KTH Engineering Sciences

Quantum error correction

JONAS ALMLÖF

Licentiate Thesis in Physics
KTH School of Engineering Sciences
Stockholm, Sweden, 2012

TRITA-FYS 2012:19
ISSN 0280-316X
ISRN KTH/FYS/–12:19–SE
ISBN 978-91-7501-317-6

KTH, Skolan för teknikvetenskap
Lindstedtsvägen 5
SE-100 44 Stockholm
Sweden

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie licentiatexamen i fysik onsdagen den 19 dec 2012 klockan 14:00 i sal FB53, AlbaNova Universitetscentrum, Kungl Tekniska högskolan, Roslagstullsbacken 21, Stockholm.

Tryck: Universitetsservice US AB

Abstract

This thesis intends to familiarise the reader with quantum error correction, and also show some relations to the well known concept of information – and the lesser known quantum information. Quantum information describes how information can be carried by quantum states, and how interaction with other systems give rise to a full set of quantum phenomena, many of which have no correspondence in classical information theory. These phenomena include decoherence, as a consequence of entanglement. Decoherence can also be understood as "information leakage", i.e., knowledge of an event is transferred to the reservoir – an effect that in general destroys superpositions of pure states.

It is possible to protect quantum states (e.g., qubits) from interaction with the environment – but not by amplification or duplication, due to the "no-cloning" theorem. Instead, this is done using coding, non-demolition measurements, and recovery operations. In a typical scenario, however, not *all* types of destructive events are likely to occur, but only those allowed by the information carrier, the type of interaction with the environment, and how the environment "picks up" information of the error events. These characteristics can be incorporated into a code, i.e., a channel-adapted quantum error-correcting code. Often, it is assumed that the environment's ability to distinguish between error events is small, and I will denote such environments "memory-less".

This assumption is not always valid, since the ability to distinguish error events is related to the *temperature* of the environment, and in the particular case of information coded onto photons, $k_{\mathrm{B}}T_{\mathrm{R}} \ll \hbar\omega$ typically holds, and one must then assume that the environment *has* a "memory". In this thesis, I describe a short quantum error-correcting code (QECC), adapted for photons interacting with a cold environment, i.e., this code protects from an environment that continuously records which error occurred in the coded quantum state.

Also, it is of interest to compare the performance of different QECCs – But which yardstick should one use? We compare two such figures of merit, namely the quantum mutual information and the quantum fidelity, and show that they can not, in general, be simultaneously maximised in an error correcting procedure. To show this, we have used a five-qubit perfect code, but assumed a channel that only cause bit-flip errors. It appears that quantum mutual information is the better suited yardstick of the two, however more tedious to calculate than quantum fidelity – which is more commonly used.

Sammanfattning

Denna avhandling är en introduktion till kvantfelrättning, där jag undersöker släktskapet med teorin om klassisk information – men också det mindre välkända området kvantinformation. Kvantinformation beskriver hur information kan bäras av kvanttillstånd, och hur växelverkan med andra system ger upphov till åtskilliga typer av fel och effekter, varav många saknar motsvarighet i den klassiska informationsteorin. Bland dessa effekter återfinns dekoherens – en konsekvens av s.k. *sammanflätning*. Dekoherens kan också förstås som "informationsläckage", det vill säga att kunskap om en händelse överförs till omgivningen – en effekt som i allmänhet förstör superpositioner i rena kvanttillstånd.

Det är möjligt att med hjälp av kvantfelrättning skydda kvanttillstånd (t.ex. qubitar) från omgivningens påverkan, dock kan sådana tillstånd aldrig förstärkas eller dupliceras, p.g.a *icke-kloningsteoremet*. Tillstånden skyddas genom att införa redundans, varpå tillstånden interagerar med omgivningen. Felen identifieras m.h.a. icke-förstörande mätningar och återställs med unitära grindar och ancilla-tillstånd. Men i realiteten kommer inte *alla* tänkbara fel att inträffa, utan dessa begränsas av vilken informationsbärare som används, vilken interaktion som uppstår med omgivningen, samt hur omgivningen "fångar upp" information om felhändelserna. Med kunskap om sådan karakteristik kan man bygga koder, s.k. kanalanpassade kvantfelrättande koder. Vanligtvis antas att omgivningens förmåga att särskilja felhändelser är liten, och man kan då tala om en *minneslös* omgivning.

Antagandet gäller inte alltid, då denna förmåga bestäms av reservoirens *temperatur*, och i det speciella fall då fotoner används som informationsbärare gäller typiskt $k_{\mathrm{B}}T_{\mathrm{R}} \ll \hbar\omega$, och vi måste anta att reservoiren faktiskt *har* ett "minne". I avhandlingen beskrivs en kort, kvantfelrättande kod som är anpassad för fotoner i växelverkan med en "kall" omgivning, d.v.s. denna kod skyddar mot en omgivning som kontinuerligt registrerar vilket fel som uppstått i det kodade tillståndet.

Det är också av stort intresse att kunna jämföra prestanda hos kvantfelrättande koder, utifrån någon slags "måttstock" – men vilken? Jag jämför två sådana mått, nämligen *ömsesidig kvantinformation*, samt *kvantfidelitet*, och visar att dessa i allmänhet inte kan maximeras samtidigt i en felrättningsprocedur. För att visa detta har en 5-qubitarskod använts i en tänkt kanal där bara bitflip-fel uppstår, och utrymme därför finns att *detektera fel*. Ömsesidig kvantinformation framstår som det bättre måttet, dock är detta mått betydligt mer arbetskrävande att beräkna, än kvantfidelitet – som är det mest förekommande måttet.

Preface

This thesis has two main parts. First I start off with a chapter called "classical coding", where a few key concepts from information theory and coding are briefly outlined. The next part is called "quantum error correction" and aims at setting up the stage for paper A, but providing only the necessary set of the theory. I will probe a little deeper on some subtle assumptions and simplifications, which are underpinning the topic, but nevertheless are essential. Some unorthodox notions which are new, or stem from other parts of quantum optics have also been added, simply due to paper A's resistance to "fit" into the more conventional theory, which is based upon SU(2)-algebra. Paper B is more related to the first part, due to its origin in classical information theory. This "wrong order" may seem odd, but was chosen because classical coding was discovered before quantum error correction (which happens to be opposite to the discoveries of paper A and paper B). A reader very familiar with information theory may largely skip chapter 2, except perhaps for the section on *mutual information*, which is very central for paper B. Readers familiar with quantum mechanics may skip section 3.1. I wish you happy reading!

The work presented in this thesis was performed under the supervision of Prof. Gunnar Björk in the Quantum Electronics and Quantum Optics group (QEO), which is part of the School of Engineering Sciences at the Royal Institute of Technology (KTH) in Stockholm.

Acknowledgements

This licentiate thesis would not have been written without the support from several people whom I would like to thank, in no particular order.

My wife Qiang and my son Alfred who had patience with me – even after learning that the quantum computer may not be built anytime in the near future.

My supervisor, professor Gunnar Björk, who I have had the privilege of working (and having fun) with, over the last decade. I owe many thanks to Jonas Söderholm, who provided a great deal of help and inspiration during my master thesis, as well as on occasional visits. Also many thanks to Aziza Surdiman and Saroosh Shabbir – my room mates, for interesting and fun discussions. Marcin Swillo, Sébastien Saugé, Christian Kothe, Isabel Sainz, Jonas Tidström, Mauritz Andersson, Maria Tengner and Daniel Ljunggren have also helped me on many occasions. Thanks also goes to David Yap, for explaining fault tolerance in space, Emil Nilsson, for explaining DNA mutations and to Lars Engström for introducing me to quantum mechanics. I want to thank my younger brothers Per, Jens, Erik, Tom, Mattis and Rasmus, for forcing me to explain what I am doing. Thanks also go to my parents.

Contents

List of papers and contributions

Papers which are part of the thesis:

Paper A

J. Almlöf and G. Björk,
A short and efficient error correcting code for polarization coded photonic qubits in a dissipative channel,
Opt. Commun. **284** (2011), 550–554.

Paper B

J. Almlöf and G. Björk,
Fidelity as a figure of merit in quantum error correction,
accepted for publication in the Jan. issue 2013 in Quant. Info. Commun.

My contributions to the papers:

Paper A

I found the $[[3, 1, 2]]_3$ QECC using an exhaustive computer search program, suggested the modulo-7 recovery logic and wrote the paper.

Paper B

I wrote part of the paper and made some of the calculations.

Paper which is not part of the thesis:

G. Björk, J. Almlöf, and I. Sainz,
On the efficiency of nondegenerate quantum error correction codes for Pauli channels,
arXiv:0810.0541.

Conference contributions:

G. Björk and J. Almlöf,
Quantum error correction - emendo noli me tangere!,
invited talk at **Optikdagarna 2010**, Lund, Sweden, October 19-20, 2010.

G. Björk and J. Almlöf,
Quantum codes, fidelity and information,
invited talk at the **18th International Laser Physics Workshop**, Barcelona, Spain, July 12-17, 2009.

I. Sainz, G. Björk, and J. Almlöf,
Efficency and success of quantum error correction,
talk at **Quantum Optics IV**, Florianópolis, Brazil, October 13-17, 2008.

G. Björk, J. Almlöf, and I. Sainz,
Efficiency of quantum coding and error correction,
invited talk at **17th International Laser Physics Workshop**, Trondheim, Norway, June 30 - July 4, 2008.

R. Asplund, J. Almlöf, J. Söderholm, T. Tsegaye, A. Trifonov, and G. Björk,
Qubits, complementarity, entanglement and decoherence,
talk at **3rd Sweden-Japan International Workshop on Quantum Nanoelectronics**, Kyoto, Japan, Dec 13-14, 1999.

Posters:

J. Almlöf and G. Björk,
A short and effcient error correcting code for polarization coded photonic qubits in a dissipative channel,
contributed poster at **International Conference on Quantum Information and Computation**, Stockholm, Sweden, October 4-8, 2010.

J. Almlöf and G. Björk,
A short and efficient quantum-erasure code for polarization-coded photonic qubits,
contributed poster at the **CLEO/Europe-EQEC**, Munich, Germany, June 14-19, 2009.

G. Björk, J. Almlöf, and I. Sainz,
Are multiple-error correcting codes worth the trouble?,
contributed poster at the **19th Quantum Information Technology Symposium**, Osaka, Japan, November 20-21, 2008.

List of acronyms and conventions

Acronyms

QEC quantum error correction

QECC quantum error-correcting code

CNOT controlled-not

CD compact disc

QND quantum non-demolition

SE Schrödinger equation

QM quantum mechanics

Conventions

The following conventions are used throughout the thesis:

$\mathbb{1}$	matrix identity
$\|\phi\rangle, \|\psi\rangle, \ldots$	states in a Hilbert space
$\|\phi\rangle_\perp$	a state orthogonal to $\|\phi\rangle$
$\|0_\mathrm{L}\rangle, \|1_\mathrm{L}\rangle \ldots$	logical qudit states
$\|0\rangle, \|1\rangle \ldots$	physical qudit states
$0, 1$	(classical) bit values
$\mathcal{O}(k)$	a term of order higher than or equal to k, i.e., $k \in \{1, x, x^2 \ldots\}$
$\propto$	proportional to
$\otimes$	tensor product
$\oplus$	addition modulo 2
$(\ldots)^\mathbf{T}$	transpose of a matrix
S	the system under consideration
A	a system kept by "Alice"
B	"Bob"'s state, usually the receiver of a message from Alice
AB	a joint system of Alice and Bob
R	a reservoir system, also known as "the environment"
$\mathcal{H}_\mathrm{S}$	Hilbert space for system S
$\mathcal{H}^{(N)}$	Hilbert space of dimension N
H	entropy
H_q	quantum entropy
$I(\mathrm{A} : \mathrm{B})$	classical mutual information between Alice and Bob
$I_\mathrm{q}(\mathrm{A} : \mathrm{B})$	quantum mutual information between Alice and Bob
F	fidelity
$\mathcal{F}$	quantum fidelity
k_B	Boltzmann's constant
T	temperature
$\|S_\kappa^{(i)}\rangle$	syndrome vector, i.e. a quantum state stemming from codeword i as a result of an error operation κ
s_κ	the eigenvalue corresponding to a syndrome vector

List of Figures

Chapter 1

Introduction

Quantum information theory is the exciting merging of two mature fields – information theory and quantum theory – which have independently been well tested over many years. When studying one in the light of the other, we see that the combined field has many interesting features, due to the microscopic scale in which it operates, and due to its quantum nature – but also drawbacks and limitations for the same reasons. While many of the ideas upon which this new field of physics are based are imported from information theory, there are also unique features in the combined theory due to the fact that quantum theory allows for superpositions, and as a result, a richer information structure. For quantum error correction, which is a sub-field of quantum information, this structure can, and must, be taken advantage of, e.g., by making use of entanglement in codes, but also accounting for more diverse types of errors. Most quantum codes existing today are based on classical codes, but there are also situations where intuition gained from classical coding theory may lead us wrong, and quantum codes may exist where there is no classical counterpart. In this thesis, I will investigate quantum error correction with the following questions in mind:

- How do we realistically harness quantum coding, i.e., how do we exploit the "quantumness" of codes, while at the same time, control the unwanted quantum effects? In particular, how are code structure, carrier, channel, environment and overall scheme complexity related?
- How is the performance of quantum codes rated? For example, how do we know if a quantum error-correcting code (QECC) is better than another one?
- What is the future for new codes? Where should we look to improve quantum codes? Does it pay to invent even longer codes than existing codes?

The smallest representation of classical information is one "bit", i.e., a bit can represent one of the two values 0 or 1. In quantum theory, the bit translates to a "qubit", which also has two elements in the form of orthogonal quantum

states in a two-dimensional Hilbert space. Even though the qubit has an infinite number of configurations in this space, it can still host at most one classical bit of information. This important fact lets us treat the concept of "information" on the same footing in the two descriptions, and we can "reuse" large parts of the classical theory due to, e.g., the results of Shannon and others. But a qubit can also exhibit other phenomena – which are forbidden in classical information theory – such as entanglement. Entanglement gives rise to an entirely new type of resource, the *e-bit*, which also has an important role to play in quantum information. A magnificent example of this is *teleportation* (of quantum states) [BBC$^+$93].

Of course, we are not restricted to represent information as bits, in fact the representation can move freely between bases, such as 2, 8 and 10 - however, some transitions of representation give rise to impractical mathematical objects (groups), such as storage of bits by means of *trits*, i.e., elements from a size three alphabet. In quantum error correction (QEC), it is of essence that we find a practical physical system that willingly can incorporate the information – an information carrier – and that the system exhibits the sought for qualities, such as a long lifetime and limited modes of decoherence. We shall see an example of how one can use a system made from qutrits to redundantly encode a qubit in paper A, however in doing so, *parity* operations for diagnosing errors will no longer use base 2, so other operations are needed that use base 3. Base 2 codes abide by the SU(2) algebra, where notably the Pauli operations provide a complete set of operations that can be performed. On the other hand, base 3 codes follow the SU(3) algebra, which is governed by 9 (including the unit matrix) *Gell-Mann matrices*. The description is further complicated, when noting that the algebra used may, in a particular physical setting, not take into account that some operations are improbable or forbidden. These restrictions involves both the carrier and the characteristics of an external reservoir, and can be adapted for in a quantum code.

Today's digital computers and media are inherently analog, in the sense that all bit values are represented using large numbers of electrons, directed magnetic dipole moments in the case of magnetic storage, or "missing matter" in the case of imprints on a music compact disc (CD). This fact has several advantages, e.g., in a computer memory there is under normal conditions no need for error correction at all. This is due to extremely stable voltage pulses (+1.5/0 Volts for a modern DDR3 memory) that are used to represent the bit values. If one were to look at a digital pulse in an oscilloscope - one would see that there are minor fluctuations due to, e.g., capacitive losses, or external fields. As modern computers tend to have smaller and smaller components, these fluctuations will one day become large enough to matter. In fact, for extreme applications, such as space satellite applications where computers are exposed to, e.g., cosmic rays, computers are set up in racks of three. Each computer routinely performs the same set of instructions, and the overall output is the result of a majority-voting of the output from these computers [WFS94]. Majority voting is also one of the simplest and most used error correction procedures. However, it is in general neither the most efficient, nor the most resilient one - as we shall see in chapter 2.

Hence, a classical computer on Earth is stable in its operation and usually does not need any error correction. However, when storing and transmitting information, usually some form of error correction is applied. The techniques used are often, if not always, based on assumptions on what kind of errors will most likely occur. One illustrative example is the case of error correction for CDs, where the imprinted information needs to be protected from scratches. A scratch has a nonzero width, that will sometimes intersect the imprinted track from a tangential direction. Thus, a probable error event is that many adjacent imprints will be damaged, i.e., a burst of errors. Therefore, a special type of encoding is used, a *Reed–Solomon* code [RS60], and it can correct up to 12 000 adjacent errors, which corresponds to a track length of 8.5 mm on a CD. In addition, the coded information is recorded in a "non-local" way, on opposing positions on the disc, to minimise the risk that the information is erased by a single scratch. The point to be retained is that in classical error correction, it is usually the probabilities for various errors that ultimately decide which error correction code will be used. This is also true for QEC, as we shall see in chapter 3.

An important advantage of computers, or other processing devices for classical information, is that the stream of information can at any time be amplified, or duplicated (using a source of power). This is something that we take for granted. However, the situation is different for a quantum computer, because it turns out that copying is a severely restricted operation for quantum states, as we shall see in chapter 3. Thus, if we cannot amplify our quantum information, it seems that the only alternative we have for processing is to continuously use error correction, in order to keep the quantum states from being distorted. Other means to protect qubits, is to encode them onto quantum states with long decoherence times, and consider channels where interaction with the surrounding environment is minimal. Also, while QECCs necessarily increase the length of an unprotected string of qubits (by introducing redundancy), each added qubit increases the influence from the environment. Therefore, any good QECC must add, loosely speaking, more protection per added qubit, than the increased need for protection per added qubit. Whether or not it really pays to have long QECCs (that correct many errors, or encodes many qubits) will be touched upon in section 3.4.5.

Feynman wrote on the topic of energy dissipation in information processing, in a paper called "Quantum mechanical computers" [Fey86]:

> However, it is apparently very difficult to make inductive elements on silicon wafers with present techniques. Even Nature, in her DNA copying machine, dissipates about 100 $k_{\mathrm{B}}T$ per bit copied. Being, at present, so very far from this $k_{\mathrm{B}}T \ln 2$ figure, it seems ridiculous to argue that even this is too high and the minimum is really essentially zero.

–Should not our DNA be a perfect example of a coding that perhaps needs error correction? And why has Nature chosen the base 4? Is it simply because of the need for splitting the double helix, or is there some other insight in this way

of coding? Outside the scope of this thesis, I have thought about these problems, and others too, see Liebovitch *et al.*, [LTTL96]. Their study did not find any such error correction code. Later studies show [SPC+03] that an enzyme called *DNA polymerase* does "proofreading" of the DNA strands, and corrects errors – thereby decreasing the error rate by a factor 100. This indicates that perhaps there is an error detecting, or error correcting code in the DNA after all. On the other hand, an error correction code in our DNA could perhaps not be a perfect one, since then, DNA variation due to, e.g., copying errors, would not exist.

Chapter 2

Classical coding

Coding deals with the problem of transmitting or storing a message in a certain form or shape – a code, so that it can be retrieved safely or efficiently. "Safely" implies that the message may be sent over a noisy channel, using some form of error correction. Error correction can be performed only if redundancy is present, and such redundancy is then typically added, to form a coded message. "Efficiently" on the other hand, means that if the message contains redundancy, e.g., this is the case for natural languages, coding also can be used to compress the message. This means that unnecessary redundancy is removed from the message, and its information density therefore increases. However, such a coded message would be difficult to decode and understand for a human, and therefore automated decoding should be performed at the receiving end. Loosely speaking, we can say that coding deals with transforming messages so that redundancy is either added or removed – typically one wants to strike a balance between the raw information and the redundancy in a form that suits the needs of the communicating parties, and the channel of communication.

There are also coding schemes where some *information* is removed, e.g., JPEG (Joint Photographic Experts Group) and MP3 (MPEG-1 Audio Layer 3) compression. Such compression coding is called destructive, and can in the MP3 case be motivated by the fact that the human ear senses sound best within a limited frequency range, so that recorded frequencies outside this band may be suppressed, or discarded. Coding can also be used in conjunction with public, shared, or private keys – to send secret messages between parties. However, I shall in this thesis mainly focus on different aspects of *quantum error correction*, and in this chapter I will give a brief background in classical information theory, from where several concepts have quantum counterparts that will be used in chapter 3.

2.1 Entropy and information

Entropy is essentially the logarithm of the number of allowed values for some parameter. If, on a combination lock, the number of possible combinations is Ω, then we may calculate the number of rotating discs, $\log_b \Omega$. But if the number of symbols written on each disc b is unknown, then the choice of logarithm base is equally unclear, and we can only qualitatively do so. For example, we can merely say that in order to increase the number of combinations to Ω^2, we need to double the number of discs, since $\log \Omega^2 = 2 \log \Omega$. A number of permitted, but unknown values for a parameter implies uncertainty, or "ignorance", while knowledge of exactly which of the values the parameter has, can be interpreted as "information". The interplay between information and ignorance, is at the heart of information theory.

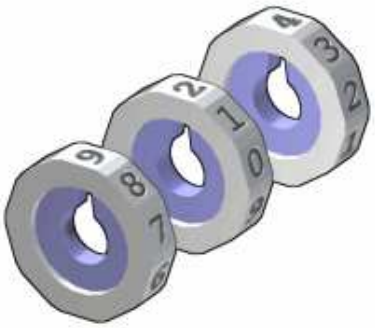

Figure 2.1: A simple combination lock with three rotating discs and 10 symbols per disc. Credit: Wapcaplet under Creative Commons License.

2.1.1 Statistical mechanics

Classically, entropy is defined (due to Boltzmann)

$$H = k_{\mathrm{B}} \log \Omega, \tag{2.1}$$

where Ω denotes the number of microstates, i.e., the number of possible configurations for a physical system, and k_{B} is known as *Boltzmann's constant*. In classical mechanics, the notion of Ω made little sense, because e.g., position and momentum can take an infinite number of values. But this problem was circumvented, particularly in thermodynamics, by assuming that Ω for a ideal gas, should *qualitatively* be proportional to the degrees of freedom in the following way:

$$\Omega \propto V^N E^{(3N-1)/2}, \tag{2.2}$$

where N is the number of particles in a gas of volume V, and an energy E. The energy dependent part of the expression is essentially the area of a $3N$-dimensional sphere, with the radius $\sqrt{E}$. Thus, the bigger sphere that is spanned by the velocity vectors of the gas particles, the more states can be fitted. Here, Eq. (2.2) should be corrected by $N!$ in the denominator to reflect that only distinguishable configurations are counted in a (bosonic) gas. However, at the time of Boltzmann, such quantum mechanical corrections for bosons and fermions were not known, and

it turns out that some important results can be extracted even without this knowledge. Taking the logarithm of Eq. (2.2) results in a property that depends much less dramatically on the degrees of freedom. Interestingly, the logarithm of the "number of possible states" $\log \Omega$, often has real physical meaning, i.e., revealing clues about the system's degrees of freedom. Such descriptions are, e.g., for the temperature and pressure of an ideal gas,

$$\frac{1}{T} = \frac{\partial H}{\partial E}, \quad \text{and} \quad P = T \cdot \frac{\partial H}{\partial V},$$

which immediately results in familiar expressions for internal energy E, and the well known ideal gas law

$$E = \frac{3}{2} k_{\mathrm{B}} N T, \quad \text{and} \quad PV = k_{\mathrm{B}} N T,$$

respectively. The Boltzmann entropy is especially suited for this purpose for several reasons, i.e., the logarithm function is the only function that scales linearly as the argument grows exponentially,

$$\log \left(\prod_i \Omega_i \right) = \sum_i \log \Omega_i.$$

Also, the logarithm function is a strictly increasing function of its argument, which implies that both Ω and $\log \Omega$ will reach their maximum value simultaneously.

2.1.2 Information theory

Also in information theory it is common to study entropy as a function of the system degrees of freedom [Weh78], but more commonly on a microscopic, rather than the macroscopic scale exhibited in the previous examples. The word *entropy* will be used here in analogy with statistical mechanics, however in the strictest sense, it is disputed if the two descriptions are identical:

> My greatest concern was what to call it. I thought of calling it "information", but the word was overly used, so I decided to call it "uncertainty". When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage."
>
> *Claude E. Shannon* [TM71]

The logarithm of the total number of states qualitatively describe the number of resources needed to represent the states, e.g., in computer science – the number

256 needs $\log_2 256 = 8$ bits for its representation. Here, we have assumed that all integers from 1 to 256 are equally probable, i.e., that we are not allowed to exclude any of those numbers.

Definition 2.1. *(Symbol, alphabet) A symbol represents an element taken from a set of distinct elements* $\{0, 1, \ldots b\}$*, called an alphabet. Binary symbols can assume only the values* $\{0, 1\}$*, thus, they have an alphabet size, or base,* $b = 2$*.*

Despite the occurrence of non-binary alphabets in this text, we shall however persist the choice of base 2 for logarithms, since this choice is generally unimportant, but will allow us to speak about an entropy that we can measure in bits.

Definition 2.2. *(String) A sequence of symbols, taken from an alphabet with base b, is called a string.*

Example: Two common types of strings:

- A binary string: "1001011111000100110000010011010010 00", from {0,1}
- A base 19 string: "The clever fox strode through the snow.", from {T, h, e, ' ', c, l, v, r, f, o, x, s, t, d, u, g, n, w, .}

The latter example raises a question – the string only uses 19 symbols, but do we need to worry about other symbols that may occur, i.e., *hypothetical* strings? The answer is that the alphabet used for communication is subject to assumptions, specified by a standard which are supposedly shared by two communicating parties. One such standard is the ASCII alphabet, which has $2^7 = 128$ symbols, and covers most of the English strings that can be written. Nowadays, a character encoding called *Unicode* is commonly used which has a 2^{16}-symbol alphabet, and includes characters from most languages, and special symbols such as the relatively new Euro currency symbol €. One may argue that it is wasteful to use such a large alphabet, since if Alice and Bob communicates in English, they do not need an alphabet supporting, e.g., all Chinese characters. Morse code is an alphabet that uses less resources, i.e., dashes and dots, for common letters in English, and for uncommon letters like "X" – it uses more. This tends to save time for Alice, as she encodes her message – since the total number of dots and dashes is on average lower compared to if all characters had the same length. If – in a long sequence of symbols – not all symbols are equally probable, a central concept is the *Shannon entropy* [Sha48], defined as

$$H = -\sum_{i}^{N} p_i \log p_i, \tag{2.3}$$

where N is the number of different values that the symbol may have, and p_i is the probability for a given value, i. The maximum entropy is reached when all probabilities are equal, i.e., the situation for a two symbol alphabet with symbol

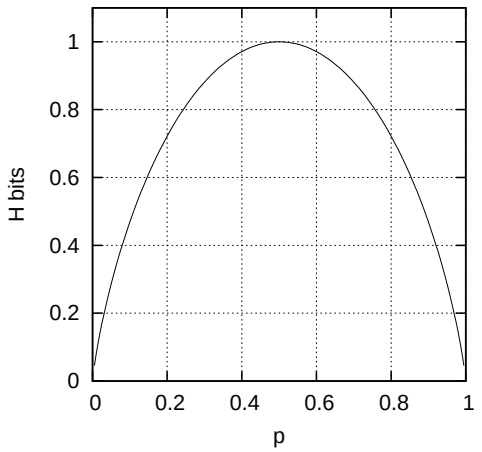

Figure 2.2: The entropy per symbol for an alphabet with two symbols. The probabilities for the first outcome is p and thus $1-p$ for the other.

probabilities p and $q = 1-p$ is illustrated in Fig. 2.2. If the character probabilities are not same, such as in natural languages, the "wastefulness" described earlier can be mitigated using source encoding, where Morse code is one example.

Consider the example of a communication line which can convey information at the rate 1000 baud, i.e., 1000 symbols per second. However the probability for one symbol is one, and all the others are zero. Can such a channel convey any information? The answer is "no", which is straightforward to calculate using the Shannon entropy $H(\mathrm{A})$, which is equal to $-1 \cdot \log 1 - 0 \cdot \log 0 \ldots = 0$ (here $0 \log 0$ is defined to be equal to 0). The situation for a two-symbol alphabet is shown for varying probabilities in Fig. 2.2.

As another example, consider the symbols A, B, C and D with relative frequencies $1/2, 1/4, 1/8, 1/8$ respectively. The source entropy per symbol will in this case be $H = -(\log(1/2)/2 + \log(1/4)/4 + \log(1/8)/4) = 7/4$, i.e., less than the optimal entropy 2 ($= \log 4$). We can in this case compress the average information sent using a code according the following scheme:

C1: A source encoding

$$\mathrm{A} \to 0, \mathrm{B} \to 10, \mathrm{C} \to 110, \mathrm{D} \to 111.$$

This coding is called block coding (with variable block length) and in this case it will restore the source entropy per symbol to its maximum value 1. To see this, we can calculate the average number of bits, $\overline{L}$, per symbol, in a **C1**-coded string,

$$\sum_i p_i L_i = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 7/4.$$

However, such perfect compression encodings are not always possible to find. An important lesson can be learned from this code – *improbable symbols should be encoded with longer strings, and vice versa.* This is evident in all languages, e.g., "if" and "it" are common words and have few letters, while "university" is longer, and not as frequent. There are of course differences between languages, e.g., in English, one has only one letter for "I" compared to "you", which implies that English-speakers prefer to talk about themselves, rather than about others. In Swedish however, the situation is reversed ("jag"/"du"), so information theory lets us draw the (perhaps dubious) conclusion that Swedish-speakers are less self-centered than English-speakers.

One can say that the amount of *surprise* in a symbol, constitute a measure of information, and should be reflected in its block length to ensure efficient source encoding. An efficient technique for coding the source, according to the relative frequencies of message symbols is *Huffman coding* [Huf52]. While recognised as one of the best compression schemes, it only takes into account single symbol frequencies and ignores any transition probabilities for sequences of symbols, which may also exist. More optimal compression codings take care of this latter situation, such as *arithmetic coding*, see e.g., [RL79], and its variants. These methods are based on Shannon's notion of n-graphs [Sha48], but also cover destructive compression techniques with applications in still imaging and video.

Finally, I must mention a celebrated result of Shannon, which sums up this section:

Theorem 1. *(Noiseless coding theorem) Let a source have entropy H (bits per symbol) and a channel have a capacity C (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate $C/H - \epsilon$ symbols per second over the channel where ϵ is arbitrarily small. It is not possible to transmit at an average rate greater than C/H.*

For a proof, see e.g., [Pre97] (chapter 5).

2.1.3 The channel

When a string of symbols is sent from a point A to a point B, different circumstances may affect the string, such as electrical interference, or other noise that may cause misinterpretation of the symbols in the string. Such effects are usually referred to as the action of the channel. Channels can conveniently be characterised by a matrix, containing probabilities for misinterpreting symbols in a string. E.g., consider the symbols $\{0, 1\}$, and the transition probabilities $\{p_{0\to0}, p_{0\to1}, p_{1\to0}, p_{1\to1}\}$. The channel matrix is then written

$$C_{\mathrm{AB}} = \begin{bmatrix} p_{0\to0} & p_{0\to1} \\ p_{1\to0} & p_{1\to1} \end{bmatrix}. \tag{2.4}$$

Such a matrix can also be illustrated as a diagram, as illustrated in Fig. 2.3.

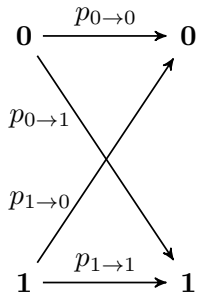

Figure 2.3: A diagram showing the symbol transition probabilities for a binary flip channel.

Definition 2.3. *(Symmetric channel) If, for a binary flip channel, the flip probabilities are equal so that $p_{0\to1} = p_{1\to0}$, the channel is said to be symmetric.*

2.1.4 Rate of transmission for a discrete channel with noise

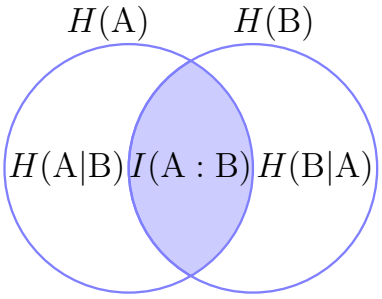

Figure 2.4: A Venn diagram showing the relation between the entropies for A and B, the conditional entropies $H(\mathrm{A}|\mathrm{B})$ and $H(\mathrm{B}|\mathrm{A})$ and the mutual information $I(\mathrm{A}:\mathrm{B})$. $H(\mathrm{A},\mathrm{B})$ is represented as the union of $H(\mathrm{A})$ and $H(\mathrm{B})$.

How is the transmission of a message, i.e., *a string of symbols*, affected by channel noise? As mentioned in the introduction, there is a subtle distinction between the arranging of symbols at the sending party, and the disordering of symbols as a result of sending them over a noisy channel. For a noisy channel, Shannon defines the *rate of transmission*

$$I(\mathrm{A}:\mathrm{B}) = H(\mathrm{A}) - H(\mathrm{A}|\mathrm{B}), \tag{2.5}$$

where $H(\mathrm{A})$ is called "entropy of the source", which constructively contributes to the transmission rate between two parties, while the conditional entropy $H(\mathrm{A}|\mathrm{B})$,

also called "equivocation", instead contributes negatively, and can be seen from Fig. 2.4 to be

$$H(\mathrm{A}|\mathrm{B}) = H(\mathrm{A},\mathrm{B}) - H(\mathrm{B}), \tag{2.6}$$

and is defined, for the (discrete) distributions $\mathrm{A} : \{a, p_{\mathrm{A}}(a)\}$, and $\mathrm{B} : \{b, p_{\mathrm{B}}(b)\}$,

$$H(\mathrm{A}|\mathrm{B}) = -\sum_a \sum_b p(a,b) \log p(a|b), \tag{2.7}$$

where $p(a|b)$ is the probability that $\mathrm{A} = a$ given that $\mathrm{B} = b$. $H(\mathrm{A})$ depends on the size of the "alphabet", i.e., how many possibilities one has to vary each symbol – but also on the relative frequencies/probabilities of those symbols. As indicated earlier, $H(\mathrm{A})$ is maximised if all probabilities are the same. $H(\mathrm{A}|\mathrm{B})$ represents errors introduced by the channel, i.e., "the amount of uncertainty remaining about A after B is known". Shannon's "rate of transmission", is lately denoted *mutual information*, because it is the information that two parties can agree upon, sitting at the two ends of a communication channel. Mutual information is the term favoured in today's literature, and it is also the term that will be used in this thesis.

2.1.5 Classical channel capacity

We now know that the mutual information between A and B sets the limit of how much information can be transmitted e.g., per time unit. But sometimes we wish to characterise the channel alone, not taking into account the encoding performed at A, we extend the definition of *channel capacity* C (in Theorem 1) in the presence of noise,

$$C = \max_{\{p(a)\}} I(\mathrm{A} : \mathrm{B}). \tag{2.8}$$

Hence, the channel capacity is defined as the mutual information maximised over all source probabilities $p(a)$, which is equivalent to the previous notion in the absence of noise.

2.2 Classical error correction

Assume that Alice sends a message to Bob, but over a symmetric bit-flip channel, so that with a non-zero probability, bits in the message will be flipped, independently of each other. The goal of error correction is to maximise the mutual information between Alice and Bob by adding redundant information to the message, that will protect the message from errors. The efficiency which this feat can be accomplished is the quotient of the actual information bits, say k bits – and the total number of bits, including the redundant ones, n. Thus, the message is divided into sequences of n bits, called *blocks*. It turns out that cleverly crafted codes can achieve a higher ratio k/n than others, but the problem of finding such codes is difficult, and no

general method exists. To make matters worse, the channel characteristics is also an important part of the problem, so that different channels have different optimal codes.

For the remainder of this chapter (but not the next!), we shall only consider the *binary symmetric channel*, i.e., errors affect bits independently of each other, and perform the bit-flip operation $0 \to 1$, and $1 \to 0$ with equal probability.

2.2.1 Linear binary codes

A linear binary (block) code C, or simply "code" from now on (if not stated otherwise), is defined as the discrete space containing 2^n *words*, whereof n of them are linearly independent. The space is assigned a norm (inner product), an addition and a multiplication operation. The nomenclature is summarised below:

Definition 2.4. *(Word) A word in a code C is n consecutive elements taken from $\{0, 1\}$.*

Example: A word in a $n = 4$ code is written, e.g., (0110).

Definition 2.5. *(Inner product) Addition and multiplication is taken modulo 2 for binary codes, so that the inner product*

$$u \cdot v = \left(\sum_i (u_i v_i \mod 2) \right) \mod 2.$$

Example:

$$(0110) \cdot (1110) = (0 \cdot 1) + (1 \cdot 1) + (1 \cdot 1) + (0 \cdot 0) = 0.$$

Definition 2.6. *(Hamming weight) The Hamming weight of a codeword u is denoted* $\mathrm{wt}\,(u)$*, and equals to the number of non-zero elements of u.*

Example:

$$\mathrm{wt}\,(1110) = 3.$$

Definition 2.7. *(Code subspace, codeword) If a code C containing 2^n words has a linear subspace C', spanned by 2^k words which are closed under addition, i.e., $u+v \in C', \forall\ u, v \in C'$, and $k < n$, then any set of linearly independent words from C' are called codewords for the code C, and are commonly denoted $0_{\mathrm{L}}, 1_{\mathrm{L}}, \ldots (2^k - 1)_{\mathrm{L}}$.*

Example: Let C be a space with 2^4 elements. Let C' be a 2^2 linear subspace of C, with elements $(0000), (0011), (1100), (1111)$. Any sum of these elements is also an element of C'. C' is spanned by two linearly independent words, e.g., $(1100), (0011)$. Such words are called codewords.

Definition 2.8. *(Distance) A subspace C' of a code C is said to have distance d, which is the minimum weight of all pairwise combinations of its codewords $i_{\mathrm{L}}, j_{\mathrm{L}}$ – i.e.,*

$$\min \mathrm{wt}\,(i_{\mathrm{L}} + j_{\mathrm{L}}), \quad i, j \in \{1, 2, \ldots k\}, \quad i \neq j.$$

Definition 2.9. *(Notation) A code C is written $[n, k, d]_b$, or simply $[n, k, d]$ if it is binary.*

So far nothing have been said about error correction, but the ability to detect or correct errors is intimately connected to the distance d. d, on the other hand is defined for a certain type of errors, namely the *bit flip* errors – which is important to remember. I state without proof a basic error correction result, which will be illustrated in a moment:

Theorem 2. *A linear binary error correcting code which uses n bits of information to encode k bits, can correct up to $t = (d-1)/2$ errors and detect up to $t+1$ errors, where d is the distance of the code.*

Since t is used to denote the number of correctable arbitrary errors, one can optionally use the code notation $[n, k, 2t + 1]$. As an illustration of the theorem, consider the code

C2: A repetition code

$$0_{\mathrm{L}} = (000), \quad 1_{\mathrm{L}} = (111).$$

Example: The distance d of **C2** is $\mathrm{wt}\,((111) + (000)) = 3$. We have $2^k = 2$ codewords – thus we denote the code $[3, 1, 3]$, and its complete space is illustrated in Fig. 2.5. From this figure, we can see that any 1 bit-flip errors in $\{0_{\mathrm{L}}, 1_{\mathrm{L}}\}$ can be identified and corrected. If errors need only be detected, we see that we can do so for up to 2 errors. Detection is therefore a powerful mechanism, and can be used to classify a block as erroneous, so that it can be subsequently re-transmitted in a communication scenario. In this coding scheme, since the code is *perfect* (see section 2.3.1), we must choose a detection strategy *or* a correction strategy – we may not do both.

Definition 2.10. *(Generator matrix, parity check matrix, syndrome) A generator matrix G is a $k \times n$ matrix containing any k words in the code subspace C', that span C'. An $(n-k) \times n$ matrix P with the property $PG^{\boldsymbol{T}} = 0$, is called a parity check matrix and is used to determine, for each received word w through the operation $Pw^{\boldsymbol{T}}$, the location of the bit that is in error and should be flipped. The result of $Pw^{\boldsymbol{T}}$ is called the* syndrome *of w.*

Example: The generator and parity check matrix in the previous example are

$$G = \begin{bmatrix} 111 \end{bmatrix}, \quad P = \begin{bmatrix} 110 \\ 101 \end{bmatrix}, \tag{2.9}$$

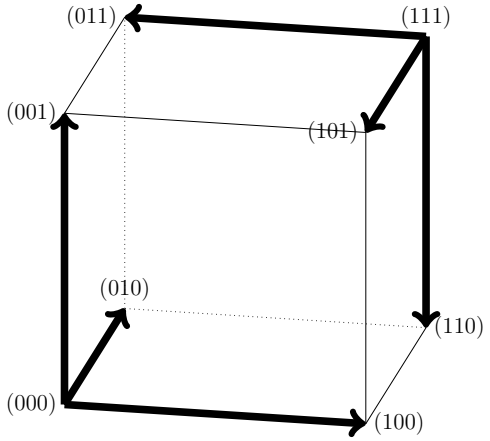

Figure 2.5: A code protects an encoded bit by separating their code words by at least a distance $2k + 1$, where k denotes the number of errors that the code can correct. The situation is shown for a 1-bit flip error correcting repetition code, denoted $[3, 1, 3]$. Clearly, this code has distance $d = 3$, which is the required distance in order to correct one arbitrary bit-flip error.

so that the syndromes can be calculated as $P \cdot (111)^{\mathbf{T}} = P \cdot (000)^{\mathbf{T}} = 00$ (do nothing), $P \cdot (110)^{\mathbf{T}} = P \cdot (001)^{\mathbf{T}} = 01$ (flip third bit), $P \cdot (101)^{\mathbf{T}} = P \cdot (010)^{\mathbf{T}} = 10$ (flip second bit), and $P \cdot (100)^{\mathbf{T}} = P \cdot (011)^{\mathbf{T}} = 11$ (flip first bit).

Note that errors in this case give rise to pairwise identical syndromes, which is a consequence of the properties of linear codes. This is advantageous from an implementation point of view, since either memory or computing capacity can be saved, compared to the situation where each error has a unique syndrome. We shall see in the next chapter, that this property is sought for also in quantum error correction, but for an entirely different reason.

2.3 Strategies for error correction and detection

Consider the code

C3: A 4-bit repetition code, $[4, 1, 4]$

$$0_{\mathrm{L}} = (0000), \quad 1_{\mathrm{L}} = (1111).$$

This code can correct all single bit-flip errors, but no 2-flip errors. In general, one would need a $d = 5$ code to be able to do so. Interestingly, all the 2-errors can be detected, and we will see in a moment what to do with these.

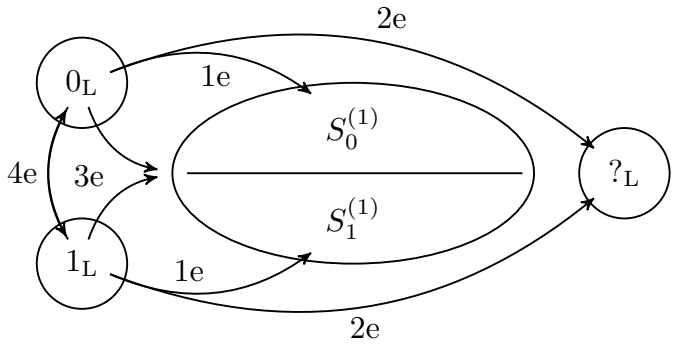

Figure 2.6: Alice sends a coded message to Bob over a noisy bit-flip channel, using the code **C3**. Each of Bob's blocks will after correction belong to one of the 3 disjoint sets $\{0_L, 1_L, ?_L\}$, where $?_L$ represents the detectable, but uncorrectable 2-error blocks. Note that blocks with 3 or 4 errors will possibly be misdiagnosed, since they represent elements in the more probable set of 0- and 1-error blocks.

Definition 2.11. *(Fidelity) The fidelity F is a measure of "sameness", which I define for two messages m_A and m_B of equal length M, consisting of logical codewords $\{0_L, 1_L, ?_L\}$, as*

$$F(m_A, m_B) = 1 - \frac{\mathrm{wt}\,(m_A + m_B)}{M}, \tag{2.10}$$

where I have extended the Hamming weight definition with addition rules for $?_L$,

$$0_L + 0_L = 1_L + 1_L = 0,$$
$$1_L + 0_L = 1_L + ?_L = 0_L + ?_L = 1.$$

Example:

$$\begin{aligned} F((11111),(11011)) &= 0.8, \\ F((11111),(11?11)) &= 0.8, \\ F((00000),(00?00)) &= 0.8. \end{aligned}$$

The introduction of detectable errors is important, since detection can be done more efficiently than correction, and can complement error correction in order to improve e.g., information transmission. Errors which can only be detected (but not corrected) typically involves re-sending the message, or part of it.

Example: Assume that Alice sends a message consisting of 100 blocks over a symmetric bit-flip channel, coded using **C3**. Bob knows which code Alice has used, thus he can correct all 1-errors in the message. However, assume that Bob will receive a 2-error block, e.g., one of $\{(1100),(1010),(1001),\ldots\}$, with probability $\gamma = 0.01$.

We note that such errors can be distinguished from the codewords and all 1-errors (detected), but cannot be corrected (because Bob cannot know whether the block was originally $0_{\rm L}$ or $1_{\rm L}$).

—*What to do once such an error is detected?*

We will contemplate two strategies, I and II:

I: Replace the block with a random logical bit $0_{\rm L}$ or $1_{\rm L}$

II: Mark the logical bit as erroneous, and do not use it

If Bob uses strategy II, the sent and received messages (after correction) $m_{\rm A}$ and $m_{\rm B}$ will differ in 1 bit out of 100, i.e., the similarity, or *fidelity*, of the two messages is $F = 1 - {\rm wt}\,(m_{\rm A} - m_{\rm B})\,/100 = 0.99$. In contrast, if Bob replaces this block randomly with $0_{\rm L}$ or $1_{\rm L}$, with equal probability, then half of the times he would be able to "correct" the error and achieve $F = 1.00$. However, half of the times, he would be unlucky, so that $F = 0.99$, but on average, he would be able to increase the fidelity to 0.995, using strategy I.

–*What does Shannon tell us about the rate of transmission (mutual information) in the two cases?*

Calculating the mutual information $I({\rm A}:{\rm B})$ for the two strategies results in $1 - (0.99\log 0.99 + 0.01\log 0.01) \cong 0.92$ for strategy I, while strategy II gives $I({\rm A}:{\rm B}) \cong 0.99$. This illustrates the seemingly odd fact that optimising similarity will result in a sub-optimal mutual information. This can mainly be attributed to the insight that strategy I erases *the location of the error*.

Example: Assume that communication between Alice and Bob is affected by strong channel noise, so that $p(a,b) = 0.25, \forall\ (a,b) \in \{0,1\}$. –What is the value of $F({\rm A},{\rm B})$ and $I({\rm A}:{\rm B})$?

The fidelity in this case becomes on average $\sum_{(a,b)=(0,0),(1,1)} p(a,b) = 0.5$, while the mutual information becomes $1 - (0.5\log 0.5 + 0.5\log 0.5) = 0$. This means that communication is not possible over the channel.

In information theory, the mutual information between A and B is the generally accepted figure of merit for data transmission, and not similarity, i.e., fidelity. In paper B, it is shown that fidelity and mutual information for a non-zero error rate cannot be simultaneously optimised, in the case of detectable-only errors, neither in classical nor in quantum error correction.

2.3.1 Bounds for linear codes

The Hamming bound sets a lower limit for how many bits n are needed to accommodate a $[n, k, 2t+1]$ code,

$$2^n \geq \sum_{j=0}^{t} \binom{n}{j} 2^k. \tag{2.11}$$

This bound is also known as the *sphere-packing bound*. For large k and n, this approaches asymptotically

$$\frac{k}{n} \leq 1 - H\left(\frac{t}{n}\right), \tag{2.12}$$

where $H(\cdot)$ is the Shannon entropy depicted in Fig. 2.2.

Definition 2.12. *(Code rate) For block coded information, where each block uses n bits to encode k logical bits, the rate of the code is defined to be k/n.*

Definition 2.13. *(Perfect codes) A perfect code has the property that it satisfy Eq. (2.11) with equality. Thus, a perfect code has a codespace just big enough to host a $[n, k, d]$-code.*

Example: One family of perfect codes is called Hamming codes. They can be written on the form

$$[2^r - 1, 2^r - r - 1, 3]_2, \tag{2.13}$$

where $r \geq 2$. The simplest example of a perfect code is the $r = 2$, three-bit repetition code **C2** on page 14. For $r = 3$, we have

C4: A [7,4,3] Hamming code

$$\begin{aligned}
0_{\mathrm{L}} &= (0000000), & 1_{\mathrm{L}} &= (1110000), & 2_{\mathrm{L}} &= (1001100), & 3_{\mathrm{L}} &= (0111100),\\
4_{\mathrm{L}} &= (0101010), & 5_{\mathrm{L}} &= (1011010), & 6_{\mathrm{L}} &= (1100110), & 7_{\mathrm{L}} &= (0010110),\\
8_{\mathrm{L}} &= (1101001), & 9_{\mathrm{L}} &= (0011001), & 10_{\mathrm{L}} &= (0100101), & 11_{\mathrm{L}} &= (1010101),\\
12_{\mathrm{L}} &= (1000011), & 13_{\mathrm{L}} &= (0110011), & 14_{\mathrm{L}} &= (0001111), & 15_{\mathrm{L}} &= (1111111).
\end{aligned}$$

This error correction code can under extreme conditions be used for memory storage, but since a practical block size in a computer is 8 bits, this Hamming code is usually extended using an extra bit, to accomplish better error detection.

Another important bound is the Gilbert-Varshamov bond, which reads

$$2^k \sum_{i=0}^{d-2} \binom{n-1}{i} < 2^n. \tag{2.14}$$

Eq. (2.14) ensures the *existence* of "good" codes, reasonably close to the Hamming bound. Shannon has shown that for a code with $k/n < C$, that by increasing n it is in principle always possible to achieve an arbitrarily low failure probability. Eq. (2.14) makes Shannon's result more powerful, by showing that such codes exist.

Chapter 3

Quantum error correction

Through the understanding that quantum mechanics (QM) is governed by unitary (and therefore reversible) operations, quantum computing emerged from the idea of reversible computation, in the early work of Bennett [Ben73], Feynman [Fey82, Fey86], Fredkin, Toffoli [FT82] and others. One particularly powerful consequence of these thoughts, later proven by Deutsch [Deu85], is that a quantum computer can compute many results simultaneously, i.e., by means of *qubits* (see section 3.2.2) and unitary operations. In contrast, a classical computer would have to perform those calculations one by one, which is clearly a disadvantage – e.g., this weakness is exploited in today's public cryptographic keys, as they rely on the exponential increase, per bit added, of computing resources needed to factor large integers. Shor showed that this "security from lack of resources", can be overthrown by a quantum computer, proving that it can perform such a factorisation at a mind boggling efficiency, i.e., polynomial time [Sho97]. While this would render today's public key distribution (based on integer factorisation) weak, at best, quantum information also offers fundamentally safe *quantum key distribution* (QKD), e.g., using the BB84 protocol, invented by Bennett and Brassard [BB84]. Such quantum cryptographic systems are today commercially available (from ID Quantique and MagiQ), however, due to imperfections in their technical implementation, they are currently not secure in the strictest sense, see e.g., Saugé *et al.* [SLA+11].

A qubit, being a pure quantum state (with one orthogonal alternative), is extremely sensitive to interactions with other, nearby states, which will ultimately cause it to become impure (when measuring only the qubit system) in a process called *decoherence*, see section 3.1.4. Such interactions will entangle the qubit with some state in the environment, and in the process destroy interference between the qubit's two basis states – thereby ruling out the possibility to perform operations on both states simultaneously, i.e., reducing the qubit to a mere bit.

It was soon realised that errors caused by decoherence in quantum states are different from those assumed in classical error correction, where coding, errors, and decoding can be seen without regard to the error mechanism. —In fact, for QEC,

every conceivable error is a result of interaction with reservoir states, thus, *our description must treat errors on coded states as the result of operations on extended states, where the reservoir states are included.* QEC picked up speed around 1990, and soon resulted in a concrete code for protecting one qubit from any type of Pauli error, i.e., a bit flip, phase-flip, or a combination of both [Sho95]. However, it was soon realised that decoherence errors, such as amplitude-damping errors, needed a different approach [FSW08].

–How can one suppress decoherence? —A qubit is often defined as a two-level system, where the actual system is not specified. Thus, a qubit can be realised in many different ways, e.g., using a spin-1/2 system. The transition probabilities of a particular state into other states, depend on the interaction of such "carrier" systems, and the characteristics of the environment. For a given carrier system, not all transitions are equally probable and in fact, some carrier states are more stable than others — one example of a stable state is the lowest state of the electromagnetic field, the vacuum state. This state exhibits fluctuations, i.e. "virtual" transitions to higher energy modes, but only for a very short time, due to energy conservation. The vacuum state could therefore prove to be a useful element in QEC.

–One may then ask if classical error correction can be used for qubits? The answer is surprisingly "no", or at least not directly. The reason is that even though errors on a coded qubit can be uniquely identified, i.e., the correct two coded states (called $|0_{\mathrm{L}}\rangle$ or $|1_{\mathrm{L}}\rangle$ in analogy with section 2.2.1) and their resulting erroneous states are all mutually orthogonal (which is the quantum meaning of "different"), the correction procedure must not directly detect such an error. If it did, then not only information of which error occurred would be gained, but also information of which original state it was, i.e., $|0_{\mathrm{L}}\rangle$ or $|1_{\mathrm{L}}\rangle$. This is sometimes referred to as "collapse of the wave function" and once this information becomes known to any observer (even a seemingly insignificant atom), the qubit will start to act like a classical bit, i.e., all interference between $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$ would vanish. This is of course unacceptable for a qubit, whose main purpose is to represent 0 and 1 simultaneously, i.e., maintain an arbitrary superposition between its components $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$. The trick, as Shor discovered in his nine-qubit code, is to *delocalise* the information in the coded qubit, and in the error detection stage, perform only measurements that *do not* distinguish between $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$ errors, instead the result of identification is an eigenvalue corresponding to two candidate states - one from each code word. Such measurements are typically done by measuring parity between constituent states of the syndromes. To actually undo the error, a unitary operator is applied that simultaneously maps both candidates back to the "no-error" state — with the help of so called *ancilla states.*

In section 3.1.4.2, we will look at the evolution of errors in a code, which will entangle the coded qubit with the environment. From there, I will go on with the formal theory behind QEC, and give some background to the QECC presented in paper A.

To correct errors in a decohering qubit is a formidable task, in fact, just exactly how the decoherence itself works is a topic of hundreds of papers, and perhaps

we will never get a description for the evolution of quantum states that satisfy everyone. Or perhaps, there are still a few secrets left for us to discover. In particular, something that has haunted quantum theory since its advent, is the *measurement problem*. It can simply be stated as the following question:

> "If we assume that quantum states can be perfectly modelled by means of unitary transformation of their wave function – how is it that our actual measurements on the same system can only be described statistically, as Born probabilities?"

3.1 Quantum mechanics

Quantum wave mechanics and the Schrödinger equation (SE) allowed for an accurate description of physical phenomena such as the spectra of single atom gases, e.g., the Lyman, Balmer and Paschen series, by realising that bound states, e.g., an atom, could only exist in *discrete "states"*, i.e., eigensolutions to the SE, later called *eigenstates*. More interestingly, the wave function $\Psi(x)$, can take the form of *any linear combination* of such solutions. While it was unclear if it had any physical meaning in itself, the wave function – a weighted sum of orthogonal, complex solutions to the SE – taken modulus squared, turned out to accurately describe *probability density*, so that e.g. the probability to find a particle in the interval $[a, b]$ is strictly positive, and equal to

$$\int_a^b ||\Psi(x)||^2 dx. \tag{3.1}$$

Here, $||\cdot||^2$ is taken to be the *complex* modulus squared, $\Psi^*(x)\Psi(x)$. Notably, *radioactive decay*, through a process known as *tunnelling* (see e.g., [GC29]), could successfully be modelled with this notion of probability density.

The orthogonality of two different solutions to the SE, labelled i and j, can be expressed

$$\int_{-\infty}^{\infty} \varphi_i^*(x)\varphi_j(x)dx = \delta_{ij}, \tag{3.2}$$

where the case $i = j$ describes the normalisation criterion; if a particle is in a definite state, the probability to find it somewhere on the x-axis equals one. The *linear* behaviour of the wave function is remarkable, and gives rise to many effects that are unique for quantum mechanics. I will for the remainder of the thesis, instead use the language of Dirac, and continue this section with some basic building blocks and terminology.

3.1.1 Quantum states

Due to the insight that distinguishable outcomes of an experiment always corresponded to orthogonal eigensolutions to the SE (given the definition above), ex-

plicitly calculating the solutions is not necessary, instead a shorthand notation has become widely used – the Dirac *bra-ket* notation.

The Dirac formalism is particularly convenient for – but not limited to – systems with a finite number of eigensolutions to the SE, which leads to a finite spectrum of measurement outcomes (for the topic of this thesis, such treatment suffices well, and will be used unless otherwise specified). These states will be called *eigenstates* $|\varphi_i\rangle$, to a linear operator, see section 3.1.3, and N such states will constitute a basis in a functional space, called *Hilbert space*, denoted $\mathcal{H}^{(N)}$. Some properties of the Hilbert space are outlined in Appendix A.1.

Quantum mechanics dictate that one important class of states, called *pure*, and their *dual*, can be written as a superposition,

$$|\psi\rangle = \sum_{i=1}^{N} a_i |\phi_i\rangle, \quad \langle\psi| = \sum_{i=1}^{N} a_i^* \langle\phi_i|, \tag{3.3}$$

where the coefficients a_i are complex, and their modulus squared $|a_i|^2$ equals the probability p_i for the corresponding measurement outcome ϕ_i. Such states, can also for finite N conveniently be written as the column and row vectors

$$|\psi\rangle \doteq \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}, \quad \langle\psi| \doteq \begin{bmatrix} a_1^*, \ldots, a_N^* \end{bmatrix}. \tag{3.4}$$

However, a state $|\psi\rangle$ is a *ray* in $\mathcal{H}^{(N)}$, and an overall phase factor does not change the state,

$$|\psi\rangle \equiv e^{i\alpha}|\psi\rangle, \quad \alpha \in \mathbb{R}. \tag{3.5}$$

The sum of the probabilities for all outcomes ϕ_i in Eq. (3.3) equals one, given that the basis $\{|\phi_i\rangle\}$ is complete, therefore

$$\sum_i p_i = \langle\psi|\psi\rangle = \sum_i |a_i|^2 = 1, \tag{3.6}$$

where we have used the *inner product* $\langle\psi|\varphi\rangle$, which is the projection of $|\psi\rangle$ onto $|\varphi\rangle$ (i.e., degree of parallellity, see Appendix A.4) – also known as "overlap integral", due to the wave function origin. Thus, due to Eq. (3.3), (3.5) and (3.6), a general pure state in $\mathcal{H}^{(N)}$ is characterised by $2N-2$, real numbers. Eq. (3.3) illustrates that a quantum state can be *superposed*, i.e., manifest itself in several physical (classical) realities *simultaneously*. If a quantum state can be written in this form, it is called pure.

Now, if one considers two independent states, prepared by Alice and Bob,

$$\begin{aligned} |\psi_\mathrm{A}\rangle &= a_0|0_\mathrm{A}\rangle + a_1|1_\mathrm{A}\rangle, \\ |\psi_\mathrm{B}\rangle &= b_0|0_\mathrm{B}\rangle + b_1|1_\mathrm{B}\rangle, \end{aligned}$$

we may without loss of generality, write the joint state

$$\begin{aligned} |\psi_{\mathrm{AB}}\rangle &= |\psi_{\mathrm{A}}\rangle \otimes |\psi_{\mathrm{B}}\rangle \\ &= c_{00}|0,0\rangle + c_{10}|1,0\rangle + c_{01}|0,1\rangle + c_{11}|1,1\rangle \\ &= \sum_{ij} c_{ij}|i,j\rangle, \end{aligned} \tag{3.7}$$

where the $\otimes$ symbol represents the *tensor product* of two states, defined in an extended Hilbert space $\mathcal{H}_{\mathrm{A}}^{(2)} \otimes \mathcal{H}_{\mathrm{B}}^{(2)} = \mathcal{H}_{\mathrm{AB}}^{(4)}$. We have assumed that the two states are independent, therefore they can be written as a product. We may also identify the coefficients $c_{ij} = a_i b_j$. However, if the two states interact in some way, so that the coefficients in $\mathcal{H}_{\mathrm{AB}}^{(4)}$ change, e.g., if the cross-terms become zero, $c_{01} = c_{10} = 0$, we may no longer write the two-(particle) state as a product of independent states. Whenever it is not possible to write a many-particle pure state as a product of single-particle pure states, we say that the larger state is *entangled*, see Werner [Wer89].

Often, one may not be able to access a complete, pure system $|\psi_{\mathrm{AB}}\rangle$, but only a part of it. *Mixedness* then occurs as a result of entanglement, because if we only measure state A in a state which is fully entangled over AB, e.g., the bipartite pure state

$$|\psi_{\mathrm{AB}}\rangle = \frac{1}{\sqrt{2}}(|0_{\mathrm{A}}\rangle \otimes |0_{\mathrm{B}}\rangle - |1_{\mathrm{A}}\rangle \otimes |1_{\mathrm{B}}\rangle), \tag{3.8}$$

we will not see any interference between the outcomes of Alice's state – Alice's state must in this case written as a *mixed state* (a motivation is given in section 3.1.5!), represented by a *density matrix*

$$\boldsymbol{\rho}_{\mathrm{A}} = \frac{1}{2}(|0_{\mathrm{A}}\rangle\langle 0_{\mathrm{A}}| + |1_{\mathrm{A}}\rangle\langle 1_{\mathrm{A}}|). \tag{3.9}$$

One may think of this construction as $|0_{\mathrm{A}}\rangle$ *or* $|1_{\mathrm{A}}\rangle$, each with probability 1/2 – in contrast to the pure state $(|0_{\mathrm{A}}\rangle + |1_{\mathrm{A}}\rangle)/\sqrt{2}$ where instead the word "and" is appropriate. The state (3.9) is a very special state, which takes the same form in all bases, i.e., $\boldsymbol{\rho}_{\mathrm{A}} = \frac{1}{2}\mathbb{1}$, see Appendix A.5.

3.1.2 Density matrices

According to the spectral theorem, any Hermitian matrix can be diagonalised, corresponding to a change of basis. The eigenvalues that then appear on the diagonal are real.

In addition to being Hermitian ($\boldsymbol{\rho}^\dagger = \boldsymbol{\rho}$), all density matrices are also positive semi-definite, i.e., their eigenvalues are non-negative. Consequently, for any normalised density matrix, its eigenvalues correspond to probabilities p_k and we have

$$\boldsymbol{\rho} = \sum_k p_k |\phi_k\rangle\langle\phi_k|, \tag{3.10}$$

where the states $|\phi_k\rangle$ are orthonormal. However, the decomposition (3.10) is generally not unique. If there is only one non-zero probability, i.e., if the density matrix has rank one, the state is pure, and $\boldsymbol{\rho} = |\psi\rangle\langle\psi|$. All other states are said to be mixed.

Any completely mixed states in a Hilbert space of dimension N has $p_k = 1/N$, for $k = 1, 2, \ldots, N$.

Partially mixed states $\boldsymbol{\rho}$ can be characterised by their *purity*, defined as $\mathrm{Tr}\left(\boldsymbol{\rho}^2\right)$. This measure satisfies

$$\frac{1}{N} \leq \mathrm{Tr}\left(\boldsymbol{\rho}^2\right) = \sum_k p_k^2 \leq 1, \tag{3.11}$$

where the lower and upper bound corresponds to completely mixed and pure states, respectively.

3.1.3 Linear operators

A linear operator $\boldsymbol{A}$ takes a quantum state to a different state

$$\boldsymbol{A}|\psi\rangle = c|\psi'\rangle, \quad c \in \mathbb{C}$$

which need not be normalised. The linearity of $\boldsymbol{A}$ means that

$$\boldsymbol{A}(c_1|\psi\rangle + c_2|\varphi\rangle) = c_1\boldsymbol{A}|\psi\rangle + c_2\boldsymbol{A}|\varphi\rangle, \tag{3.12}$$

and for two linear operators $\boldsymbol{A}$ and $\boldsymbol{B}$

$$(\boldsymbol{A} + \boldsymbol{B})|\psi\rangle = \boldsymbol{A}|\psi\rangle + \boldsymbol{B}|\psi\rangle. \tag{3.13}$$

A linear operation on a state can be expanded into a complete basis $\{|\phi_i\rangle\}$

$$\boldsymbol{A}|\psi\rangle = \sum_i a_i|\phi_i\rangle,$$

and therefore

$$\boldsymbol{A} = \sum_{ij} c_{ij}|\phi_i\rangle\langle\phi_j|, \tag{3.14}$$

where $|\phi_i\rangle\langle\phi_j|$ is the *outer product* and $c_{ij} = a_i a_j^*$.

3.1.4 Unitary and non-unitary operations

A postulate of QM says that all evolution of quantum states is governed by *unitary operations*, $\boldsymbol{U}$, defined as the linear operators fulfilling

$$\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1} \Leftrightarrow \boldsymbol{U}^\dagger\boldsymbol{U} = \mathbb{1} \quad (= \boldsymbol{U}\boldsymbol{U}^\dagger), \tag{3.15}$$

where the "dagger" operation means conjugation and transposition, $\boldsymbol{U}^\dagger = \left(\boldsymbol{U}^{\mathbf{T}}\right)^*$. This remarkable property means that unitary operations are *reversible*, and there is no *preferred direction* of this evolution. Eq. (3.15) also ensures that the probability density of a pure state $|\psi_0\rangle$ is conserved,

$$\langle\psi|\psi\rangle = \langle\psi_0|\boldsymbol{U}^\dagger\boldsymbol{U}|\psi_0\rangle = 1. \tag{3.16}$$

If we assume that unitary operations "push the universe forward in time", by operating on pure states (which can always be found in a large enough system), there seems to be a contradiction in that the direction of time itself has no meaning in this context. The evolution of quantum states should reflect the increase in entropy that we observe, i.e., in agreement with the second law of thermodynamics. For a discussion on this dilemma, see [Mac09] and references therein. In section 3.2.4 we shall see that the quantum entropy of a pure state is zero, and quantum entropy emerges as a consequence of the inability of measuring this complete state.

It can be shown (see Stinespring theorem [Sti55]), that if a particular state $\boldsymbol{\rho}$ cannot be described as a pure quantum state on its own, it is always possible to include hypothetical *reservoir* states, so that a pure state $|\mathrm{R}, \boldsymbol{\rho}\rangle$ can be found. This high-dimensional Hilbert space state is called a *purification* of the smaller state. One advantage of this view is that the evolution of pure states is governed by unitary operators, see section 3.1.4. Conversely, the mixed state $\boldsymbol{\rho}$ can be recovered from the purification by averaging over all possible outcomes in the the reservoir states,

$$\boldsymbol{\rho} = \mathrm{Tr}_\mathrm{R}\left(|\mathrm{R}, \boldsymbol{\rho}\,\rangle\langle\,\mathrm{R}, \boldsymbol{\rho}|\right), \tag{3.17}$$

where $\mathrm{Tr}_\mathrm{R}(\cdot)$ is the *partial trace*, i.e., a trace operation performed only over the reservoir basis states – see Appendix A.3.2 for details. It is also assumed that such hypothetical reservoir states *actually exist*, and that the occurrence of mixed states is only due to observation of an incomplete state, which is part of the purification.

Eq. (3.17) describes a process known as *decoherence* – the evolution of a pure state into a mixed state, as a result of entanglement between the state and states in the environment. Completely mixed states, i.e., where no interference exists between outcomes of a measurement, can be seen as classical, since they behave like we are used to. That is, they evolve according to our common perception of every-day life objects.

In the next sections I will list two types of operators that are important for qubits, as they represent *errors*, i.e., transformation of qubit states that can either be in the form of unitary rotations – or in the form of decoherence due to entanglement with a reservoir. In section 3.3.4 I will show how the effects of such operations can be counter-acted by means of other unitary operators, called *quantum gates*.

3.1.4.1 The Pauli operators

An *operator basis* $\{\boldsymbol{A}_i\}$ in $\mathcal{H}^{(N)}$ has $N \times N$ elements, and a linear combination of operators $\boldsymbol{A}_i$ can transform a pure state $|\psi\rangle$ into any other state

$$|\psi'\rangle = (\sum_i a_i \boldsymbol{A}_i)|\psi\rangle,$$

where $\sum |a_i|^2 = 1$, $a_i \in \mathbb{C}$ and $|\psi'\rangle$ needs not be normalised. If the operators $\boldsymbol{A}_i$ are *unitary*, they will be *amplitude conserving*, i.e., in $\mathcal{H}^{(2)}$, $|\det A_i| = 1$ and in $\mathcal{H}^{(2)}$ it is then convenient to use the 4 *Pauli operators*

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Since Pauli operators are unitary, they are useful for describing a general unitary qubit error. $\boldsymbol{X}$ represents a *bit-flip*,

$$\boldsymbol{X}|k\rangle \rightarrow |k \oplus 1\rangle,$$

where $\oplus$ is the modulo 2 operation and $k \in \{0, 1\}$. The $\boldsymbol{Z}$ operator is called *phase-flip*,

$$\boldsymbol{Z}|k\rangle \rightarrow (-1)^k |k\rangle,$$

i.e., it will flip the sign of the $|1\rangle$ probability amplitude. Furthermore, the $\mathbb{1}$ operation does nothing, and thus corresponds to the "no-error" case, while $\boldsymbol{Y}$ is identical to the combined operation $i\boldsymbol{X}\boldsymbol{Z}$.

3.1.4.2 The Kraus operators

Consider a system S, assumed to be in a pure state for simplicity. If a unitary operator acts *solely* on S, we can write

$$\boldsymbol{\rho}'_{\mathrm{S}} = \boldsymbol{U}_{\mathrm{S}} \boldsymbol{\rho}_{\mathrm{S}} \boldsymbol{U}_{\mathrm{S}}^{\dagger}.$$

Or, if the operations describe the interaction between S and a reservoir R,

$$\boldsymbol{\rho}'_{\mathrm{SR}} = \boldsymbol{U}_{\mathrm{SR}} \boldsymbol{\rho}_{\mathrm{SR}} \boldsymbol{U}_{\mathrm{SR}}^{\dagger}.$$

Also in the latter case we may assume, for some R, that $\boldsymbol{\rho}_{\mathrm{SR}}$ is pure. However, one may not have access to R, and the state for S can only be obtained as the partial trace over R,

$$\boldsymbol{\rho}'_{\mathrm{S}} = \mathrm{Tr}_{\mathrm{R}} \left(\boldsymbol{U}_{\mathrm{SR}} |\mathrm{R}, \boldsymbol{\rho}_{\mathrm{S}}\rangle\langle \mathrm{R}, \boldsymbol{\rho}_{\mathrm{S}}| \boldsymbol{U}_{\mathrm{SR}}^{\dagger} \right). \tag{3.18}$$

I will now mention an equivalent procedure, which can be performed on S only, assuming that R is inaccessible, but often with some assumption on an initial state in R,

$$\boldsymbol{\rho}'_{\mathrm{S}} = \sum_{\mu} \boldsymbol{K}_{\mu} \boldsymbol{\rho}_{\mathrm{S}} \boldsymbol{K}_{\mu}^{\dagger}, \tag{3.19}$$

where $\boldsymbol{K}_{\mu}$ are so-called *Kraus operators* [Kra83], which are in general *non-unitary*, and fulfil $\sum_{\mu} \boldsymbol{K}_{\mu}^{\dagger} \boldsymbol{K}_{\mu} = \mathbb{1}$. If the evolution of $\boldsymbol{\rho}_{\mathrm{S}}$ is unitary, then there can be only *one* term in Eq. (3.19), but if there are more terms, a pure state in S would in general become entangled with R after unitary evolution of the system SR.

Example: *Amplitude damping* qubit errors can be modelled in a two-qubit system initially prepared in $|Q_{\mathrm{S}}\rangle \otimes |0_{\mathrm{R}}\rangle = (a|1\rangle + b|0\rangle) \otimes |0_{\mathrm{R}}\rangle$, using a unitary operator with the following effect,

$$\begin{aligned} |1\rangle \otimes |0\rangle &\rightarrow \sqrt{p}|0\rangle \otimes |1\rangle + \sqrt{1-p}|1\rangle \otimes |0\rangle, \\ |0\rangle \otimes |0\rangle &\rightarrow |0\rangle \otimes |0\rangle. \end{aligned}$$

For $\boldsymbol{U}_{\mathrm{SR}}$ to be unitary we also need to take into account the reverse effect $|0\rangle \otimes |1\rangle \rightarrow |1\rangle \otimes |0\rangle$, for R initially in $|1\rangle$. The operation *entangles* the combined SR state, and can be written in the basis $\{|i\rangle_{\mathrm{S}} \otimes |j\rangle_{\mathrm{R}}\} = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$,

$$\boldsymbol{U}_{\mathrm{SR}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p} & -\sqrt{p} & 0 \\ 0 & \sqrt{p} & \sqrt{1-p} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.20}$$

with the property $\boldsymbol{U}_{\mathrm{SR}}^{\dagger} = \boldsymbol{U}_{\mathrm{SR}}^{-1}$, needed for unitarity. Now, we can calculate $\boldsymbol{\rho}'_{\mathrm{S}}$ in two equivalent ways, either as $\mathrm{Tr}_{\mathrm{R}}\left(\boldsymbol{U}_{\mathrm{SR}}|Q_{\mathrm{S}}\rangle \otimes |0_{\mathrm{R}}\rangle\right)$, or as in Eq. (3.19), using the two Kraus operators

$$\boldsymbol{K}_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \quad \boldsymbol{K}_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}. \tag{3.21}$$

In both cases we get for, e.g., the initial state $\boldsymbol{\rho}_{\mathrm{SR}} = |10\rangle\langle 10|$,

$$\boldsymbol{\rho}'_{\mathrm{S}} = \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix},$$

which is clearly a mixed state, as a consequence of the entanglement of SR.

3.1.5 Observables are Hermitian

Measurements on quantum states are represented by *observables*, which are a special type of linear operators. Such operators $\boldsymbol{O}$ are *Hermitian*, or *self-adjoint*, which means they have the property $\boldsymbol{O} = \boldsymbol{O}^{\dagger}$, or equivalently,

$$c_{ij} = c_{ji}^{*}, \tag{3.22}$$

where c_{ij} are the coefficients in Eq. (3.14). This property ensures that any observable has real eigenvalues. Thus, for an eigenbasis of an observable $\boldsymbol{O}^{(\phi)}$, $\{|\phi_i\rangle\}$, there exists a corresponding set of real numbers $\{\phi_i\}$, so that

$$\boldsymbol{O}^{(\phi)}|\phi_i\rangle = \phi_i|\phi_i\rangle.$$

As a consequence of Eq. (3.22), eigenstates to Hermitian operators with different eigenvalues are orthogonal.

The *projection postulate* states that a measurement of a pure state $|\psi\rangle$ in the $\{|\phi_i\rangle\}$ basis, i.e.,

$$\boldsymbol{O}^{(\phi)} = \sum_i \phi_i|\phi_i\rangle\langle\phi_i|,$$

can only result in *one* of the values ϕ_i, and this particular value will be recorded with probability $|\langle\psi|\phi_i\rangle|^2$ (*Born rule*), in which case the measurement will cause a preparation of the state $|\phi_i\rangle$.

Since a measurement involves a *meter*, which typically after interaction with some state $|\psi\rangle$ becomes entangled with it, one cannot easily analyse this situation, and the above statement remains a postulate in QM. We also note that $\boldsymbol{O}^{(\phi)}$ is not *unitary*, which indicates that part of the system is ignored (the meter).

The emergence of mixed states in section 3.1.1, can be seen as a result of a joint operation $\boldsymbol{O}_\mathrm{A} \otimes \mathbb{1}_\mathrm{B}$ that corresponds to a measurement on Alice's system with $\boldsymbol{O}_\mathrm{A}$, and on Bob's system with the identity operation. If we consider a pure, fully entangled, two-qubit state shared by Alice and Bob

$$|\psi\rangle = a|0_\mathrm{A}\rangle \otimes |0_\mathrm{B}\rangle + b|1_\mathrm{A}\rangle \otimes |1_\mathrm{B}\rangle,$$

and form the expectation value

$$\langle\psi|\boldsymbol{O}_\mathrm{A} \otimes \mathbb{1}_\mathrm{B}|\psi\rangle,$$

we get

$$|a|^2\langle 0_\mathrm{A}|\boldsymbol{O}_\mathrm{A}|0_\mathrm{A}\rangle + |b|^2\langle 1_\mathrm{A}|\boldsymbol{O}_\mathrm{A}|1_\mathrm{A}\rangle,$$

due to the orthogonality of $|0_\mathrm{B}\rangle$ and $|1_\mathrm{B}\rangle$. We can write the expectation value in a different form

$$\langle\psi|\boldsymbol{O}_\mathrm{A} \otimes \mathbb{1}_\mathrm{B}|\psi\rangle = \mathrm{Tr}\left(\boldsymbol{O}_\mathrm{A}\boldsymbol{\rho}_\mathrm{A}\right), \tag{3.23}$$

where $\boldsymbol{\rho}_\mathrm{A}$ is Alice's *reduced density matrix*

$$\boldsymbol{\rho}_\mathrm{A} = |a|^2|0_\mathrm{A}\rangle\langle 0_\mathrm{A}| + |b|^2|1_\mathrm{A}\rangle\langle 1_\mathrm{A}|.$$

3.1.6 Collective QND measurements

A QND measurement [GLP98] is a type of observable that does not change the measured state, i.e., there is no *back-action* effect. The only way to do this deterministically is to measure eigenstates to the observable, which naturally results in the eigenvalues for that observable.

For QEC, it is common to use a *collective* QND measurement $\boldsymbol{S}$, that does not distinguish between pairs of elements, taken from mutually orthogonal sets of bases $\{|\psi_i\rangle_{(0)}\}$ and $\{|\psi_i\rangle_{(1)}\}$, whose elements are also eigenstates of $\boldsymbol{S}$. Thus, such base sets have a common set of eigenvalues s_i, and a collective QND operator can be written

$$\boldsymbol{S} = \sum_i s_i \left(|\psi_i\rangle_{(0)\ (0)}\langle\psi_i| + |\psi_i\rangle_{(1)\ (1)}\langle\psi_i| \right). \tag{3.24}$$

Despite the simple form of Eq. (3.24), it is in general non-trivial to construct such a grouping of eigenstates, i.e., to implement the operation. In QEC, the measurement outcomes s_i are called *syndromes*, and the operation $\boldsymbol{S}$ is in the qubit case called a *parity operation*.

3.2 Quantum information

I will start out this section by an important theorem, discovered by Wootters and Zurek [WZ82], which simply states that a general single quantum state cannot be duplicated.

3.2.1 No-cloning theorem

Theorem 3. *(No-cloning) There is no quantum operation that takes a state $|\psi\rangle$ to $|\psi\rangle \otimes |\psi\rangle$ for all states $|\psi\rangle$.*

Proof. Suppose $|\psi\rangle$ and $|\varphi\rangle$ are orthogonal. Then, the cloning operation must be able to perform the following operations

$$|\psi\rangle \quad \rightarrow \quad |\psi\rangle|\psi\rangle, \tag{3.25}$$
$$|\varphi\rangle \quad \rightarrow \quad |\varphi\rangle|\varphi\rangle. \tag{3.26}$$

QM operations are linear (see section 3.1.3), therefore we may add Eq. (3.25) and (3.26) to calculate the operational result for a superposition of $|\psi\rangle$ and $|\varphi\rangle$,

$$\frac{1}{\sqrt{2}}(|\psi\rangle + |\varphi\rangle) \rightarrow \frac{1}{\sqrt{2}}(|\psi\rangle|\psi\rangle + |\varphi\rangle|\varphi\rangle). \tag{3.27}$$

But this contradicts the assumption that a general state can be cloned, which would require that

$$\frac{1}{\sqrt{2}}(|\psi\rangle + |\varphi\rangle) \rightarrow \frac{1}{2}(|\psi\rangle + |\varphi\rangle)(|\psi\rangle + |\varphi\rangle). \tag{3.28}$$

Since Eq. (3.27) and (3.28) are different, we have proven a general state cannot be cloned by a quantum operation. □

3.2.2 The classical bit (cbit), the qubit and the ebit

Unlike a classical bit, i.e., an element taken from $\{0, 1\}$ – the basic building block in quantum information is an element residing in $\mathcal{H}^{(2)}$, called a *qubit*. Thus, it is a quantum state with two orthogonal constituent states that we will denote $|Q\rangle$ and $|Q_\perp\rangle$. A single qubit prepared by Alice, but unknown to Bob, is from Bob's perspective a mixture of the two states (assuming the two states are sent with equal probability)

$$\boldsymbol{\rho} = \frac{1}{2}\left(|Q\rangle\langle Q| + |Q_\perp\rangle\langle Q_\perp|\right) = \frac{1}{2}\mathbb{1}. \tag{3.29}$$

In general, these states are unknown, but can be written

$$\begin{aligned} |Q\rangle &= \sin\alpha|0_{\mathrm{L}}\rangle + e^{i\beta}\cos\alpha|1_{\mathrm{L}}\rangle, \\ |Q_\perp\rangle &= e^{-i\beta}\cos\alpha|0_{\mathrm{L}}\rangle - \sin\alpha|1_{\mathrm{L}}\rangle, \end{aligned} \tag{3.30}$$

so that $\langle Q|Q_\perp\rangle = 0$. Since QEC utilises redundancy, it is assumed that the pure and mutually orthogonal basis states $\{|0_{\mathrm{L}}\rangle, |1_{\mathrm{L}}\rangle\}$ occupy a 2-dimensional *subspace* of the Hilbert space spanned by n constituent quantum states (see section 3.3.1). I will reserve the use of "qubit" as an information entity, in the above sense.

The qubit, despite having an infinite number of configurations, can only store one classical bit (also called *cbit*) of information, due to its two orthogonal states. However, it may represent the two bit-values *simultaneously*, allowing for simultaneous calculations [Deu85] – which is why it has attracted much attention as a means for doing quantum computing.

A qubit can also be used as a resource of entanglement. A *Bell state*, e.g.,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|Q\rangle \otimes |Q\rangle + |Q_\perp\rangle \otimes |Q_\perp\rangle\right), \tag{3.31}$$

is a *maximally entangled* two-qubit state, and this amount of entanglement is defined to equal one *ebit*, irrespective of basis. While two qubits may be used to form one ebit, this comes at a cost, namely that the state (3.31) cannot simultaneously be used to represent two cbits, since it cannot be written as a product of two states on the form (3.30). The ebit has many uses in quantum information, one prominent example being *teleportation* of qubit states [BBC$^+$93]. Teleportation here (unlike in some science-fiction movies), does not mean that a piece of matter is teleported, but rather that an unknown qubit state can be moved to a different point in space, while destroying the original qubit. The "carrier" of the qubit, e.g., a spin-1/2 particle, is not teleported – only the unknown superposition. Therefore, an equally sized Hilbert space must be in place at the destination point of the teleportation.

3.2.3 Alice and Bob

In order to characterise the information effects of QEC for a given channel and code, I will consider a joint density matrix for a single qubit, before and after the effects from the channel and any error correction is applied. I will label the initial qubit state A (Alice), and the final qubit state B (Bob), alluding to a scenario where Alice prepares two identical qubits and sends one to Bob. For the sake of reasoning, Alice keeps the other qubit as a reference state which can be written as a density matrix in the basis $\{|Q\rangle, |Q_\perp\rangle\}$ as

$$\boldsymbol{\rho}_{\mathrm{A}} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

assuming, for simplicity, that Alice sends $|Q\rangle$ and $|Q_\perp\rangle$ with the same probability.

The joint density matrix for the AB system allows us to quantify how well the qubit has "survived" a single pass through the channel, but $\boldsymbol{\rho}_{\mathrm{AB}}$ depends strongly on what measurement basis Bob uses to measure his qubit. If the channel interaction is absent, or errors caused by the channel can be perfectly corrected, the best basis Bob may use is $\{|Q\rangle, |Q_\perp\rangle\}$, which results in the joint density matrix

$$\begin{aligned} \boldsymbol{\rho}_{\mathrm{AB}} &= \frac{1}{2}\begin{bmatrix} 1\cdot\boldsymbol{\rho}_{\mathrm{B}} & 0 \\ 0 & 1\cdot\boldsymbol{\rho}_{\mathrm{B}}^{\perp} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \end{aligned} \tag{3.32}$$

in the basis $\{|Q\ Q\rangle, |Q\ Q_\perp\rangle, |Q_\perp\ Q\rangle, |Q_\perp\ Q_\perp\rangle\}$, where e.g., $|Q\ Q\rangle$ is shorthand for $|Q\rangle_{\mathrm{A}} \otimes |Q\rangle_{\mathrm{B}}$.

In contrast, the worst possible measurement basis Bob can use will result in, firstly, a rotation of basis $\boldsymbol{R} = \mathbb{1}_{\mathrm{A}} \otimes \boldsymbol{R}_{\mathrm{B}}$, i.e., a *pre-measurement*

$$\boldsymbol{\rho}_{\mathrm{AB}}^{\mathrm{pre}} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \tag{3.33}$$

followed by a von Neumann measurement (that erases any off-diagonal elements), i.e.,

$$\boldsymbol{\rho}_{\mathrm{AB}} = \frac{1}{4}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.34}$$

using e.g., $\{(|Q\rangle + |Q_\perp\rangle)/\sqrt{2}, (|Q\rangle - |Q_\perp\rangle)/\sqrt{2}\}$ as measurement basis.

Shortly, I will show that the states (3.32) and (3.34) are really the best and worst cases, respectively, as measured by a quantity called *quantum mutual information*, i.e., the quantum analogue of Eq. (2.5).

3.2.4 Quantum entropy

For quantum states, their entropy is defined due to von Neumann [von32]

$$H_{\mathrm{q}}(\boldsymbol{\rho}) = -k_{\mathrm{B}}\mathrm{Tr}\left(\boldsymbol{\rho}\log\boldsymbol{\rho}\right). \tag{3.35}$$

Density matrices can always be diagonalised, since they are a sum of dyadic products. In addition, the eigenvalues λ_i of the diagonal form are always real and non-negative. These properties of the density matrix ensure that $\mathrm{Tr}\left(\boldsymbol{\rho}\log\boldsymbol{\rho}\right)$ is well-defined and it is convenient to write

$$H_{\mathrm{q}}(\boldsymbol{\rho}) = -\sum_i \lambda_i \log\lambda_i, \tag{3.36}$$

where $0\cdot\log 0$ is defined to be zero. I have in Eq. (3.36) omitted k_{B}, since I shall express quantum entropy in units of Boltzmann's constant, in better analogy with Shannon entropy in Eq. (2.3).

It should be noted that the quantum entropy, as defined in Eq. (3.35), is zero for all pure states. Take e.g., a pure state in $\mathcal{H}^{(2)}$, with density matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Its von Neumann entropy is then calculated as $-0\cdot\log 0 - 1\cdot\log 1 = 0$.

3.2.5 Quantum mutual information

In analogy to classical mutual information, quantum mutual information is defined

$$I_{\mathrm{q}}(\mathrm{A}:\mathrm{B}) = H_{\mathrm{q}}(\boldsymbol{\rho}_{\mathrm{A}}) + H_{\mathrm{q}}(\boldsymbol{\rho}_{\mathrm{B}}) - H_{\mathrm{q}}(\boldsymbol{\rho}_{\mathrm{AB}}), \tag{3.37}$$

where $\boldsymbol{\rho}_{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\left(\boldsymbol{\rho}_{\mathrm{AB}}\right)$ and $\boldsymbol{\rho}_{\mathrm{B}} = \mathrm{Tr}_{\mathrm{A}}\left(\boldsymbol{\rho}_{\mathrm{AB}}\right)$.

This expression gives immediately for the perfect transmission and measurement, Eq. (3.32),

$$I_{\mathrm{q}}(\mathrm{A}:\mathrm{B}) = 1 + 1 - 1 = 1.$$

However, for the case when Bob performs the worst measurement, Eq. (3.34),

$$I_{\mathrm{q}}(\mathrm{A}:\mathrm{B}) = 1 + 1 - 2 = 0,$$

which tells us that no information is transmitted in the latter case.

Since the action of the channel, and the error correction is intimately connected to how the logical states $|0_{\rm L}\rangle$ and $|1_{\rm L}\rangle$ are coded, and the quantum computer itself is agnostic about the qubit parameters α and β in Eq. (3.30), it is fruitful to change basis to $\{|Q\ 0_{\rm L}\rangle, |Q\ 1_{\rm L}\rangle, |Q_\perp\ 0_{\rm L}\rangle, |Q_\perp\ 1_{\rm L}\rangle\}$, in which $\boldsymbol{\rho}_{\rm AB}$ takes the general form

$$\boldsymbol{\rho}_{\rm AB} = \begin{bmatrix} a & b & 0 & 0 \\ b^* & c & 0 & 0 \\ 0 & 0 & d & -b \\ 0 & 0 & -b^* & e \end{bmatrix}. \tag{3.38}$$

A remarkable quality of $I_{\rm q}({\rm A:B})$ is that it is *invariant under local unitary operations*, such as a rotation of basis performed by Bob, $\boldsymbol{R}\boldsymbol{\rho}_{\rm AB}\boldsymbol{R}^{-1}$, where $\boldsymbol{R} = \mathbb{1}_{\rm A}\otimes\boldsymbol{R}_{\rm B}$. Therefore, operations on qubits can be performed in e.g., the basis $\{|0_{\rm L}\rangle, |1_{\rm L}\rangle\}$, without harm, as long as the operations are unitary. One such rotation of Bob's basis is the optimal pre-measurement, and will unitarily transform Eq. (3.38) into its diagonal form. Bob's rotation of basis can be found as ${\rm Tr_A}\,(\boldsymbol{R})$, where the columns of $\boldsymbol{R}$ constitute the eigenvectors of $\boldsymbol{\rho}_{\rm AB}$. The maximal quantum mutual information possible to extract from $\boldsymbol{\rho}_{\rm AB}$ is therefore given by the eigenvalues of Eq. (3.38),

$$\begin{aligned}
\lambda_1 &= \frac{1}{2}\left(-\sqrt{(a-c)^2+4bb^*}+a+c\right), \\
\lambda_2 &= \frac{1}{2}\left(\sqrt{(a-c)^2+4bb^*}+a+c\right), \\
\lambda_3 &= \frac{1}{2}\left(-\sqrt{(d-e)^2+4bb^*}+d+e\right), \\
\lambda_4 &= \frac{1}{2}\left(\sqrt{(d-e)^2+4bb^*}+d+e\right).
\end{aligned}$$

The mutual information is then

$$I_{\rm q}({\rm A:B}) = 2 + \sum_{i=1}^{4}\lambda_i \log\lambda_i.$$

Here, I have used

$${\rm Tr_A}\,(\boldsymbol{\rho}_{\rm AB}) = {\rm Tr_B}\,(\boldsymbol{\rho}_{\rm AB}) = \frac{1}{2}\mathbb{1}^{(2)},$$

assuming that operations are trace-preserving, and

$$H_{\rm q}(\frac{1}{2}\mathbb{1}^{(2)}) = 1.$$

3.2.6 Is fidelity an information measure?

A quantity, commonly used to quantify "sameness" between two states, is the *fidelity*, defined for pure states as

$$\mathcal{F}(|\psi\rangle, |\varphi\rangle) = |\,\langle\psi|\varphi\rangle\,|^2, \tag{3.39}$$

i.e., the probability of preparing $|\varphi\rangle$ when measuring $|\psi\rangle$. Fidelity thus takes the value 0 for two orthogonal states, and the value 1 for identical states. For mixed states, fidelity is defined through averaging, see [Joz94],

$$\mathcal{F}(\boldsymbol{\rho},\boldsymbol{\sigma}) = \mathcal{F}(\boldsymbol{\rho},\boldsymbol{\sigma}) = \left(\mathrm{Tr}\left(\sqrt{\sqrt{\boldsymbol{\rho}}\boldsymbol{\sigma}\sqrt{\boldsymbol{\rho}}}\right)\right)^2 = \left(\mathrm{Tr}\left(\sqrt{\sqrt{\boldsymbol{\sigma}}\boldsymbol{\rho}\sqrt{\boldsymbol{\sigma}}}\right)\right)^2. \quad (3.40)$$

For the situation when Alice sends a qubit $\frac{1}{2}\mathbb{1}$, and Bob receives it in the worst basis, we therefore get, by substituting $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ with the joint density matrix in Eq. (3.32) and Eq. (3.33) respectively,

$$\mathcal{F}(\mathrm{A},\mathrm{B}) = 1/2. \quad (3.41)$$

As seen in the previous sections, (quantum) fidelity and quantum mutual information both become 1 in the ideal situation, but for the worst situation, where no information can be transmitted between Alice and Bob,

$$I_{\mathrm{q}}(\mathrm{A}:\mathrm{B}) = 0, \quad \mathcal{F}(\mathrm{A},\mathrm{B}) = 1/2. \quad (3.42)$$

Therefore, one may ask if fidelity is a good figure of merit in QEC? In paper B, I study a QECC consisting of 5 qubits, in a particular channel, and show that fidelity and quantum mutual information in general will not be simultaneously optimised – the optimisation of fidelity can only be done at the expense of quantum mutual information, and vice versa.

We know from classical information theory that we may expect some sameness between two random strings, i.e., non-zero fidelity as defined in Eq. (2.10), and that *zero* fidelity between two random strings will only occur *very* rarely. Instead, entropic measures such as mutual information, are used to characterise classical information transmission. This also affects classical coding strategies, as we saw in section 2.3, so it is not surprising that the same holds also for QEC.

3.3 Error correction procedure

The goal of QEC is to protect a qubit from decoherence and small unitary errors, by introducing redundancy to it, i.e., encoding the logical states $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$ from Eq. (3.30) onto several physical qubits, so that despite errors, the coded qubit can be recovered. This section will take the reader through the different phases of QEC.

3.3.1 Preliminaries

Firstly, I would like to mention an error effect which is subtle, and needs extra attention – the *reservoir memory effect*.

3.3.1.1 Reservoir memory effect

From section section 3.1.4, we know that errors on a system S are best described as a unitary operation over SR, i.e., the system and a reservoir. This will in general cause S to become entangled with R, which leads to decoherence in S if the reservoir is inaccessible. We then describe the evolution of S by tracing out the reservoir system, or equivalently, by application of so-called Kraus operators on S – i.e., *non-unitary* operations which are trace preserving only in the operator sum sense, see Eq. (3.19).

However, it is illustrative to express the effect of $\boldsymbol{U}_{\mathrm{SR}}$ on a combined state of a qubit $\boldsymbol{\rho}_Q = \mathbb{1}/2$ and an initially pure reservoir state $|\phi_0\rangle_{\mathrm{R}}$ in terms of Pauli operators on S, and some unknown back-action on R. Then, for any state $|Q\rangle$,

$$\begin{aligned} \boldsymbol{U}_{\mathrm{SR}} : |Q\rangle \otimes |\phi_0\rangle_{\mathrm{R}} \rightarrow \\ \mathbb{1}|Q\rangle \otimes |\phi_1\rangle_{\mathrm{R}} + \boldsymbol{X}|Q\rangle \otimes |\phi_2\rangle_{\mathrm{R}} + \boldsymbol{Y}|Q\rangle \otimes |\phi_3\rangle_{\mathrm{R}} + \boldsymbol{Z}|Q\rangle \otimes |\phi_4\rangle_{\mathrm{R}}, \end{aligned} \tag{3.43}$$

where the reservoir states $\{|\phi_i\rangle_{\mathrm{R}}\}$ are *not* necessarily mutually orthogonal or normalised. There will be two extreme cases, depending on the characteristics of the reservoir:

Example: Assume *all the reservoir states are parallel,* $|\phi_i\rangle = a_i|\phi_{\|}\rangle$, then

$$|\psi\rangle \otimes |\phi_0\rangle_{\mathrm{R}} \rightarrow (a_0\mathbb{1} + a_1\boldsymbol{X} + a_2\boldsymbol{Y} + a_3\boldsymbol{Z})\,|\psi\rangle \otimes |\phi_{\|}\rangle_{\mathrm{R}}, \tag{3.44}$$

where $a_i \in \mathbb{C}$, determined by $\boldsymbol{U}_{\mathrm{SR}}$, $\sum_i |a_i|^2 = 1$ and in this case the system is a pure state after the operation. With knowledge of the coefficients a_i, one can restore the state $|\psi\rangle$ perfectly, and no memory of the event is retained in the reservoir. To support this claim, I will calculate the leaked information, i.e., the mutual information between S and R, and show that it vanishes,

$$I_{\mathrm{q}}(\mathrm{S}:\mathrm{R}) = \underbrace{H_{\mathrm{q}}\left(\frac{1}{2}\mathbb{1}\right)}_{1} + \underbrace{H_{\mathrm{q}}\left(\boldsymbol{\rho}_{\mathrm{R}}\right)}_{0} - \underbrace{H_{\mathrm{q}}\left(\frac{1}{4}\begin{bmatrix} 1\cdot\boldsymbol{\rho}_{\mathrm{R}} & 0 \\ 0 & 1\cdot\boldsymbol{\rho}_{\mathrm{R}} \end{bmatrix}\right)}_{1} = 0. \tag{3.45}$$

I will call such reservoirs *memory-less*, due to this property. Some macroscopic states, e.g., the *thermal states* $\boldsymbol{\rho}_{\mathrm{T}}(\overline{n})$, consisting on the average of $\overline{n}$ photons, could constitute an almost memory-less reservoir due to $\mathrm{Tr}\left(\boldsymbol{\rho}_{\mathrm{T}}(\overline{n})\boldsymbol{\rho}_{\mathrm{T}}(\overline{n}+1)\right) \approx 1$, for $\overline{n} \gg 1$. Thus, for a photonic quantum code, transferring one photon to a reservoir thermal state with $\overline{n} \gg 1$, can hardly be distinguished from not doing so.

In contrast, a more realistic view is that the interaction with a reservoir results in, to some degree, distinguishable states in R. As an extreme example, consider the *Fock states*, or number states, which are simply labelled by the number of photons in the state, and have the property $\langle n|n+1\rangle = 0$, i.e., the state corresponding to one added photon is perfectly distinguishable to the case when no photon is added.

Example: Assume *all the reservoir states are orthogonal,* $\langle\phi_i|\phi_j\rangle = a_i^* a_j \delta_{ij}$, then we must describe S as the partial trace of Eq. (3.43) over R

$$\boldsymbol{\rho}'_{\mathrm{S}} = |a_0|^2 \boldsymbol{\rho}_{\mathrm{S}} + |a_1|^2 \boldsymbol{X}\boldsymbol{\rho}_{\mathrm{S}}\boldsymbol{X}^{-1} + |a_2|^2 \boldsymbol{Y}\boldsymbol{\rho}_{\mathrm{S}}\boldsymbol{Y}^{-1} + |a_3|^2 \boldsymbol{Z}\boldsymbol{\rho}_{\mathrm{S}}\boldsymbol{Z}^{-1}. \tag{3.46}$$

In this case the reservoir retains a memory of which of the events occurred. The operators $\mathbb{1}, \boldsymbol{X}, \boldsymbol{Y}$, and $\boldsymbol{Z}$ are in this context called *Pauli Kraus operators*, due to the resemblance to Eq. (3.19).

I shall call errors in the form Eq. (3.44) *unitary rotations* (on S), and errors stemming from Eq. (3.46) will be referred to as *decoherence*.

In the examples above, since only one system qubit is considered, it is not possible to distinguish 4 such Pauli errors, but if *a string* of n physical qubits are considered, the situation is different, as we will see shortly. To clarify what was just said, I will use the following new terms:

Definition 3.1. *(Physical qubit, carrier) A physical qubit is a quantum state spanning a 2-dimensional Hilbert space. The explicit realisation of this Hilbert space, is called a carrier. For example, the two linear polarisation states of a photon span* $\mathcal{H}^{(2)}$*, therefore a photon can "carry" a qubit.*

Definition 3.2. *(Logical qubit, codeword, code) A block of n physical qubits are typically used for the encoding of a single logical qubit. The full Hilbert space is then* $\mathcal{H}^{(2^n)}$*, and it has a 2-dimensional subspace spanned by the codewords* $|0_{\mathrm{L}}\rangle$ *and* $|1_{\mathrm{L}}\rangle$*. This subspace is called a code.*

3.3.1.2 Motivation for the memory-less condition

A memory-less reservoir can be motivated by the following argument: Let a photon impinge on a mirror at a normal angle. If the mirror is heavy, the momentum of the photon will change with $2\hbar\omega/c$, where $\hbar\omega$ is the photon energy, while the energy will be almost unchanged (elastic collision). If we do not care to measure in what direction the photon travels, we may say that a qubit encoded on the photon in the number basis $\{|0\rangle, |1\rangle\}$ is not affected. If, on the other hand, the mirror is very light, there will be some degree of entanglement between the mirror and the qubit. If one were to perform a partial trace over the mirror, the qubit would be in a mixed state. A similar argument can be made if instead, the photon is encoded in the spin basis $\{|S = -1\rangle, |S = 1\rangle\}$, or the polarisation basis $\{|H\rangle, |V\rangle\}$.

If the photon is instead *absorbed* in a reservoir, the distinguishability of the event depends on the reservoir temperature T_{R}. When a "cold" reservoir macrostate absorbs a "hot" photon, the initial and final macrostates are almost orthogonal, whereas a "hot" reservoir cannot distinguish this event, i.e., the initial and final macrostates are almost parallel. The latter assumption is valid for $k_{\mathrm{B}}T_{\mathrm{R}} \gg \hbar\omega$.

3.3.1.3 Simple codes

One may be tempted to try to simply copy a qubit three times, in analogy to the perfect classical three-bit code:

Example: Consider the equivalent of the classical repetition code **C2**, which is formed by simply repeating a general qubit three times,

QC1: Repeating a general qubit as a code

$$|Q\rangle \to |QQQ\rangle, \quad |Q_\perp\rangle \to |Q_\perp Q_\perp Q_\perp\rangle.$$

This code is possible to construct, since it is possible to clone orthogonal qubits in a known basis. However, it is usually assumed that encoding of the qubit is independent of the preparation. Then, the encoder has no knowledge of the parameters α and β in Eq. (3.30), and therefore the no-cloning theorem (Theorem 3) forbids such operations.

Instead, the encoding is done in a particular basis dictated by the gates in the encoder. If qubit gates are used, I will write the basis $\{|0\rangle, |1\rangle\}$, denoting individual physical qubit states in a logical qubit, as well as the input and output states of the gate. While encoding a logical qubit on n physical qubits expands the Hilbert space of the code, corresponding gates will account for this. The logical basis then becomes $\{|0\rangle, |1\rangle\}^{\otimes n}$, since all gates are assumed to use the same basis.

The parameters α and β in Eq. (3.30) will in general affect how well a particular QECC performs, for protecting a single qubit. However, the variation in effectiveness due to these parameters is, for "small errors" negligible. In what follows, I will therefore without loss of generality focus on studying the codewords $|0_\mathrm{L}\rangle$ and $|1_\mathrm{L}\rangle$.

Example: In the basis $\{|0\rangle, |1\rangle\}^{\otimes 3}$, we can write a three-qubit code

QC2: Three qubit repetition code

$$|0_\mathrm{L}\rangle = |000\rangle, \quad |1_\mathrm{L}\rangle = |111\rangle.$$

If we consider single bit-flips as the only source of error, and that such errors affect the physical qubits independently with probability $\gamma/3$, we firstly note that

$$\begin{aligned}
\mathbb{1}\mathbb{1}\mathbb{1}|0_\mathrm{L}\rangle &= |000\rangle = |S_0^{(0)}\rangle, & \mathbb{1}\mathbb{1}\mathbb{1}|1_\mathrm{L}\rangle &= |111\rangle = |S_0^{(1)}\rangle,\\
\boldsymbol{X}\mathbb{1}\mathbb{1}|0_\mathrm{L}\rangle &= |100\rangle = |S_1^{(0)}\rangle, & \boldsymbol{X}\mathbb{1}\mathbb{1}|1_\mathrm{L}\rangle &= |011\rangle = |S_1^{(1)}\rangle,\\
\mathbb{1}\boldsymbol{X}\mathbb{1}|0_\mathrm{L}\rangle &= |010\rangle = |S_2^{(0)}\rangle, & \mathbb{1}\boldsymbol{X}\mathbb{1}|1_\mathrm{L}\rangle &= |101\rangle = |S_2^{(1)}\rangle,\\
\mathbb{1}\mathbb{1}\boldsymbol{X}|0_\mathrm{L}\rangle &= |001\rangle = |S_3^{(0)}\rangle, & \mathbb{1}\mathbb{1}\boldsymbol{X}|1_\mathrm{L}\rangle &= |110\rangle = |S_3^{(1)}\rangle.
\end{aligned}$$

These 8 states $|S_i^{(j)}\rangle$, called *syndrome vectors*, are all mutually orthogonal, and the code is therefore said to be *non-degenerate*. Thus,

$$\left\langle S_i^{(j)} | S_k^{(l)} \right\rangle = \delta_{jl}\delta_{ik}, \tag{3.47}$$

which is a sufficient error correction criterion, in the case of a memory-less reservoir, i.e., when errors are unitary operations on S. I will come back to this criterion in section 3.4.4. Also, the resulting state from, e.g. $|0_{\rm L}\rangle$, will be pure

$$|0_{\rm L}\rangle \to \sqrt{1-\gamma}|S_0^{(0)}\rangle + \sqrt{\gamma/3}\left(|S_1^{(0)}\rangle + |S_2^{(0)}\rangle + |S_3^{(0)}\rangle\right).$$

This example shows how Pauli operators affecting a single physical qubit can be described in the case of a string of n physical qubits, i.e., by forming the unitary operators

$$\{\boldsymbol{E}_i\} = \{\mathbb{1}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}^{\otimes n}, \tag{3.48}$$

e.g., for a three-qubit state, there exist $4^3 = 64$ such operations. It is common to use a quantum version of the classical term Hamming weight from section 2.2.1,

Definition 3.3. *(Weight) The number of non-trivial single qubit Pauli operations in an n-qubit operation is called weight.*

Example: For a three-qubit state, there exists 10 operators of weight ≤ 1, e.g., $\mathbb{1} \otimes \mathbb{1} \otimes \boldsymbol{Y}$, which encompasses the situation that at most one Pauli error has occurred.

The code **QC2** was experimentally realised by Chiaverini *et al.*, [CLS+04] using trapped ions and simulated bit-flip noise.

3.3.1.4 Ancilla states – a reservoir that we can control

Consider again the example code **QC2**. The fact that all syndrome vectors (up to maximum one bit-flip error) are different, and that errors resulting from $|0_{\rm L}\rangle$ can be distinguished from those resulting from $|1_{\rm L}\rangle$, indicates that all such errors can be corrected. Naïvely, one would now like to map all the syndrome vectors $\{|S_i^{(0)}\rangle\} \to |0_{\rm L}\rangle$ and $\{|S_i^{(1)}\rangle\} \to |1_{\rm L}\rangle$, $i \in \{0 \ldots 3\}$, but such an operation is non-unitary, and therefore impossible to perform using only the code Hilbert space. Unitary operations need to map orthogonal states "one-to-one", and we can achieve this by introducing *ancilla states*.

Definition 3.4. *(Ancilla) An accessible set of qubits, initially prepared in a known state in the basis $\{|0\rangle, |1\rangle\}$, which takes part in unitary operations where it acts as an entropy storage. In the recovery stage, ancilla states are disentangled from the system, but will typically need to be reset before a new correction can take place.*

Since we have 4 syndrome vectors for each codeword, the ancilla states needs to span $\mathcal{H}^{(4)}$. Two ancilla qubits will suffice for this task, and in principle, we can

unitarily transform the syndrome vectors

$$\left(\sqrt{1-\gamma}|S_0^{(0)}\rangle + \sqrt{\gamma/3}\sum_{i=1}^{3}|S_i^{(0)}\rangle\right)\otimes|00\rangle,$$
$$\left(\sqrt{1-\gamma}|S_0^{(1)}\rangle + \sqrt{\gamma/3}\sum_{i=1}^{3}|S_i^{(1)}\rangle\right)\otimes|00\rangle,$$

into e.g.,

$$|000\rangle\otimes\left(\sqrt{1-\gamma}|00\rangle + \sqrt{\gamma/3}|10\rangle + \sqrt{\gamma/3}|01\rangle + \sqrt{\gamma/3}|11\rangle\right)$$

and

$$|111\rangle\otimes\left(\sqrt{1-\gamma}|00\rangle + \sqrt{\gamma/3}|10\rangle + \sqrt{\gamma/3}|01\rangle + \sqrt{\gamma/3}|11\rangle\right),$$

respectively. The ancilla qubits occupy positions 4 and 5, and we see that a partial trace over the ancilla states will now preserve superpositions of $|0_{\rm L}\rangle$ and $|1_{\rm L}\rangle$. With this I want to show that the ancilla states are needed for keeping error recovery unitary, and constitute an important ally in fighting errors. I will come back to how errors are reversed (undone) in section 3.3.4.

3.3.1.5 Quantum gates

Operations on qubits are performed by quantum gates, joined together in a quantum circuit. It is important to note that the action of a gate is related to one particular basis, usually written $\{|0\rangle, |1\rangle\}^{\otimes m}$ for an m-qubit gate.

Definition 3.5. *(Quantum gate) A quantum gate is a unitary operation that transforms an input m-qubit state into an output m-qubit state.*

Example: A CNOT gate operates on two qubits simultaneously, and can be completely described by the "truth values" in Table 3.1. In a quantum circuit, the CNOT gate is denoted with a symbol shown in Fig. 3.1.

Input qubits		Output qubits	
Control(•)	Target(⊕)	Control	Target
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$
$\|1\rangle$	$\|1\rangle$	$\|1\rangle$	$\|0\rangle$

Table 3.1: Truth table for the CNOT-gate.

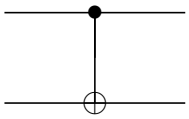

Figure 3.1: A CNOT qubit gate with two inputs (left); one control input ($\bullet$) and one target input ($\oplus$). The gate has the property that applying it twice is equivalent to the identity operator.

The CNOT gate can be written as a unitary operator in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$,

$$\boldsymbol{U}_{\mathrm{CNOT}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3.49}$$

Example: Quantum gates are usually made for qubits, in which case CNOT gates and so called *Hadamard rotations* give a complete "toolbox", i.e., allowing for the creation of all possible quantum circuits. For qutrits, I use a special gate in paper A, which is unitary and whose operation is listed in Table 3.2. Its symbolic representation in a circuit is shown in Fig. 3.2.

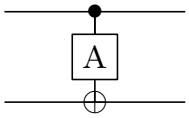

Figure 3.2: A qutrit gate with two inputs; one control input ($\bullet$) and one target input ($\oplus$), which also serves as output. The gate has the property that applying it twice is equivalent to the identity operator.

In the next section, I will show how gates can be combined to perform operations on several physical qubits.

3.3.2 Encoding

A *quantum circuit* is a sequence of unitary gates, which together form a more complex unitary operation. Fig. 3.1 shows the representation of a CNOT gate, which can be used to form a simple encoding circuit, shown in Fig. 3.3. This circuit takes a state $|\psi\rangle = a|0\rangle + b|1\rangle$ in the top of the figure, and extends it to a logical qubit with three physical qubits, creating the quantum code **QC2**, i.e., $|\psi\rangle \rightarrow a|000\rangle + b|111\rangle$.

Input qutrits		Output qutrits	
Control(•)	Target(⊕)	Control	Target
$\|0\rangle$	$\|H\rangle$	$\|0\rangle$	$\|V\rangle$
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|V\rangle$	$\|0\rangle$	$\|H\rangle$
$\|H\rangle$	$\|H\rangle$	$\|H\rangle$	$\|0\rangle$
$\|H\rangle$	$\|0\rangle$	$\|H\rangle$	$\|H\rangle$
$\|H\rangle$	$\|V\rangle$	$\|H\rangle$	$\|V\rangle$
$\|V\rangle$	$\|H\rangle$	$\|V\rangle$	$\|H\rangle$
$\|V\rangle$	$\|0\rangle$	$\|V\rangle$	$\|V\rangle$
$\|V\rangle$	$\|V\rangle$	$\|V\rangle$	$\|0\rangle$

Table 3.2: Truth table for the A-gate.

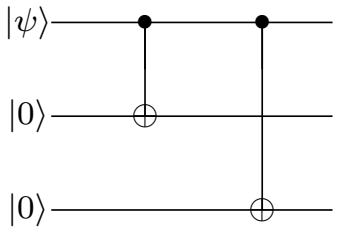

Figure 3.3: Two CNOT gates are used to encode a general qubit into three physical qubits, forming a quantum code.

3.3.3 The action of the channel

Definition 3.6. *(Channel) The channel incorporates the operational effects from interactions between* S *and* R*, allowing calculation of code state evolution. However, if* R *is inaccessible, it will be "traced out" and typically, memory effects are not accounted for in the channel description.*

QEC should ideally protect from both unitary rotations and decoherence, however many codes assume a memory-less channel, so that the reservoir states corresponding to errors are parallel, i.e., independent of S after the error occurred. Thus, while codes adapted from the classical domain typically will only correct unitary rotations, and not decoherence errors due to entanglement with R, this depends strongly on the channel. We have seen example of the memory-less bit-flip channel earlier, now I will consider the action of a bit-flip channel with memory, using the same code **QC2**.

It is easy to see that **QC2** will not protect a qubit from decoherence errors, in case the reservoir is entangled with a physical qubit in the codeword after the error. From Eq. (3.30), we know that a general qubit $|Q\rangle$ or $|Q_\perp\rangle$ are superpositions of the logical states $|0_\mathrm{L}\rangle$ and $|1_\mathrm{L}\rangle$. Suppose that after a bit-flip on the first qubit of

the state $|Q\rangle$, we have

$$|Q\rangle \to a|100\rangle \otimes |0\rangle_{\mathrm{R}} + b|011\rangle \otimes |1\rangle_{\mathrm{R}}. \tag{3.50}$$

If we rotate the state back in S only, using $\boldsymbol{E}_1^\dagger \otimes \mathbb{1}_{\mathrm{R}} = \boldsymbol{X}^{-1}\mathbb{1}\mathbb{1} \otimes \mathbb{1}_{\mathrm{R}}$, we would get

$$a|000\rangle \otimes |0\rangle_{\mathrm{R}} + b|111\rangle \otimes |1\rangle_{\mathrm{R}}, \tag{3.51}$$

i.e., the entanglement with the reservoir remains. In fact, it is impossible to remove the entanglement between S and R using only local unitary operations (on S). The error destroys any superposition between $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$ so that the best we can do is to "recover" (see section 3.3.4 for details) into the mixed state

$$|a|^2|000\rangle\langle 000| + |b|^2|111\rangle\langle 111|. \tag{3.52}$$

The fidelity between this state and $|Q\rangle$ is $|a|^4 + |b|^4$. For $a = b = 1/\sqrt{2}$, we have $\mathcal{F} = 1/2$, and a full calculation will show that $I_{\mathrm{q}}(\mathrm{S}:\mathrm{R}) = 1$, i.e., one bit of information was dissipated into the reservoir. Thus, the coded qubit completely loses its information if this error occurs.

3.3.3.1 Amplitude damping channel

Now, consider the channel usually described by the operator sum $\boldsymbol{\rho}'_{\mathrm{S}} = \sum_\kappa \boldsymbol{K}_\kappa \boldsymbol{\rho}_{\mathrm{S}} \boldsymbol{K}_\kappa^\dagger$, where $\{\boldsymbol{K}_\kappa\}$ are the two Kraus operators from Eq. (3.21):

$$\boldsymbol{K}_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \quad \boldsymbol{K}_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}.$$

In essence, this channel is a SR bit-flip channel, so that a bit-flip in S has a corresponding back-action on R. The reservoir is supposed to be in the $|0\rangle_{\mathrm{R}}$ state, so that it can "pick up" excitation from the system, i.e., $|1\rangle_{\mathrm{S}} \to |0\rangle_{\mathrm{S}}$, but not the other way around. Similarly to multi-qubit Pauli errors, see Def. 3.3, I will define the *weight* of an n qubit (qudit) amplitude damping error as the number of non-trivial single Kraus operators in $\{\boldsymbol{E}_i\} = \{\boldsymbol{K}_0, \boldsymbol{K}_1\}^{\otimes n}$, where the "trivial" Kraus operator is $\boldsymbol{K}_0$. Then, a QECC that can correct all such errors up to weight $\leq t$ is said to correct t errors. A QECC was developed for this channel in [FSW08], which also has the nice property that it can protect the qubit from decoherence, as we shall see.

Example:

QC3: Four qubit amplitude damping code

$$|0_{\mathrm{L}}\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle),$$
$$|1_{\mathrm{L}}\rangle = \frac{1}{\sqrt{2}}(|0011\rangle + |1100\rangle).$$

This code can be shown to correct one error (all Kraus operators of weight ≤ 1), despite the fact that $|S_0^{(i)}\rangle \neq |i_{\rm L}\rangle$ in general, i.e., its "no error" operator $\boldsymbol{K}_0$ is not the identity operator in this case.

In order to show that this code can, *in addition*, protect from decoherence (in this channel), I will consider the full unitary operation $\boldsymbol{U}_{\rm S_4R}$ acting on the 4:th physical qubit and the reservoir R, assuming that R is initially in the state $|0\rangle_{\rm R}$ and has physical qubit position 5. Making use of $\{|i\rangle_{\rm S_4} \otimes |j\rangle_{\rm R_5}\} = \{|00\rangle, |10\rangle, |01\rangle, |11\rangle\}$, we have

$$(\mathbb{1}_{\rm S_1} \otimes \mathbb{1}_{\rm S_2} \otimes \mathbb{1}_{\rm S_3} \otimes \boldsymbol{U}_{\rm S_4R_5})\,|0_{\rm L}\rangle \otimes |0\rangle_{\rm R} =$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & -\sqrt{\gamma} & 0 \\ 0 & \sqrt{\gamma} & \sqrt{1-\gamma} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4,5} \left(|000\rangle \otimes \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \end{bmatrix} + |111\rangle \otimes \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} \right)$$

$$= \frac{1}{\sqrt{2}} \Big[|0000\rangle + \sqrt{1-\gamma}|1111\rangle \Big] \otimes |0\rangle_{\rm R} + \frac{\sqrt{\gamma}}{\sqrt{2}} |0111\rangle \otimes |1\rangle_{\rm R}.$$

Similarly, for the second logical codeword,

$$(\mathbb{1}_{\rm S_1} \otimes \mathbb{1}_{\rm S_2} \otimes \mathbb{1}_{\rm S_3} \otimes \boldsymbol{U}_{\rm S_4R_5})\,|1_{\rm L}\rangle \otimes |0\rangle_{\rm R}$$

$$= \frac{1}{\sqrt{2}} \Big[|0011\rangle + \sqrt{1-\gamma}|1100\rangle \Big] \otimes |0\rangle_{\rm R} + \frac{\sqrt{\gamma}}{\sqrt{2}} |0010\rangle \otimes |1\rangle_{\rm R}.$$

Here we see that if the reservoir is found in the state $|1\rangle_{\rm R}$, there is no way to tell if this state was due to an error in $|0_{\rm L}\rangle$ or $|1_{\rm L}\rangle$, thus the *pairwise* superpositions of syndrome vectors stemming from the two codewords in S survive. A complete calculation (where all weight-0 and weight-1 amplitude damping errors are taken into account), will show that **QC3** will avoid decoherence, regardless of the the memory retained in R. N.B. that this was not so for the code **QC2** previously mentioned, which requires that R is memory-less.

3.3.3.2 Dissipative channel

In a dissipative channel, the *dissipation* (or loss of excitation) of code states is implied, resulting in emission, or scattering of quanta (e.g., photons) which are possible to detect "in-flight" and possibly reveal clues of their origin. Such clues, i.e., information transfer to R, are as we have seen detrimental to the performance of QECCs since they may cause decoherence. Through coding, one may protect a qubit from dissipation errors by eliminating clues in the same way as for the amplitude damping channel, see **QC3**. A photon with energy $\hbar\omega$, has two degrees of freedom which allows the encoding of one bit, e.g., using the orthogonal polarisation states $|H\rangle$ and $|V\rangle$. The dissipation of a photon results in the highly stable vacuum state, which I will denote $|0\rangle$, and which is orthogonal to both $|H\rangle$ and

$|V\rangle$. Transitions between these states are assumed to take place with probability γ, according to Fig. 3.4.

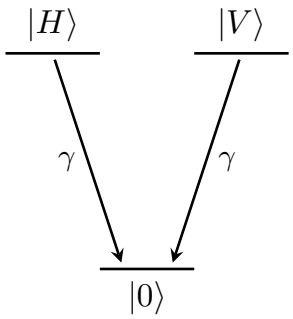

Figure 3.4: At probability rate γ, the doubly energy-degenerate states $|H\rangle$ and $|V\rangle$ can decay to the vacuum state $|0\rangle$ through the loss of one photon with energy $\hbar\omega$. The state $|0\rangle$ is orthogonal to both $|H\rangle$ and $|V\rangle$.

More precisely, such a physical qutrit in S evolves according to three Kraus operators in the $\{|0\rangle, |H\rangle, |V\rangle\}$ basis,

$$\boldsymbol{K}_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 \\ 0 & 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad \boldsymbol{K}_1 = \begin{pmatrix} 0 & \sqrt{\gamma} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{K}_2 = \begin{pmatrix} 0 & 0 & \sqrt{\gamma} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.53}$$

where γ is the probability of a jump from either $|H\rangle$ or $|V\rangle$ to $|0\rangle$. $\boldsymbol{K}_0$ and $\boldsymbol{K}_{1,2}$ are Kraus operators for the "no-jump" and "jump" events, respectively. We can also verify that $\boldsymbol{K}_0^\dagger \boldsymbol{K}_0 + \boldsymbol{K}_1^\dagger \boldsymbol{K}_1 + \boldsymbol{K}_2^\dagger \boldsymbol{K}_2 = \mathbb{1}_{\mathrm{S}}$, which is a consequence of that $\boldsymbol{U}_{\mathrm{SR}}$ conserves trace in S, R and SR simultaneously, however shifting relative probabilities between eigenstates.

The typical case for photons in a lab setup, is that the reservoir is "cold", i.e., $k_{\mathrm{B}} T_{\mathrm{R}} \ll \hbar\omega$, where $\hbar\omega$ is the energy of the photon and T_{R} is the temperature of the reservoir. In this scenario, the reservoir may be able to distinguish an emitted (scattered) photon's polarisation, and after tracing out the reservoir, superpositions in the coded qubit may be lost. It is preferable to protect a qubit from such reservoir memory effects – thus the code must *hide the origin of an emitted photon.* A QECC that does exactly this is reported in paper A, and its logical code words consist of the highly entangled states given by **QC4** below.

Example:

QC4: Three photon code - with the feature that R may without restriction measure an emitted photon due to a dissipation error

$$|0_{\mathrm{L}}\rangle \to \frac{1}{\sqrt{3}}\left(|0\mathrm{VH}\rangle + |\mathrm{H0V}\rangle + |\mathrm{VH0}\rangle\right),$$
$$|1_{\mathrm{L}}\rangle \to \frac{1}{\sqrt{3}}\left(|000\rangle + |\mathrm{HHH}\rangle + |\mathrm{VVV}\rangle\right).$$

Some intuitive understanding of why this code protects against decoherence may be found in Fig. 3.5.

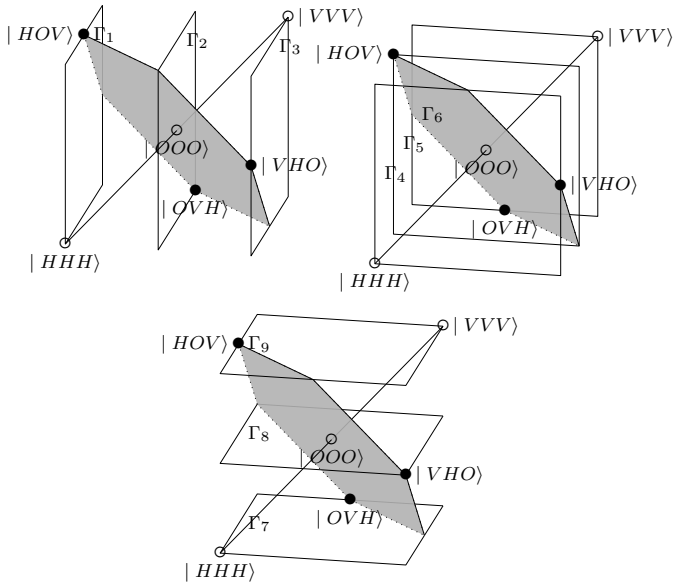

Figure 3.5: $|0_{\mathrm{L}}\rangle$ and $|1_{\mathrm{L}}\rangle$ are marked with dots and circles respectively. Note that each of the 9 planes representing the photon state of a given mode contains exactly two kets – one circle from $|1_{\mathrm{L}}\rangle$ and one dot from $|0_{\mathrm{L}}\rangle$. The 6 planes $\Gamma_1, \Gamma_3, \Gamma_4, \Gamma_6, \Gamma_7, \Gamma_9$ represent the modes $|\mathrm{H}\rangle$ and $|\mathrm{V}\rangle$ which can dissipate. Therefore any one dissipated photon will not reveal if it came from the $|0_{\mathrm{L}}\rangle$ or $|1_{\mathrm{L}}\rangle$ codeword.

3.3.4 Syndrome measurement and recovery

The unitary operation envisioned in section 3.3.1.4 can conveniently be realised in two steps, where firstly a *syndrome measurement* is performed, i.e., an operation on the form Eq. (3.24), which gives the same result for syndrome vectors stemming from different codewords, but distinct results for different errors $\boldsymbol{E}_\kappa$.

$$\boldsymbol{S} = \sum_\kappa s_\kappa \left(|S_\kappa^{(0)}\rangle\langle S_\kappa^{(0)}| + |S_\kappa^{(1)}\rangle\langle S_\kappa^{(1)}| \right).$$

Once κ is known, one can in principle always apply a corresponding recovery operator

$$\boldsymbol{R}_\kappa = |0_{\rm L}\rangle\langle S_\kappa^{(0)}| + |1_{\rm L}\rangle\langle S_\kappa^{(1)}|. \tag{3.54}$$

Example: If we go back to the simple "quantum repetition code" **QC2**, which only protects against bit-flip errors, I will show how error correction can proceed. We have seen in section 3.3.1.4 that we will need 2 ancilla states to be able to recover the superposition between the logical codewords. We also know that we must require the reservoir to "erase" the outcomes of the S − R interaction, i.e., R may not store the outcomes as orthogonal states $|i\rangle_{\rm R}$. A suitable syndrome measurement is to perform pairwise CNOT operations on the physical qubits and the ancilla states, a kind of *parity* operation, similar to the one used in classical error correction. The net effect of the two first CNOT gates is to set the upper ancilla state in $|i \oplus j\rangle$, i.e., addition modulo 2. The lower ancilla state will become $|i \oplus k\rangle$, see Fig. 3.6. As an illustration, assume that the first qubit has flipped, so that e.g., $|Q'\rangle = a|100\rangle + b|011\rangle$. Then the ancillæ will read $a_1 = 1$ and $a_2 = 1$, and we can correct the error by applying $\boldsymbol{X}\mathbb{1}\mathbb{1}$ on the logical qubit. Thus we will recover $|Q\rangle = a|000\rangle + b|111\rangle$ perfectly. After the correction, we must reset the ancilla states to their original states ($|0\rangle$).

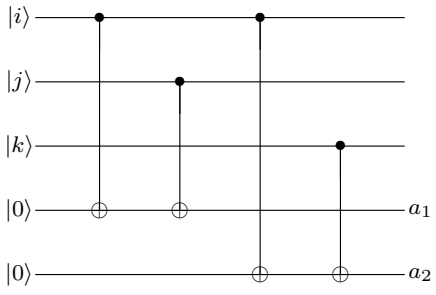

Figure 3.6: A syndrome measurement circuit for **QC2**. The ancilla values $\{a_1 a_2\}$ will take the values $\{00, 10, 01, 11\} = \{s_\kappa\}$, and these will determine which of the operations $\{\mathbb{1}\mathbb{1}\mathbb{1}, \mathbb{1}\boldsymbol{X}\mathbb{1}, \mathbb{1}\mathbb{1}\boldsymbol{X}, \boldsymbol{X}\mathbb{1}\mathbb{1}\}$ will be applied to the three output states.

The parity operation $i \oplus j$ in the example deserves an extra comment; in QEC we may in general not directly measure the "bit values" of the physical qubits i, j, k. Doing so may harm superpositions by excluding states incompatible with the measurement outcome. Instead, Eq. (3.24) requires that we perform a *collective* measurement which gives the same outcome (s_κ) for the same kind of error in the two codewords simultaneously. We have seen the exact same principle in use, in the case of classical error correction, where parity is a good choice in order to save either memory or processing time. In QEC, we are not *allowed* to choose any other strategy.

3.4 More on quantum codes

3.4.1 Notation

In analogy with classical codes, we can classify QECCs using a notation,

Definition 3.7. *(Notation) An $[[n, k, d]]_2$ quantum error correction code uses n physical qubits to encode k logical qubits, i.e., using 2^k logical codewords, and has a distance d, the minimum weight of a Pauli operation $\boldsymbol{E}_a$ that does not fulfil $\langle i_\mathrm{L}|\boldsymbol{E}_a|j_\mathrm{L}\rangle = C_a\delta_{ij}$, irrespective of i and j.*

Example: We see that **QC2** is denoted $[[3, 1, 1]]$, since e.g.,
$\langle 0_\mathrm{L}|\mathbb{1}\mathbb{1}\boldsymbol{Z}|0_\mathrm{L}\rangle \neq \langle 1_\mathrm{L}|\mathbb{1}\mathbb{1}\boldsymbol{Z}|1_\mathrm{L}\rangle$.

For Pauli errors, a code that can correct t errors has distance $d = 2t + 1$. N.B. that the relation between t and d becomes more diffuse in the case of e.g., codes adapted for the amplitude damping channel. Crépeau *et al.* comment on this peculiarity in [CGS05]:

> ...It demonstrates that the connection between correcting general errors and erasure errors breaks down for approximate QECCs. This calls into question some of the basic foundations of the theory of quantum error correction, as it it suggests there is no sensible notion of distance for an approximate quantum error correcting code.

3.4.2 The information carrier

The carrier for a qubit, i.e., the manifestation of a $\mathcal{H}^{(2)}$ quantum state has implications for the stability of the qubit, and also limits the transitions between its states. One suggested way of encoding qubits, is by means of trapped ions [LBMW03]. In a qubit transmission scenario, coding qubits on "flying" carriers, such as the photon's two polarisation states, may be a good way to protect exchanged qubits. Since dissipation involves transition to the electromagnetic vacuum mode $|0\rangle$, one may use this state already in the coding – effectively encoding a qutrit.

But the general theory for qutrits tells us that if the carrier can represent three orthogonal states, all conceivable operations on a general qutrit is e.g., given by the 9 Gell-Mann matrices,

$$\boldsymbol{\lambda}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\lambda}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\lambda}_2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{\lambda}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\lambda}_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\lambda}_5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{\lambda}_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{\lambda}_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad \boldsymbol{\lambda}_8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}. \quad (3.55)$$

One can then show that if all syndrome vectors are mutually orthogonal, the needed dimensionality of the Hilbert space put a limit on a $[[n, k, 2t + 1]]_3$-code, i.e., it must satisfy the bound

$$2^k \sum_{i=0}^{t} 8^i \binom{n}{i} \leq 3^n. \quad (3.56)$$

A similar bound for qubits is found in section 3.4.4. The minimum number of qutrits needed for satisfying Eq. (3.56), while correcting one error and encoding one bit is 4 – thus a $[[4, 1, 3]]_3$ code may exist. In this case the number of unique syndrome vectors is 66, which can be fitted into $\mathcal{H}^{(81)}$. Hypothetically, a perfect such QECC may exist, i.e., a code that saturates Eq. (3.56) is $[[10, 6 \log 3, 3]]_3$, using $3^{10} = 59049$ syndrome vectors and encoding 6 qutrits (or $6 \log 3$ qubits).

One realises that not all of the qutrit operations in Eq. (3.55) are likely or even possible, for the three photon states depicted in Fig. 3.4. For example, the $|\mathrm{V}\rangle$ state is not likely to change into $|\mathrm{H}\rangle$ in this channel model, and vice versa. Moreover, the state $|0\rangle$, the electromagnetical vacuum state, can be seen as a *decoherence-free subspace*, see [LCW98], i.e., a "quiet corner" of the system Hilbert space, unaffected by errors. This is also the reason why a **QC4** may exist, in spite of violating the bound Eq. (3.56), that general non-degenerate Pauli error codes should abide by.

3.4.3 Where is the information stored?

To illustrate how QEC can benefit from entanglement, consider the code **QC2**. In this code, a measurement of any single constituent physical qubits will immediately exclude one of the logical codewords, incompatible with the outcome. This destroys a superposition of the logical states, e.g., $|Q\rangle = a|0_\mathrm{L}\rangle + b|1_\mathrm{L}\rangle$.

To remedy this, and since errors are usually assumed to affect single physical qubits independently in a bit-flip channel, we may apply a three-qubit rotation on the codewords in **QC2**, so that each logical codeword is fully entangled:

QC5: "Non-local" repetition code for the bit-flip channel

$$|0_{\rm L}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle),$$
$$|1_{\rm L}\rangle = \frac{1}{\sqrt{2}}(|000\rangle - |111\rangle).$$

Now, a measurement of a *single* physical qubit will not exclude any logical codeword, since both codewords are compatible with any of the outcomes 0 or 1.

This principle is at work in Shor's code [Sho95],

QC6: Shor's [[9,1,3]] code

$$|0_{\rm L}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)^{\otimes 3},$$
$$|1_{\rm L}\rangle = \frac{1}{\sqrt{2}}(|000\rangle - |111\rangle)^{\otimes 3}.$$

Here an entangled block is repeated three times, in order to also correct one phase-flip error.

3.4.4 Error correction criteria

Criteria that QECCs need to abide by include those in [EM96] and [KL97]. These criteria had in mind a certain type of QECCs, namely those adapted for channels which are unitary over S, such as the Pauli channel. However, it was later realised that much of the underpinnings of such rules were overthrown for non-unitary (on S) channels, such as the amplitude damping channel – where other codes could successfully be used. In particular, the "trivial operation" (with weight 0) is typically not the identity operation $\mathbb{1}$ (c.f., the Kraus operator $\boldsymbol{K}_0$ for the amplitude damping channel), but correction can nevertheless be performed to an acceptable degree, in a recovery procedure called *approximate error correction*. The amplitude damping channel models the situation where the system S becomes entangled with a reservoir R and therefore suffers from decoherence. For these latter channels, the codes adapted for Pauli channels did not work well, and Leung *et al.*[LNCY97] found a new set of rules for these channels. In addition to these requirements, there is also one criterion (Theorem III.5 in [KL97]) that relates to the discussion on "reservoir memory" in section 3.3.1.1.

Recall that for the "repetition code" **QC2**, in section 3.3.1.3, all the error operators $\boldsymbol{E}_i$ up to weight 1 resulted in mutually orthogonal syndrome vectors. We may express Eq. (3.47) more generally,

$$\langle j_{\rm L}|\boldsymbol{E}_i\boldsymbol{E}_k|l_{\rm L}\rangle = \delta_{jl}\delta_{ik}. \tag{3.57}$$

But it turns out that Eq. (3.57) is a too strict condition, a sufficient and necessary criterion for QEC, (for error operators unitary over S) is

$$\langle j_{\rm L}|\boldsymbol{E}_i\boldsymbol{E}_k|l_{\rm L}\rangle = C_{ik}\delta_{jl}, \tag{3.58}$$

where $C_{ik} = \langle m_{\rm L}|\boldsymbol{E}_i\boldsymbol{E}_k|m_{\rm L}\rangle$ is a Hermitian matrix that must not depend on m.

These criteria on QECCs immediately leads to the following important classifications, which have implications on how densely one may encode information,

Definition 3.8. *(Non-degenerate code) A non-degenerate QECC has mutually orthogonal syndrome vectors* $\{|S_i^{(j)}\rangle\}$.

Definition 3.9. *(Degenerate code) A degenerate QECC has at least two pairs of parallel syndrome vectors.*

3.4.4.1 Non-degenerate codes

Non-degenerate codes are similar to classical codes, in the sense that they will need as many dimensions in the code space as there exist syndrome vectors. The Hilbert space consisting of n qubits, can thus accommodate a maximum of 2^n syndrome vectors. Assuming a Pauli channel code that can correct t errors and encode k qubits, we can write this as an inequality,

Definition 3.10. *(Quantum Hamming bound)*

$$2^k \sum_{i=0}^{t} 3^i \binom{n}{i} \leq 2^n. \tag{3.59}$$

See [KL97, BDSW96, EM96, Got96] for details.

Definition 3.11. *(Perfect quantum code) QECCs that fulfil Eq. (3.59) with equality are called perfect.*

Example: The *five qubit quantum code* discovered independently in [LMPZ96] and [BDSW96], is the shortest QECC that can correct one Pauli error of any kind. One implementation of this code uses the following encoding [LMPZ96],

QC7: 5 qubit code, $[[5, 1, 3]]$

$$\begin{aligned}|0_{\rm L}\rangle = \frac{1}{\sqrt{8}}(&-|00000\rangle + |01111\rangle - |10011\rangle + |11100\rangle + \\ &|00110\rangle + |01001\rangle + |10101\rangle + |11010\rangle), \\ |1_{\rm L}\rangle = \frac{1}{\sqrt{8}}(&-|11111\rangle + |10000\rangle + |01100\rangle - |00011\rangle + \\ &|11001\rangle + |10110\rangle - |01010\rangle - |00101\rangle).\end{aligned}$$

Since **QC7** is perfect, it leaves no room at all for correcting, or even detecting weight 2 errors. Instead such errors will be misdiagnosed.

3.4.5 Short versus long codes

Codes can be made longer than **QC7** (by increasing n) for two purposes, one is to correct more errors (increase t), while another reason is to encode more qubits in a code block (increase k). As an example of the latter, there exists an $[[85, 77, 3]]$ perfect code. As a rule of thumb, the former codes are more difficult to find.

For non-degenerate qubit QECCs, we know from the quantum Hamming bound Eq. (3.59), that any code is a tradeoff between the number of encoded qubits k and the number of correctable Pauli errors t, for a given n. Such QECCs also need to fulfil other bounds, such as the *singleton bound* [KL97], and a bound presented in [CRSS85].

Thus, to correct additional errors one has to increase the dimensionality of the Hilbert space by adding extra physical qubits, however while doing so, the channel interaction increases at the same time as such higher order errors become more and more unlikely. Also, the number of syndromes needed to correct a weight $k + 1$ error is significantly larger than to correct a weight k error. These "counter-acting forces" led us to study the "efficiency" of long codes [BAS08]. To do so, we used tabulated codes from [Gra07], but also *hypothetical codes* (which have not been found, but fulfill the quantum Hamming bound). We also assumed a *depolarising channel*, which is a channel where (the nontrivial) Pauli errors are assumed to occur independently, each with a probability of $p/3$.

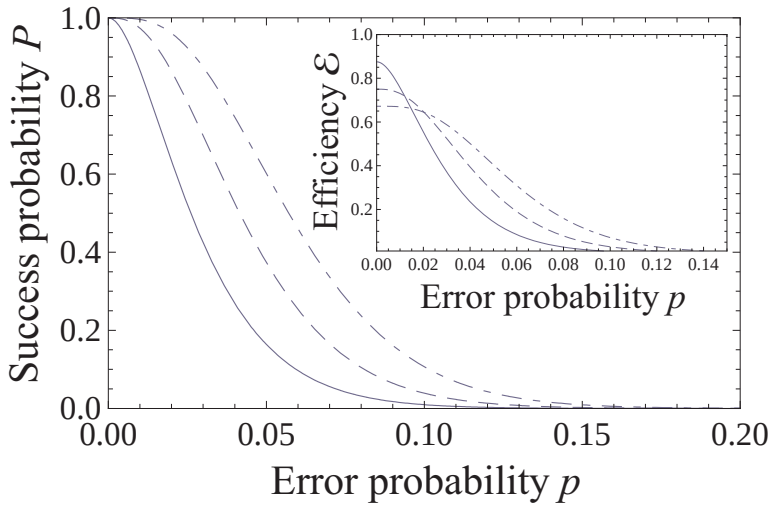

Figure 3.7: The probability that the error corrected state is identical to the original state for different codes. The codes are assumed to have parameters [[64,56,3]] (solid), [[64,48,5]] (dashed), and [[64,43,7]] (dot-dashed). Inset, the corresponding code efficiency $\mathcal{E}$ is plotted.

Then, the probability P that no more than t errors occur, so that a $[[n, k, 2t+1]]$ QEC can correct such errors, is given for one code block, by

$$P = \sum_{i=0}^{t} (1-p)^{n-i} p^i \binom{n}{i}. \tag{3.60}$$

This probability is plotted in Fig. 3.7 for three $n = 64$ codes, where one can see that when more errors are corrected (at the expense of less encoded qubits), the success probability increases, which is well known.

If one instead asks – *"how do we quantify efficient use of a fixed number of physical qubits"*, i.e., in order to maximise the number of correctly transmitted qubits over the channel?

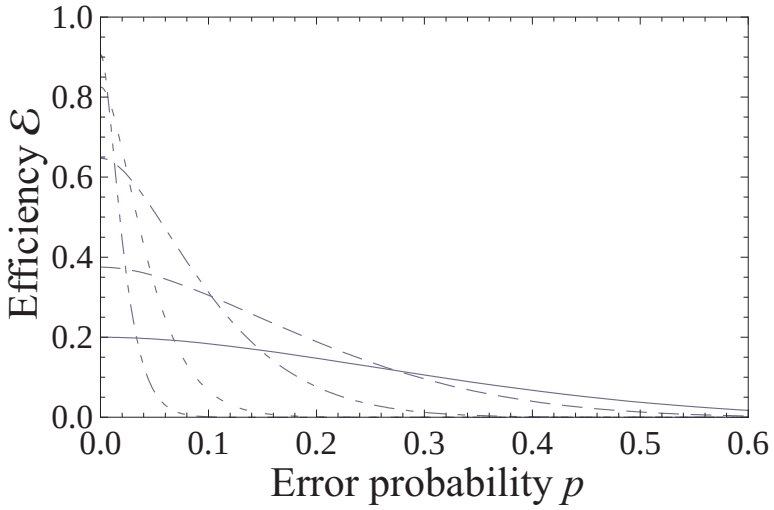

Figure 3.8: The efficiency for codes with assumed parameters [[5,1,3]] (solid), [[8,3,3]] (dashed), [[17,11,3]] (dot-dashed), [[40,33,3]] (small-dashed), and [[85,77,3]] (dot-dot-dashed).

There are several ways to define "efficiency" and a measure based on mutual information appears most natural. However, we have opted to define a measure of efficiency which is the average number of correctly transmitted qubits, per physical qubit and per code block,

Definition 3.12. *(Efficiency) We define the efficiency of a Pauli channel, non-degenerate* $[[n, k, d]]$ *QECC, with independent errors*

$$\mathcal{E} = \frac{Pk}{n}. \tag{3.61}$$

Some intuitive understanding of the efficiency $\mathcal{E}$ as a function of p can be gained from Fig. 3.8, for some existing, and some hypothetical QECCs. Keeping $t = 1$ fixed, we see that the efficiency is high for large n, and small p, while for higher p, the short codes are more efficient. A more exhaustive discussion on this can be found in [BAS08].

3.5 Discussion and open questions

An interesting, and alternative viewpoint on QEC is to investigate properties of the "carriers" of qubits, with respect to their stability, i.e., their tendency to interact with reservoir states. The vacuum state for a photon carrier, $|0\rangle$, is one such interesting state that is very stable, and could be used for QEC. When making maximum use of particular carriers, and their typical resulting states after interaction with the environment, the most practical choice of base b in a $[[n, t, d]]_b$ code may be different from 2.

One may also view QEC from an informational viewpoint, and I have given some examples of how to quantify "informational leakage" through the concept of quantum mutual information.

While the set of Pauli operators completely describe errors on qubits in contact with a memory-less reservoir, this description may be overly pessimistic in that for many channels, not all such errors will occur. Also this description is overly optimistic, in the sense that not all reservoirs are memory-less. A "hot" reservoir is less inclined to save a record of interaction with the code system, than a "cold" reservoir. But on the other hand, a hot reservoir typically increases the overall noise, and would make error-correction harder.

It appears that given a limited number of information-carrying resources, it is still an open question how to use such resources efficiently, i.e., choosing the code parameters k, d and n in a $[[n, t, d]]_b$ code.

Appendix A

Useful identities in quantum mechanics

A.1 Functional analysis

Associated with the complex vector space $\mathcal{H}^{(N)}$ is a real valued *norm*, denoted $||\ \cdot\ ||$, with the properties

$$\begin{aligned} ||\ v\ || &\geq 0, \\ ||\ v\ || = 0 &\Leftrightarrow v = 0, \\ ||\ cv\ || &= |c|\ ||\ v\ ||, \\ ||\ v + w\ || &\leq ||\ v\ || + ||\ w\ ||, \end{aligned}$$

where $v, w \in \mathcal{H}^{(N)}$, and $c \in \mathbb{C}$. From the norm, we also introduce a measure of distance,

$$d(v, w) = ||\ v - w\ ||, \quad \forall\ v, w \in \mathcal{H}^{(N)},$$

called the *metric* induced by the norm.

The *inner product* is a map from $\mathcal{H}^{(N)} \times \mathcal{H}^{(N)}$ to the scalar K, i.e., for every pair $(|\psi\rangle, |\phi\rangle)$, there is an associated scalar, called inner product, with the property

$$\langle\psi|\psi\rangle \geq 0.$$

A quantum state $|\psi\rangle$ is defined in Hilbert space, denoted $\mathcal{H}^{(N)}$ of dimension N, which is a complex valued space with a dual state and a metric defined on it.

A.2 Notation

I will list an alternative "tensor notation", which will sometimes be used due to its compactness. It is implied that when two indices appear exactly twice, they should

be summed over (Einstein summation convention), e.g.,

$$A_{ij}b_j \equiv \sum_j A_{ij}b_j.$$

Some of the basic notation elements are

$$\begin{aligned}
\mathbb{1}\boldsymbol{A} &\doteq \delta_{ij}A_{jk} = A_{ik},\\
\boldsymbol{AB} &\doteq A_{ij}B_{jk},\\
\boldsymbol{Ab} &\doteq A_{ij}b_j,\\
\boldsymbol{a}\cdot\boldsymbol{b} &\doteq a_ib_i,\\
\mathrm{Tr}\,(\boldsymbol{A}) &\doteq A_{ij}\delta_{ij} = A_{ii}\delta_{ii} = A_{ii},
\end{aligned}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices, $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors. We also have

$$\boldsymbol{a}\times\boldsymbol{b} \doteq \varepsilon_{ijk}a_ib_j\hat{\boldsymbol{e}}_k,$$

where $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^3$. The *Levi-Civita tensor* ε_{ijk} is zero for $i=j$, $j=k$, or $i=k$, but

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = -\varepsilon_{213} = -\varepsilon_{132} = -\varepsilon_{321} = 1.$$

The dyadic product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is written

$$\boldsymbol{ab} \doteq a_ib_j.$$

The trace of any dyadic product per definition equals the scalar product

$$\mathrm{Tr}\,(\boldsymbol{ab}) = \boldsymbol{a}\cdot\boldsymbol{b} \doteq a_ib_i.$$

The transpose of a matrix is simply obtained by switching indices

$${A_{ij}}^{\mathbf{T}} = A_{ji}.$$

A.3 Density matrices

Tensor notation is useful for representing quantum states, especially when those are defined in a Hilbert space of larger dimension than two. For the important case when no correlations between states $\boldsymbol{\rho}_\mathrm{A}$ in $\mathcal{H}_\mathrm{A}^{(n_\mathrm{A})}$ and $\boldsymbol{\rho}_\mathrm{B}$ in $\mathcal{H}_\mathrm{B}^{(n_\mathrm{B})}$, we may write

$$\boldsymbol{\rho}_\mathrm{AB} = \boldsymbol{\rho}_\mathrm{A}\otimes\boldsymbol{\rho}_\mathrm{B} \doteq \rho^\mathrm{A}_{ij}\rho^\mathrm{B}_{kl}.$$

In this case, the number of unique elements is $n_\mathrm{A}^2 + n_\mathrm{B}^2$, and assuming $n_\mathrm{B} = 2$, we explicitly have

$$\boldsymbol{\rho}_\mathrm{AB} = \begin{bmatrix} \boldsymbol{\rho}_\mathrm{A}\rho^\mathrm{B}_{11} & \boldsymbol{\rho}_\mathrm{A}\rho^\mathrm{B}_{12} \\ \boldsymbol{\rho}_\mathrm{A}\rho^\mathrm{B}_{21} & \boldsymbol{\rho}_\mathrm{A}\rho^\mathrm{B}_{22} \end{bmatrix}.$$

In general, however, a bipartite state $\boldsymbol{\rho}_\mathrm{AB}$ has $(n_\mathrm{A}\cdot n_\mathrm{B})^2$ elements and can be written

$$\boldsymbol{\rho}_\mathrm{AB} \doteq \rho^\mathrm{AB}_{ijkl}.$$

A.3.1 Trace operations

Let $\boldsymbol{\rho}_{\mathrm{AB}}$ be a pure state in the systems A and B. To perform a measurement $\boldsymbol{M}^{\mathrm{A}}$ on only A is equivalent to perform $\boldsymbol{M}^{\mathrm{A}} \otimes \mathbb{1}^{\mathrm{B}}$, which yields an expectation value in tensor notation

$$\langle \boldsymbol{M}^{\mathrm{A}} \rangle = \rho^{\mathrm{AB}}_{ijkl} M^{\mathrm{A}}_{ki} \delta_{jl} = \rho^{\mathrm{AB}}_{ijkj} M^{\mathrm{A}}_{ki}. \tag{A.1}$$

This is usually written in matrix notation as

$$\langle \boldsymbol{M}^{\mathrm{A}} \rangle = \mathrm{Tr}\left(\boldsymbol{M}^{\mathrm{A}} \boldsymbol{\rho}_{\mathrm{A}} \right),$$

where

$$\boldsymbol{\rho}_{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\left(\boldsymbol{\rho}_{\mathrm{AB}} \right),$$

meaning the trace is only performed over the system B. In tensor notation we can identify

$$\mathrm{Tr}_{\mathrm{B}}\left(\boldsymbol{\rho}_{\mathrm{AB}} \right) = \rho^{\mathrm{AB}}_{ijkl} \delta_{jl} = \rho^{\mathrm{AB}}_{ijkj}. \tag{A.2}$$

A.3.2 Partial trace (procedure)

Take the Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$:

$$\boldsymbol{\rho}_{12} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11| \right).$$

Trace out the second qubit:

$$\begin{aligned}
\boldsymbol{\rho}_1 = \mathrm{Tr}_2\left(\boldsymbol{\rho}_{12} \right) &= \frac{1}{2}\left(\mathrm{Tr}_2\left(|00\rangle\langle 00| \right) + \mathrm{Tr}_2\left(|11\rangle\langle 00| \right) + \mathrm{Tr}_2\left(|00\rangle\langle 11| \right) + \mathrm{Tr}_2\left(|11\rangle\langle 11| \right) \right) \\
&= \frac{1}{2}\left(|0\rangle\langle 0| \langle 0|0\rangle + |1\rangle\langle 0| \langle 1|0\rangle + |0\rangle\langle 1| \langle 0|1\rangle + |1\rangle\langle 1| \langle 1|1\rangle \right) \\
&= \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1| \right) = \mathbb{1}/2.
\end{aligned}$$

This state is a mixed state even though the composite state was pure.

A.4 Parallellity and orthogonality

In quantum theory and in most scientific fields the notion of *projection* is commonly used, so that the projection of a vector $\mathbf{a}$ on another vector $\mathbf{b}$ is calculated as $\mathbf{a} \cdot \mathbf{b}$ and vice versa. We note that this measure depends on the length of the vectors, so that if we want a measure that just quantifies how parallel $\mathbf{a}$ and $\mathbf{b}$ are, but says nothing about their length, we could define *parallellity* as

$$\mathbb{P}(\mathbf{a}, \mathbf{b}) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \hat{\mathbf{a}} \cdot \hat{\mathbf{b}}.$$

This measure has the nice properties that it takes the values one for parallel vectors, and zero for orthogonal vectors. There is another "complementary" property for the vectors, namely how *orthogonal* they are. We would like to define a similar measure for this property, which has the property that it is zero for parallel vectors, and is one for orthogonal vectors. We would then like to define a relation between them, namely

$$\mathbb{P}^2(\mathbf{a},\mathbf{b}) + \mathbb{O}^2(\mathbf{a},\mathbf{b}) = 1.$$

–But what would such an $\mathbb{O}$ look like, does it even exist in general? The Schwartz inequality translates in our language to

$$\mathbb{P}^2(\mathbf{a},\mathbf{b}) \leq 1,$$

but gives no hint of what needs to be added for the equality to hold. However, there surely must be some simple cases where we can find an explicit form for our orthogonality measure – if it exists? Yes, we have for example in $\mathbb{R}^3$:

$$\mathbb{P}^2(\mathbf{a},\mathbf{b}) + \mathbb{O}^2(\mathbf{a},\mathbf{b}) = \frac{|\boldsymbol{a}\cdot\boldsymbol{b}|^2 + |\boldsymbol{a}\times\boldsymbol{b}|^2}{|\boldsymbol{a}|^2|\boldsymbol{b}|^2} = 1,$$

where $\boldsymbol{a}\times\boldsymbol{b} = \varepsilon_{ijk}a_i b_j \hat{\boldsymbol{e}}_k$. Actually $\boldsymbol{a}\times\boldsymbol{b}$ is only defined in $\mathbb{R}^3$, but I will now define in $\mathbb{C}^N$,

$$\boldsymbol{M} = \boldsymbol{ab} - \boldsymbol{ba} = a_i b_j^* - b_i a_j^*,$$

and

$$|\boldsymbol{a}\times\boldsymbol{b}|^2 = \sum_{i>j} |M_{ij}|^2.$$

Then, for pure states $|a\rangle$ and $|b\rangle$ in $\mathcal{H}^{(N)}$,

$$\mathbb{O}^2(|a\rangle,|b\rangle) = \frac{\sum_{i>j}|a_i b_j^* - b_i a_j^*|^2}{\langle a|a\rangle\,\langle b|b\rangle},$$
$$\mathbb{P}^2(|a\rangle,|b\rangle) = \frac{\langle a|b\rangle^2}{\langle a|a\rangle\,\langle b|b\rangle}.$$

A.5 Completely mixed states

A completely mixed state $\boldsymbol{\rho} = \frac{1}{2}\mathbb{1}$, in some basis $\{|0\rangle,|1\rangle\}$ takes the same form in *any* basis. To see this, a different general basis can be written $\{|\phi\rangle,|\phi_\perp\rangle\}$, where

$$\begin{aligned} |0\rangle &= \cos\theta|\phi\rangle + e^{i\phi}\sin\theta|\phi_\perp\rangle, \\ |1\rangle &= \sin\theta|\phi\rangle - e^{i\phi}\cos\theta|\phi_\perp\rangle, \end{aligned}$$

so that

$$\begin{aligned}
\boldsymbol{\rho} &= \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right), \\
&= \frac{1}{2}((\cos\theta|\phi\rangle + e^{i\phi}\sin\theta|\phi_\perp\rangle)(\cos\theta\langle\phi| + e^{-i\phi}\sin\theta\langle\phi_\perp|) \\
&\quad +(\sin\theta|\phi\rangle - e^{i\phi}\cos\theta|\phi_\perp\rangle)(\sin\theta\langle\phi| - e^{-i\phi}\cos\theta\langle\phi_\perp|)) \\
&= \frac{1}{2}((\cos^2\theta + \sin^2\theta)|\phi\rangle\langle\phi| + (\cos^2\theta + \sin^2\theta)|\phi_\perp\rangle\langle\phi_\perp|) \\
&= \frac{1}{2}\left(|\phi\rangle\langle\phi| + |\phi_\perp\rangle\langle\phi_\perp|\right).
\end{aligned}$$

□

Bibliography

[BAS08] G. Björk, J. Almlöf, and I. Sainz, *On the efficiency of non-degenerate quantum error correction codes for Pauli channels*, arXiv:0810.0541v4 [quant-ph] (2008).

[BB84] C. H. Bennett and G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, Proc. IEEE Int. Conference on Computers, Systems and Signal Processing, Bangalore, India (1984), 175–179.

[BBC+93] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, Phys. Rev. Lett. **70** (1993), 1895–1899.

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

[Ben73] C. H. Bennett, *Logical reversibility of computation*, IBM J. Res. Dev. (USA) **17** (1973), no. **6**, 525 – 532.

[CGS05] C. Crépeau, D. Gottesman, and A. Smith, *Approximate quantum error-correcting codes and secret sharing schemes*, Advances in Cryptology - EUROCRYPT 2005, Lecture Notes in Computer Science, vol. 3494, Springer Berlin / Heidelberg, 2005, pp. 285–301.

[CLS+04] J. Chiaverini, D. Leibried, T. Schaetz, M. D. Barret, B. R. B., J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, and D. J. Wineland, *Realization of quantum error correction*, Nature **432** (2004), 602–605.

[CRSS85] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, *Quantum error correction via codes over GF*(4), IEEE Trans. Info. Theory **44** (1985), 1369–1387.

[Deu85] D. Deutsch, *Quantum theory, the Church-Turing principle and the universal quantum computer*, Proc. R. Soc. Lond. **A400** (1985), 97–117.

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

[Fey82] R. P. Feynman, *Simulating physics with computers*, Int. J. Theor. Phys. **21** (1982), 467–488.

[Fey86] R. P. Feynman, *Quantum mechanical computers*, Found. Phys. **16** (1986), 507–531.

[FSW08] A. S. Fletcher, P. W. Shor, and M. Z. Win, *Channel-adapted quantum error correction for the amplitude damping channel*, IEEE Trans. Inf. Theory **54** (2008), 5705–5718.

[FT82] E. Fredkin and T. Toffoli, *Conservative logic*, Int. J. Theor. Phys. **21** (1982), 219–253.

[GC29] R. W. Gurney and E. U. Condon, *Quantum mechanics and radioactive disintegration*, Phys. Rev. **33** (1929), 127–140.

[GLP98] P. Grangier, J. A. Levenson, and J. P. Poizat, *Quantum non-demolition measurements in optics*, Nature **396** (1998), 537–542.

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

[Gra07] M. Grassl, tabulated codes online available at `http://www.codetables.de`, 2007, Accessed on 2008-10-07.

[Huf52] D. A. Huffman, *A method for construction of minimum redundancy codes*, Proc. IRE **40** (1952), 1098–1101.

[Joz94] R. Jozsa, *Fidelity for mixed quantum states*, J. Mod. Opt. **41** (1994), 2315–2323.

[KL97] E. Knill and R. Laflamme, *Theory of quantum error-correcting codes*, Phys. Rev. A **55** (1997), 900–911.

[Kra83] K. Kraus, *States, effects, operations: Fundamental notions of quantum theory*, Springer-Verlag, Berlin, 1983.

[LBMW03] D. Leibfried, R. Blatt, C. Monroe, and D. Wineland, *Quantum dynamics of single trapped ions*, Rev. Mod. Phys. **75** (2003), 281–324.

[LCW98] D. A. Lidar, I. L. Chuang, and K. B. Whaley, *Decoherence-free subspaces for quantum computation*, Phys. Rev. Lett. **81** (1998), 2594–2597.

[LMPZ96] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, *Perfect quantum error correcting code*, Phys. Rev. Lett. **77** (1996), 198–201.

[LNCY97] D. W. Leung, M. A. Nielsen, I. L. Chuang, and Y. Yamamoto, *Approximate quantum error correction can lead to better codes*, Phys. Rev. A **56** (1997), 2567–2573.

[LTTL96] L. S. Liebovitch, Y. Tao, A. T. Todorov, and L. Levine, *Is there an error correcting code in the base sequence in DNA?*, Biophys. J. **71** (1996), 1539–1544.

[Mac09] L. Maccone, *Quantum solution to the arrow-of-time dilemma*, Phys. Rev. Lett. **103** (2009), 080401–080404.

[Pre97] J. Preskill, *Quantum computing*, `http://theory.caltech.edu/preskill/ph229/` (1997).

[RL79] J. J. Rissanen and G. G. Langdon, Jr., *Arithmetic coding*, IBM J. Res. Dev. (USA) **23** (1979), 149–162.

[RS60] I. S. Reed and G. Solomon, *Polynomial codes over certain finite fields*, SIAM J. Appl. Math. **8** (1960), 300–304.

[Sha48] C. E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. **27** (1948), 378–423 and 623–656.

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

[Sho97] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput. **26** (1997), 1484–1509.

[SLA+11] S. Saugé, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, *Controlling an actively-quenched single photon detector with bright light*, Opt. Express **19** (2011), 23590–23600.

[SPC+03] P. V. Shcherbakova, Y. I. Pavlov, O. Chilkova, I. B. Rogozin, E. Johansson, and T. A. Kunkel, *Unique error signature of the four-subunit yeast DNA polymerase ϵ*, J. Biol. Chem. **278** (2003), 43770–43780.

[Sti55] W. F. Stinespring, *Positive functions on C^*-algebras*, Proc. Amer. Math. Soc. **6** (1955), 211–216.

[TM71] M. Tribus and E. C. McIrvine, *Energy and information*, Sci. Am. **225** (1971), 179–188.

[von32] J. von Neumann, *Mathematische grundlagen der quantenmechanik*, Springer, Berlin, 1932.

[Weh78] A. Wehrl, *General properties of entropy*, Rev. Mod. Phys. **50** (1978), 221–260.

[Wer89] R. F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, Phys. Rev. A **40** (1989), 4277–4281.

[WFS94] P. S. Winokur, D. M. Fleetwood, and F. W. Sexton, *Radiation-hardened microelectronics for space applications*, Radiat. Phys. Chem. **43** (1994), 175–190.

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