{j_s}\rangle\langle v_{j_s}|$ then

$$\begin{aligned}
\mathrm{tr}\,[\mathcal{E}(\rho_i)\mathcal{E}(\rho_j)] &= \mathrm{tr}\left[\sum_a \sum_r \lambda_{i_r} E_a |v_{i_r}\rangle\langle v_{i_r}| E_a^\dagger \sum_b \sum_s \lambda_{j_s} E_b |v_{j_s}\rangle\langle v_{j_s}| E_b^\dagger \right] \\
&= \mathrm{tr}\left[\sum_a \sum_r \sum_b \sum_s \lambda_{i_r}\lambda_{j_s} E_a |v_{i_r}\rangle\langle v_{i_r}| E_a^\dagger E_b |v_{j_s}\rangle\langle v_{j_s}| E_b^\dagger \right] \\
&= \sum_{a,r,b,s} \lambda_{i_r}\lambda_{j_s} ||\langle v_{i_r}| E_a^\dagger E_b |v_{j_s}\rangle||^2. \qquad (6.14)
\end{aligned}$$

Without loss of generality (w.l.o.g), define a new set $\mathcal{S}' = \{|v_{1_1}\rangle, \ldots, |v_{l_1}\rangle\}$, where $|v_{i_1}\rangle \in \mathrm{supp}\,\rho_i$ is a pure state in the support of ρ_i. Call $\mathcal{G}'$ the characteristic graph due to $\mathcal{S}'$. Our aim is to demonstrate that replacing ρ_i with $|v_{i_1}\rangle$ does not create new adjacencies. To visualize this, note that

$$\begin{aligned}
\mathrm{tr}\,[\mathcal{E}(|v_{i_1}\rangle)\mathcal{E}(|v_{j_1}\rangle)] &= \mathrm{tr}\left[\sum_a E_a |v_{i_1}\rangle\langle v_{i_1}| E_a^\dagger \sum_b E_b |v_{j_1}\rangle\langle v_{j_1}| E_b^\dagger \right] \\
&= \mathrm{tr}\left[\sum_a \sum_b E_a |v_{i_1}\rangle\langle v_{i_1}| E_a^\dagger E_b |v_{j_1}\rangle\langle v_{j_1}| E_b^\dagger \right] \\
&= \sum_{a,b} ||\langle v_{i_1}| E_a^\dagger E_b |v_{j_1}\rangle||^2. \qquad (6.15)
\end{aligned}$$

It is known that if $\rho_i \perp_{\mathcal{E}} \rho_j$ then $\text{tr}\,[\mathcal{E}(\rho_i)\mathcal{E}(\rho_j)] = 0$. This means that $\langle v_{i_r}|E_a^\dagger E_b|v_{j_s}\rangle = 0$ for all indexes r and s in Equation (6.14). Therefore, $\text{tr}\,[\mathcal{E}(|v_{i_1}\rangle)\mathcal{E}(|v_{j_1}\rangle)] = 0$ and $|v_{i_1}\rangle \perp_{\mathcal{E}} |v_{j_1}\rangle$. It is clear that the characteristic graph $\mathcal{G}'$ can be obtained from $\mathcal{G}$ by (probably) adding a number of edges but never deleting edges. In addition, adding edges never decreases (and may increase) the clique number of a graph [29], i.e., $\omega(\mathcal{G}) \leq \omega(\mathcal{G}')$. Therefore,

$$\sup_n \frac{1}{n} \log \omega(\mathcal{G}^n) \leq \sup_n \frac{1}{n} \log \omega(\mathcal{G}'^n).$$

Because $\mathcal{S}$ attains the supremum in Equation (6.4),

$$C_0(\mathcal{E}) = \sup \frac{1}{n} \log \omega(\mathcal{G}^n) \geq \sup_n \frac{1}{n} \log \omega(\mathcal{G}'^n),$$

which means that $\mathcal{S}'$ does attain and the result follows. ■

It is clear that adjacency relations between input states play a crucial role in calculating the quantum error-free capacity. By definition, if two input states $|v_i\rangle, |v_j\rangle \in \mathcal{S}$ are non-adjacent, then the Hilbert subspaces spanned by the eigenvectors in the support of $\mathcal{E}(|v_i\rangle)$ and $\mathcal{E}(|v_j\rangle)$ are orthogonal. Moreover, as we show below, if $|v_i\rangle \perp_{\mathcal{E}} |v_j\rangle$ then $|v_i\rangle$ and $|v_j\rangle$ are essentially orthogonal. To demonstrate this, we make use of the trace distance between quantum states σ_1 and σ_2 [2, pp.403],

$$D(\sigma_1, \sigma_2) = \frac{1}{2} \text{tr}\, |\sigma_1 - \sigma_2|\,.$$

The trace distance is maximum and equal to one if and only if σ_1 and σ_2 have orthogonal supports. Assuming that $|v_1\rangle$ and $|v_2\rangle$ are non-adjacent pure states, the trace distance between their images is $D(\mathcal{E}(|v_1\rangle), \mathcal{E}(|v_2\rangle)) = 1$. Because quantum channels are contractive [2, pp. 406], i.e., $D(|v_1\rangle, |v_2\rangle) \geq D(\mathcal{E}(|v_1\rangle), \mathcal{E}(|v_2\rangle))$,

$$1 \geq D(|v_1\rangle, |v_2\rangle) \geq D(\mathcal{E}(|v_1\rangle), \mathcal{E}(|v_2\rangle)) = 1, \tag{6.16}$$

which means that $D(|v_1\rangle, |v_2\rangle) = 1$ and $|v_1\rangle$ is orthogonal to $|v_2\rangle$. Intuitively, this means that quantum channels can not take confoundable states into non-confoundable ones.

Consider a qubit channel and an orthonormal basis for the 2-dimensional Hilbert space. Our results allow for the analysis of such channels in a zero-error context: either the zero-error capacity is equal to one bit per use or to zero. This is because these channels have at most two pairwise orthogonal input states, $|v_1\rangle, |v_2\rangle$, and if we take any other state $|v_3\rangle$, it will be non-orthogonal to $|v_1\rangle$ and $|v_2\rangle$ and therefore adjacent.

The above discussions might give the impression that the quantum error-free capacity would be a trivial generalisation of the classical zero-error capacity. By trivial, we mean that

- the capacity is archived using a error-free quantum block code of length one, and

- the supremum in Equation (6.10) can always be achieved by a set $\mathcal{S}$ of mutually orthogonal quantum states.

Surprisedly, there are quantum channels for which the number of non-adjacent codewords behaves unexpectedly when the length of the quantum block code is increased. For a quantum channel exhibited in Section 6.5.5, we claim that the QZEC can only be reached by a set of non-orthogonal quantum states.

6.3.1 The cardinality of the set $\mathcal{S}$ achieving the QZEC

Our next result shows that the quantum zero-error capacity can always be achieved by a set $\mathcal{S}$ of at most d pure states, where d is the dimension of the input Hilbert space. In order to prove this, we need before demonstrate an interesting and useful result to both classical and quantum zero-error information theory.

Let $G = (V, E)$ be an undirected graph such that $V = \{0, \ldots, l-1\}$ and $E \subset \{(i,j); i,j \in V; i \neq j\}$. As we have already seen, the Shannon's n-product of G is defined as follows:

$$\begin{aligned} V(G^n) &= \{0, \ldots, l-1\}^n \\ E(G^n) &= \{(i_1 \ldots i_n, j_1 \ldots j_n); (i_k, j_k) \in E(G) \text{ for at least one } k, \\ &\quad 1 \leq k \leq n\}. \end{aligned} \tag{6.17}$$

For each vertex $i \in V(G)$, we denote by $N(i)$ the set of neighbours of i:

$$N(i) = \{j \in V(G); (i,j) \in E(G)\}. \tag{6.18}$$

Let $\omega(G^n)$ be the clique number of G^n , i.e., the size of the largest clique in G^n. We are interested in determining the clique number of a graph G_k obtained from G in a special way:

Definition 27 *The k-Extended-by-cloning graph (EbC) of G, denoted by G_k, is a graph with $l+1$ vertices which is obtained from G by "cloning" the vertex k of G:*

1. *$V(G_k) = \{0, \ldots, l\}$, where l stands for the label of the "cloned" vertex;*
2. *$E(G_k) = E(G) \cup \{(l,j); j \in N(k)\}$, i.e., both vertices l and k have the same neighbours.*

As an example, let G be the graph illustrated in Figure 6.3(a). Note that in the 3-EbC graph G_3 of Figure 6.3(b), the cloned vertex 5 has the same neighbours of the original vertex 3.

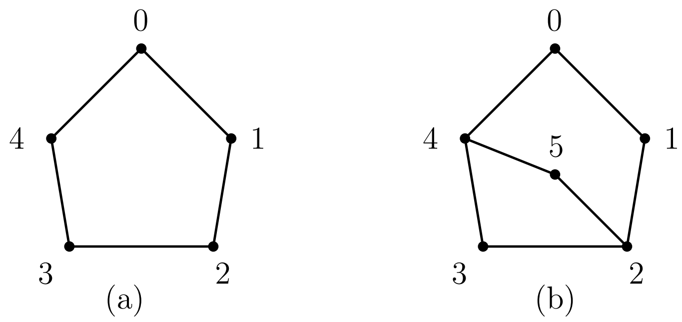

Figure 6.3: (a) A graph G. (b) The 3-extended-by-cloning graph G_3

Theorem 18 *For any n, $\omega(G^n) = \omega(G_k^n)$.*

The theorem implies that the zero-error capacity of a (classical or quantum) channel associated with the graph G_k is equal to the zero-error capacity of a channel associated with G.

Proof. Let $S' \subseteq \{0, \ldots, l\}^n$ be the vertex set of a maximal order clique in G_k^n. By definition, vertices in S' are n-tuples elements of $V(G_k)$ such that, for any two sequences in S', there exists at least one position where the corresponding vertices in G_k are neighbours.

From S', we construct a subset of vertices S of G^n as follows. For any sequence in S' containing the vertex l in one or more positions, we replace l by the original vertex k. An observation shows that all new sequences of S are pairwise distinct, otherwise there would exist at least two sequences belonging to S' for which, in each position, either they are equal or one has l and the other has k. However, from item 2 of Definition 27, l and k are not connected in G_k.

To accomplish the proof, we just need to show that S forms a clique in G^n. Any two sequences in S, say a and b, come from the corresponding sequences a' and b' in S', whose corresponding vertices are connected in G_k^n, since S' forms a clique. Therefore, there is at least one index i for which the vertex a'_i is connected to b'_i in G_k. Moreover, it turns out that either both a'_i and b'_i are different from l – and hence $a_i = a'_i$ and $b_i = b'_i$ so a and b are connected in S – or w.l.o.g. $a'_i = l$ and $a_i = k$ from which we conclude that a and b are connected in S.

Finally, we can write $\omega(G^n) \geq \omega(G_k^n)$. Since the inverse inequality is trivial, the equality holds. ■

Given a graph $G = (V, E)$, a vertex-induced subgraph H of G (often called induced subgraph) is a subset of vertices of G together with all edges whose endpoints are both in this subset. There are two important results which are immediately consequences of Theorem 18.

Corollary 19 *Suppose that instead of cloning a vertex of G we clone any vertex-induced* *subgraph of G to produce a new graph G'. By cloning the subgraph we means that vertices of the subgraph in the cloned graph has the same corresponding neighbours in the original graph. Then, $\omega(G'^n) = \omega(G^n)$ for every n.*

The proof of Corollary 19 is analogous to the proof of Theorem 20.

Corollary 20 *In Definition 27, if we maintain $V(G_{k^*}) = \{0, \dots, l\}$ but replace the statement (2.) with*

2* *$E(G_{k^*}) = E(G) \cup \{(l, j); j \in N(l)\}$, where $N(l) \subseteq N(k)$. i.e., the vertex l in G_{k^*} has the same neighbours of the vertex k in G, but the latter is allowed to have more. (Note that vertices l and k should never be connected).*

Then, $\omega(G^n_{k^*}) = \omega(G^n)$ still holds.

Proof. Note that the graph G_{k^*} can be obtained from the k-EbC G_k of G by probably deleting some edges. Then $\omega(G^n_{k^*}) \leq \omega(G^n_k) \leq \omega(G^n)$. The inverse inequality is trivial. ■

Theorem 18, together with Corollary 20, gives a simple criterion to analyze the zero-error behavior of a quantum channel when a quantum state is "appended" to the set $\mathcal{S}$, since adding a state to $\mathcal{S}$ is equivalent to add a vertex on the corresponding characteristic graph. Below, we show that the zero-error capacity of a quantum channel can always be reached by a set of at most d quantum states, where d is the dimension of the quantum channel.

Proposition 21 *The zero-error capacity of a d-dimensional quantum channel can always be achieved by a set of at most d pure quantum states.*

We first note that there are channels that need exactly d quantum states to reach the capacity, e.g., the identity channel. In order to demonstrate Proposition 21, we only need to prove that, give any set $\mathcal{S}$ containing d quantum states, we cannot do better if we add a state to the set $\mathcal{S}$. The only assumption we make about the set $\mathcal{S}$ is that $\mathcal{S}$ is a set of linearly independent states. Therefore, we do not assume that $\mathcal{S}$ is a set of pairwise orthogonal quantum states.

The main idea of the proof is the following. We add a new state $|\sigma\rangle$ to $\mathcal{S}$. Then, we investigate adjacency relations between $|\sigma\rangle$ and states in $\mathcal{S}$.

Let $\mathcal{S} = \{|\psi_1\rangle, \dots, |\psi_d\rangle\}$ be a linearly independent set of quantum pure states. Because $\mathcal{S}$ is a basis for the Hilbert space of dimension d, the added state $|\sigma\rangle$ is a superposition of states in $\mathcal{S}$. W.l.o.g, let

$$|\sigma\rangle = \sum_{i}^{k} a_i |\psi_i\rangle \tag{6.19}$$

be a superposition state of the first k states of $\mathcal{S}$. Clearly, $|\sigma\rangle$ is non-orthogonal to $|\psi_i\rangle$, $i \leq k$, and therefore it is adjacent to such states. Consider a quantum state $|\psi_m\rangle$, $m > k$.

Lemma 22 *$|\psi_m\rangle \perp_{\mathcal{E}} |\psi_i\rangle$, $i = 1, \ldots, k$, if and only if $|\sigma\rangle \perp_{\mathcal{E}} |\psi_m\rangle$.*

Proof. of Lemma 22. We first prove the direct part. For all i,

$$\mathrm{tr}\,[\mathcal{E}(|\psi_i\rangle\langle\psi_i|)\mathcal{E}(|\psi_m\rangle\langle\psi_m|)] = 0. \tag{6.20}$$

Consider the spectral decomposition $\mathcal{E}(|\psi_m\rangle\langle\psi_m|) = \sum_x \lambda_x |x\rangle\langle x|$. Then,

$$\begin{aligned}
\mathrm{tr}\,[\mathcal{E}(|\psi_i\rangle\langle\psi_i|)\mathcal{E}(|\psi_m\rangle\langle\psi_m|)] &= \mathrm{tr}\left[\mathcal{E}(|\psi_i\rangle\langle\psi_i|)\sum_x \lambda_x |x\rangle\langle x|\right] && (6.21)\\
&= \sum_x \lambda_x \langle x|\mathcal{E}(|\psi_i\rangle\langle\psi_i|)|x\rangle && (6.22)\\
&= 0. && (6.23)
\end{aligned}$$

Because $\mathcal{E}(|\psi_i\rangle\langle\psi_i|)$ is positive,

$$\lambda_x \langle x|\mathcal{E}(|\psi_i\rangle\langle\psi_i|)|x\rangle = 0 \tag{6.24}$$

for all x. Moreover, for all a and $i = 1, \ldots, k$,

$$\begin{aligned}
\lambda_x \langle x|\mathcal{E}(|\psi_i\rangle\langle\psi_i|)|x\rangle &= \lambda_x \langle x|\sum_a E_a|\psi_i\rangle\langle\psi_i|E_a^\dagger)|x\rangle && (6.25)\\
&= \sum_a \lambda_x \langle x|E_a|\psi_i\rangle\langle\psi_i|E_a^\dagger)|x\rangle && (6.26)\\
&= \sum_a \lambda_x ||\langle x|E_a|\psi_i\rangle||^2 && (6.27)\\
&= 0, && (6.28)
\end{aligned}$$

which means that $\lambda_x ||\langle x|E_a|\psi_i\rangle|| = 0$ for all a, x and $i = 1, \ldots, k$. Finally,

$$\begin{aligned}
\mathrm{tr}\,[\mathcal{E}(\sigma)\mathcal{E}(|\psi_m\rangle\langle\psi_m|)] &= \sum_x \lambda_x \langle x|\mathcal{E}(\sigma)|x\rangle && (6.29)\\
&= \sum_x \lambda_x \langle x|\sum_a E_a \sum_{i,j=1}^{k} a_i a_j^* |\psi_i\rangle\langle\psi_i|E_a^\dagger|x\rangle && (6.30)\\
&= \sum_x \sum_a \sum_{i,j} a_i a_j^* \lambda_x \langle x|E_a|\psi_i\rangle\langle\psi_i|E_a^\dagger|x\rangle && (6.31)\\
&= 0, && (6.32)
\end{aligned}$$

since all (complex) numbers $\lambda_x \langle x|E_a|\psi_i\rangle$ have real and imaginary parts equal to zero. The converse part is straightforwardly obtained by developing Equation (6.29). ■

Proof. **(of Proposition 21)** Let $|\sigma\rangle = \sum_{i=1}^{k} a_i|\psi_i\rangle$ be the "appended" state to the set $\mathcal{S} = \{|\psi_1\rangle, \ldots, |\psi_d\rangle\}$. Let G' be the characteristic graph related to $\{|\psi_1\rangle, \ldots, |\psi_d\rangle, |\sigma\rangle\}$. Then, by the Lemma 22, the set of neighbours of $|\sigma\rangle$ is given by

$$N(\sigma) = \{j; |\psi_j\rangle \perp_{\mathcal{E}} |\psi_i\rangle \forall i = 1, \ldots, k; j \in \{k+1, \ldots, d\}\}. \tag{6.33}$$

Therefore, the result follows, since $N(\sigma) \subset N(1)$ and G' can be viewed as G_{1^*} in the sense of Corollary 20. ■

6.4 Measurements reaching the capacity

We discuss in this section quantum measurements attaining the quantum zero-error capacity. As it was defined, the quantum error-free capacity is the maximum transmission rate $R = \frac{1}{n}\log K_n$ of any error-free quantum code of length n and alphabet $\mathcal{S} = \{\rho_1, \ldots, \rho_l\}$. This implies that, for a given n attaining the supremum in Equation (6.4), there exists an error-free quantum code whose codebook contains K_n codewords of length n, $\{\overline{\rho}_1, \overline{\rho}_2, \ldots, \overline{\rho}_{K_n}\}$, such that

$$\begin{aligned}
\mathcal{E}(\overline{\rho}_1) &= \underbrace{\mathcal{E}(\rho_{1_1}) \otimes \mathcal{E}(\rho_{1_2}) \otimes \cdots \otimes \mathcal{E}(\rho_{1_n})}_{P_1}, \\
\mathcal{E}(\overline{\rho}_2) &= \underbrace{\mathcal{E}(\rho_{2_1}) \otimes \mathcal{E}(\rho_{2_2}) \otimes \cdots \otimes \mathcal{E}(\rho_{2_n})}_{P_2}, \\
\vdots \quad & \quad \vdots \\
\mathcal{E}(\overline{\rho}_{K_n}) &= \underbrace{\mathcal{E}(\rho_{K_{n_1}}) \otimes \mathcal{E}(\rho_{K_{n_2}}) \otimes \cdots \otimes \mathcal{E}(\rho_{K_{n_n}})}_{P_{K_n}}
\end{aligned} \tag{6.34}$$

are pairwise orthogonal quantum states in the output Hilbert space of dimension d^n. Define P_i the projector onto the Hilbert subspace spanned by quantum states in the support of $\mathcal{E}(\overline{\rho}_i)$. It is clear that

$$\mathcal{P} = \{P_1, \ldots, P_{K_n}, P_{K_n+1}\}, \tag{6.35}$$

$P_{K_n+1} = \mathbb{1} - \sum_{i=1}^{K_n} P_i$, is a von Neumann measurement allowing of the distinguishibility of the K_n classical messages. Therefore, collective measurements are sufficient to decode any error-free quantum code. It is well-known that measurements performed between several channel outputs are required in order to achieve the Holevo-Schumacher-Westmoreland capacity [2]. Essentially, this means that the mutual information between the input and the output may increase if we make collective measurements instead of individual measurements. A natural question is whether or not individual measurements are sufficient to decode an error-free quantum code. Equivalently, we ask if Bob can always distinguish between the K_n orthogonal tensor product sequences $\mathcal{E}(\overline{\rho}_i) = \bigotimes_{k=1}^{n} \mathcal{E}(\rho_{i_k})$ by means of individual measurements $\mathcal{P}^{(1)}$ on each state $\mathcal{E}(\rho_{i_k})$. As we argue below, the answer is not.

Quantum state discrimination is an important branch of quantum information theory. The general problem consists in determining, with maximum accuracy, the state of a given quantum system chosen from a finite set of quantum states. A variant on the main problem consists in distinguishing multipartite orthogonal quantum states, in a scenario where the compound quantum system, composed of several parts, is held by separated observers [49, 50]. Participants are only allowed to perform individual measurements but they can exchange an arbitrary amount of classical information in order to discriminate the given quantum state. We are interested in the case where global multipartite states are restricted to be tensor products of each shared state [49, 50].

The individual-measurements based decode scheme for a quantum zero-error block code can be viewed as a particular case of the discrimination protocol studied in [49, 50], where all individual measurement on the states $\mathcal{E}(\rho_{i_k})$ should be performed using the same POVM $\mathcal{P}^{(1)}$. Bennett *et. al.* [51] analyzed an example in which two participants, Alice and Bob, are each given a three-state particle and their goal is to distinguish which of nine orthogonal product states in $\{|\psi_1\rangle, \ldots, |\psi_9\rangle\}$, $|\psi_i\rangle = |\alpha_i\rangle \otimes |\beta_i\rangle$, the composite quantum system was prepared in. Because the nine joint quantum states were pairwise orthogonal, they could be reliably distinguished by a collective measurement on both particles. However, the nine states were not orthogonal as individually seen by Alice and Bob. Bennett *et. al.* showed that such joint states could not be reliably distinguished by any sequence of individual measurements, even allowing an arbitrary amount of classical communication between Alice and Bob. This example shows that we cannot always distinguish between states of an orthogonal set of tensor product states using individual measurements. Therefore, individual measurements are not sufficient to attain the quantum zero-error capacity of Definition 24.

6.5 Examples

6.5.1 Bit flip channel

The bit flip channel is a 2-dimensional quantum channel which leaves an input state ρ intact with probability p, and invert the qubit with probability $1-p$.

$$\mathcal{E}(\rho) = p\rho + (1-p)X\rho X. \tag{6.36}$$

This channel has two orthogonal, non-adjacent input states given by

$$\begin{aligned}|v_1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle),\\ |v_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).\end{aligned}$$

The zero-error capacity is achieved by $\mathcal{S} = \{|v_1\rangle, |v_2\rangle\}$, which implies that the zero-error capacity is trivially calculated: $C^0(\mathcal{E}) = \frac{1}{1}\log(2) = 1$ bits per use.

6.5.2 Depolarizing channel

The depolarizing channel in a d-dimensional Hilbert space models a scenario where an input state ρ is either carried out intact with probability p or it is replaced by the completely mixed state $\frac{1}{d}\mathbb{1}_d$ with probability $1-p$ [2]:

$$\mathcal{E}(\rho) = p\frac{1}{d}\mathbb{1}_d + (1-p)\rho, \tag{6.37}$$

where $\mathbb{1}_d$ is the identity operator of dimension d. For this channel, any two input states ρ_i and ρ_j are adjacent for a given $0 < p < 1$. To demonstrate this, we write

$$\begin{aligned}\mathrm{tr}\left[\mathcal{E}(\rho_i)\mathcal{E}(\rho_j)\right] &= \mathrm{tr}\left[\left(p\rho_1 + (1-p)\frac{1}{d}\mathbb{1}_d\right)\left(p\rho_2 + (1-p)\frac{1}{d}\mathbb{1}_d\right)\right]\\ &= \mathrm{tr}\left[p^2\mathrm{tr}\left[\rho_1\rho_2\right] + \frac{p(1-p)}{d}\mathrm{tr}\left[\rho_1+\rho_2\right] + \frac{(1-p)^2}{d}\right]\\ &> 0\end{aligned} \tag{6.38}$$

since $0 < p < 1$. Therefore, the error-free capacity of the d-dimensional depolarizing channel is zero.

6.5.3 Zero-error capacity of classical-quantum channels

In the literature, a quantum channel $\mathcal{E}$ for which the quantum state $(\mathbb{1}\otimes\mathcal{E})(\Gamma)$ is always separable (even for entangled Γ) is called entanglement breaking channel [24]. This important class of quantum channel was first introduced by Holevo [23]. Horodecki *et. al* [24] showed that any entanglement breaking channel can be written in the Holevo form:

$$\mathcal{E}(\rho) = \sum_i \sigma_i \mathrm{tr}\left[\rho X_i\right], \tag{6.39}$$

where $\{\sigma_i\}$ is a fixed family of quantum states and $\{X_i\}$ defines a POVM measurement. The channel is called classical-quantum (c-q) if $X_i = |\psi_i\rangle\langle\psi_i|$, where $\{|\psi_i\rangle\}$ is an orthonormal basis, i.e., POVM elements are one dimensional projectors. In contrast, if $\sigma_i = |\psi_i\rangle\langle\psi_i|$ then it is called a quantum-classical (q-c) channel.

Classical-quantum channels have the property that interference due to superpositions at the channel input are never destroyed at the channel output. To see this, consider a c-q channel defined by an ensemble $\{\sigma_i\}$ and a POVM with operators $X_i = |\psi_i\rangle\langle\psi_i|$. Suppose that a superposition state $|v\rangle = \sum_i v_i|\psi_i\rangle$ is sent through the channel. The density operator at the channel input is $\rho_v = \sum_{ij} v_i v_j^*|\psi_i\rangle\langle\psi_j|$. The output state will be

$$\begin{aligned}\mathcal{E}(\rho_v) &= \sum_i \sigma_i \mathrm{tr}\left[\rho_v|\psi_i\rangle\langle\psi_i|\right] \\ &= \sum_i \langle\psi_i|\rho_v|\psi_i\rangle\sigma_i \\ &= \sum_i \sum_{jk} \langle\psi_i|v_j v_k^*|\psi_j\rangle\langle\psi_k||\psi_i\rangle\sigma_i \\ &= \sum_i ||v_i||^2\sigma_i. \end{aligned} \tag{6.40}$$

Remember that to find the quantum zero-error capacity, one needs to maximize over all sets of input states $\mathcal{S}$. We show below that the zero-error capacity of d-dimensional classical-quantum channels can be attained by the set

$$\mathcal{S} = \{|\psi_1\rangle, \ldots, |\psi_d\rangle\}, \tag{6.41}$$

where $\{|\psi_i\rangle\}$ is an orthonormal basis whose one-dimensional projectors define the POVM of the c-q channel.

Given an arbitrary set $\mathcal{S}$ of input states for a c-q channel $\mathcal{E}_{cq}$, we can construct a characteristic graph G, and the maximum information transmission rate $R_{\mathcal{S}}$ using zero-error quantum codes with alphabet $\mathcal{S}$ is given by:

$$R_{\mathcal{S}} = \sup_n \frac{1}{n} \log \omega(G^n). \tag{6.42}$$

Straightforwardly, the zero-error capacity of $\mathcal{E}_{cq}$ is given by

$$C_0(\mathcal{E}) = \sup_{\mathcal{S}} R_{\mathcal{S}}. \tag{6.43}$$

In order to show that $\mathcal{S}$ in Equation (6.41) attains the capacity, we need to show the following:

Proposition 23 *For a d-dimensional c-q channel defined by $\{\sigma_i\}$ and $\{X_i = |\psi_i\rangle\langle\psi_i|\}_{i=1}^d$,*

$$\sup_{\mathcal{S};|\mathcal{S}|\leq d} R_{\mathcal{S}} \tag{6.44}$$

can always be archived by the set

$$\mathcal{S} = \{|\psi_1\rangle, \ldots, |\psi_d\rangle\}. \tag{6.45}$$

First of all, note that for any state belongs to $\mathcal{S}$,

$$\mathcal{E}(|\psi_i\rangle) = \sigma_i, \tag{6.46}$$

whereas if $|v\rangle$ is a linear combination of $\{|\psi_i\rangle\}$, then the output is given by Equation (6.40). Second, we remember that two vertices u and v are connected in the characteristic graph if and only if $\text{tr}\,[\mathcal{E}(|u\rangle)\mathcal{E}(|v\rangle)] = 0$, i.e., the corresponding output states have orthogonal supports.

Proof. The result follows by construction. Let k be the maximum number of pairwise orthogonal states in $\{\sigma_i\}$, say $\{\sigma_1, \ldots, \sigma_k\}$, $k \leq d$. Due to Equation (6.46), the maximum rate R_{S_k} for any code with $|\mathcal{S}| \leq k$ is achieved by the set $\mathcal{S}_k = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$, since the characteristic graph $G_{(k)}$ due to $\mathcal{S}_k$ is a complete graph. If $k < d$, we should append another pure state $|v\rangle$ to $\mathcal{S}_k$ until $k = d$. The state to be added must lead to a graph $G_{(k+1)}$ with as more connected vertices as possible, i.e, $\mathcal{E}(|v\rangle)$ must have its support orthogonal to as many supp σ_i, $i \leq k$, as possible. Suppose that $|v\rangle$ is a linear combination of $\{|\psi_i\rangle\}$. Then, $\mathcal{E}(|v\rangle) = \sum_i p_i\sigma_i$. If $p_i > 0 \forall i$ then $|v\rangle$ is adjacent to all states in $\mathcal{S}_k$. Because interference due to superpositions of $\{|\psi_i\rangle\}$ are never destroyed at channel output, the state $|v\rangle$ must be any of the $|\psi_m\rangle$, $m > k$, belonging to $\mathcal{S}\backslash\mathcal{S}_k$ such that the set $\{j; |\psi_m\rangle \perp_{\mathcal{E}} |\psi_j\rangle; 1 \leq j \leq k\}$ has maximum cardinality, since $E(G_{(k+1)}) = E(G_{(k)}) \cup \{(i,j); |\psi_i\rangle \perp_{\mathcal{E}} |\psi_j\rangle; 1 \leq j \leq k\}$. The new set will be $\mathcal{S}_{k+1} = \{|\psi_1\rangle, \ldots, |\psi_{k+1}\rangle\}$, where the appended state $|\psi_m\rangle$ has index $k+1$ in $\mathcal{S}_{k+1}$. Clearly, $R_{\mathcal{S}_{k+1}} \geq R_{\mathcal{S}_k}$. Repeating this process will give $\mathcal{S}_d = \mathcal{S}$. ■

What this means is that finding the quantum zero-error capacity of c-q channels is a completely classical problem: we just need to explicit adjacency relation between states in $\mathcal{S}$ in order to determine the characteristic graph $\mathcal{G}$. Then, a maximization is taken over all n: $C^0(\mathcal{E}) = \sup_n \frac{1}{n} \log \omega(\mathcal{G}^n)$. Moreover, the zero-error capacity of a c-q channel can always be reached by a set of pairwise orthogonal states, since $\mathcal{S} = \{|\psi\rangle\}$ is an orthonormal basis for the d-dimensional Hilbert space.

6.5.4 A particular classical-quantum channel

Consider the 5-dimensional c-q channel defined by

$$|\sigma_i\rangle = \frac{|i\rangle + |i+1 \mod 5\rangle}{\sqrt{2}}, \sigma_i = |\sigma_i\rangle\langle\sigma_i| \qquad \text{and} \qquad X_i = |i\rangle\langle i|, \quad 0 \leq i \leq 4, \tag{6.47}$$

where $\{|0\rangle, \ldots, |4\rangle\}$ is the computational basis for the Hilbert space of dimension 5. The set $\mathcal{S}$ that achieves the zero-errro capacity is given by

$$\mathcal{S} = \{|0\rangle, \ldots, |4\rangle\}. \tag{6.48}$$

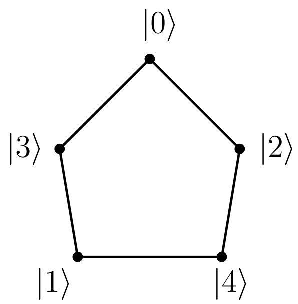

Figure 6.4: Characteristic graph corresponding to the set $\mathcal{S}$ attaining the zero-error capacity of the c-q channel.

The corresponding output states are

$$\begin{aligned}\mathcal{E}(|i\rangle) &= \sum_{j=0}^{4} \sigma_j ||\langle i|j\rangle||^2. \\ &= \sigma_i. \end{aligned} \tag{6.49}$$

Now we can write down the adjacency relations between states in $\mathcal{S}$. The state $|0\rangle$ is non-adjacent to states $|2\rangle$ and $|3\rangle$. To see this note that

$$\mathcal{E}(|0\rangle) = \sigma_0 = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| + \langle 1|}{\sqrt{2}}\right) \tag{6.50}$$

$$= \frac{1}{2}(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|) \tag{6.51}$$

and

$$\mathcal{E}(|2\rangle) = \sigma_2 = \left(\frac{|2\rangle + |3\rangle}{\sqrt{2}}\right)\left(\frac{\langle 2| + \langle 3|}{\sqrt{2}}\right) \tag{6.52}$$

$$= \frac{1}{2}(|2\rangle\langle 2| + |2\rangle\langle 3| + |3\rangle\langle 2| + |3\rangle\langle 3|) \tag{6.53}$$

have orthogonal supports, as well as $\mathcal{E}(|0\rangle)$ and $\mathcal{E}(|3\rangle)$. Therefore,

$$|0\rangle \perp_{\mathcal{E}} |2\rangle, \qquad |0\rangle \perp_{\mathcal{E}} |3\rangle. \tag{6.54}$$

Straightforwardly, one can verify that

$$|1\rangle \perp_{\mathcal{E}} |3\rangle, \qquad |1\rangle \perp_{\mathcal{E}} |4\rangle \text{ and } \qquad |2\rangle \perp_{\mathcal{E}} |4\rangle. \tag{6.55}$$

The characteristic graph related to $\mathcal{S}$ is shown in Figure 6.4(a).

Surprisedly, the $\mathcal{S}$ attaining the capacity gives rise to the pentagon as characteristic graph. Therefore, the capacity of the corresponding c-q channel is

$$C^{(0)}(\mathcal{E}) = C_0(G_5) = \frac{1}{2}\log 5 \text{ bits/use}. \tag{6.56}$$

Although the capacity is reached by a set of pairwise orthogonal states, it is necessary two or more uses of the channel in order to attain the zero-error capacity. A quantum code of length two reaching the capacity is presented below:

$$\begin{aligned} \overline{\rho}_1 = |0\rangle|0\rangle, \quad \overline{\rho}_2 = |1\rangle|2\rangle, \quad \overline{\rho}_3 = |2\rangle|4\rangle \\ \overline{\rho}_4 = |3\rangle|1\rangle, \qquad \overline{\rho}_5 = |4\rangle|3\rangle. \end{aligned} \tag{6.57}$$

The next example presents a mathematically motivated channel that we claim the capacity can only be attained by a set of non-orthogonal states.

6.5.5 Non-orthogonal states attaining the QZEC

We discuss in this section an example of a quantum channel whose zero-error capacity is conjectured to be non-trivial. By non-trivial we mean that the supremum in Equation (6.4) is attained for $n > 1$ and states in the set $\mathcal{S}$ reaching the QZEC contains non-orthogonal states. The following example is mathematically motivated, and has no physical meaning. However, it is interesting because the quantum channel we constructed gives rise to the pentagon as the characteristic graph for a set $\mathcal{S}$ containing non-orthogonal quantum states. Moreover, if the conjecture holds then the capacity cannot be reached by using a set of mutually orthogonal quantum states.

Let $\mathcal{E}$ be a quantum channel with Kraus operators $\{E_1, E_2, E_3\}$ given by

$$E_1 = \begin{bmatrix} 0.5 & 0 & 0 & 0 & \frac{\sqrt{49902}}{620} \\ 0.5 & -0.5 & 0 & 0 & 0 \\ 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -\frac{\sqrt{457}}{50} & \frac{\sqrt{457}}{50} \\ 0 & 0 & 0 & -0.62 & -\frac{289}{1550} \end{bmatrix}, \quad E2 = \begin{bmatrix} 0.5 & 0 & 0 & 0 & -\frac{\sqrt{49902}}{620} \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & \frac{\sqrt{457}}{50} & -\frac{\sqrt{457}}{50} \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix},$$

$$E3 = 0.3|4\rangle\langle 4|,$$

where $\beta = \{|0\rangle, \ldots, |4\rangle\}$ is the computational basis for the Hilbert space of dimension five, as in the example of Section 6.5.4. It is easy to see that $\sum_a E_a^\dagger E_a = \mathbb{1}$, which means that $\mathcal{E}$ is a completely positive trace-preserving quantum operation representing a physical process. The quantum channel was constructed using Matlab®, wherein the .m file is given at Appendix 6.A.

Consider the following set $\mathcal{S}$ of input states for $\mathcal{E}$:

$$\mathcal{S} = \left\{ |v_1\rangle = |0\rangle, |v_2\rangle = |1\rangle, |v_3\rangle = |2\rangle, |v_4\rangle = |3\rangle, |v_5\rangle = \frac{|3\rangle + |4\rangle}{\sqrt{2}} \right\}. \tag{6.58}$$

In order to construct the characteristic graph $\mathcal{G}$, we need to explicit all adjacency relations between states in $\mathcal{S}$. If the channel output $\mathcal{E}(|v_i\rangle)$ is calculated for every $|v_i\rangle \in \mathcal{S}$, one

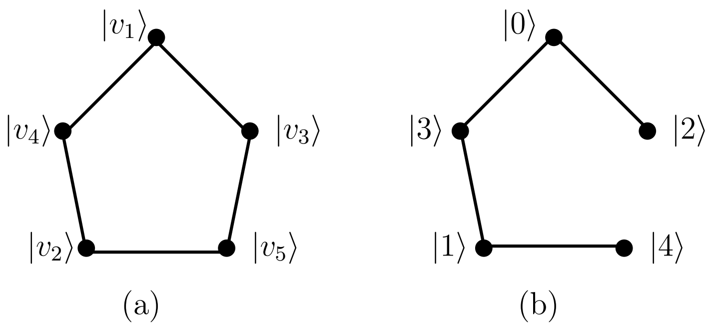

Figure 6.5: (a) Characteristic graph $\mathcal{G}$ for the subset $\mathcal{S}$ containing non-adjacent input states. (b) Characteristic graph for a subset $\mathcal{S}'$ of mutually orthogonal input states. In this case the transmission rate is less than $C^{(0)}$(pentagon) for any zero-error quantum code with alphabet $\mathcal{S}'$.

can verify that

$$|v_1\rangle \perp_{\mathcal{E}} |v_3\rangle, \qquad |v_1\rangle \perp_{\mathcal{E}} |v_4\rangle, \qquad |v_2\rangle \perp_{\mathcal{E}} |v_4\rangle,$$
$$|v_2\rangle \perp_{\mathcal{E}} |v_5\rangle, \qquad \text{and} \qquad |v_3\rangle \perp_{\mathcal{E}} |v_5\rangle.$$

Surprisedly, these relations give rise to the pentagon as characteristic graph, as it is illustrated in Figure 6.5(a).

Note that if we make use of codewords of length one, we can only transmit at most two error-free classical messages through this quantum channel, e.g., by choosing $|v_1\rangle$ and $|v_3\rangle$ or $|v_2\rangle$ and $|v_4\rangle$. Moreover, following the initial Shannon construction, we can construct a quantum error-free codebook of length two containing five non-adjacent codewords:

$$\overline{\rho}_1 = |v_1\rangle|v_1\rangle, \quad \overline{\rho}_2 = |v_2\rangle|v_3\rangle, \quad \overline{\rho}_3 = |v_3\rangle|v_5\rangle$$
$$\overline{\rho}_4 = |v_4\rangle|v_2\rangle, \qquad \overline{\rho}_5 = |v_5\rangle|v_4\rangle. \tag{6.59}$$

The quantum channel discussed above behaves very interestingly because the pentagon is obtained using a set of non-orthogonal quantum states at the channel input. Suppose that we replace the state $|v_5\rangle$ in $\mathcal{S}$ with the state $|4\rangle$ in order to construct a new set $\mathcal{S}' = \beta$ of pairwise orthogonal states. In this case, a calculation shows that the states $|2\rangle$ and $|4\rangle$ are adjacent and the corresponding characteristic graph is given in Figure 6.5(b). The Shannon capacity of this graph is already known [12] and equal to 1 bit per use. Therefore, the maximum rate of any zero-error quantum code with alphabet $\mathcal{S}'$ is one and hence less than the capacity of the pentagon. Finally, we conjecture that one can not do better by taking another set $\mathcal{S}'$, specially if it is composed of pairwise orthogonal states.

6.6 Zero-error capacity and HSW capacity

Quantum channels have a number of capacities that depends fundamentally on the kind of information to be carried (classical or quantum) and on the communication protocol. For example, suppose that Alice and Bob agree on a protocol where codewords are tensor products of quantum states, and decoding is performed using measurements entangled across multiple uses of the channel. In this case, the capacity of the quantum channel for transmitting classical information with a negligible probability of error is given by the Holevo-Schumacher-Westmoreland theorem [7, 8]. Bennett *et. al.* [9, 10] showed that Alice and Bob can do better if they make use of an arbitrary amount of shared entanglement. The so called entanglement-assisted capacity is proved to be an upper bound of the HSW capacity [9].

We demonstrate below that the error-free capacity of a given quantum channel is upper bounded by the HSW capacity $C_{1,\infty}(\mathcal{E})$, i.e.,

$$C^{(0)}(\mathcal{E}) \leq C_{1,\infty}(\mathcal{E}) \equiv \max_{\{p_i,\rho_i\}} \chi_{\{p_i,\rho_i\}},$$

where

$$\chi_{\{p_i,\rho_i\}} = S\left(\mathcal{E}\left(\sum_i p_i\rho_i\right)\right) - \sum_i p_i S(\mathcal{E}(\rho_i)) \tag{6.60}$$

stands for the χ quantity.

The HSW protocol states that codewords are composed of signal states ρ_i, where the probability of using ρ_i is p_i. Note that the maximum is taken over all ensembles $\{p_i, \rho_i\}$ of possible input states ρ_i to the channel. The coding theorem says that if Alice and Bob agree on a quantum code with rate less than or equal to the HSW capacity, it is possible to transmit classical information reliably through a quantum channel with a probability of error *asymptotically* zero (not actually zero).

Let R be the rate of any error-free quantum code. We assume that Alice sends to Bob messages chosen randomly and uniformly from the set $\{1, \ldots, 2^{nR}\}$, i.e., if we define $\mathbf{X}$ as a random variable representing indexes of classical messages, then $\mathbf{X}$ is uniformly distributed over $\{1, \ldots, 2^{nR}\}$. As a straightforward consequence we have

$$H(\mathbf{X}) = nR, \tag{6.61}$$

where H stands for the classical Shannon entropy [20]. Now we take $\mathbf{Y}$ as a random variable representing the output when Bob performs measurements described by a POVM $\{M_i\}$. By the definition of mutual information,

$$nR = H(\mathbf{X}) = H(\mathbf{X}|\mathbf{Y}) + I(\mathbf{X}, \mathbf{Y}). \tag{6.62}$$

Because we are making use of an error-free quantum code, there are no decoding errors. Then, given an output word y, there is no uncertainty about the classical message actually sent, i.e., $H(\mathbf{X}|\mathbf{Y}) = 0$. Suppose that Alice encodes the message i as $\overline{\rho}_i = \rho_{i_1} \otimes \cdots \otimes \rho_{i_n}$. Applying the Holevo bound we get

$$\begin{aligned} nR &= I(\mathbf{X}, \mathbf{Y}) && (6.63) \\ &\leq S\left(\sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} \mathcal{E}(\overline{\rho}_i)\right) - \sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} S(\mathcal{E}(\overline{\rho}_i)). && (6.64) \end{aligned}$$

Remember that $\mathcal{E}(\overline{\rho}_i) = \mathcal{E}(\rho_{i_1}) \otimes \cdots \otimes \mathcal{E}(\rho_{i_n})$. Hence, we can apply the subadditivity of the entropy, $S(A, B) \leq S(A) + S(B)$ [2, pp. 515]:

$$\begin{aligned} nR &\leq \sum_{j=1}^{n} S\left(\sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} \mathcal{E}(\rho_{i_j})\right) - \sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} \sum_{j=1}^{n} S(\mathcal{E}(\rho_{i_j})) && (6.65) \\ &= \sum_{j=1}^{n} \left[S\left(\sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} \mathcal{E}(\rho_{i_j})\right) - \sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} S(\mathcal{E}(\rho_{i_j}))\right]. && (6.66) \end{aligned}$$

Because the capacity in Eq. (6.60) is calculated by taking the ensemble that gives the maximum, we can conclude that each term on the right side of (6.66) is less than or equal to $C_{1,\infty}(\mathcal{E})$. Then,

$$nR \leq nC_{1,\infty}(\mathcal{E}) \tag{6.67}$$

and the inequality follows for all zero-error quantum block codes of length n and rate R. This is an intuitive result, since one would expect to increase the information transmission rate whenever a small probability of error is allowed.

Example 3 *Consider the quantum channel of Section 6.5.5 and the set $\mathcal{S}$ of non-orthogonal states giving rise to the pentagon as characteristic graph. Obviously, we do not know if $\mathcal{S}$ attains the supremum in Equation (6.10). However, if $\mathcal{S}$ does attain then the zero-error capacity of $\mathcal{E}$ is $\frac{1}{2}\log 5$. In this case, a simple calculation shows that the χ quantity for the family $\{\mathcal{S}, p_i = 1/5\}$ is greater than $C_0(G_5)$, i.e,*

$$\begin{aligned} \chi_{\{\mathcal{S},1/5\}} &= \frac{1}{5}\left[S\left(\mathcal{E}\left(\sum_{i=1}^{5} |v_i\rangle\langle v_i|\right)\right) - \sum_{i=1}^{5} S(\mathcal{E}(|v_i\rangle\langle v_i|))\right] \\ &= 1.53 \\ &\geq C_0(G_5) \\ &= 1.16. && (6.68) \end{aligned}$$

6.7 Conclusions

We have introduced in this chapter a new kind of capacity of quantum channels. The quantum zero-error capacity was defined as the least upper bound of rates at which classical information can be transmitted through a noisy quantum channel with a probability of error equal to zero. The communication protocol is essentially the same protocol of the Holevo-Schumacher-Westmoreland capacity [7, 8], except that no transmission errors are allowed. The quantum zero-error capacity is a generalisation of the classical zero-error capacity defined by Shannon [12].

Appendix

6.A Matlab m-file

The matlab m-file below was used to find the quantum channel of the example in Section 6.5.5.

```
%
% Find a quantum channel whose zero-error capacity is
% reached by using a set of non-orthogonal input states.
%

% Clear the workspace
clear all;

% Define variables
syms a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 a12 zero real;
syms c2 n real;

% Computational basis for the 5-dimensional Hilbert space
v1 = [1;0;0;0;0];
v2 = [0;1;0;0;0];
v3 = [0;0;1;0;0];
v4 = [0;0;0;1;0];
v5 = [0;0;0;0;1];
u5 = 1/sqrt(2)*[0;0;0;1;1];

% ... and the corresponding density matrices
P1=v1*v1';
P2=v2*v2';
```

```
P3=v3*v3';
P4=v4*v4';
P5=v5*v5';
U5=u5*u5';

% Projectors with some desired properties
P12 = P1 + P2;
P23 = P2 + P3;
P34 = P3 + P4;
P45 = P4 + P5;
P51 = P5 + P1;

% Variable initializations in order to simplify system of equation
E3 = zeros(5);
E2 = zeros(5);
n = 0.5;
a5= 0.62;
a3= sqrt((1-a5^2 - n^2)/2);
c2 = 0.3;
zero = 0;

% Kraus operators
E1 = [ a1 0 0 0 a2; a1 a1 0 0 0; 0 a1 a1 0 0; ...,
    0 0 a1 -a3 a3; 0 0 0 a4 a1];
E2 = [ a1 0 0 0 -a2; a1 -a1 0 0 0; 0 a1 -a1 0 0; ...,
    0 0 a1 a3 -a3; 0 0 0 a5 a6];
E3 = [0 0 0 0 0; 0 0 0 0 0; 0 0 0 0 0; ...,
     0 0 0 0 0; 0 0 0 0 c2];

% Avoid the channel to own 3 pairwise non-adjacent input states
Condicoes(:,:,1) = E1*P1*E1' + E2*P1*E2' + E3*P1*E3' - ...,
    P12*(E1*P1*E1' + E2*P1*E2' + E3*P1*E3')*P12;
Condicoes(:,:,2) = E1*P2*E1' + E2*P2*E2' + E3*P2*E3' - ...,
    P23*(E1*P2*E1' + E2*P2*E2' + E3*P2*E3')*P23;
Condicoes(:,:,3) = E1*P3*E1' + E2*P3*E2' + E3*P3*E3' - ...,
    P34*(E1*P3*E1' + E2*P3*E2' + E3*P3*E3')*P34;
```

```
Condicoes(:,:,4) = E1*P4*E1' + E2*P4*E2' + E3*P4*E3' - ...,
    P45*(E1*P4*E1' + E2*P4*E2' + E3*P4*E3')*P45;
Condicoes(:,:,5) = E1*U5*E1' + E2*U5*E2' + E3*U5*E3' - ...,
    P51*(E1*U5*E1' + E2*U5*E2' + E3*U5*E3')*P51;

% Completude condition  \sum_k E_k'E_k = I
Condicoes(:,:,6) = E1'*E1 + E2'*E2 + E3'*E3 - eye(5);



% OrdemC ->  Order of matrices E_i e P_i
% nCondicoes -> Number of conditions
[i OrdemC nCondicoes] = size (Condicoes);

% ArgumentoSolve groups all nonzero conditions
ArgumentoSolve = 'solve(';

% Prepare solve argument
for k = 1:nCondicoes
    for i=1:OrdemC
        for j=1:OrdemC
            if Condicoes(i,j,k) ~= zero
                ArgumentoSolve = [ArgumentoSolve, 'Condicoes(',...,
                num2str(i),',',num2str(j),',',num2str(k),'),'];
            end
        end
    end
end

% Replace last comma with a parenthesis
i = length (ArgumentoSolve);
ArgumentoSolve (1,i) = ')';

% Now solve the system equation
Sol = eval (ArgumentoSolve);

% Format the output
```

```
Variaveis = fieldnames(Sol);

for i=1:length(Variaveis)
    ArgumentoEval = [char(Variaveis(i)) '= Sol.',...,
        char(Variaveis(i)),'(7);'];
    eval (ArgumentoEval);
end
```

Chapter 7

Conclusions and Perspectives

7.1 Conclusions

In this work we have proposed a new kind of capacity for quantum channel, namely, the quantum zero-error capacity, which was defined as the least upper bound of rates at which classical information can be transmitted without error through a noisy quantum channel. The quantum zero-error capacity is a generalisation of the zero-error capacity of classical discrete memoryless channels. The error-free capacity can also be viewed as a particular case of the Holevo-Schumacher-Westmoreland capacity [7, 8], in a scenario where no transmission errors are allowed.

Initially, we formally defined an error-free quantum code and the concept of non-adjacent input states. We have established a necessary and sufficient condition for a quantum channel to have a positive zero-error capacity. We also reformulated the problem of finding the quantum zero-error capacity in the language of graph theory, and we have shown that the two definitions are equivalent. This equivalence in the definitions led to an interpretation of the quantum zero-error capacity in terms of zero-error capacities of DMCs.

Next, we have studied quantum states and measurements attaining the quantum zero-error capacity. We have shown that the channel capacity can be reached by using an ensemble of pure states. In the literature, there exists a similar result about the HSW capacity [2, pp. 555]. We also defined the concept of k-extended-by-cloning graph and we have demonstrated that the Shannon capacity of a k-EbC graph is equal to the Shannon capacity of the original graph. This was used to show that the quantum zero-error capacity can always be reached by a set of at most d pure states. Concerning measurements, we have shown that collective von Neumann measurements are sufficient to attain the quantum zero-error capacity. Next, we investigated the error-free capacity of some quantum channels. For classical-quantum (c-q) channels, which are a class of entanglement-breaking channels, we have determined the set $\mathcal{S}$ achieving the capacity; an example of a particular c-q channel was given for which we were able to calculate the capacity. We also have exhibited a quantum channel whose zero-error capacity is claimed to be non-trivial, in the sense that the quantum zero-error capacity can only be reached by using a set of non-orthogonal quantum states, and we need to make two or more uses of the channel in order to attain the capacity. Furthermore, the quantum channel we have exhibited gives rise to the pentagon as characteristic graph for the ensemble of non-adjacent quantum states.

Finally, we have related the quantum zero-error capacity to the HSW capacity, by showing that the former is upper bounded by the latter.

7.2 Perspectives

We give below a (non-exhaustive) list of topics that can be investigated in the quantum zero-error scenario.

7.2.1 A generalisation of the Lovász's theta function

Lovász's theta [21] is a polynomially computable functional which is an upper bound of the zero-error capacity of discrete memoryless channels. It would be interesting to verify the existence of a generalisation of such functional to quantum information theory. Classically, Lovász's theta function is defined as being the *value* of an *orthonormal vector representation* of the adjacency graph associated with a DMC. A non-trivial generalisation should consider an *orthonormal representation* obtained (in some way) from quantum channel operators $\{E_i\}$.

Recently, Beigi and Shor [46] studied the complexity of computing the zero-error capacity of quantum channels. The authors showed that the quantum zero-error capacity belongs to a class of problems called QMA-complete. QMA is the class of problems that can be solved by a quantum algorithm in polynomial time given that a quantum witness is available. Authors restricted themselves to entanglement-breaking channels [24]. A polynomial computable generalisation of the Lovázs theta function would be an interesting tool to investigate the zero-error capacity of quantum channels.

7.2.2 Variations in the communication protocol

In a recent paper, Duan and Shi [48] have showed an interesting feature of quantum channels concerning the quantum zero-error capacity. Initially, senders and receivers share an arbitrary amount of entanglement. In a scenario where m senders want to transmit information to n receivers, authors described a protocol that enable, for a particular quantum channel, two senders and two receivers to exchange information with zero probability of error. What is interesting is that senders can only transmit information if they make two or more uses of the channel, i.e, no information can be transmitted with a single use of the channel. This behaviour contrasts significantly with the classical case, where information can be transmitted in a single use if and only if it can be transmitted in multiple uses.

Another possibility is investigate feedback channels as an extra resource. Because classical feedback can increase the zero-error capacity of classical DMC [12], one may expect that the same is true in the quantum case. We remember that the Shannon's feedback protocol described in Section 5.4 requires the transmission (from the receiver to the sender) of each actual received symbol, which is used to choose the next symbol to be transmitted. Nevertheless, this feedback protocol cannot be directly employed in the quantum case because measurements must be performed collectively on the whole received quantum codeword. Therefore, a different feedback strategy must be adopted in order to investigate the quantum zero-error capacity with feedback. We could also investigate the scenario where an arbitrary amount of shared entanglement among the sender and the receiver is available.

7.2.3 Decoherence-free subspaces and noiseless subsystems

Apart from studies of quantum error-correction codes, some researches allowed for the development of an alternative "passive" error *prevention* scheme, in which logical qubits are encoded within subspaces which do not decohere for reasons of symmetry [37, 36]. The existence of such Decoherence-Free Subspaces (DFS) has been shown by projection onto the symmetric subspace of multiple copies of a quantum computer [52], and by use of a group-theoretic argument [36]. Further works suggested that universal quantum computation is possible within these subspaces [40, 42]. The so-called Noiseless Subsystems (NS) [44] are a generalisation of DFS, in which quantum information is encoded in a specific sector of a given quantum system. This sector remains invariant to decoherence. We should study relations between the theory of noiseless subsystems (including noiseless quantum codes) and the zero-error capacity of quantum channels.

7.2.4 Graph states

The quantum error-free capacity has a nice formulation in terms of graph. It would be interesting to investigate whether there exist connections between the zero-error capacity and other areas of quantum information whose properties can be stated in terms of graphs, e.g., quantum Fourier transform in a one-way computer [53, 54] and quantum error correction codes [55]. We should pay a special attention to the theory of *graph states* [56, 57, 58]. A graph state is a pure multipartite quantum state of a distributed quantum system that corresponds to a graph, where vertices take the role of quantum spin systems (qubits) and edges represent Ising interactions between pairs of such quantum systems.

Acknowledgements

I would like to thank the Programme Alβan, the European Union Programme of High Level Scholarships for Latin America, scholarship no. E05D051893BR, for financial support. This work is supported in Brazil by the Brazilian National Council for Scientific and Technological Development (CNPq) (CT-INFO Quanta, grants # 552254/02-9). This work has been partially supported by EC under project SECOQC (contract # IST-2003-506813).

List of publications

[1] R. A. C. Medeiros and F. M. de Assis, Quantum zero-error capacity, *Int. J. Quant. Inf.*, vol. 3, no. 1, pp. 135–139, 2005.

[2] R. A. C. Medeiros and F. M. de Assis, Quantum zero-error capacity and HSW capacity, in *Proceedings of the The Seventh International Conference on Quantum Communication, Measurement and Computing QCMC'04*. AIP Conference Proceedings, vol. 734, no. 1. American Institute of Physics, 2004, pp. 52–54.

[3] R. A. C. Medeiros and F. M. de Assis, Capacidade erro-zero de canais quânticos e estados puros, in *Anais do XXII Simpósio Brasileiro de Telecomunicações*, ser. XXII Simpósio Brasileiro de Telecomunicações - SBrT'05, Campinas-SP, Brazil, 2005.

[4] R. A. C. Medeiros, R. Alléaume, G. Cohen, and F. M. de Assis. Zero-error capacity of quantum channel and noiseless subsystems. In *IEEE International Telecommunications Symposium ITS2006*, Brazil, 2006.

[5] R. A. C. Medeiros, R. Alléaume, G. Cohen, and F. M. de Assis. On the zero-error capacity of quantum channels and noiseless subsystems. In *Accepted to the 8th International Conference on Quantum Communication, Measurement and Computing (QCMC 2006)*, Japan, 2006.

[6] R. A. C. Medeiros, R. Alléaume, G. Cohen, and F. M. de Assis. Quantum states characterization for the zero-error capacity. quant-ph/0611042, 2006.

[7] R. A. C. Medeiros, R. Alléaume, G. Cohen, and F. M. de Assis. Quantum states characterization for the zero-error capacity. In *2007 IEEE Winter School on Coding and Information Theory*, La Colle sur Loup, France, 2007.

[8] R. A. C. Medeiros, F. M. de Assis, R. Alléaume and G. Cohen. Capacidade erro-zero de canais quânticos com medições coletivas. In *Anais do XXV Simpósio Brasileiro de Telecomunicações (SBrT07)*, Recife, Brasil, 2007.

Bibliography

[1] C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE Trans. Info. Theory*, 44(6):2724–2755, October 1998.

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

[3] L. B. Levitin. Physical information, part II: Quantum systems. In D. Matzke, editor, *Proceedings of the Workshop on Physics and Computation: PhysComp'92*, pages 215–219, Los Alamitos, CA, 1993. IEEE Computer Society.

[4] L. B. Levitin. *Optimal Quantum Measurements for Two Pure and Mixed State*, pages 439–448. Quantum Communications and Measurement. Plenum Press, New York, 1995.

[5] C. A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. PhD thesis, University of New Mexico, Albuquerque, 1995.

[6] P. W. Shor. The adaptive classical capacity of a quantum channel, or information capacities of three symmetric pure states in three dimensions. *IBM. J. Res. & Dev.*, 48(1):115–137, 2004.

[7] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Info. Theory*, 44(1):269–273, 1998.

[8] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56(1):131–138, 1997.

[9] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83:3081–3084, 1999.

[10] 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. *IEEE Trans. Info. Theory*, 48:2637–265l, 1998.

[11] C. Macchiavello and G. M. Palma. Entanglement-enhanced information transmission over a quantum channel with correlated noise. *Phys. Rev. A*, 65(5):050301, Apr 2002.

[12] C. E. Shannon. The zero error capaciy of a noisy channel. *IRE Trans. Inform. Theory*, IT-2(3):8–19, 1956.

[13] C. E. Shannon. A mathematical theory of communication. *The Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[14] W. Haemers. An upper bound for the Shannon capacity of a graph. In *Colloq. Math Soc.*, Algebraic Methods in Graph Theory, pages 267–272, Szeged, Hungary, 1978.

[15] W. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Info. Theory*, 25:231–232, 1979.

[16] N. Alon. The Shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[17] A. W. Naylor and G. R. Sell. *Linear Operator Theory in Engineering and Science*. Applied Math Sciences. Springer-Verlag New York Inc., New York, 2nd. edition, 1982.

[18] A. S. Holevo. Information theoretical aspects of quantum measurements. *Probl. Info. Transm.*, 9(2):31–42, 1973.

[19] P. W. Shor. Capacities of quantum channels and how to find them. *Mathematical Programming*, 97(1):311–335, 2003.

[20] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons Inc., New York, 1991.

[21] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Info. Theory*, 25(1):1–7, 1979.

[22] M. Grötschel and L. Lovász. *Geometric Algorithms and Combinatorial Optimization*. Springer, Berlin, Germany, 1988.

[23] A. S. Holevo. Coding theorems for quantum channels. *quant-ph/9809023*, 1998.

[24] M. Horodecki, P. W. Shor, and M. B. Ruskai. Entanglement breaking channels. *Rev. Math. Phys*, 15:629–641, 2003.

[25] H. Barnum, M. A. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Phys. Rev. A*, 57:4153–4175, 1998.

[26] J. Körner and A. Orlitsky. Zero-error information theory. *IEEE Trans. Info. Theory*, 44(6):2207–2229, 1998.

[27] E. L. Lima. *Espaços Métricos*. Projeto Euclides. Instituto de Matemática Pura e Aplicada, CNPQ, Rio de Janeiro, 1nd. edition, 1977.

[28] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge University Press, Cambridge, second edition edition, 2001.

[29] B. Bollobás. *Modern graph theory*. Springer-Verlag New York, Inc., New York, 1998.

[30] M. Rosenfeld. On a problem of Shannon. In *Proc. Amer. Math. Soc.*, volume 18, pages 318–319, 1967.

[31] E. R. Scheinerman and D. H. Ullman. *Fractional graph theory*. Wiley-Interscience, New York, 1997.

[32] P. Elias. List decoding for noisy channels. *WESCON Conv. Rec.*, P-II:94–104, 1957.

[33] P. Elias. Zero-error capacity under list decoding. *IEEE Trans. Info. Theory*, 34:1070–1074, 1988.

[34] D. Gottesman. Class of quantum error-correcting codes saturating the quantum Hamming bound. *Phys. Rev. A*, 54(3):1862–1868, Sep 1996.

[35] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.*, 78(3):405–408, Jan 1997.

[36] P. Zanardi and M. Rasetti. Error avoiding quantum codes. *Mod. Phys. Lett. B*, 11(25):1085, 1997.

[37] P. Zanardi and M. Rasetti. Noiseless quantum codes. *Phys. Rev. Lett.*, 79:3306, 1997.

[38] D.W. Kribs, R. Laflamme, D. Poulin, and M. Lesosky. Operator quantum error correction. *quant-ph/0504189*, 2005.

[39] D. W. Kribs, R. Laflamme, and D. Poulin. Unified and generalized approach to quantum error correction. *Phys. Rev. Lett.*, 94:180501, 2005.

[40] D. A. Lidar, I.L. Chuang, and K. B. Whaley. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.*, 81:2594, 1998.

[41] D. Bacon, J. Kempe, D. A. Lidar, and K. B. Whaley. Universal fault-tolerant quantum computation on decoherence-free subspaces. *Phys. Rev. Lett.*, 85(8):1758–1761, Aug 2000.

[42] J. Kempe, D. Bacon, D. A. Lidar, and K. B. Whaley. Theory of decoherence-free fault-tolerant universal quantum computation. *Phys. Rev. A*, 63:042307, 2001.

[43] E. Knill, R. Laflamme, and L. Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, 84:2525, 2000.

[44] P. Zanardi. Stabilizing quantum information. *Phys. Rev. A*, 63:12301, 2001.

[45] R. A. C. Medeiros and F. M. de Assis. Quantum zero-error capacity. *Int. J. Quant. Inf.*, 3(1):135–139, 2005.

[46] S. Beigi and P. W. Shor. On the complexity of computing zero-error and Holevo capacity of quantum channels. *quant-ph/0709.2090v2*, 2007.

[47] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi. Classical and quantum computation. In *Graduate Studies in Mathematics*, volume 47. American Mathematical Society, 2002.

[48] R. Duan and Y. Shi. Entanglement between two uses of a noisy multipartite quantum channel enables perfect transmission of classical information, 2007.

[49] W. K. Wootters. Distinguishing unentangled states with an unentangled measurement. *quant-ph/0506149*, 2005.

[50] J. Walgate, A. J. Short, L. Hardy, and V. Vedral. Local distinguishability of multipartite orthogonal quantum states. *quant-ph/0007098*, 2000.

[51] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59(2):1070–1091, Feb 1999.

[52] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello. Stabilization of quantum computations by symmetrization. *SIAM J. Comp.*, 26:1541, 1997.

[53] R. Raussendorf, D. E. Browne, and H. J. Briegel. Measurement-based quantum computation on cluster states. *Phys. Rev. A*, 68(2):022312, Aug 2003.

[54] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86(22):5188–5191, May 2001.

[55] D. Schlingemann and R. F. Werner. Quantum error-correcting codes associated with graphs. *Phys. Rev. A*, 65(1):012308, Dec 2001.

[56] D. Schlingemann. Cluster states, algorithms and graphs. *quant-ph/0305170*, 2003.

[57] W. Dür, H. Aschauer, and H.-J. Briegel. Multiparticle entanglement purification for graph states. *Phys. Rev. Lett.*, 91(10):107903, Sep 2003.

[58] M. Hein, J. Eisert, and H. J. Briegel. Multiparty entanglement in graph states. *Phys. Rev. A*, 69:062311, 2004.

Index

accessible information, 24, **49**, 50
adjacency
  between classical symbols, **59**
  between quantum states, **77**
  graph, *see* graph, adjacency
  matrix, **64**, 80
adjacency-reducing mapping, **61**, 66

Bell basis, **41**, 53

capacity
  classical
    information capacity, **56**
    Shannon capacity of G, **63**
    zero-error, 25, 58, **60**, 63
  quantum, 45
    adaptive, 24, **52**
    entanglement-assisted, 52, **53**
    Holevo-Schumacher-Westmoreland, 24, 50, **51**, 53, 80, 95
    one-shot, 24, **50**
    zero-error, 25, 75, **76**
    zero-error equivalent definition, 80
Cauchy sequence, **31**
channel
  G_5 channel, **57**, 61
  binary erasure, **56**
  bit flip, 88
  classical DMC, 25, **55**, 61, 62
  classical-quantum, **89**
  contractive, 82
  depolarizing, **48**, 51, 89
  entanglement breaking, 89
  product, 71
  quantum channel, 47, **48**
  quantum-classical, 89
  sum, 71
channel coding theorem, **58**
chromatic number, 26, **63**
clique number, 26, **62**, 79, 82
codes
  classical (M, n) block code, **57**
  classical (M, n) error-free block code, **58**
  quantum (K_n, n) error-free block code, **76**
  quantum zero-error, 25
collective measurements, 51, 87
communication protocol
  error-free, 75
complete feedback, 69
computational basis, *see* vector space, computational basis

density operator, 42
  characterization, 42
Dirac's notation, 29
distinguishable sequences, 63
distinguishable states, 77

eigenspace, **34**
eigenvalue, **34**
eigenvalues, 51
eigenvector, **34**

entanglement, 41
entropy
  Shannon, 95
    binary entropy, **49**, 56
  von Neumann, **45**
    conditional entropy, 47
    joint entropy, 46
    properties, 46
    relative entropy, 46
    subadditivity, **47**, 96
EPR pair, 41, 52

fractional chromatic number, **65**

graph, 25
  k-extended-by-cloning, **83**
  Petersen, 64
  adjacency, 26, **64**
  characteristic, 26, **62**, 64, **79**, 92
  complete subgraph, 79
  covered, 65
  neighbours, 83
  orthonormal representation, **66**
  pentagon, 26, 62, 68, 92
  perfect, **64**
  Shannon n-product of $\mathcal{G}$, **79**
  Shannon capacity, 25
  Shannon product G, 62
  value of a representation, 66
  vertex-induced subgraph, 84

Hilbert space, *see* vector space, Hilbert space
Holevo bound, **49**, 96
Holevo quantity, **50**
HSW theorem, *see* capacity, Holevo-Schumacher-Westmoreland

information transmission rate, 90
inner product, **30**, 32
interference, **90**

ket, 30

linear algebra, 29
linearly independent set, **30**, 85
list decoding, 70
  zero-error capacity, 70
Lovász theta function, 61, **67**
Lovázs theta function, 26

measurements reaching the QZEC, 87
metric, **31**
mixed state, *see* quantum state, mixed
mutual information
  classical, 49, 53
  quantum, **47**, 53

non-orthogonal quantum states, 93

operation elements, 48
operator
  density, *see* density operator
  Hermitian, **34**
  linear, 33
  normal, 35
  Pauli, **34**, 49, 53
  positive, **35**
  projector, **35**, 87
  trace, 42, 81
  unitary, **35**, 47
operator-sum representation, *see* channel, quantum channel

postulate, 37
  composite systems, 41
  evolution, 38
  in terms of density operators, 43
  measurements, 38

POVM, 40
projective, 39
state space, 37
POVM, *see* postulate, measurements, POVM
pure state, *see* quantum state, pure

quantum channels, *see* channel, quantum channel
quantum codewords, 76
quantum mechanics, 29
quantum operation, 47
quantum operations
completely positive, 47
completely positive trace-preserving, **48**, 75, 93
quantum state
depolarized, 48
entangled, 42
achieving the QZEC, **80**
discrimination, 88
mixed, **42**, 49
pure, **42**, 81
qubit, *see* vector, qubit

spectral decomposition, 35
superdense coding, 52
superposition, 37

teleportation, 52
tensor product, **35**, 53
trace, *see* operator, trace
trace distance, **82**

vector, **29**
orthogonal, **31**
qubit, 37
unitary, **31**
vector space, **29**
basis, **30**
complete, **32**
computational basis, **33**, 91
Hilbert space, **32**
metric space, **31**
orthogonal basis, **31**
orthonormal basis, **31**, 33
vector subspace, **30**
vertex-induced subgraph, *see* graph, vertex-induced subgraph
von Neumann entropy, *see* entropy, von Neumann