Symmetric Functionals over Tensor Product Spaces in the Context of Quantum Information Theory

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Moritz Fabian Ernst

aus Kleve

2014

Berichterstatter: PD Dr. Rochus Klesse

Prof. Dr. Simon Trebst

Tag der letzten mündlichen Prüfung: 03.12.2014

Kurzzusammenfassung

Die Arbeit besteht im Wesentlichen aus drei Teilen. Wir beginnen unsere Untersuchung mit einer knappen Einführung in offene Quantensysteme. Daraufhin erklären wir verschiedene Maße zur Unterscheidbarkeit von Dichteoperatoren, insbesondere die Uhlmann-Fidelity, welche als Grundlage für das in dieser Arbeit studierte Modell dienen wird.
Detaillierter widmen wir uns dann der zeitlichen Entwicklung offener Systeme und führen einen wichtigen Begriff der Quanteninformationstheorie ein, den des Quanteninformationskanals. Dieser ermöglicht uns, ein zentrales Problem der Quanteninformationstheorie zu formulieren: Die Bestimmung der Quanteninformationskapazität eines Quanteninformationskanals. Als wesentliche Schwierigkeit dabei stellt sich das Maximieren der Coherent Information heraus, eine auf dem n-fachen Tensorprodukt eines Hilbertraumes definierte Funktion, die auf Grund ihrer Nichtadditivität im Limes $n \to \infty$ zu betrachten ist.
Um einen Zugang zu diesem derzeit noch ungelösten Problem zu bekommen, studieren wir in dieser Arbeit eine ebenfalls auf dem n-fachen Tensorproduktraum definierte Funktion, die Channel-Fidelity eines Quantenkanals. Diese ist ebenfalls nicht additiv und hat somit eine wesentliche Eigenschaft mit der Coherent Information gemein. Sie ist allerdings im Gegensatz zu dieser mathematisch zugänglicher. Wir fassen die Channel-Fidelity als Modell der Coherent Information auf und studieren ihre Eigenschaften.
Im zweiten Teil geben wir nach einer kurzen Einführung in die Darstellungstheorie symmetrischer und unitärer Gruppen eine konkrete Anleitung zu Collins' und Śniadys Formel zur Integration von Funktionen von Matrixelementen unitärer Gruppen über das Haar'sche Maß. Schließlich vereinfachen wir diese Formel auf ein für die Untersuchung der Channel-Fidelity optimales Niveau.
Im dritten Teil berechnen wir allgemeine Momente der Channel-Fidelity-Verteilung für beliebige n. Um konkretere Ergebnisse zu erzielen beschränken wir uns auf Pauli-Kanäle. Für diese diskutieren wir für Mittelwert und Varianz den Übergang von kleinen n zum Limes $n \to \infty$ und können für beide eine explizite Formel für diesen angeben. Insbesondere stellt sich heraus, dass für eine große Anzahl n von Pauli-Kanälen die Verteilung sehr stark um den Mittelwert konzentriert ist. Weiterhin ermöglicht uns die vereinfachte Formel aus dem zweiten Teil unter gewissen Voraussetzungen die konkrete Berechnung höherer Momente.
Schließlich vergleichen wir unsere neuen Resultate mit unseren Ergebnissen aus einer älteren Arbeit. In dieser hatten wir nach Maxima der Channel-Fidelity gesucht und haben für Pauli-Kanäle Zustände gefunden, die die Channel-Fidelity zumindest lokal maximieren. Da diese lokalen Maxima weit oberhalb des Mittelwertes der stark

konzentrierten Verteilung liegen, folgern wir, dass diese mit einem gewöhnlichen Maximierungsverfahren kaum zu finden sind. Sollte die Channel-Fidelity in dieser Hinsicht ein gutes Modell für die Coherent Information darstellen, stellt deren Maximierung damit ein sehr schwieriges Problem dar.

Abstract

This thesis consists of three parts. We begin our investigation with a brief introduction into open quantum systems. Then we explain different measures of distinguishability of density operators, especially the Uhlmann Fidelity, which will be the basis for the model function we investigate in this work. We continue by explaining the time evolution of open quantum systems in more detail, and introducing the important notion of quantum channels as a concept in quantum information theory. This allows us to state a central problem of quantum information theory: the characterization of the quantum information capacity of a given quantum channel. The major challenge is the maximization of the coherent information, a function defined on a n-fold tensor product of a Hilbert space, which is non-additive and thus has to be considered in the limit as $n \to \infty$.

To gain an insight into this unsolved problem, we study the channel fidelity of a quantum channel, which is a simpler function, also defined on n-fold tensor product spaces. It shares an essential feature with the coherent information in being non-additive. However, in contrast to the coherent information it is mathematically accessible. We establish the channel fidelity as a model for the coherent information and study its properties.

In the second part, a short introduction to the representation theory of symmetric and unitary groups is followed by concrete instructions for Collins' and Śniady's formula for the integration of functions of matrix elements of unitary groups with respect to the Haar measure. This exposition culminates in a simplification of the general formula that is optimal for investigating the channel fidelity.

In the third part, we calculate channel fidelity moments for arbitrary n. In order to obtain more concrete results, we restrict ourselves to the study of Pauli channels. For these we discuss the transition of the average and variance from small n to the limit $n \to \infty$ and give an explicit formula for both in this limit. In particular, we find that for a large number n of Pauli channels, the channel fidelity distribution is peaked very strongly. Additionally, under certain restrictions, the simplified formula from part two also allows us to give concrete expressions for higher moments in the limit $n \to \infty$.

We conclude by comparing our new results with results from a former work, where, in the search for maximizing states of the channel fidelity, we found states that maximize the channel fidelity of Pauli channels, at least locally. Because these local maxima have a much higher fidelity than the average of the very strongly peaked distribution, we infer that these states would not be found by a standard numerical maximization procedure. If the channel fidelity models the coherent information accurately in this regard, its maximization thus poses a very hard problem.

Contents

1. Introduction

1.1. Quantum Information — Open Quantum Systems

To study quantum information, it is essential to understand the concepts of open quantum systems. We will now begin with the basic notions of quantum mechanics and explain how we can interpret pure state quantum mechanics as the quantum mechanics of closed systems, and how by going over to open quantum systems, the density operator formalism arises naturally. All the following concepts can be found in any advanced quantum mechanics course or an introductory text on quantum computing, for example [26].

The most basic structure of relevance is the Hilbert space of a physical system.

Definition 1.1.1 (Hilbert Space)**.** *A complex vector space $\mathcal{H} \cong \mathbb{C}^d$ equipped with the standard hermitian inner product is called a Hilbert space. Very often, when we have to consider multiple Hilbert spaces, we will differentiate them with an index $\mathcal{H}_j$ and denote their normalized basis as $\{|i\rangle_j\}_{i=1}^d$. If there is no confusion possible, sometimes we will drop the index j on the basis.*

Typically the inner product of a Hilbert space is denoted in bra-ket notation: For two elements $|\phi\rangle$ and $|\psi\rangle$ in $\mathcal{H}$ we write $\langle\phi|\,|\psi\rangle$, where the left braket means the adjoint vector: $\langle\phi| = |\phi\rangle^\dagger$.
If we have full knowledge of the system, the system will be in a pure state, which can be described as a vector from the Hilbert space.

Definition 1.1.2 (Pure States)**.** *The normalized elements of a Hilbert space ($|\phi\rangle \in \mathcal{H}$) or to be more precise, their induced projectors, $|\phi\rangle\langle\phi|$, are called pure states.*

More generally however, the state of the system will not be so simple and it has to be described by normalized, self-adjoint positive semi-definite operators on $\mathcal{H}$.

Definition 1.1.3. *The set of linear operators from $\mathcal{H}$ to itself is denoted by $\mathcal{L}(\mathcal{H})$.*

Definition 1.1.4 (Density Operator)**.** *An operator $\rho \in \mathcal{L}(\mathcal{H})$ on $\mathcal{H}$ is called a density operator iff it is positive semi-definite, $\rho \geq 0$; hermitian, $\rho = \rho^\dagger$; and its trace is one, $Tr(\rho) = 1$.*

In agreement with the standard physics notation we usually use single small Greek letters to denote density operators, e.g. ρ, σ. A density operator will not generally be a one-dimensional projector.

Fact 1.1.5. *A positive semi-definite hermitian operator ρ can be diagonalized as $\rho = \sum_i p_i |i\rangle \langle i|$, for some orthonormal basis, $\{|i\rangle\}_i$, and positive real numbers $\{p_i\}_i$.*

If the state of a system is described by a density operator, we usually understand it as a statistical ensemble of systems with isomorphic Hilbert spaces each in a different pure state. The system is in the pure state $|i\rangle$ with probability p_i. The p_i are a probability distribution over the states.
It is obvious that an arbitrary state cannot be assumed to be pure, but sometimes it can be useful to have a strong contrast to pure states.

Definition 1.1.6 (Mixed State)**.** *Let ρ be the density operator of an arbitrary physical system. We call the state of the system mixed iff more than one eigenvalue is larger than zero.*

It is a bit surprising that for describing the state of one system we have to think of more than one system. Alternatively, a mixed state can be interpreted as a pure state of a larger system, where the observer restricts herself or himself to only look at a subsystem. The restriction to a subsystem is defined next.

Definition 1.1.7 (Partial Trace)**.** *For a combined Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and a density operator $\rho_{1,2}$ on the combined space, we call the mapping*

$$\begin{aligned} \mathrm{Tr}_2 : \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2) &\to \mathcal{L}(\mathcal{H}_1), \\ \rho_{1,2} \mapsto \rho_1 = \mathrm{Tr}_2(\rho_{1,2}) &= \sum_i \mathbb{1}_2 \otimes \langle i|_2 (\rho_{1,2}) \, \mathbb{1}_2 \otimes |i\rangle_2 \end{aligned}$$

the partial trace over $\mathcal{H}_2$. Alternatively one might say, we trace out $\mathcal{H}_2$.

The name partial trace and also the notation suggest that there is a strong connection with the well known notion of a trace. We will now see that the partial trace basically goes halfway towards the trace.

Proposition 1.1.8 (Partial Trace)**.** *For a combined Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ it is equivalent to trace out $\mathcal{H}_1$ first and $\mathcal{H}_2$ second or the other way round. Furthermore, tracing out both systems is equal to the normal trace,*

$$\mathrm{Tr}_1 \left(\mathrm{Tr}_2(\rho_{1,2}) \right) = \mathrm{Tr}_2 \left(\mathrm{Tr}_1(\rho_{1,2}) \right) = \mathrm{Tr}(\rho_{1,2}).$$

Proof. In Definition 1.1.7 we see that:

$$\mathrm{Tr}_1 \left(\mathrm{Tr}_2(\rho_{1,2}) \right) = \sum_j \langle j|_1 \left(\sum_i \mathbb{1}_1 \otimes \langle i|_2 (\rho_{1,2}) \, \mathbb{1}_1 \otimes |i\rangle_2 \right) |j\rangle_1 ,$$

which is certainly equivalent to

$$\mathrm{Tr}_2 \left(\mathrm{Tr}_1(\rho_{1,2}) \right) = \sum_j \langle j|_2 \left(\sum_i \langle i|_1 \otimes \mathbb{1}_1 (\rho_{1,2}) \, |i\rangle_1 \otimes \mathbb{1}_1 \right) |j\rangle_2 .$$

Moreover, both expressions can be simplified to:

$$\mathrm{Tr}_2\left(\mathrm{Tr}_1(\rho_{1,2})\right) = \sum_{i,j} \left(\langle j|_1 \otimes \langle i|_2\right)(\rho_{1,2})\left(|j\rangle_1 \otimes |i\rangle_2\right) = \mathrm{Tr}\left(\rho_{1,2}\right).$$

□

Using this new tool we can now show that every mixed state can be represented by a pure state of a larger system.

Lemma 1.1.9 (Purification)**.** *For $\rho = \sum_i \lambda_i |i\rangle_1 \langle i|_1$, a positive semi-definite hermitian operator on $\mathcal{H}_1 = \mathbb{C}^{d_1}$, with normalized eigenstates (if necessary, completed to a basis of $\mathcal{H}_1$) $\{|i\rangle_1\}_{i=1}^{d_1}$, and eigenvalues $\{\lambda_i\}_i$; and a second Hilbert space $\mathcal{H}_2 = \mathbb{C}^{d_2}$ with $d_2 \geq d_1$ and basis $\{|j\rangle_2\}_{j=1}^{d_2}$, there exists a pure state*

$$|\phi\rangle = \sum_{i=1}^{d_1} \sqrt{\lambda_i}\, |i\rangle_1 \otimes |i\rangle_2 \in \mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2,$$

where the sum is over all basis vectors of $\mathcal{H}_1$ and a suitable subset of arbitrary basis vectors of $\mathcal{H}_2$ such that

$$\mathrm{Tr}_2\left(|\phi\rangle\langle\phi|\right) = \rho.$$

In this way, all states can be purified.

Proof. By tracing out $\mathcal{H}_2$:

$$\begin{aligned}\mathrm{Tr}_2\left(|\phi\rangle\langle\phi|\right) &= \mathrm{Tr}_2\left(\sum_{i,j} \sqrt{\lambda_i\lambda_j}\, |i\rangle_1 \otimes |i\rangle_2 \langle j|_1 \otimes \langle j|_2\right) \\ &= \sum_{i,j} \sqrt{\lambda_i\lambda_j}\, |i\rangle_1 \langle j|_1 \,\delta_{i,j} = \sum_i \lambda_i |i\rangle_1 \langle i|_1\end{aligned}$$

□

In Lemma 1.1.9 we saw that we can choose arbitrary basis vectors of $\mathcal{H}_2$ for the purification. Thus a purification is not unique.

Lemma 1.1.10. *In the same setting as in Lemma 1.1.9, if*

$$|\phi\rangle = \sum_i \sqrt{\lambda_i}\, |i\rangle_1 \otimes |i\rangle_2 \in \mathcal{H}_1 \otimes \mathcal{H}_2$$

is a purification of ρ and $\mathcal{U}_2$ is a unitary transformation on $\mathcal{H}_2$, then

$$|\phi'\rangle = \sum_i \sqrt{\lambda_i}\, |i\rangle_1 \otimes \mathcal{U}_2 |i\rangle_2$$

is also a purification of ρ.

Proof. The trace is invariant under cyclic permutations, which extends to partial traces in the obvious way:

$$\begin{aligned}\mathrm{Tr}_2\left(|\phi'\rangle\langle\phi'|\right) &= \mathrm{Tr}_2\left(\sum_{i,j}\sqrt{\lambda_i\lambda_j}\,|i\rangle_1 \otimes \mathcal{U}_2\,|i\rangle_2\,\langle j|_1 \otimes \langle j|_2\,\mathcal{U}_2^\dagger\right)\\ &= \mathrm{Tr}_2\left(\sum_{i,j}\sqrt{\lambda_i\lambda_j}\,|i\rangle_1 \otimes |i\rangle_2\,\langle j|_1 \otimes \langle j|_2\right).\end{aligned}$$

□

Together, Lemma 1.1.9 and Lemma 1.1.10 show that the auxiliary system needs only the same dimension as the system to be purified, since we make no use of a larger space: we can use $\mathcal{U}_2$ to adjust the order of the basis vectors and then forget about the unused basis vectors.

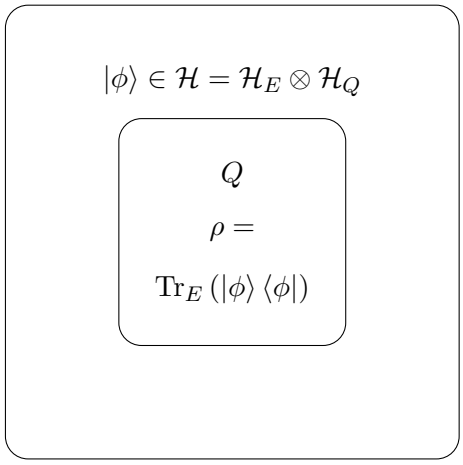


Figure 1.1.: Open Quantum System

In Figure 1.1 we see how we can interpret a mixed state ρ of a system Q as a pure state of a "complete" system where we traced out an environment E about which we do not care. This interpretation is more natural, since we only think of one system. It is only our lack of knowledge (about the environment) that makes it seem to be an ensemble of systems. The system is open in the sense that there can be hidden interactions with the environment. The implications of this will be discussed in the next section.

Definition 1.1.11 (Open Quantum System)**.** *If we know that a quantum system is in a mixed state, it is an open quantum system and as such, part of a larger, closed quantum system.*

One special quantum system is the smallest non-trivial system — the system on a Hilbert space with dimension two.

Definition 1.1.12 (Qubit)**.** *A quantum system where $\mathcal{H} = \mathbb{C}^2$ is called a qubit.*

Since a qubit is the smallest non-trivial system, it is central to quantum information theory. We will focus on this system, and especially on its time evolution.

Another useful tool, and also a neat property of tensor products, is the Schmidt decomposition.

Theorem 1.1.13 (Schmidt Decomposition)**.** *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces, where $\dim \mathcal{H}_1 = n$ and $\dim \mathcal{H}_2 = m$ and let n be larger than m.*
For an arbitrary $|\phi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$, there are sets of orthonormal vectors $\{|1\rangle_1, \ldots, |m\rangle_1\} \in \mathcal{H}_1$, and $\{|1\rangle_2, \ldots, |m\rangle_2\} \in \mathcal{H}_2$, and, up to ordering, a unique set of $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_0^+$, such that

$$|\phi\rangle = \sum_{i=1}^{m} \lambda_i |i\rangle_1 \otimes |i\rangle_2 .$$

As a result of this section we can — as long as we are not restricted to a specific system — choose to work with pure states. Furthermore, if we need an auxiliary system to purify, it will always be sufficient for the auxiliary system to have the same dimension as the original smaller system.

1.2. Distinguishing Quantum States

For pure states there is a very intuitive and natural way of comparing them: the aforementioned inner product. The inner product is equal to one if and only if the states are equal, and zero only when they are orthogonal. Further, it can be used to compute an overlap of two states, which is the probability to measure the first when the system is actually in the second or vice versa. A high probability indicates that the compared states are close.
In contrast, it is a priori not clear how to find out if two general states ρ_1 and ρ_2 are close to each other.
Typically there are only two main concepts [25]. The first, the *trace distance*, focuses on the operator property, and the second, the *fidelity*, comes from a more vector-like interpretation of general states. We will use the latter to construct a symmetric functional and analyse the effect of a quantum channel, which will be introduced in the next section, in later parts of this work, Chapter 3.

Before we can get into discussing the two measures, we need to understand that $\mathcal{L}(\mathcal{H})$ is actually a Hilbert space itself. It is quite obvious that the set of linear operators (or endomorphisms) is a vector space, and we can equip this space with an inner product.

Definition 1.2.1 (Hilbert-Schmidt inner product)**.** *Consider two operators A and B on a Hilbert space $\mathcal{H}$. We call*

$$(A, B)_{HS} = Tr\Big(A^\dagger\ B\Big)$$

their Hilbert-Schmidt inner product.

Proposition 1.2.2. *The Hilbert-Schmidt inner product is an inner product on the space $\mathcal{L}(\mathcal{H})$.*

Proof. It is sesquilinear since the trace is linear and taking the adjoint is conjugation for complex numbers. For operators A, B, C and D, and complex numbers λ and μ, we have:

$$\begin{aligned}(\lambda A + B, \mu C + D)_{HS} &= \mathrm{Tr}\left(\bar{\lambda}\,\mu\, A^\dagger C + \mu\, B^\dagger C + \bar{\lambda}\, A^\dagger\, D + B^\dagger\, D\right)\\ &= \bar{\lambda}\mu\,(A, C)_{HS} + \mu\,(B, C)_{HS}\\ &\quad + \bar{\lambda}\,(A, D)_{HS} + (B, D)_{HS}\,.\end{aligned}$$

It is hermitian:

$$(A, B)_{HS} = \mathrm{Tr}\left(A^\dagger\, B\right) = \overline{\mathrm{Tr}\,(B^\dagger\, A)} = \overline{(B, A)_{HS}}.$$

It is positive semi-definite,

$$(A, A)_{HS} = \mathrm{Tr}\left(A^\dagger\ A\right) \geq 0,$$

since $A^\dagger A$ is a positive operator. □

As an inner product, it satisfies the Cauchy-Schwarz inequality.

Fact 1.2.3 (Cauchy-Schwarz Inequality)**.**

$$(\rho_1, \rho_2)_{HS} \leq \sqrt{(\rho_1, \rho_1)_{HS}\ (\rho_2, \rho_2)_{HS}}$$

1.2.1. Trace Distance

Equipped with the Hilbert-Schmidt inner product, we can get into exploring the trace distance. First we can generalize the notion of an absolute value to operators.

Definition 1.2.4 (Absolute Value of Operators)**.** *Given an arbitrary operator A we call the square root of the inner product with itself its absolute value:*

$$|A| = \sqrt{(A, A)_{HS}} = \sqrt{A^\dagger\ A}.$$

Now the trace distance is defined as the absolute value of the difference of two operators.

Definition 1.2.5 (Trace Distance)**.** *Given two operators A and B, their trace distance $D(A, B)$ is defined as:*

$$D(A, B) = \frac{1}{2} Tr\left(|A - B|\right).$$

The trace distance is a nice measure since it can be used as metric on quantum states ρ. Furthermore, there is a direct physical interpretation.

Example 1.2.6 (Trace Distance)**.** *A physical system is prepared in state ρ_1 with probability $\frac{1}{2}$ and in state ρ_2 also with probability $\frac{1}{2}$.*

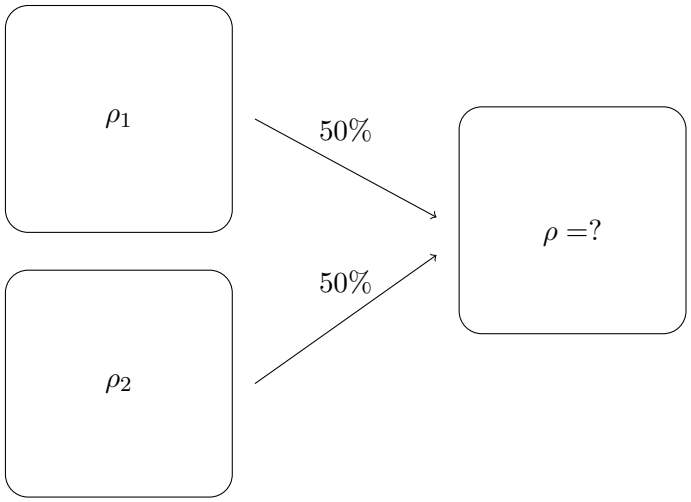


Figure 1.2.: Distinguishability

A measurement aiming to identify which state has been prepared then has the probability $\frac{1}{2} + \frac{D(\rho_1, \rho_2)}{2}$ to succeed [25].

1.2.2. Uhlmann Fidelity

The fidelity will be the basis of the symmetric functional that we will investigate extensively in later parts of this thesis. Hence we will take more time to introduce it.

Let us get back to the overlap of two unit vectors.

Example 1.2.7 (Fidelity for Pure States)**.** *Take two vectors $|\phi_1\rangle, |\phi_2\rangle$ from an arbitrary Hilbert space $\mathcal{H}$. It is clear that their overlap*

$$F(\phi_1, \phi_2) = |\langle\phi_1| |\phi_2\rangle|^2$$

is a good way of identifying whether they are identical, $F = 1$, or orthogonal, $F = 0$. The concrete value for F is the probability to measure ϕ_1 if the system is in ϕ_2 and vice versa. Furthermore it allows the interpretation of an angle between the two.

Now we seek to extend this functional to general states. On first glance the Hilbert Schmidt inner product seems to be a good candidate. Considering that we are usually interested only in density operators, the inner product of two general states would be 1 only if they are identical and 0 if their supports are mutually orthogonal. However we struggle to interpret the area in between. It is certainly possible to define an angle between two operators using the Hilbert-Schmidt inner product but there is a more natural way:

Definition 1.2.8 (Uhlmann Fidelity)**.** *For two operators ρ_1 and ρ_2 on $\mathcal{H}$ we define their fidelity as the maximal overlap their purifications $|\phi_1\rangle$ and $|\phi_2\rangle$ can have:*

$$F(\rho_1, \rho_2) = \max_{\phi_1, \phi_2} |\langle \phi_1 | \, | \phi_2 \rangle|^2.$$

Theorem 1.2.9 (Uhlmann's Theorem)**.** *Maximization is achieved in the following expression:*

$$F(\rho_1, \rho_2) = Tr\left(\sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}} \right)^2 .$$

Though it certainly does not look that way, the fidelity is indeed symmetric in its inputs.

The expression in the theorem was considered before by Bures [5], [3], however the interpretation as a transition probability, an overlap between two quantum states, and the proof of equality Theorem 1.2.9 were first done by Uhlmann [35]. The fidelity expression for two general states is quite difficult to handle. Fortunately, comparing a mixed and a pure state is much easier.

Proposition 1.2.10. *Let ρ be a mixed state and $\phi = |\phi\rangle \langle \phi|$ be pure. Their fidelity is simply:*

$$F(\phi, \rho) = Tr\left(\sqrt{|\phi\rangle \langle \phi| \, \rho \, |\phi\rangle \langle \phi|} \right)^2 = \langle \phi | \, \rho \, | \phi \rangle \, .$$

This is why the fidelity is a more natural measure for comparing quantum states. It allows a direct physical interpretation: Imagine an experimenter trying to prepare ϕ, but he actually prepares ρ. The fidelity describes the probability of success [25].

1.3. Quantum Channel — Completely positive maps

Now we want to transmit quantum information. It is well known that time evolution in the quantum world is governed by unitary evolution. But this is true only for the evolution of closed systems. In this section we will first create an intuition about what we expect a proper quantum channel to be and later describe this mathematically and give a full characterization. The presented matter is standard material for quantum information courses and can be found, for example, in [26] or [22].

A quantum channel will describe the time evolution of an open quantum system. We do not make a mistake if we define it as a mapping from one open quantum system (the input) to another quantum system (the output), however, to be a physical map there will additional restrictions and these restrictions will arise naturally in the physical context.

Definition 1.3.1 (Quantum Channel)**.** *Consider an input density operator ρ_I on the input Hilbert space $\mathcal{H}_I$, a quantum channel $\mathcal{N}$ and an output Hilbert space $\mathcal{H}_O$.*
A quantum channel is a linear map that maps density operators on $\mathcal{H}_I$ to density operators on $\mathcal{H}_O$:

$$\mathcal{N} : \mathcal{L}(\mathcal{H}_I) \to \mathcal{L}(\mathcal{H}_O), \rho_I \mapsto \mathcal{N}(\rho_I).$$

Let us now assume we have an open system Q with quantum information ρ_Q stored in it. We want to understand how the time evolution of the open system Q develops over time.
In quantum mechanics, time evolution of a closed system is described by unitary evolution. An open quantum system is always a part of a larger system. We call the rest of the system the environment E, with density operator ρ_E. The process will be influenced by interactions with the E. However the combined system is closed and has the usual unitary time evolution.

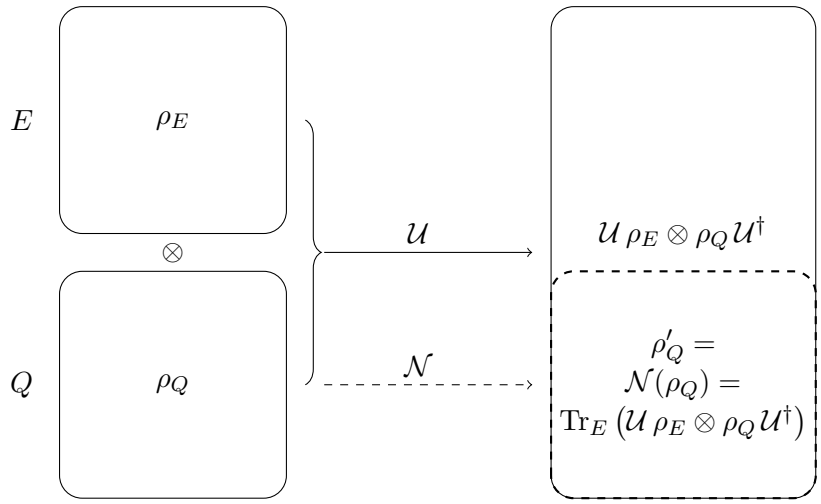


Figure 1.3.: Time evolution of an Open Quantum System

Since in the end we only care about the evolution of Q, we trace out E and get the evolved density operator ρ'_Q. We could easily overlook this as it is obvious that we want the output ρ'_Q to be a proper density operator, but certainly that is a restriction on $\mathcal{N}$: it has to preserve the inputs trace and its positivity.

It is a non-essential restriction that input system and output system should always be the same. Sometimes it is too restrictive, for instance in a computation of a precise question the input is often much more complicated then the answer or in an experiment the setup can require a rather large system, where in the end the experimenter only cares about properties of a small part. Let us explore these more general settings in two examples.

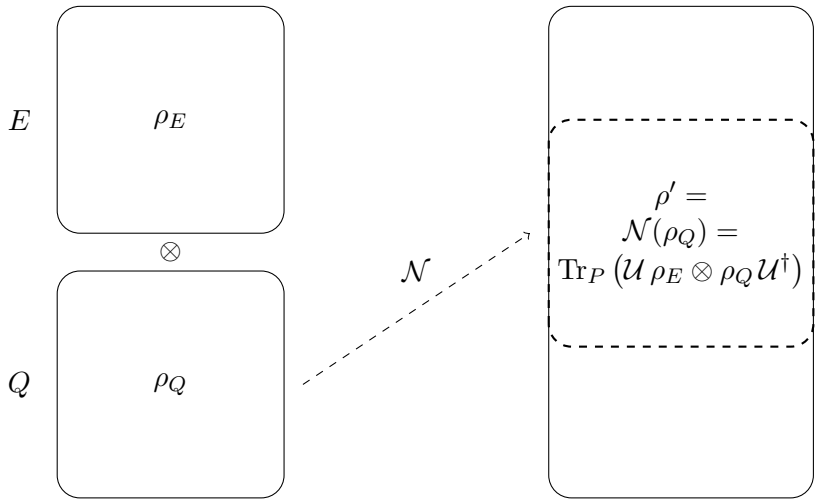


Figure 1.4.: General Quantum Channel

Example 1.3.2 (Grover's Algorithm)**.** *In Grover's algorithm [15] the aim is to find a marked state among a set of orthogonal states. As the input the set of states are stored in one system, the function that marks the state in another system. Over the course of the computation other systems might be added, however in the end, the output is only over the system containing the states and an additional qubit indicating if a state is marked or not. We can interpret the action of the algorithm as a quantum channel where input and output systems are inherently different.*

Example 1.3.3 (Three Part System)**.** *Let us now take a more physical perspective and consider a Hilbert space that is the tensor product of three systems, maybe three qubits, with Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$.*
We can declare any combination of those the input or the output systems. To be clear let us make the example more concrete by choosing $\mathcal{H}_1$ to be the input system Q with Hilbert space $\mathcal{H}_Q$ thus the environment E with Hilbert space $\mathcal{H}_E = \mathcal{H}_2 \otimes \mathcal{H}_3$ and the output P with Hilbert space $\mathcal{H}_P = \mathcal{H}_1 \otimes \mathcal{H}_2$ which makes the new environment $\mathcal{H}_{E'} = \mathcal{H}_3$.

As an overview, and because we will have a similar structure when we construct a quantum channel precisely, we give a list of the different combinations of the three

Hilbert spaces:

$$\begin{aligned}
\mathcal{H} &= \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \\
\mathcal{H}_Q &= \mathcal{H}_1 \\
\mathcal{H}_E &= \mathcal{H}_2 \otimes \mathcal{H}_3 \\
\mathcal{H}_P &= \mathcal{H}_1 \otimes \mathcal{H}_2 \\
\mathcal{H}_{E'} &= \mathcal{H}_3 \\
\mathcal{H} &= \mathcal{H}_Q \otimes \mathcal{H}_E, \text{ and} \\
\mathcal{H} &= \mathcal{H}_P \otimes \mathcal{H}_{E'}.
\end{aligned}$$

Each of these combinations can be a valid interpretation of $\mathcal{H}$. Certainly there are even more complicated tensor product structures possible. By tracing out E' we map quantum information from system Q to system P, as illustrated in Figure 1.4.

In the examples, we have seen that a general quantum channel demands the possibility to send quantum information from one system to another system. We will see that this is not just possible but arises naturally in the physical construction.

Before we get into the mathematical characterization of a quantum channel, we expect another general property: physical extensibility.

Definition 1.3.4 (Extensibility)**.** *Consider a Hilbert space that is a tensor product $\mathcal{H} = \mathcal{H}_Q \otimes \mathcal{H}_A$ and a quantum channel $\mathcal{N}_Q$ that is defined for density operators on $\mathcal{H}_Q$ and another quantum channel $\mathcal{N}_A$ that is defined for density operators on $\mathcal{H}_A$. If for all $\mathcal{H}_A$ and $\mathcal{N}(A)$ the combined quantum channel $\mathcal{N}_Q \otimes \mathcal{N}_A$ is still a proper quantum channel, in the sense that it maps a density operator ρ on $\mathcal{H}$ to a valid density operator $\mathcal{N}_Q \otimes \mathcal{N}_A(\rho)$ on $\mathcal{H}$, $\mathcal{N}_Q$ is called extensible.*

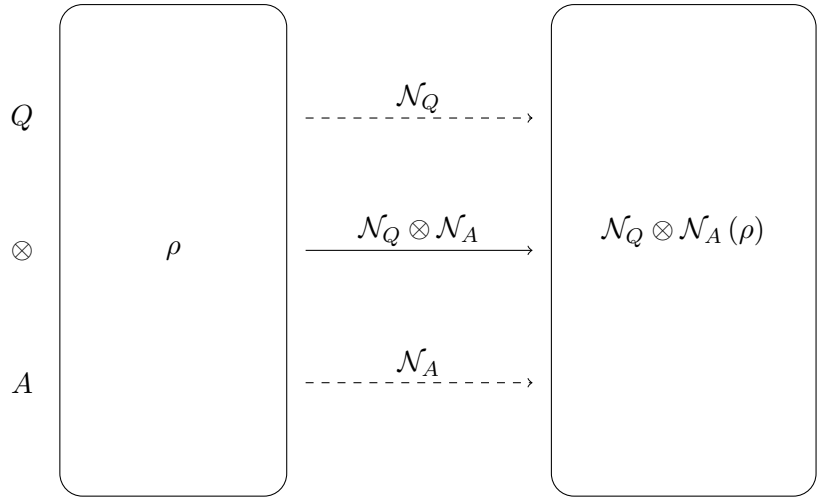


Figure 1.5.: Extensibility

Note that the combined system $\mathcal{H}_Q \otimes \mathcal{H}_A$ is in general still an open quantum system.

We require extensibility, because we want to be free in combining quantum channel with each other without any further restriction. To understand why extensibility is an issue we should formalize our thoughts.

We already mentioned that the essential properties of a density operator, that they are semi definite positive operators with trace one, have to be preserved. It follows that the set of quantum channels is clearly not equivalent to the set of linear maps of operators to operators. We will now see that extensibility shrinks the set further.

Example 1.3.5. *For operators on $\mathbb{C}^2$ consider the transposition map T.*

$$T : \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2), A \mapsto A^T,$$

which clearly is trace preserving and also positivity preserving since A and A^T have the same eigenvalues.

Now we extend the map trivially — the trivial map is obviously positivity and trace preserving:

$$\mathbb{1} \otimes T : \mathcal{L}(\mathbb{C}^2) \otimes \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2) \otimes \mathcal{L}(\mathbb{C}^2), A \mapsto \mathbb{1} \otimes T(A).$$

Now we take a positive semi-definite operator that is not normalized, which makes things easier to read.

$$\begin{aligned} A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Apply the extended maps:

$$\begin{aligned}\mathbb{1} \otimes T(A) = &\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},\end{aligned}$$

and get an operator that has negative determinant:

$$\det\left(\mathbb{1} \otimes T(A)\right) = -1.$$

Thus we see that the extended map is not positivity preserving.

We need to distinguish maps that will, even with extension, preserve positivity, from those that do not. This concept is called complete positivity.

Definition 1.3.6 (Complete Positivity)**.** *Consider $\mathcal{L}(\mathcal{H})$ for $\mathcal{H}$ and an auxiliary system of arbitrary size $\mathcal{H}_A$. A map $\mathcal{N}$ is called completely positive if it not only maps positive (semi-definite) maps to positive (semi-definite) maps but also all extensions of $\mathcal{N}$ as $\mathbb{1}_A \otimes \mathcal{N}$ conserve positivity.*

The next proposition follows immediately.

Proposition 1.3.7. *Complete positivity is necessary and sufficient for physical extensibility.*

Proof. The necessity is obvious. For the sufficiency consider four systems $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$ and $\mathcal{H}_D$ and two quantum channel $\mathcal{N}_1$ and $\mathcal{N}_2$ with:

$$\begin{aligned}\mathcal{N}_1 : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B), \text{ and} \\ \mathcal{N}_2 : \mathcal{L}(\mathcal{H}_C) \to \mathcal{L}(\mathcal{H}_D).\end{aligned}$$

Then the combined map $\mathcal{N}_1 \otimes \mathcal{N}_2$,

$$\mathcal{N}_1 \otimes \mathcal{N}_2 : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_C) \to \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_D),$$

is positivity preserving, since we can write the map as a composition of completely positive maps:

$$\mathcal{N}_1 \otimes \mathcal{N}_2 = (\mathcal{N}_1 \otimes \mathbb{1}_A) \circ (\mathbb{1}_D \otimes \mathcal{N}_2) = (\mathbb{1}_B \otimes \mathcal{N}_2) \circ (\mathcal{N}_1 \otimes \mathbb{1}_C).$$

The input and output systems are not necessarily equivalent. We have to be careful that the tensors actually act on the correct spaces. □

We have seen that a realistic physical quantum channel has to meet certain requirements which we conclude in the next proposition.

Proposition 1.3.8. *A quantum channel $\mathcal{N}$ is trace preserving and completely positive.*

We will get a handy standard form for quantum channels once we have a more detailed look at the construction.

As in Figure 1.4, consider a system Q in state ρ_Q and an environment E in a pure state $\Phi = |\phi\rangle\langle\phi|$. We can assume the environment to be in a pure state because otherwise it could easily be purified. Furthermore, keep in mind that we have an underlying structure as in Example 1.3.3.

The full input density operator is given by:

$$\rho = \rho_Q \otimes \Phi.$$

We want to trace out an arbitrary environment E' of the whole system.

$$\mathcal{N}(\rho_Q) = \sum_{i=1}^{\dim E'} \mathbb{1}_P \overset{P,E'}{\otimes} \langle i| \left[\mathcal{U}\left(\rho_Q \overset{Q,E}{\otimes} \phi\right)\mathcal{U}^\dagger\right] \mathbb{1}_P \overset{P,E'}{\otimes} |i\rangle$$

Note that the tensor products are not necessarily combining the same spaces as indicated by the superscripts.

Now we use that the environment is in a pure state.

$$\begin{aligned}\mathcal{N}(\rho_Q) &= \sum_{i=1}^{\dim E'} \left(\mathbb{1}_P \overset{P,E'}{\otimes} \langle i|\right)\left[\mathcal{U}\left(\rho_Q \overset{Q,E}{\otimes} |\phi\rangle\langle\phi|\right)\mathcal{U}^\dagger\right]\left(\mathbb{1}_P \overset{P,E'}{\otimes} |i\rangle\right) \\ &= \sum_{i=1}^{\dim E'} \left(\mathbb{1}_P \overset{P,E'}{\otimes} \langle i|\right)\left[\mathcal{U}\left(\mathbb{1}_Q \overset{Q,E}{\otimes} |\phi\rangle\right)\rho_Q\left(\mathbb{1}_Q \overset{Q,E}{\otimes} \langle\phi|\right)\mathcal{U}^\dagger\right]\left(\mathbb{1}_P \overset{P,E'}{\otimes} |i\rangle\right)\end{aligned}$$

The second step looks a bit surprising. This is a weakness of the bra-ket notation. We are tempted to naively write a $\mathbb{1}$ beside ρ_Q. We can clear the mess up by precisely looking at the mappings.

Consider:

$$\rho_Q \overset{Q,E}{\otimes} |\phi\rangle\langle\phi| : \mathcal{H}_Q \otimes \mathcal{H}_E \to \mathcal{H}_Q \otimes \mathcal{H}_E, \text{ and}$$
$$\rho_Q : \mathcal{H}_Q \to \mathcal{H}_Q,$$

and define:

$$W_\phi := \mathbb{1}_Q \overset{Q,E}{\otimes} |\phi\rangle : \mathcal{H}_Q \to \mathcal{H}_Q \otimes \mathcal{H}_E \text{ and}$$
$$W_\phi^\dagger := \mathbb{1}_Q \overset{Q,E}{\otimes} \langle\phi| : \mathcal{H}_Q \otimes \mathcal{H}_E \to \mathcal{H}_Q.$$

We see that only $W_\phi \rho_Q W_\phi^\dagger = \rho_Q \overset{Q,E}{\otimes} |\phi\rangle\langle\phi|$ is well defined.

A hypothetical $W_\phi \rho_Q \overset{Q,E}{\otimes} \mathbb{1}_E W_\phi^\dagger =?$ would not make sense.

Finally we define the operators:

$$A_i = \left(\mathbb{1}_P \overset{P,E'}{\otimes} \langle i| \right) \mathcal{U} W_\phi, \tag{1.1}$$

which simplifies the quantum channel to:

$$\mathcal{N}(\rho_Q) = \sum_{i=1}^{\dim E'} A_i \rho_Q A_i^\dagger.$$

The latter is called the Kraus representation or operator sum representation of a quantum channel.

Definition 1.3.9 (Kraus Representation)**.** *A completely positive map $\mathcal{N}$ has a Kraus representation or is in Kraus form, if there exist operators A_i such that:*

$$\mathcal{N}(\rho) = \sum_i A_i \rho A_i^\dagger.$$

The A_i are called Kraus operators.

Theorem 1.3.10 (Quantum Channel)**.** *For a completely positive map $\mathcal{N}$ the existence of a Kraus form is necessary and sufficient. The map $\mathcal{N}$ with Kraus operators A_i is also trace preserving iff*

$$\sum_i A_i^\dagger A_i = \mathbb{1}.$$

A proof for this theorem can be found in [21, 7].

It is easy to show that the quantum channel we constructed, is trace preserving.

Proposition 1.3.11. *For Kraus operators defined as in Equation 1.1 we have:*

$$\sum_i A_i^\dagger A_i = \mathbb{1}_Q$$

and furthermore

$$Tr(\mathcal{N}(\rho_Q)) = Tr\left(\sum_i A_i \rho_Q A_i^\dagger \right) = Tr\,(\rho_Q).$$

Proof. The trick is basically to pull the sum through, so that $|i\rangle\langle i|$ acts as an identity map:

$$\begin{aligned}\sum_i A_i^\dagger A_i &= \sum_i \left(\mathbb{1}_Q \overset{Q,E}{\otimes} \langle\phi|\right) \mathcal{U}^\dagger \left(\mathbb{1}_P \overset{P,E'}{\otimes} |i\rangle\right) \left(\mathbb{1}_P \overset{P,E'}{\otimes} \langle i|\right) \mathcal{U} \left(\mathbb{1}_Q \overset{Q,E}{\otimes} |\phi\rangle\right) \\ &= \left(\mathbb{1}_Q \overset{Q,E}{\otimes} \langle\phi|\right) \mathcal{U}^\dagger \sum_i \left(\mathbb{1}_P \overset{P,E'}{\otimes} |i\rangle\langle i|\right) \mathcal{U} \left(\mathbb{1}_Q \overset{Q,E}{\otimes} |\phi\rangle\right) \\ &= \left(\mathbb{1}_Q \overset{Q,E}{\otimes} \langle\phi|\right) \mathcal{U}^\dagger \mathcal{U} \left(\mathbb{1}_Q \overset{Q,E}{\otimes} |\phi\rangle\right) \\ &= \mathbb{1}_Q.\end{aligned}$$

Then using the cyclic invariance and linearity of the trace, the second statement follows quickly:

$$\begin{aligned}\mathrm{Tr}\left(\mathcal{N}(\rho_Q)\right) &= \mathrm{Tr}\left(\sum_{i=1,\dim E'} A_i\, \rho_Q\, A_i^\dagger\right) = \sum_{i=1,\dim E'} \mathrm{Tr}\left(A_i^\dagger A_i \rho_Q\right) \\ &= \mathrm{Tr}\left(\mathbb{1}_Q \rho_Q\right) = \mathrm{Tr}\left(\rho_Q\right).\end{aligned}$$

□

In the next section we will introduce a class of quantum channel that will be relevant in later parts of this work. For now we want to give a few examples.

Example 1.3.12 (Amplitude-Damping Channel [26])**.** *A very intuitive interaction with the environment is energy leakage. One system which we can model is the emission of a photon [26]. Consider a two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$ and the system in the pure state of a superposition of none and a single photon: $|\phi\rangle = a\,|0\rangle + b\,|1\rangle$.*

The density operator in the $|0\rangle$,$|1\rangle$ basis is then:

$$\rho = |\phi\rangle\langle\phi| = \begin{pmatrix} a\bar{a} & a\bar{b} \\ b\bar{a} & b\bar{b} \end{pmatrix}.$$

The amplitude damping channel $\mathcal{N}_\gamma$ has Kraus operators

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \quad \textit{and} \quad A_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix},$$

which describe a process that lowers the probability of the system to be in state $|1\rangle$, as

it transports the system to

$$\mathcal{N}_\gamma(\rho) = \begin{pmatrix} a\bar{a} & a\bar{b}\sqrt{1-\gamma} \\ b\bar{a}\sqrt{1-\gamma} & b\bar{b}(1-\gamma) \end{pmatrix} + \begin{pmatrix} b\bar{b}\gamma & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a\bar{a}+\gamma b\bar{b} & a\bar{b}\sqrt{1-\gamma} \\ b\bar{a}\sqrt{1-\gamma} & b\bar{b}(1-\gamma) \end{pmatrix}.$$

We see that the probability to be in the zero photon state is increased by $\gamma\,|b|^2$. We can think of γ as the probability of leaking a photon to the environment.

Note that for $\gamma = 1$ we get, for arbitrary input density operators, the pure output state $|0\rangle\langle 0|$.

Example 1.3.13 (Von-Neumann Measurement [26])**.** *Consider a system Q in state ρ_Q. We will now explain how a measurement of this state can be interpreted as a quantum operation $\mathcal{N}_M$.*

An observable O has to be the weighted sum of projection operators P_m, where the P_m are mutually orthogonal projectors i.e. O is a self-adjoint operator:

$$O = \sum_m o_m P_m.$$

According to Born's rule, an outcome o_m is measured with probability p_m:

$$p_m = Tr\left(P_m\,\rho_Q\right).$$

The measured state ρ_m of the system is then:

$$\rho_m = \frac{P_m \rho_Q P_m}{p_m}.$$

Effectively we can see the measurement process as the mapping with the projectors as Kraus operators:

$$\mathcal{N}_M(\rho_Q) = \sum_m p_m \frac{P_m \rho_Q P_m}{p_m} = \sum_m P_m \rho_Q P_m.$$

Note that since the P_m are projectors they are self-adjoint, square to themselves and sum up to the identity:

$$\sum_m P_m P_m^\dagger = \sum_m P_m P_m = \sum_m P_m = \mathbb{1},$$

and hence fulfill the requirements of for a quantum channel.

Definition 1.3.14 (Complementary Channel [14, 36])**.** *We have seen that a quantum channel can be interpreted as unitary time evolution of a system Q and a suitable environment E, where the environment's degrees of freedom are traced out. We can exchange the interpretation of Q and E for the output and construct a map from the operators on Q to the operators on E by tracing out Q instead of E. This map is certainly closely related to $\mathcal{N}$ and is called the complementary channel $\mathcal{N}_C$.*

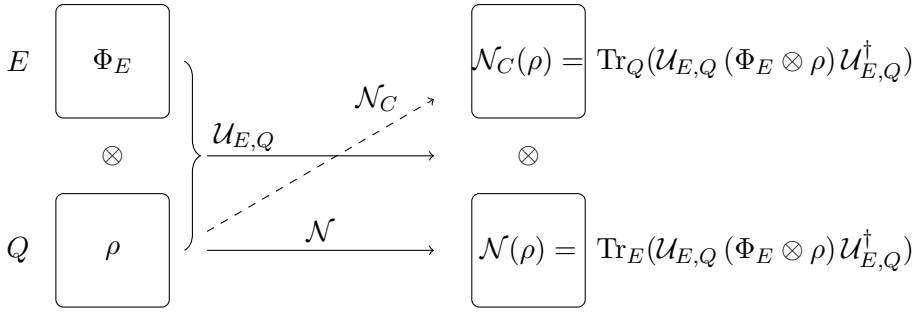
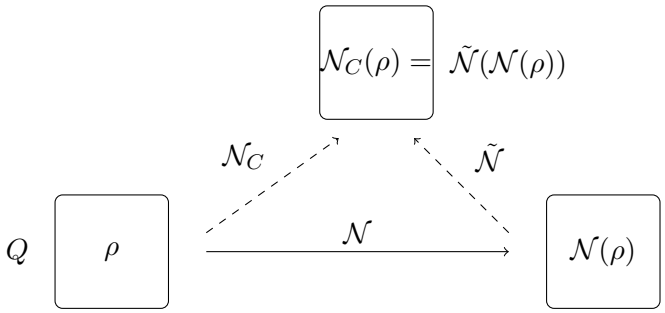


Figure 1.6.: Complementary Channel

The complementary channel is essentially unique [24].

Example 1.3.15 (Degradable Channel [14, 36, 24])**.** *An interesting class of channels are those that can be degraded to their complementary channel. That means, that there is another quantum channel $\tilde{\mathcal{N}}$ such that*

$$\mathcal{N}_C(\rho) = \tilde{\mathcal{N}}(\mathcal{N}(\rho)).$$



If the complementary channel is degradable, $\mathcal{N}$ is called anti-degradable.

1.4. Unital Qubit Channel

In our characterization of quantum channels, we ignored that we actually have the freedom to choose the bases for input and output system. Let us consider now that in

an experimental setting we have a quantum channel with Kraus operators A_i. Then at the input side we can choose a specific input basis by using a unitary operator W and also on the output side a specific output basis by applying a unitary operator V. This will allow us to transform the channel:

$$\mathcal{N}(\rho) = \sum_i A_i \rho A_i^\dagger \mapsto \mathcal{N}'(\rho) = \sum_i V A_i W \rho W^\dagger A_i^\dagger V^\dagger = \sum_i A'_i \rho A_i'^\dagger.$$

The channels $\mathcal{N}$ and $\mathcal{N}'$ are closely related, since their outputs are unitarily similar. Still the freedom to choose input and output bases can come in handy for specific channels. A very nice example is the unital qubit channel.

Definition 1.4.1. *A quantum channel $\mathcal{N}$ is called **unital** iff it maps the identity of the input system to the identity of the output system, i.e.*

$$\mathcal{N}(\mathbb{1}) = \sum_i A_i \mathbb{1} A_i^\dagger = \mathbb{1}.$$

Corollary 1.4.2. *The adjoint map $\mathcal{N}^\dagger$ of a unital channel $\mathcal{N}$ is also trace preserving and therefore a quantum channel.*

Proof. Let $\{A_i\}_i$ be the Kraus operators of $\mathcal{N}$. Then the Kraus operators of $\mathcal{N}^\dagger$ are by definition $B_i \equiv A_i^\dagger$. Furthermore, since $\mathcal{N}$ is unital we have:

$$\mathbb{1} = \sum_i A_i \mathbb{1} A_i^\dagger = \sum_i B_i^\dagger B_i.$$

□

Theorem 1.4.3 (Pauli Channel)**.** *For unital qubit channel $\mathcal{N}$, where $\dim Q = 2$, there exists a canonical form [2], $\mathcal{N}_{\{p_i\}_{i=0}^3}$:*

$$\mathcal{N}_{\{p_i\}_{i=0}^3}(\rho) = \sum_{i=0}^3 p_i \sigma_i \rho \sigma_i, \quad \sum_i p_i = 1,$$

with the Pauli operators σ_i and the identity $\sigma_0 = \mathbb{1}_2$. The standard Kraus form can be obtained with $A_i = \sqrt{p_i}\sigma_i$. A unital qubit channel that is in canonical form is called a Pauli channel, $\mathcal{N}_{\{p_i\}_{i=0}^3}$, with probabilities, $\{p_i\}$.

The proof is rather involved, and furthermore the simple structure is special for $d = 2$. Unital channels in higher dimension can still be written as an affine combination of unitaries, however these do not have the nice structure that only one operator has a trace different to 0 [23, 10].
We will focus on qubit channels. Whenever we have the choice of input and output bases we will use the canonical form.

One special Pauli channel that is easy to understand is the bit flip channel. It has only one parameter. We will sometimes use it for plots.

Example 1.4.4 (Bit Flip Channel)**.** *We call the Pauli channel where $p_2 = p_3 = 0$, $p_0 = p$ and $p_1 = (1-p)$ the bit flip channel $\mathcal{N}_B$, as it flips the bit with probability $(1-p)$.*

However, since the bit flip channel is degradable, [14], it is already well understood. We will see what this means in the next section.

Depolarizing Channel

As we want to plot functionals that depend on error probabilities, it will be useful to only have a single parameter. A special quantum channel with this property is the depolarizing channel $\mathcal{N}_D$. It maps an input state to a linear combination of the input and the maximally mixed state.

Definition 1.4.5 (Depolarizing Channel)**.** *The depolarizing channel is defined by the action:*

$$\mathcal{N}_\lambda^D(\rho) = \lambda\rho + \frac{1-\lambda}{d}\mathbb{1}_d.$$

It is non-trivial but well known [26, 3] that for $d = 2$, the depolarizing channel can be written as a Pauli channel.

Proposition 1.4.6 (Depolarizing Channel)**.** *The depolarizing channel as a Pauli channel with parameter p is given by:*

$$\mathcal{N}_p(\rho) = \mathcal{N}_{\{p,\frac{1-p}{3},\frac{1-p}{3},\frac{1-p}{3}\}}(\rho) = p\,\sigma_0\rho\sigma_0 + \frac{(1-p)}{3}\sum_{i=1}^{3}\sigma_i\rho\sigma_i.$$

We will prefer the interpretation as a Pauli channel, since most of our results apply only to such channels. The depolarizing channel is not degradable, which makes it more interesting to study than the bit flip channel.

A special case of the depolarizing channel is the channel that maps all inputs to the completely mixed state, which is equivalent to $\lambda = 0$ in Definition 1.4.5.

Definition 1.4.7 (Completely Depolarizing Channel)**.** *The completely depolarizing channel is the depolarization channel with $\lambda = 0$, which can be expressed as a Pauli channel as:*

$$\mathcal{N}_{\frac{3}{4}}(\rho) = \sum_{i=0}^{3}\frac{1}{4}\sigma_i\rho\sigma_i = \frac{\mathbb{1}}{2}.$$

1.5. Channel Capacity: a Challenge in Quantum Information Theory

Now that we have introduced the notion of quantum information transmission the next logical step is to characterize it.

The name quantum channel directly suggests that it will be used to send quantum information and thus it is a an obvious question, how much information can be reliably transmitted.

Definition 1.5.1 (Quantum Capacity of a Quantum Channel)**.** *The quantum capacity $\mathcal{Q}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is its ability to transmit quantum information. It is measured in qubits per channel use.*

This question has been partially answered by the quantum Shannon theorem.

Theorem 1.5.2 (Quantum Shannon)**.** *For a quantum channel $\mathcal{N}$ its quantum capacity is the maximal regularized coherent information in the limit of $n \to \infty$ channel,*

$$\mathcal{Q}(\mathcal{N}) = \lim_{n\to\infty} I_n(\mathcal{N}).$$

The proof for this theorem is rather involved [17, 19] and we will not present it here, however we will try to create an intuition for it and in the end compare it to the classical capacity of a classical channel.

In some sense information that is stored in a system is the opposite of the uncertainty about the system. Therefore its information content can be measured by its von-Neumann entropy.

Definition 1.5.3 (Von-Neumann Entropy)**.** *For a density operator ρ the von-Neumann entropy $S(\rho)$ is given by:*

$$S(\rho) = -Tr\left(\rho \log_2 \rho\right).$$

With this measure of information, we can define a measure for the information a quantum channel can transport, the coherent information.

Definition 1.5.4 (Coherent Information)**.** *For quantum information stored in a density operator ρ and a quantum channel $\mathcal{N}$ that can transport ρ we define their coherent information $I(\rho, \mathcal{N})$ as:*

$$I(\rho, \mathcal{N}) = S(\mathcal{N}(\rho)) - S_e(\rho, \mathcal{N}).$$

The coherent information is the information of the output $S(\mathcal{N}(\rho))$ minus the information lost to the environment $S_e(\rho, \mathcal{N})$, called the entropy exchange, and to be defined shortly. Before we get into explaining the latter, we have to understand that in the quantum world entanglement effects can occur, and these make it possible that multiple copies of the same channel can actually transport more information per channel than a single channel, [9, 28, 13], thus we have to consider the coherent information for n channels, but per channel.

Definition 1.5.5 (Regularized Coherent Information)**.**

$$I_n(\mathcal{N}) = \frac{1}{n} \max_{\rho} I(\rho, \mathcal{N}^{\otimes n})$$

If the regularized coherent information is equal to the coherent information, the capacity can be calculated easily. This is the case for example for degradable channels, [13], [36]. This is why the bit flip channel, Example 1.4.4, is not at the center of our attention, whereas the depolarizing channel, Proposition 1.4.6, is.
However if the regularized coherent information is not equal to the coherent information, we can understand why the question is only partially answered. Since the dimension of $\mathcal{H}$ or ρ for that matter grows exponentially with the number of channels, a maximization for only $n = 10$ would already be very difficult and that is not even close to a limit of $n \to \infty$.
Ideally one would find an explicit n-dependent expression for I_n, such that one can analytically maximize it and then take the limit of $n \to \infty$. These expressions are called single letter formulas.

Definition 1.5.6 (Single Letter Formula)**.** *For a functional F over a Hilbert space that can naturally be extended to a tensor product of n Hilbert spaces, an explicit, n-dependent expression is called a single letter formula.*

1.5.1. Entropy Exchange

We will now motivate the concept of entropy exchange, $S_e(\rho, \mathcal{N})$.

The following small proposition is essential.

Proposition 1.5.7 (Entropy of Partial Traces)**.** *Consider two density operators ρ_1 on $\mathcal{H}_1$ and ρ_2 on $\mathcal{H}_2$ that can be purified to the same state $\Psi = |\psi\rangle\langle\psi|$:*

$$\mathrm{Tr}_1(\Psi) = \rho_2, \quad \textit{and} \quad \mathrm{Tr}_2(\Psi) = \rho_1.$$

Then ρ_1 and ρ_2 have the same entropy.

Proof. Since $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is a state from a product space, it has a Schmidt decomposition, Theorem 1.1.13, $|\psi\rangle = \sum_i \sqrt{p_i}\, |i\rangle \otimes |i\rangle$, which makes it easy to calculate:

$$S(\rho_1) = S(\rho_2) = \sum_i p_i \log_2(p_i).$$

□

We will now see that a quantum channel can lose information to the environment; it can exchange information between the open system and the environment.

Definition 1.5.8 (Entropy Exchange)**.** *Consider a quantum system Q with environment E and an auxiliary space R such that an operator ρ on $\mathcal{H}_Q$ can be purified to a state $\Psi_{Q,R} = |\psi\rangle\langle\psi|$ with $|\psi\rangle \in \mathcal{H}_Q \otimes \mathcal{H}_R$. Furthermore $\mathcal{N}$ shall be a quantum channel with unitary evolution over $\mathcal{H}_E \otimes \mathcal{H}_Q$. The entropy exchange S_e between Q and E is given by:*

$$S_e(\rho, \mathcal{N}) = S\left([\mathbb{1}_R \otimes \mathcal{N}](\Psi_{Q,R})\right).$$

We will explain this definition further in the following example.

Example 1.5.9 (Entropy Exchange)**.** *Consider a situation as in the above definition or the following picture.*

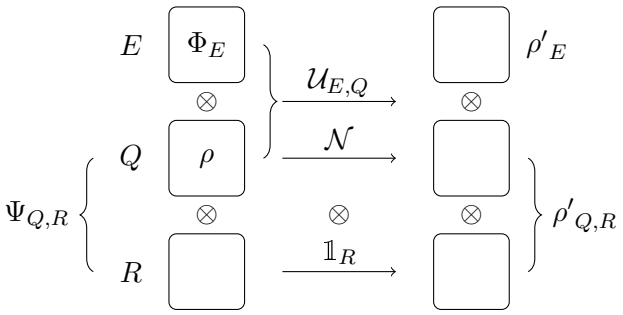


Figure 1.7.: Entropy Exchange

Quantum information ρ is stored in system Q. ρ is purified to $\Psi_{Q,R}$ using the auxiliary system R. A quantum channel $\mathcal{N}$ describes a time evolution of ρ with an underlying unitary interaction $\mathcal{U}_{E,Q}$ between Q and an environment E. In the beginning the state of the environment is the pure state Φ_E. Since the auxiliary system R is not involved in the unitary interaction its time evolution is given as the identity $\mathbb{1}_R$. Considering all of this, the full input state is $\rho_{E,Q,R} = \Phi_E \otimes \Psi_{Q,R}$ and the full time evolution can be written in the following way:

$$[\mathcal{U}_{E,Q} \otimes \mathbb{1}_R](\rho_{E,Q,R}) = \rho'_{E,Q,R}.$$

Note that both $\rho_{E,Q,R}$ and $\rho'_{E,Q,R}$ are pure states and thus their entropies are zero. The same is true for the environment alone.
Now we take the partial trace to get output operators for E and $Q \otimes R$ alone:

$$\begin{aligned}\rho'_E &= \mathrm{Tr}_{Q,R}(\rho'_{E,Q,R}),\\ \rho'_{Q,R} &= \mathrm{Tr}_E(\rho'_{E,Q,R}) = [\mathbb{1}_R \otimes \mathcal{N}](\Psi_{Q,R}).\end{aligned}$$

These are in general mixed states with the same entropy Proposition 1.5.7: This means that the entropy of E has changed from zero to a finite value. This explains the name entropy exchange: Entropy has been exchanged between the combined system $Q \otimes R$ and the environment E.

1.5.2. Comparison to Classical Capacity

The following is based on Shannon's original paper, [27].
The information content of a random variable is measured by its Shannon entropy.

Definition 1.5.10 (Shannon Entropy)**.** *For a random variable X with possible values $x_1, \ldots, x_n$ and their probabilities $p(x_i)$, its Shannon Entropy is*

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2(p(x_i)).$$

Since we seek to compare two random variables, namely the output of a channel given a certain input, we need the concept of conditional entropy.

Definition 1.5.11 (Conditional Entropy)**.** *For two random variables X with possible values $x_1, \ldots, x_n$ and their probabilities $p(x_i)$ and Y with possible values $y_1, \ldots, y_m$ and their probabilities $p(y_i)$ as in Definition 1.5.10 the conditional entropy of Y given X is*

$$H(Y|X) = \sum_{i=1}^{n} p(x_i) H(Y|X = x) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \left(\frac{p(x_i, y_j)}{p(x_i)} \right).$$

With this tool we can describe how much two probability distributions have in common.

Definition 1.5.12 (Mutual Information)**.** *For two random variables X and Y their mutual information is given by:*

$$I(X, Y) = H(Y) - H(Y|X).$$

Similar to the coherent information, Definition 1.5.4, the mutual information is the information content of the output minus a relation between output and input.

It is, however, rather difficult to completely motivate these statistical measures. Mutual information and coherent information as concepts have to show and have shown their usefulness in the characterization of (quantum) information transmission [1] and especially (quantum) channel capacities.
The validity of mutual information has been shown by [27], as it provides a formula to obtain the capacity of a classical channel.

Theorem 1.5.13 (Shannon)**.** *For a classical channel N that maps a random variable X to another random variable $N(X)$, its classical capacity $C(N)$ is the maximum over all possible inputs,*

$$C(N) = \max_X I(X, N(X)).$$

On first glance the two capacities look very similar, especially because they characterize the capacity of a channel as the capacity for an optimal input. The big difference is that for a feasible input system, the classical capacity can easily be computed numerically, whereas the quantum capacity is much more difficult to handle; even a small input system leads to an unsolvable optimization problem over a vast space.

1.5.3. Subjects for this Thesis

Finding the general capacity of a quantum channel, especially for most Pauli channels, is still an open question. Ultimately we are interested in a solution to this problem, however we do not have a direct approach and will be content with studying the effect of high tensor powers of quantum channels more generally.

We will investigate a quantity, the channel fidelity, that is much simpler than the coherent information, in the sense that it is much easier to handle mathematically, however complicated enough such that it will show entanglement dependence. For us the channel fidelity only allows pure inputs to a quantum channel and then compares this input to the output of the quantum channel. A more precise definition will be found in the beginning of Chapter 3.
We begin our investigation by computing moments of the channel fidelity distribution, in particular the average and the variance. We will even find single letter formulas (see Definition 1.5.6) for these. This will allow us to study the effect of multiple quantum channels on the channel fidelity distribution. Later, we analyze the structures that lead to single letter formulas. After explaining the structures we can use them to find single letter formulas for higher moments, under certain restrictions. In the end we will compare our insights about the distribution with results of previous work, where we used an iterative algorithm to find maximizing states.

For the moment calculation we will need additional mathematical tools. We will introduce these in Chapter 2.

2. Integration with Respect to the Haar Measure on Unitary Groups

When studying a physical system, it is often useful to find the behavior of an average state. Without defining a measure it is not clear what an average state is.
If there are no constraints — that is, if no symmetry is broken — the physical states are the normalized vectors of a Hilbert space $\mathcal{H} = \mathbb{C}^n$. The set of those states is often called the unit sphere, S^{n-1},

$$S^{n-1} = \{|\phi\rangle \in \mathcal{H} = \mathbb{C}^n, \||\phi\rangle\| = 1\}.$$

As a manifold S^{n-1} is rather tedious to parametrize, especially for large n. Instead we can view S^{n-1} as the set generated by the unitary group $\mathcal{U}(n)$,

$$S^{n-1} = \mathcal{U}(n)\,|\phi_0\rangle\,,$$

with an arbitrary $|\phi_0\rangle \in \mathbb{C}^n$ with length one. Naturally, on the unitary group $\mathcal{U}(n)$, we have the unitarily invariant Haar measure $d\mathcal{U}$.
Now given a functional F over $\mathcal{H}$ we want to find an average value of F by averaging over all physical states. Our established interpretation of S^{n-1} together with the Haar measure allow us to calculate this average:

$$\int_{\text{physical states}} d\phi\, F(|\phi\rangle) = \int_{S^{n-1}} d\phi\, F(|\phi\rangle) = \int_{\mathcal{U}(n)} d\mathcal{U}\, F(\mathcal{U}\,|\phi_0\rangle).$$

Integration over S^{n-1} as a manifold involves complicated angle parametrization using high dimensional sphere elements, whereas integration over the unitary group is solved in general by Collins and Śniady [8]. Their method relies on the famous Schur-Weyl duality, a relation between the representations of general linear and symmetric groups. Unfortunately, the typical theoretical physicist will not be familiar with all the necessary mathematical background to use their result. The aim of this chapter is to explain these concepts in a direct-to-use way.
We will begin by discussing symmetric groups, especially cycles, on a basic level, but we also define the dimension of a permutation — a concept that will be essential later. Next we review some group theoretical concepts, especially orbits and symmetry groups.
Afterwards we give an introduction to the necessary concepts of representation theory and especially Schur-Weyl duality. Then we will learn how to characterize the representations of symmetric groups. Eventually we will give concrete examples as a guideline for Collins' and Śniady's formula.

Finally we derive a simplified version of the general integration formula, well suited for investigating the channel fidelity in the next chapter. To achieve an efficient notation we define the notion of a trace product according to a permutation.

2.1. Some Group Theory

The following basic group theoretical concepts were taken from [30] but can be found in any introduction to group theory or algebra text book.

2.1.1. On Permutations and Cycles

It will be crucial for this work to gain an understanding of the symmetric group. The symmetric group S_n is the set of permutations of a set of n elements. There are different notations characterizing elements and structures of S_n, all of which correspond to useful interpretations. We will show how cycles and permutations are related.

Definition 2.1.1 (Permutation)**.** *A permutation is a bijective function from* $\{1, \ldots, n\}$ *to itself.*

Example 2.1.2. *Consider a permutation* $\sigma_1 : \{1,2,3,4\} \to \{1,2,3,4\}$ *defined*

$$\sigma_1(1) = 1, \quad \sigma_1(2) = 3, \quad \sigma_1(3) = 2, \quad \text{and} \quad \sigma_1(4) = 4.$$

We can write this as:

$$\sigma_1 = \begin{pmatrix} 1\,2\,3\,4 \\ 1\,3\,2\,4 \end{pmatrix}.$$

In the first line there are the positions of elements in their original order and in the second the position the above indicated element will be mapped to.
It is quite obvious that the first line does not really contain any useful information. It would very well be possible to just use the second line for a full description:

$$\sigma_1 = [1,3,2,4]\,.$$

Definition 2.1.3 (Symmetric Group)**.** *For a natural number* n*, we call the set of all permutations on* $\{1, \ldots, n\}$ *the symmetric group* S_n*. In every symmetric group, we find the trivial permutation, the identity. We will universally denote the trivial element as* e*.*

Naturally, for every element σ in S_n we have an action on $\{1, \ldots, n\}$ with

$$\sigma : i \in \{1, \ldots, n\} \mapsto \sigma(i).$$

We can formalize this idea of an action.

Definition 2.1.4 (Group Action)**.** *A group action of a group G on a set X is a set of functions $\{\lambda_g : X \to X\}_{g \in G}$ such that*

$$\lambda_e : x \mapsto \lambda_e(x) = x, \ \forall x \in X, \ \textit{and}$$
$$\lambda_{g_1} \circ \lambda_{g_2} = \lambda_{g_1 \, g_2}, \ \forall g_1, g_2 \in G.$$

When the group action is clear, we will sometimes write $\lambda_g(x)$ as $g(x)$.

In addition to the simple group action of S_n on $\{1, ..., n\}$ defined by identifying σ with its natural function, we can think about the action of S_n on sets of more complicated objects.

Definition 2.1.5 (Sequence)**.** *A sequence s is an ordered list. It is a function from the set $\{1, \ldots, n\}$ to a finite set of symbols. The number n is called the length of s.*

The definition is a bit dry but becomes clear with a simple example.

Example 2.1.6 (Sequence)**.** *The ordered list $[a, b, a]$ is a sequence $s : \{1, 2, 3\} \to \{a, b\}$, with*

$$s(1) = a, \quad s(2) = b \quad \textit{and} \quad s(3) = a.$$

With the notion of a sequence we can define permutations of ordered lists by composition.

Example 2.1.7 (Permutation of a Sequence)**.** *The group S_n acts on the set of sequences $s : \{1, \ldots, n\} \to \Sigma$, for any finite set Σ, by composition: $\sigma(s) \mapsto s \circ \sigma$. Concretely, a permutation $\sigma \in S_n$ permutes a sequence s with length n in the following way:*

$$\sigma : s(i) \mapsto s(\sigma(i)).$$

In Example 2.1.2 we saw that not all elements are affected by the permutation. We see that the position of the first and the fourth element are not changed due to σ_1. We call these elements invariants of σ_1.

Definition 2.1.8 (Invariant Elements)**.** *Given a permutation $\sigma \in S_n$, and a set on which S_n acts, X, an element $i \in X$ for which $\sigma(i) = i$ is called an invariant element of σ.*

In the description of a permutation used in Example 2.1.2, it seems inefficient to even mention its invariants. How can we leave them out without losing the structure of the permutation itself?

Example 2.1.9 (Cycle)**.** *Let*

$$\sigma_2 = \begin{pmatrix} 1\,2\,3\,4 \\ 3\,1\,2\,4 \end{pmatrix}.$$

We see that the first element is mapped to the third position and the third is mapped to the second position, which itself is mapped to the first position, where the cycle closes. The fourth element is an invariant. It remains where it is. In cycle notation we can write $\sigma_2 = (132)(4)$*. We can now drop the invariants, represented by a cycle with only one element and thus write:* $\sigma_2 = (132)$*.*

The example can only give us an idea how to use the cycle notation, we will give a proper definition later, in Definition 2.1.27, after introducing more group theoretical background.
The cycle notation is very efficient in the description of a permutation. It does not, however, contain information about the set that is to be permuted. This can be confusing at times, but mostly it is an advantage: We can talk about transpositions (cycles with two elements) for S_2 or S_{200}.
We now have three ways of describing a concrete permutation; first listing the set that is to be permuted and where it is permuted to, second only listing the permuted set and third only the non-trivial cycles.
The last two can be confused, as we will see in the following example.

Example 2.1.10. *Let* σ *be*

$$\sigma = \begin{pmatrix} 1\,2\,3\,4 \\ 4\,1\,2\,3 \end{pmatrix}.$$

Since all elements are permuted its cycle notation is (1234)*, that could a priori be interpreted as the identity element, since it lists all elements in their original order. We avoid the confusion by writing the permuted set using square brackets, here* $[4, 1, 2, 3]$*.*

The cycle notation will turn out to be very efficient in characterizing symmetric groups. Because of their importance, we will discuss a few more of their properties.

Definition 2.1.11 (Length of a Cycle)**.** *For a cycle* c*, we call the number of elements in* c *the length* $L(c)$*.*

This is a very easy concept, as we can see in the following example.

Example 2.1.12. *The permutation* σ_1 *from Example 2.1.2 consists of three cycles:*

$$\sigma_1 = (1)(23)(4).$$

The first and third have length 1*, whereas the second has length* 2*.*

The length of a cycle and the cycle itself have the nice relation that a cycle taken to the power of its length is the identity.

Proposition 2.1.13 (Trivial Property of Cycles)**.** *For a sequence* s *of length at least* k*, and a cycle* σ *of length* $L(\sigma) = k$*, the cycle has the property* $\sigma^k s = s$*.*

Cycles are the center of our attention here, because it is a well known fact that all permutations can be written as a product of its disjoint cycles.

Theorem 2.1.14 (Permutations are Products of Disjoint Cycles)**.** *Every permutation $\sigma \in S_n$ can be written as a product of disjoint cycles, and every product of disjoint cycles describes a unique permutation.*

Proof. We have seen this in Example 2.1.12 and it is also not difficult to imagine. A permutation permutes every element at most once. If an element is not moved, it is in a cycle by itself, and if it is moved, it is in a cycle preceding the integer labeling the location to where it is moved. So every element has its place and the permutation is completely characterized. □

As we will later see, the number of distinct cycles is a deciding property of a permutation. For this reason we will give it its own name.

Definition 2.1.15 (Dimension of a Permutation)**.** *For a permutation σ in disjoint cycle notation, we call the number of cycles δ the dimension of σ.*

Note that the trivial cycles do count towards the dimension. It is not standard to interpret the number of cycles of a permutation as a dimension. Usually it is pretty clear why something is called a dimension, but here it is not. We have to put off this problem to a later point, Subsection 2.3.6.

2.1.2. Orbits and Symmetries

Now we will discuss how we can use group theoretical methods to describe symmetries and furthermore divide a set according to these symmetries.
First we need to remember the notion of a subgroup.

Definition 2.1.16 (Subgroup)**.** *A subset H of a group G is called subgroup of G if H is a group itself under the operation of G.*

Obviously a group is always a subgroup of itself.

Example 2.1.17 (Subgroup)**.** *Consider S_3, the permutations of three elements. The group of permutations of two elements, S_2, is clearly a subgroup of S_3.*

$$\begin{aligned} S_3 &= \{e, (12), (13), (23), (123), (132)\} \\ S_2 &= \{e, (12)\} \\ S_2' &= \{e, (13)\} \\ S_2'' &= \{e, (23)\}. \end{aligned}$$

There are even three copies of S_2 in S_3.

Given an element or elements of a group G we can generate a subgroup of G by combining the elements in all possible ways.

Definition 2.1.18 (Generated Group)**.** *For a group G consider a set $M = \{g_1, g_2, \ldots, g_n\}$ of elements of G. The smallest subgroup of G that contains M is called the group generated by M.*

Example 2.1.19 (Generated Group)**.** *Take* (12) *from S_3. By itself it does not create a group since it is not closed under the action*

$$(12)(12) = e.$$

However if we add e to the set. We have the smallest subgroup of S_3 that contains (12)*.*

Once we see how a group can be used to split a set into subsets according to specific properties (symmetries), it will become clear why we need the concept of generated groups.

Definition 2.1.20 (G-Set)**.** *For a group G, a set X is called G-set if it is compatible with the groups action as in Definition 2.1.4.*

Trivially given a subgroup H of G, any G-set is also an H-set.

Example 2.1.21. *For the general linear group $Gl(\mathbb{C}^n)$ we can have it act with its normal multiplication on vectors in $\mathbb{C}^n$, in other words $\mathbb{C}^n$ is a $Gl(\mathbb{C}^n)$-set.*

Every element of X is left unaffected by the identity element, however there can be more elements of G that leave certain elements unaffected.

Definition 2.1.22 (Isotropy Group - Symmetries)**.** *For a group G and an element $x \in X$ from a G-set X, we define the isotropy group G_x:*

$$G_x = \{g \in G | gx = x\}.$$

We call the elements of G_x symmetries of x. The isotropy group is often called the stabilizer of x.

After all these dry definitions we can now give a nice example of why all these concepts are useful. We can use the isotropy group to characterize symmetries.

Example 2.1.23 (Symmetries I)**.** *It is clear that we have an S_5 action on S. Consider now the sequence, $s = [a, a, b, b, a]$. It is easy to see that s is invariant under* (12),(34) *and* (15)*. It is however not too easy to find all elements of S_5 that leave s invariant.*
While with an educated guess, we certainly could work out the isotropy group by hand, it is easier to use a computer to find the subgroup H generated by $(12), (34), (15)$*:*

$$\begin{aligned} H = \{&e, (12), (15), (25), (34), (125), (152), (12)(34), \\ &(15)(34), (25)(34), (125)(34), (152)(34)\}, \end{aligned}$$

which is the full isotropy group of s as a G-set. Computers will be especially helpful when the sets are larger.

The next concepts will allow us to specify how much an element that is not in a certain isotropy group violates the described symmetry.

Definition 2.1.24 (Orbit)**.** *For a group G and an element $x \in X$ from a G-set X, the set of elements x can be moved to is called the orbit of x,*

$$Gx = \{gx | g \in G\}.$$

Lemma 2.1.25 (Disjoint Orbits)**.** *For a group G and a G-set X, the orbits for two $x, y \in X$ elements of X are either identical or disjoint.*

It is further possible to decompose any G-set into orbits of G.

Theorem 2.1.26 (Orbit-Stabilizer)**.** *For a group G and a G-set X, we have:*

$$|G| = |G_x||Gx|.$$

More precisely, let x_i be a representative of orbit X_i. Then we have:

$$|X| = |X_1| + \ldots + |X_n| = \frac{|G|}{|G_{x_1}|} + \ldots + \frac{|G|}{|G_{x_n}|}.$$

Let us return to permutations. We have seen that a permutation can be completely characterized by its cycle notation.

Definition 2.1.27 (Cycle)**.** *A permutation is called a cycle when it has only one orbit with more than one element.*

Let us get back to the symmetry example, Example 2.1.23, and understand how we can use these extra concepts in its context.

Example 2.1.28 (Symmetries II)**.** *Consider again the set of all sequences of the following kind, $S = \{s | s : \{1, 2, 3, 4, 5\} \to \{a, b\}\}$ and the specific sequence $s = [a, a, b, b, a]$. The element (13) does violate the symmetry of s as it maps s to another sequence in S, $(13)(s) = \tilde{s} = \{b, a, a, b, a\}$. We can find all elements that do the same violation by splitting S_5 into orbits under the isotropy group H. The first orbit, O_1, will be the orbit that contains the identity element, thus it is the isotropy group itself; the second orbit, O_2, will be the orbit that contains all elements that create the same symmetry violation as (13),*

$$O_2 = \{(13), (14), (23), (24), (35), (45)\}.$$

As there $\frac{5!}{3!\,2!} = 10$ ways to order s, there will be eight more orbits. It is not very illuminating to mention them all explicitly.

2.2. An Introduction to Representation Theory

In representation theory we study how group elements are represented as linear transformations when acting on sets or vector spaces. We will be especially interested in group actions, and thus, the representations of the symmetric group.
We will now give an explanatory review of the concepts of representation theory that are crucial for the understanding of our work. We will not provide any proofs, but illustrative examples. More details and proofs can be found in, for example, [12]. The explanation of Schur-Weyl duality has been taken from [32]. Where no other work is cited, it is according to this reference.

Definition 2.2.1 (Representation)**.** *For a vector space V and a finite group G, it is a simple exercise to show that every group action of G on V is a group homomorphism $\lambda : G \to GL(V)$. The homomorphism λ is called a representation of G.*

Corresponding to the representation λ, we have, for elements $g \in G$, linear transformations of V given by:

$$\lambda(g) : V \to V, \;\; v \mapsto \lambda(g)\, v.$$

Let us explain the mapping λ in the very easy example of the group with only two elements, and understand it as the symmetric group S_2.

Example 2.2.2. *We choose $G = S_2 = \{e, \sigma\}$ and a vector space V. Keeping the group structure, there are two different choices for homomorphisms: the trivial representation, λ_t, where both group elements are mapped to identity operator on V,*

$$\lambda_t(g) = \mathbb{1}_V;$$

and the alternating representation, λ_a, where every element is mapped similarly, but keeping track of the element's signum:

$$\lambda_a(g) = \mathrm{signum}(g)\, \mathbb{1}_V.$$

Note that λ_t and λ_a are different homomorphisms!

The operators in Example 2.2.2, though equipped with the standard product of operators, do not have a particularly interesting effect on V. They act one-dimensionally, by multiplication with a single number only. We will now see that if V has dimension greater than one, then V is too large to capture the essence of these simple representations.

Definition 2.2.3 (Subrepresentation)**.** *A subrepresentation of a representation λ on V is a representation $\tilde{\lambda}$ on W where W is a λ-invariant subspace of V, that is, $\lambda(g)W = W$ for all $g \in G$.*

In Example 2.2.2 we can see that every subspace of V is left invariant under both representations, thus it is a subrepresentation of V. In the definition of subrepresentation we have not distinguished between the mapping as a representation and the vector space where the group is acting on as a representation. While on first glance this looks a bit confusing it becomes a lot clearer once we come to irreducible representations. These only capture the essence of the group that is to be represented.

Definition 2.2.4 (Irreducible)**.** *We call a representation irreducible, or an irrep, if there is no subrepresentation but itself and the zero-dimensional subspace.*

Back in Example 2.2.2 we should rather talk about the irreducible representation called trivial representation, $\lambda_t(g) = 1$, and the irreducible representation called alternating representation, $\lambda_a(g) = \text{signum}(g)$, both acting on one-dimensional subspaces of V. There will be d mutually orthogonal copies of these representations in V.
Naturally, we extend the dimension of the vector space to the representation itself.

Definition 2.2.5 (Dimension of a Representation)**.** *We call the dimension of a subspace where a representation acts irreducibly the dimension of the (sub)representation.*

Let us now investigate a more interesting group action on a tensor product space and find the irreducible representations.

Example 2.2.6. *The space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ shall have bases $\{|i\rangle\}_{i=1}^2$ and $\{|j\rangle\}_{j=1}^2$, and let S_2 be the symmetric group acting on the indices. We describe this group action as the representation λ, defined for $g \in S_2$, $v = v_1 \otimes v_2 \in \mathcal{H}$ by:*

$$\lambda(g)\, v \equiv v_{g^{-1}(1)} \otimes v_{g^{-1}(2)}.$$

The action on entangled v is then defined by linear extension.
The group S_2 has two elements, the identity e and the transposition $\sigma = (12)$. We see immediately that the subspace spanned by $|1\rangle \otimes |1\rangle$ is invariant under S_2, as is the case for the subspace spanned by $|2\rangle \otimes |2\rangle$.
Furthermore, we have the same for the space spanned by the entangled state $|1\rangle \otimes |2\rangle + |2\rangle \otimes |1\rangle$. On all these spaces all elements of S_2 act trivially.
In addition, we find the space spanned by $|1\rangle \otimes |2\rangle - |2\rangle \otimes |1\rangle$ where the transposition σ does not act trivially.
We have found four subrepresentations of λ:

$$\begin{aligned} W_1 &= \text{span}(|1\rangle \otimes |1\rangle), \\ W_2 &= \text{span}(|2\rangle \otimes |2\rangle), \\ W_3 &= \text{span}(|1\rangle \otimes |2\rangle + |2\rangle \otimes |1\rangle), \text{ and} \\ W_4 &= \text{span}(|1\rangle \otimes |2\rangle - |2\rangle \otimes |1\rangle). \end{aligned}$$

On the first three, S_2 essentially acts trivially, whereas on the fourth it acts as the alternating representation. In contrast to Example 2.2.2, where we had two different homomorphisms, here only **one** *homomorphism (representation) gives rise to different irreducible (sub)representations.*

Finding the different irreducible representations is our main point of interest. We will no longer focus on all possible representations, but on irreducible representations only. The following theorems justify their importance.

Theorem 2.2.7 (Complementary Subspaces)**.** *If λ_W is a subrepresentation on W of a finite group representation λ on V then there is a complementary invariant subspace W' of V, such that $V = W \oplus W'$.*

Theorem 2.2.8 (Decomposition into Irreducibles)**.** *Any representation is a direct sum of irreducible representations λ_i. Precisely: For any representation λ on V of a finite group G, there is a decomposition*

$$V = V_1^{\oplus a_1} \oplus \cdots \oplus V_k^{\oplus a_k}, \tag{2.1}$$

where the V_i are distinct λ_i invariant subspaces. The decomposition of V into a direct sum of the k factors is unique, as are the V_i that occur and their multiplicities a_i.

Let us now use Theorem 2.2.7 and Theorem 2.2.8 for our previous examples. In Example 2.2.2 we can decompose V into the irreducible representations W_λ

$$V = W_{\lambda_t}^{\oplus d} \quad \text{or} \quad V = W_{\lambda_a}^{\oplus d}$$

and in Example 2.2.6, we can decompose $\mathcal{H}$ as

$$\mathcal{H} = W_{\lambda_t}^3 \oplus W_{\lambda_a}^1.$$

So far all (irreducible) representations have been one-dimensional. That is no coincidence.

Theorem 2.2.9 (Representations of Abelian Groups)**.** *All irreducible representations of abelian groups are one-dimensional.*

Since S_3 is the smallest non-abelian group, we now look into its irreps.

Example 2.2.10 (3-dimensional irrep)**.** *Again we shall have a vector space that is a tensor product $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$. Now let $g \in S_3$ act on the index:*

$$\lambda(g)\, v_1 \otimes v_2 \otimes v_3 = v_{g^{-1}(1)} \otimes v_{g^{-1}(2)} \otimes v_{g^{-1}(3)}.$$

The subspace

$$W = \operatorname{span}(|1\rangle \otimes |1\rangle \otimes |2\rangle, |1\rangle \otimes |2\rangle \otimes |1\rangle, |2\rangle \otimes |1\rangle \otimes |1\rangle)$$

has no further subrepresentations.

An important tool in representation theory is the concept of a character function on a group.

Definition 2.2.11 (Character Function)**.** *For an element σ of a finite group and a specific representation λ, we call*

$$\chi^{\lambda}(\sigma) = Tr(\lambda(\sigma))$$

the character of σ in λ.

The character can be used to identify the dimension of a certain representation.

Proposition 2.2.12. *The character of the identity element is equal to the dimension of the representation, that is,*

$$\chi^{\lambda}(e) = \dim(\lambda).$$

2.2.1. Schur-Weyl Duality

So far we have only used two or three factors in the tensor product spaces we have considered. We will now come closer to our original setting with n-particle Hilbert spaces, and will learn a very nice way to find irreducible representations of both the symmetric and the general linear group, and thus the unitary group.

Let the Hilbert space for a single particle be $\mathcal{H}_1 = \mathbb{C}^d$. We consider the n-fold tensor product $\mathcal{H} = \mathbb{C}^d \otimes \overset{n\text{-times}}{\cdots} \otimes \mathbb{C}^d$. Now we have the action of S_n on $\mathcal{H}$ as in Example 2.2.6 and the action of $Gl(\mathcal{H}_1)$ on $\mathcal{H}$ with $g \in Gl(\mathcal{H}_1)$ and $v = v_1 \otimes \cdots \otimes v_n \in \mathcal{H}$ as

$$gv = gv_1 \otimes \cdots \otimes gv_n.$$

$Gl(\mathcal{H}_1)$ acts on entangled states by linear extension.
For these two groups we have a nice version of Theorem 2.2.8 called Schur-Weyl Duality.

Theorem 2.2.13 (Schur Weyl Duality)**.** *Let a vector space $\mathcal{H}$ be defined $\mathcal{H} = \mathcal{H}_1^{\otimes n} = (\mathbb{C}^d)^{\otimes n}$ with the general linear group $Gl(\mathcal{H}_1)$ acting on each factor in the tensor product and the symmetric group S_n acting on the index. We have the decomposition:*

$$\mathcal{H} = \oplus_i G_i \otimes F_i = \oplus_i W_i,$$

where G_i are subspaces on which $Gl(\mathcal{H}_1)$ acts irreducibly and F_i are subspaces on which S_n acts irreducibly.

In Theorem 2.2.13 we learn that the representations of $Gl(d)$ and S_n are closely related. It is very well possible that if a particular representation of S_n does not occur, then the dual representation will not occur either. We see that the dimension of G_i is the multiplicity of F_i and vice versa.

On first glance only the existence of Schur Weyl Duality alone does not seem very useful. But there is a very handy method for finding and identifying the different representations.

2.2.2. Young Diagrams: Hooks and Contents

There is a close relation between young diagrams, partitions and the representations we are looking for.

Definition 2.2.14 (Partition)**.** *A partition λ of n is a non-ascending set of positive numbers λ_i, such that $\sum_i \lambda_i = n$.*

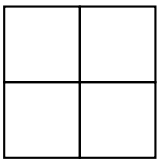
Theorem 2.2.15 (Partitions and Representations)**.** *Given a number n, for each partition λ of n, there is an irreducible representation of S_n.*

Definition 2.2.16. *Since we can identify a representation λ of S_n with a partition λ, we write for*

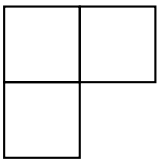
$$\lambda(S_n) = S_n^\lambda.$$

To learn what it means that partitions and representations are closely related, we have to be a bit patient and understand the following concepts. We start by creating Young diagrams.

Definition 2.2.17 (Young Diagrams)**.** *For a number n, if we have a partition $\sum_i \lambda_i = n$ with $\lambda_i \geq \lambda_{i+1}$, then the young diagram for λ is the 2-dimensional array of boxes such that the first row has λ_1 boxes, the second λ_2, and so on.*



$n = 4, \lambda = (2, 2)$



$n = 3, \lambda = (2, 1)$

Figure 2.1.: Young Diagrams

It it is not necessary to distinguish between a diagram and its corresponding partition — we will refer to both with λ.

Proposition 2.2.18 (Schur-Weyl Duality)**.** *Let again vector space $\mathcal{H}$ be defined $\mathcal{H} = \mathcal{H}_1^{\otimes n} = (\mathbb{C}^d)^{\otimes n}$ with the general linear group $Gl(\mathcal{H}_1)$ acting on each factor in the tensor product and the symmetric group S_n acting on the index. We have now learned that we can identify an irrep with a partition and sum over all partitions, to obtain a decomposition:*

$$\mathcal{H} = \oplus_\lambda G^\lambda \otimes F^\lambda = \oplus_\lambda W^\lambda,$$

where G^λ are the subspaceson which $Gl(\mathcal{H}_1)$ acts irreducibly and F^λ are subspaces where S_n does.

For a direct usage of these diagrams we introduce two very easy concepts.

Definition 2.2.19 (Hook Length)**.** *Define the hook length $h(u)$ of a box in a Young diagram as the number of boxes right of it plus the number of boxes below it plus one.*

Figure 2.2.: Hook Length

Definition 2.2.20 (Content)**.** *The content $c(u)$ of a box in a Young diagram is defined as its column index minus its row index.*

Figure 2.3.: Content

2.2.3. Dimensions

With the handy concepts from the previous section, we can finally make use of Schur Weyl Duality, Theorem 2.2.13, and decompose tensor product spaces into irreps.

Theorem 2.2.21 (Hook Length Formula [12])**.** *Given a partition λ, the dimension of S_n^λ is given by*

$$\dim S_n^\lambda = \frac{n!}{\prod_{u\in\lambda} h(u)} = \chi^\lambda(e).$$

Second, there is also an easy formula for the dimension of the corresponding Gl representation.

Theorem 2.2.22 (Stanley's Hook Content Formula [31, 20])**.** *Given a partition λ, we have*

$$\dim Gl^\lambda = \prod_{u \in \lambda} \frac{d + c(u)}{h(u)}.$$

Corollary 2.2.23. *For W^λ defined as in Proposition 2.2.18, the dimension of W^λ is:*

$$\dim W^\lambda = \dim Gl^\lambda * \dim S_n^\lambda.$$

2.2.4. Further characterization of Symmetric Groups

In the last section we learned how to associate partitions with young diagrams and their application to Schur-Weyl Duality. We will now see that partitions are even more useful in characterizing symmetric groups.
For every group we can define an action on itself called conjugation.

Definition 2.2.24 (Conjugation)**.** *Two elements g_1 and g_2 from a group G are called conjugate, if there exists a third (not necessarily different) element g_3 of G such that*

$$g_1 = g_3\, g_2\, g_3^{-1}.$$

The action of g_3 on g_2 is called conjugation.

As we can decompose any G-set into orbits under G, we can decompose G itself under conjugation.

Definition 2.2.25 (Conjugation Classes)**.** *For an element g_1 from a group G we define its conjugation class $C(g_1)$ as the orbit of g_1 under conjugation:*

$$C(g_1) = \{g \in G | \exists \tilde{g} \in G : g = \tilde{g}\, g_1 \tilde{g}^{-1}\}.$$

Elements from the same conjugation class share many properties. As a first example we see that for all groups the character function is constant on conjugation classes.

Proposition 2.2.26 (Character of Conjugation Classes)**.** *Let G be a group and $a, b \in C(a)$ elements of G in the same conjugation class. Furthermore let λ be a representation of G. Then*

$$\chi^\lambda(a) = \chi^\lambda(b).$$

Proof. There exists a $g \in G$ such that $b = g\, a\, g^{-1}$, thus with cyclic invariance of the trace we have:

$$\chi^\lambda(b) = \chi^\lambda(g\, a\, g^{-1}) = \chi^\lambda(a).$$

□

We will now see that with partitions we can not only find representations of S_n, but also all its conjugation classes. In fact, there is a particularly nice identification of partitions with conjugation classes.

Theorem 2.2.27 (Structure of Symmetric Groups)**.** *Given a symmetric group S_n, for each partition of n, there exists a corresponding conjugation class. More precisely, given a partition λ of n:*

$$\lambda = (a_1, a_1, \ldots, a_1, a_2, \ldots, a_2, a_3, \ldots),$$

where a_1 appears k_1 times, a_2 appears k_2 times and so on, λ corresponds to the conjugation class with k_1 a_1-cycles, k_2 a_2-cycles, and so on.

In other words, two elements of S_n are in the same conjugation class, if and only if they have the same number of cycles of each size.

We have explained how we can find representations of S_n using the partitions of n. Now we have this extra application and it can be confusing to differentiate the two. Let us clear things up with an example.

Example 2.2.28 (Structure of S_3)**.** *For S_3 we have $n = 3$. The partitions of n are $(3), (2,1), (1,1,1)$, which means there are three different irreducible representations of S_3. We can find their dimensions easily using the hook-length formula:*

$$\begin{aligned} \dim S_3^{(3)} &= \frac{6}{6} = 1, \\ \dim S_3^{(2,1)} &= \frac{6}{2} = 3, \text{ and} \\ \dim S_3^{(1,1,1)} &= \frac{6}{6} = 1. \end{aligned}$$

We see that there are two one-dimensional representations. Obviously these are commutative representations, again the trivial and the alternating representation. Unfortunately we have no means of differentiating them up to now. At least we see that there is a third representation with a bit more structure.
Furthermore, in S_3 there are three conjugation classes: the neutral element; the exchange of two elements — transpositions or two-cycles; and the permutation of three elements — three-cycles:

- $(1,1,1)$ *corresponds to three one-cycles. That means nothing happens. It is the conjugation class of the neutral element.*
- $(2,1)$ *corresponds to one two-cycle and one one-cycle — a transposition.*
- (3) *corresponds to one three-cycle.*

Note that all conjugation classes will appear in all representations, however they are not necessarily mapped to different operators.

We have seen that all elements of a certain conjugation class share some properties, in particular, for S_n, members of the same conjugation class have very similar structure. We can study the properties of all members of a conjugation class on one particular member.

Definition 2.2.29 (Representative)**.** *For each conjugation class (or any orbit) we can choose an element from the class. It is called a representative of the conjugation class.*

The concept of a representative is not restricted to a conjugation action. For any orbit the members of that orbit are related and can be represented by one of them.

Let us return to the symmetric group. Given a partition and a representative of the corresponding conjugation class, there is a nice formula for the size of the conjugation class.

Proposition 2.2.30 (Size of a Conjugation Class)**.** *For a partition λ as in Theorem 2.2.27 and its conjugation class C, the number of elements of C, $|C|$, is given by:*

$$|C| = \frac{n!}{\prod_i (a_i)^{k_i} (k_i)!}.$$

A dry formula is best explained with a non-trivial example.

Example 2.2.31. *Let us consider the conjugation class C with partition $(2, 2, 1)$. We have $n = 5$, $a_1 = 2$, $a_2 = 1$ and $k_1 = 2$, $k_2 = 1$. Then:*

$$|C| = \frac{5!}{((2)^2 2!)\,((1)^1 1!)} = 15.$$

2.2.5. Frobenius Character Formula

Sometimes the character of a certain conjugation class in a given representation is not at all obvious. The Frobenius character formula [12], to be defined shortly, is a handy tool to find it. Unfortunately it is a bit tricky to use. We try to explain it in a way that is accessible to physicists.

Definition 2.2.32 (Alternative Characterization of Conjugation Classes)**.** *We saw before that we can characterize an element of S_n in cycle notation. Furthermore we saw that it can be seen as a representative of all elements with the same number of cycles of a specific length. This way we see that it is sufficient for identifying a conjugation class if we use a tuple $(i_1, i_2, \ldots)_A$ where i_1 is the number of 1-cycles, i_2 is the number of 2-cycles and so on.*

Example 2.2.33. *Let $n = 4$ and let $\sigma = (12)(34)$ be a representative of C, which we characterize by the partition $(2, 2)$. Then the alternative characterization of C is $(0, 2, 0, 0)_A$.*

we will need several more definitions from the theory of representations that will be useful for our applications.

Definition 2.2.34 (Power Sums)**.** *For a k-tuple $x = (x_1, \ldots, x_k)$, we define its power sum $P_j(x)$ according to $j \in \mathbb{Z}$ by*

$$P_j(x) = x_1^j + x_2^j + \ldots + x_k^j.$$

Definition 2.2.35 (Discriminant)**.** *For a k-tuple $x = (x_1, \ldots, x_k)$, we define the discriminant:*

$$\Delta(x) = \prod_{i<j}(x_i - x_j).$$

Definition 2.2.36 (Coefficient Bracket)**.** *Let $f(x) = f(x_1, x_2, \ldots, x_k)$ be a polynomial and $(l_1, \ldots, l_k)$ be a k-tuple of non-negative integers. Define $[f(x)]_{(l_1,\ldots,l_k)}$ as the coefficient of $x^{l_1} \cdots x^{l_k}$ in f.*

Definition 2.2.37 (Coefficient Tuple)**.** *Given a partition λ of n, with $\lambda = (\lambda_1, \ldots, \lambda_k)$ we define the coefficient tuple:*

$$l_1 = \lambda_1 + k - 1, \quad l_2 = \lambda_2 + k - 2, \quad \ldots, \quad l_k = \lambda_k.$$

Theorem 2.2.38 (Frobenius Formula)**.** *For a representation λ, a conjugation class C, and its representative σ, the character of C (or σ) in λ is given by:*

$$\chi_\lambda(C) = \chi_\lambda(\sigma) = \left[\Delta(x)\prod_j P_j(x)^{i_j}\right]_{l_1,\ldots,l_k}.$$

Let us choose a very easy example, where we even know the answer. For S_2 we know that there are two partitions, (2) and $(1,1)$, so there are two representations, each with two conjugation classes. For the representation associated with (2) — the trivial representation — we know that all characters are equal to 1. In the alternating representation we have the conjugation class belonging to the transposition (12) with a character equal to -1. Let us elaborate on this.

Example 2.2.39. *Let $n = 2$. We are wondering about the character of the transposition* (12) *in the alternating representation which belongs to the partition* $(1,1)$*. The coefficient tuple is easy to calculate: $l_1 = 1 + 2 - 1 = 2$ and $l_2 = 1$. The alternative representation of C is $(0,1)_A$:*

$$\begin{aligned}\chi_{(1,1)}((12)) &= \left[(x_1 - x_2)(x_1^1 + x_2^1)^0(x_1^2 + x_2^2)^1\right]_{2,1} \\ &= \left[x_1^3 + x_1x_2^2 - \mathbf{x_1^2x_2} - x_2x_2^2\right]_{2,1} \\ &= -1.\end{aligned}$$

We have seen that using the Frobenius formula, we can easily compute characters. This particular character will come up in future calculations. However, the formula can also be used to find more complicated and less obvious characters, which is essential for the use of Collins and Śniady's formula, which we will introduce in the next section.

2.3. Integrals over Unitary Groups

After introducing these abstract concepts we can finally state Collins and Śniady's formula [8]. It makes use of the close relation between representations of general linear and symmetric groups we have seen in the last section. An integral over matrix elements of unitary groups with respect to the Haar measure can be replaced by a summation over the different representations of symmetric groups.

Theorem 2.3.1 (Collins and Śniady)**.** *For a unitary group $\mathcal{U}(d)$, an integral over entries of its unitary matrices with respect to the Haar measure is given by:*

$$\begin{aligned}&\int_{\mathcal{U}(d)} d\mathcal{U}\, \mathcal{U}_{i_1\tilde{i}_1}\cdots\mathcal{U}_{i_k\tilde{i}_k}\,\overline{\mathcal{U}_{j_1\tilde{j}_1}}\cdots\overline{\mathcal{U}_{j_k\tilde{j}_k}}\\ &= \sum_{\sigma,\tau\in\mathcal{S}_k}\delta_{i_1 j_{\sigma(1)}}\cdots\delta_{i_k j_{\sigma(k)}}\,\delta_{\tilde{i}_1\tilde{j}_{\tau(1)}}\cdots\delta_{\tilde{i}_k\tilde{j}_{\tau(k)}}W(\tau\sigma^{-1},k)\end{aligned}$$

Where $W(\sigma,k)$ denotes the Weingarten function of the permutation $\sigma\in S_k$:

$$W(\sigma,k)=\frac{1}{(k!)^2}\sum_{\lambda\vdash k}\frac{\left(\dim S_k^\lambda\right)^2}{\dim Gl^\lambda(d)}\chi^\lambda(\sigma).$$

In this formula we already supressed that all integrals with a different number of unitary elements and adjoint elements will be equal to zero.

Now we will present our applications of Theorem 2.3.1. We will explain all concepts in a detailed, instructive, direct-to-use manner. In the presentation it will become clear why it is useful to invest some time to learn the elegant formulas instead of parameterizing the spherical integrals oneself.

Later we will see that for specific settings, in particular for our investigation of the channel fidelity moments in Chapter 3, we could simplify the general formula drastically and find an elegant notation that will allow a compact description.

2.3.1. Illustrative Examples

We will now slowly become familiar with calculating Weingarten functions.

Example 2.3.2 (Overlap of unit vectors)**.** *We begin by asking how much a complex vector in a d-dimensional Hilbert space $\mathcal{H}=\mathbb{C}^d$ overlaps another vector of length one, on average. The answer to this question is equal to the absolute value of an entry of a d-dimensional unitary matrix:*

$$\int_{|\phi|=1} d\phi\,|\langle 1|\,|\phi\rangle|^2=\int_{\mathcal{U}(d)} d\mathcal{U}\,|U_{1,1}|^2=\int_{\mathcal{U}(d)} d\mathcal{U}\,\mathcal{U}_{1,1}\overline{\mathcal{U}_{1,1}}=\sum_{\sigma,\tau\in S_1}\delta_{1,\sigma(1)}\,\delta_{1,\tau(1)}W(\tau\sigma^{-1},1).$$

We could have chosen any other matrix element since the right-hand side only depends on the number of matrix elements!
The right-hand side is fairly easy to calculate. Since S_1 is the most trivial group, with only the identity element, we only need to evaluate $W(e,1)$. Since $k=1$ there are none but the trivial partition of k and the young diagram has only one box.

$h(u)$ $\qquad$ $c(u)$

Figure 2.4.: Hooks and Contents for S_1

Then we have

$$\dim S_1^{(1)} = \frac{1!}{1} = 1, \text{ and}$$
$$\dim Gl^{(1)} = \frac{d+0}{1} = d.$$

This allows us to calculate the Weingarten function:

$$W(e,1) = \frac{1}{1^2} \times \frac{1^2}{d} \times 1 = \frac{1}{d}.$$

Because everything else is equal to one, this is also the answer to our question.

We can gain more from this easy example:

Example 2.3.3 (Average Expectation Value)**.** *Let A be an operator acting on $\mathcal{H}$. What is the average expectation value of A? We can compute:*

$$\int_{|\phi|=1} d\,\phi\ \langle\phi|\,A\,|\phi\rangle = \int_{\mathcal{U}(d)} d\mathcal{U}\ \langle 1|\,U^\dagger A U\,|1\rangle = \sum_{i_1,j_1} a_{j_1,i_1} \int_{\mathcal{U}(d)} d\mathcal{U}\, U_{i_1,1}\overline{U_{j_1,1}}$$
$$= \sum_{i_1,j_1} a_{j_1,i_1} \sum_{\sigma,\tau\in S_1} \delta_{i_1,j_{\sigma(1)}}\ \delta_{1,\tau(1)} W(\tau\sigma^{-1},1).$$

In this case we see that the Kronecker delta in the second component stays trivial but in the first component it gives δ_{i_1,j_1} which together with the sum gives the trace of A. The Weingarten function stays the same, giving:

$$\int_{|\phi|=1} d\,\phi\ \langle\phi|\,A\,|\phi\rangle = \frac{Tr\,(A)}{d}.$$

The next logical step is to consider the 2-correlator of two operators.

Example 2.3.4. *Let A and B be operators acting on $\mathcal{H}$. We consider the integral:*

$$\begin{aligned} I(A,B) &= \int\limits_{|\phi|=1} d\ \phi \langle\phi| A |\phi\rangle \langle\phi| B |\phi\rangle = \int\limits_{\mathcal{U}(d)} d\mathcal{U}\ \langle 1| U^\dagger A U |1\rangle \langle 1| U^\dagger B U |1\rangle \\ &= \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{i_2,j_2} \int\limits_{\mathcal{U}(d)} d\mathcal{U}\, U_{i_1,1} U_{i_2,1} \overline{U_{j_1,1}}\, \overline{U_{j_2,1}}. \end{aligned}$$

With Collins and Śniady's formula we get:

$$I(A,B) = \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \sum_{\sigma,\tau\in S_2} \delta_{i_1,j_{\sigma(1)}} \delta_{i_2,j_{\sigma(2)}} \delta_{\tilde{i}_1,\tilde{j}_{\tau(1)}} \delta_{\tilde{i}_2,\tilde{j}_{\tau(2)}} W(\tau\sigma^{-1}, 2).$$

Expanding the double sum gives:

$$\begin{aligned} I(A,B) = \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \Big(& \delta_{i_1,j_1}\, \delta_{i_2,j_2}\, \delta_{1,1}\, \delta_{1,1} W(e \times e^{-1}, 2) \\ & + \delta_{i_1,j_1}\, \delta_{i_2,j_2}\, \delta_{1,1}\, \delta_{1,1} W((12) \times e^{-1}, 2) \\ & + \delta_{i_1,j_2}\, \delta_{i_2,j_1}\, \delta_{1,1}\, \delta_{1,1} W(e \times (12)^{-1}, 2) \\ & + \delta_{i_1,j_2}\, \delta_{i_2,j_1}\, \delta_{1,1}\, \delta_{1,1} W((12) \times (12)^{-1}, 2) \Big). \end{aligned}$$

We can see that the Kronecker deltas will collapse the sums to all possible products of traces with the different Weingarten functions as weights. We will explain this later in more detail. Next we use the fact that the inverse of (12) *is* (12) *itself:*

$$\begin{aligned} I(A,B) =& Tr(A)\ Tr(B)\, W(e,2) + Tr(A)\ Tr(B)\, W((12),2) \\ & + Tr(A\,B)\, W((12),2) + Tr(A\,B)\, W(e,2) \\ =& \left(Tr(A)\ Tr(B) + Tr(A\,B)\right) \left(W(e,2) + W((12),e)\right). \end{aligned}$$

Now all that is left to do is to evaluate the Weingarten functions. We will do this with the young diagrams of S_2 and the neat formulas.
The group $S_2 = \{e, \sigma\}$ consists only of the neutral element and the self-inverse exchange of two elements. The only two partitions of 2 *are $\lambda_1 = (2,0)$ and $\lambda_2 = (1,1)$ belonging to the following young diagrams:*

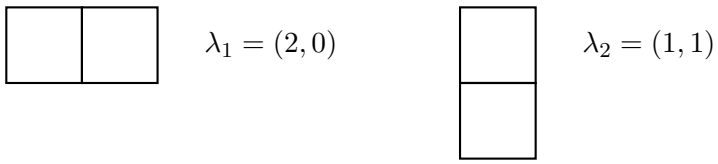


Figure 2.5.: Partitions and Young Diagrams for S_2

The partition λ_1 represents the 1-dimensional trivial representation, where everything is mapped to the identity; and λ_2 represents the 1-dimensional alternating representation.
First, using the hook length formula (Theorem 2.2.21), we easily calculate that

$$\dim S_2^{\lambda_1} = \dim S_2^{\lambda_2} = \frac{2!}{2 * 1} = 1.$$

Then Stanley's formula (Theorem 2.2.22) gives us

$$\prod_{u \in \lambda_1} \frac{d + c(u)}{h(u)} = \frac{d(d+1)}{2}$$

and

$$\prod_{u \in \lambda_2} \frac{d + c(u)}{h(u)} = \frac{d(d-1)}{2}.$$

Finally, we can calculate Weingarten functions, first for the identity e. Note that the character of e is, in all 1-dimensional representations, equal to 1:

$$\begin{aligned} W(e,2) &= \frac{1}{4}\left(\frac{\chi^{\lambda_1}(e)^2}{\prod_{u\in\lambda_1}\frac{d+c(u)}{h(u)}}\chi^{\lambda_1}(e) + \frac{\chi^{\lambda_2}(e)^2}{\prod_{u\in\lambda_2}\frac{d+c(u)}{h(u)}}\chi^{\lambda_2}(e)\right) \\ &= \frac{1}{4}\left(\frac{2}{d(d+1)} + \frac{2}{d(d-1)}\right) = \frac{1}{d^2-1}. \end{aligned}$$

Second for the other element $\sigma_1 \in S_2$, we calculated its character in the sign representation before in the example of the Frobenius formula. It is -1. Thus:

$$\begin{aligned} W(\sigma_1,2) &= \frac{1}{4}\left(\frac{\chi^{\lambda_1}(e)^2}{\prod_{u\in\lambda_1}\frac{d+c(u)}{h(u)}}\chi^{\lambda_1}(\sigma_1) + \frac{\chi^{\lambda_2}(e)^2}{\prod_{u\in\lambda_2}\frac{d+c(u)}{h(u)}}\chi^{\lambda_2}(\sigma_1)\right) \\ &= \frac{1}{4}\left(\frac{2}{d(d+1)} - \frac{2}{d(d-1)}\right) = \frac{-1}{d(d^2-1)}. \end{aligned}$$

We compute their sum:

$$W(e,2) + W(\sigma_1,2) = \frac{1}{d(d+1)}. \tag{2.2}$$

At last, we have:

$$I(A,B) = \frac{Tr(A)\ Tr(B) + Tr(A\,B)}{d(d+1)}.$$

The last example was very detailed. We will see that for integrals of this simple kind we will not have to do all the different steps. Nevertheless the example should be seen as a general explanation how to use Collins and Śniady's formula.

2.3.2. Sums of Weingarten Functions

We might have noticed that all examples where trivial in the second component. Furthermore, Equation 2.2 looks surprisingly simple. With some more representation theory we can show that this is indeed no coincidence.
First we notice that the Weingarten function is proportional to the character function and inherits a deciding property.

Proposition 2.3.5. *Weingarten functions are constant on conjugation classes.*

Proof. We see that, by definition, Weingarten functions are functions of characters and we already know that characters are constant on conjugation classes. □

Then we see that the trivial representation has a unique property.

Lemma 2.3.6 (Proposition 2.30 in [12])**.** *Let S_n be the group of permutations of n elements. Then for all representations V but the trivial representation V_t, the sum of the character function over the whole group vanishes:*

$$\sum_{g \in S_n} \chi^V(g) = 0.$$

We can extend this statement immediately.

Proposition 2.3.7. *Let λ be an arbitrary representation of S_n. Then the sum of the character function over the whole either vanishes or λ is the trivial representation. We denote this using a Kronecker δ: Either the partition λ is equal to (n) or it is equal to zero:*

$$\sum_{g \in S_n} \chi^\lambda(g) = \sum_{g \in S_n} \chi^\lambda(g)\, \delta_{\lambda,(n)}. \tag{2.3}$$

This allows us to simplify the trivial sum over Weingarten functions.

Theorem 2.3.8 (Sum over Weingarten functions with trivial weights)**.** *The sum over Weingarten functions*

$$W(\sigma, k) = \frac{1}{(k!)^2} \sum_{\lambda \vdash k} \frac{\left(\dim S_k^\lambda\right)^2}{\dim Gl^\lambda(\mathcal{H}_1)} \chi^\lambda(\sigma)$$

simplifies to

$$\sum_{\sigma \in S_n} W(\sigma, n) = \frac{1}{\prod_{c=0}^{n-1}(d+c)}.$$

Proof.

$$\sum_{\sigma\in S_n} W(\sigma,n) = \sum_{\sigma\in S_n}\frac{1}{(n!)^2}\sum_{\lambda\vdash n}\frac{\chi^\lambda(e)^2}{\prod_{c\in\lambda}\frac{d+c(u)}{h(u)}}\chi^\lambda(\sigma) = \frac{1}{(n!)^2}\sum_{\lambda\vdash n}\frac{\chi^\lambda(e)^2}{\prod_{c\in\lambda}\frac{d+c(u)}{h(u)}}\sum_{\sigma\in S_n}\chi^\lambda(\sigma)$$
$$= \frac{1}{(n!)^2}\sum_{\lambda\vdash n}\frac{\chi^\lambda(e)^2}{\prod_{c\in\lambda}\frac{d+c(u)}{h(u)}}\sum_{\sigma\in S_n}\chi^\lambda(\sigma)\delta_{\lambda,(n)} = \frac{1}{(n!)^2}\frac{1^2}{\prod_{c\in(n)}\frac{d+c(u)}{h(u)}}n!$$

□

2.3.3. Integration over only one row

We can now use our previous result to simplify Collins and Śniady's formula for the specific case where we only integrate over one row, that is, if all the $\tilde{i}_i$ and $\tilde{j}_i$ are equal:

$$\int_{\mathcal{U}(d)} d\mathcal{U}\,\mathcal{U}_{i_1\tilde{i}_1}\cdots\mathcal{U}_{i_k\tilde{i}_k}\overline{\mathcal{U}_{j_1\tilde{j}_1}}\cdots\overline{\mathcal{U}_{j_k\tilde{j}_k}} = \int_{\mathcal{U}(d)} d\mathcal{U}\,\mathcal{U}_{i_1\tilde{i}}\cdots\mathcal{U}_{i_k\tilde{i}}\overline{\mathcal{U}_{j_1\tilde{i}}}\cdots\overline{\mathcal{U}_{j_k\tilde{i}}}.$$

Theorem 2.3.9 (Integration over one Row)**.** *An integral over elements of entries of unitary matrices with respect to the Haar measure restricted to only one row is given by:*

$$\int_{\mathcal{U}(d)} d\mathcal{U}\,\mathcal{U}_{i_1\tilde{i}}\cdots\mathcal{U}_{i_k\tilde{i}}\overline{\mathcal{U}_{j_1\tilde{i}}}\cdots\overline{\mathcal{U}_{j_k\tilde{i}}} = \frac{1}{\prod_{c=0}^{k-1}(d+c)}\sum_{\sigma\in\mathcal{S}_k}\delta_{i_1 j_{\sigma(1)}}\cdots\delta_{i_k j_{\sigma(k)}}.$$

Proof. The general formula simplifies to:

$$\int_{\mathcal{U}(d)} d\mathcal{U}\,\mathcal{U}_{i_1\tilde{i}}\cdots\mathcal{U}_{i_k\tilde{i}}\overline{\mathcal{U}_{j_1\tilde{i}}}\cdots\overline{\mathcal{U}_{j_k\tilde{i}}} = \sum_{\sigma,\tau\in\mathcal{S}_k}\delta_{i_1 j_{\sigma(1)}}\cdots\delta_{i_k j_{\sigma(k)}}\,\delta_{\tilde{i},\tilde{i}}\cdots\delta_{\tilde{i},\tilde{i}}W(\tau\sigma^{-1},k),$$

where we easily see that the $\delta_{\tilde{i},\tilde{i}}$ do not restrict anything and are just always 1.
Next we see that since all the δs are independent of the τ-summation, we can pull the τ-summation through and are effectively left with:

$$\sum_{\tau\in S_k} W(\tau\sigma^{-1},k),$$

where σ is an arbitrary but fixed element of S_k.
If we multiply all elements of S_k with a fixed element of S_k we will still get all elements of S_k — only in a different order. Since we sum over all elements and the order does not matter, we are left with:

$$\sum_{\tau\in S_k} W(\tau\sigma^{-1},k) = \sum_{\tau\in S_k} W(\tau,k) = \frac{1}{\prod_{c=0}^{k-1}(d+c)}.$$

The last equal sign follows from Theorem 2.3.8. Note that the result is *independent* of σ. □

2.3.4. Cycles and Traces, a First Observation

Let us get back to the example of the 2-correlator:

$$I(A,B) = \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \int\limits_{\mathcal{U}(d)} d\mathcal{U}\, U_{i_1,1} U_{i_2,1} \overline{U_{j_1,1}}\, \overline{U_{j_2,1}}.$$

With the simplified formula, we can evaluate the integral much more quickly:

$$I(A,B) = \frac{1}{d\,(d+1)} \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \left(\sum_{\sigma \in S_2} \delta_{i_1,j_{\sigma(1)}} \delta_{i_2,j_{\sigma(2)}} \right).$$

All that is left to do is evaluate the simple summations. Again we see that we will get all possible products of traces, but we want to investigate this more carefully. The group S_2 has only two elements. We will denote them in cycle notation and write $[(1)(2)]$ and $[(12)]$. The brackets around the cycles shall only enhance readability and have no further meaning. We can compute:

$$\begin{aligned} I(A,B) &= \frac{1}{d(d+1)} \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \left(\delta_{i_1,j_{[(1)(2)](1)}} \delta_{i_2,j_{[(1)(2)](2)}} + \delta_{i_1,j_{[(1,2)](1)}} \delta_{i_2,j_{[(1,2)](2)}} \right) \\ &= \frac{1}{d(d+1)} \sum_{i_1,j_1} a_{j_1,i_1} \sum_{i_2,j_2} b_{j_2,i_2} \left(\delta_{i_1,j_1} \delta_{i_2,j_2} + \delta_{i_1,j_2} \delta_{i_2,j_1} \right) \\ &= \frac{1}{d(d+1)} \left(\mathrm{Tr}\,(A)\,\mathrm{Tr}\,(B) + \mathrm{Tr}\,(AB) \right). \end{aligned}$$

We see here for the first time that the operators are paired within the traces according to the pairings within the cycles. We will now investigate integrals of a very similar kind with more operators and find the same outcome.

2.3.5. Q-Correlator

Consider the correlator of q operators, given by:

$$I(A^1, \ldots, A^q) = \int\limits_{\mathcal{U}(d)} d\mathcal{U}\, \langle 1|\, \mathcal{U}^\dagger A^1 \mathcal{U}\, |1\rangle \cdots \langle 1|\, \mathcal{U}^\dagger A^q \mathcal{U}\, |1\rangle\,.$$

Similarly to before, we can rewrite this in matrix elements:

$$I(A^1, \ldots, A^q) = \sum_{i_1,j_1} a^1_{j_1,i_1} \ldots \sum_{i_q,j_q} a^n_{j_q,i_q} \int\limits_{\mathcal{U}(d)} d\mathcal{U}\, U_{i_1,1} \ldots U_{i_q,1}\, \overline{U_{j_1,1}} \ldots \overline{U_{j_q,1}}$$

and again use the one row formula:

$$I(A^1, \ldots, A^q) = \frac{1}{\prod_{c=0}^{q-1}(d+c)} \sum_{i_1,j_1} a^1_{j_1,i_1} \ldots \sum_{i_q,j_q} a^n_{j_q,i_q} \sum_{\sigma \in S_q} \delta_{i_1,j_{\sigma(1)}} \cdots \delta_{i_q,j_{\sigma(q)}},$$

and then realize that there is no reason for the sum over S_q to be on the right; all elements to the left of it are independent of the permutation, so we can move the sum over to the left:

$$I(A^1,\dots,A^q)=\frac{1}{\prod_{c=0}^{q-1}(d+c)}\sum_{\sigma\in S_q}\sum_{i_1,j_1}a^1_{j_1,i_1}\cdots\sum_{i_q,j_q}a^n_{j_q,i_q}\delta_{i_1,j_{\sigma(1)}}\cdots\delta_{i_q,j_{\sigma(q)}}.$$

This way we understand that the actually interesting part is the summation over σ. To get a better perspective let us assign names to the σ-dependent summands.

Definition 2.3.10.

$$I(A^1,\dots,A^q)=\frac{1}{\prod_{c=0}^{q-1}(d+c)}\sum_{\sigma\in S_q}W(A^1,\dots,A^q)_\sigma$$
$$W(A^1,\dots,A^q)_\sigma=\sum_{i_1,j_1}a^1_{j_1,i_1}\cdots\sum_{i_q,j_q}a^n_{j_q,i_q}\delta_{i_1,j_{\sigma(1)}}\cdots\delta_{i_q,j_{\sigma(q)}}$$

Once we understand how W depends on the different σ, we understand I.

Example 2.3.11. *Let us set $q=3$ and calculate $W(A^1,A^2,A^3)_\sigma$ for $\sigma_1=(1,2)(3)$, $\sigma_2=(1,2,3)$ and $\sigma_3=(1,3,2)$.*

1. *First we compute:*

$$W(A^1,A^2,A^3)_{\sigma_1}=\sum_{i_1,j_1}\sum_{i_2,j_2}\sum_{i_3,j_3}a^1_{j_1,i_1}\,a^2_{j_2,i_2}\,a^3_{j_3,i_3}\,\delta_{i_1,j_2}\delta_{i_2,j_1}\delta_{i_3,j_3}.$$

Now we collect the different elements together so we can evaluate the sums:

$$W(A^1,A^2,A^3)_{\sigma_1}=\left(\sum_{i_1,j_1}\sum_{i_2,j_2}a^1_{j_1,i_1}a^2_{j_2,i_2}\delta_{i_1,j_2}\delta_{i_2,j_1}\right)\left(\sum_{i_3,j_3}a^3_{i_3,j_3}\delta_{i_3,j_3}\right).$$

We see how the summations can be grouped together as they are group together in distinct cycles. The summations we are left with are very easy:

$$W(A^1,A^2,A^3)_{\sigma_1}=\left(\sum_{i_1}\sum_{i_2}a^1_{i_2,i_1}a^2_{i_1,i_2}\right)\left(\sum_{i_3}a^3_{i_3,i_3}\right)=Tr\left(A^1\,A^2\right)\,Tr\left(A^3\right).$$

2. *Again, we begin with:*

$$W(A^1,A^2,A^3)_{\sigma_2}=\sum_{i_1,j_1}\sum_{i_2,j_2}\sum_{i_3,j_3}a^1_{j_1,i_1}\,a^2_{j_2,i_2}\,a^3_{j_3,i_3}\,\delta_{i_1,j_2}\delta_{i_2,j_3}\delta_{i_3,j_1}.$$

This time it is not possible to group different sums together — as expected, there are no distinct cycles. We get:

$$W(A^1,A^2,A^3)_{\sigma_2}=\sum_{i_1}\sum_{i_2}\sum_{i_3}a^1_{i_3,i_1}\,a^2_{i_1,i_2}\,a^3_{i_2,i_3}=Tr\left(A^1\,A^2\,A^3\right).$$

3. *Finally, for $\sigma_3 = (1,3,2)$, we have:*

$$W(A^1, A^2, A^3)_{\sigma_3} = \sum_{i_1,j_1} \sum_{i_2,j_2} \sum_{i_3,j_3} a^1_{j_1,i_1}\, a^2_{j_2,i_2}\, a^3_{j_3,i_3}\, \delta_{i_1,j_3} \delta_{i_2,j_1} \delta_{i_3,j_2}.$$

What looks like a minor difference and is actually hard to tell will result in a different order in the trace:

$$W(A^1, A^2, A^3)_{\sigma_3} = \sum_{i_1} \sum_{i_2} \sum_{i_3} a^1_{i_2,i_1}\, a^2_{i_3,i_2}\, a^3_{i_1,i_3} = Tr\left(A^1\, A^3\, A^2\right).$$

In the example we saw that $W(A^1, \ldots, A^q)_\sigma$ preserves the structure of σ completely. It seems appropriate to drop the W and define the following:

Definition 2.3.12 (Trace Product for a Permuation)**.** *For a list of operators $A^1, \ldots, A^q$ and a permutation $\sigma \in S_q$, the trace product is defined:*

$$\left(A^1, \ldots, A^q\right)_\sigma = W\left(A^1, \ldots, A^q\right)_\sigma.$$

This allows us to state the very handy theorem:

Theorem 2.3.13 (q-correlator)**.** *For a list of operators $A^1, \ldots, A^q$,*

$$I(A^1, \ldots, A^q) = \frac{1}{\prod_{c=0}^{q-1}(d+c)} \sum_{\sigma \in S_q} \left(A^1, \ldots, A^q\right)_\sigma.$$

Proof. One might find it naive that we claim that with Example 2.3.11 all the work is already done. But consider the following.
If the statement is true for a specific q_0, evaluate it for $q_0 + 1$ operators. For all elements of S_{q_0+1} where the $(q+1)$th element is in a distinct cycle by itself, it is true trivially.
For the other elements the A_{q+1} operator will be somewhere in a cycle and be paired with the neighbouring operators while the structure is preserved as shown in 2. and 3. of Example 2.3.11. □

2.3.6. Afterthought: Dimension of a Permutation

With the trace product for a permutation (Definition 2.3.12) we can finally explain the notion of the dimension of a permutation (Definition 2.1.15). Consider the situation that all operators are the identity operator $\mathbb{1}_d$ on a vector space $\mathbb{C}^d$. Then for a permutation σ with $\dim \sigma = \delta$ we have

$$(\mathbb{1}_d, \ldots, \mathbb{1}_d)_\sigma = d^\delta,$$

because each operator product will, in the end, still be the identity operator with trace d. Since for each cycle there is a product, we collect, for each cycle, one factor of d.

3. Channel Fidelity

In Chapter 1, we explained why the study of tensor product spaces is relevant for the study of quantum information transmission. In Chapter 2, we showed how to calculate averages over unitary groups. Now we will introduce the channel fidelity as a relevant quantity in the context of quantum information theory.
First we will investigate it using the averaging methods from Chapter 2, and later, in continuation of former work [11], we will show a classical maximization algorithm and address a suggested improvement to it.

3.1. Motivation

Imagine an experiment where we can prepare the input state as a pure state, $|\phi\rangle\langle\phi|$, which we then transmit using the channel $\mathcal{N}$. In the end we check the performance, or rather, the effect of the channel, by comparing input and output state using the Uhlmann fidelity (see Proposition 1.2.10). This motivates us to define the Channel Fidelity as the Uhlmann Fidelity of the input and the output.

Definition 3.1.1 (Channel Fidelity)**.** *For an input state $|\phi\rangle$ and a quantum channel $\mathcal{N}$, their channel fidelity $F(|\phi\rangle, \mathcal{N})$ is given by:*

$$F(|\phi\rangle, \mathcal{N}) = \langle\phi| \mathcal{N}(|\phi\rangle\langle\phi|) |\phi\rangle .$$

In Chapter 1 we motivated that in the context of quantum information theory it is actually not sufficient to study a single channel. It is rather necessary to consider the n-fold tensor product of identical and independent copies of the channel, which will allow entanglement effects to occur.

Definition 3.1.2 (n-Shot Channel Fidelity)**.** *For a quantum channel $\mathcal{N}$, acting on the linear operators on $\mathcal{H}$, we will now want to consider its n-fold tensor product $\mathcal{N}^{\otimes n}$. By definition this will act on the linear operators on $\mathcal{H}^{\otimes n}$. For a channel $\mathcal{N}$, we define the n-shot channel fidelity by:*

$$F_n(|\phi\rangle, \mathcal{N}) = F(|\phi\rangle, \mathcal{N}^{\otimes n}).$$

The input state must be in the corresponding Hilbert space: $|\phi\rangle \in \mathcal{H}^{\otimes n}$.

For product states $|\phi\rangle = |\phi_1\rangle^{\otimes n}$ the n-shot channel fidelity is simply the nth power of the channel fidelity for $|\phi_1\rangle$.

An advantage of the channel fidelity is that it is easy enough to handle, and yet complicated enough that it will show non-trivial entanglement dependent behavior. However, to avoid misunderstandings, we mention that the channel fidelity does not say much on the ability of a channel for quantum communication, although it might look like that at first glance. The capability of a channel to transmit quantum information is not so much about preserving a specific state, but rather about the leakage of information to the environment (see Theorem 1.5.2).

Example 3.1.3 (Gate Fidelity)**.** *Other authors have found another relevant motivation of the same functional [4, 18]: It has been shown that one way of realizing a universal quantum computer is using quantum gates [16].*
A quantum gate as an ideal quantum mechanical process is a unitary transformation. Though engineers would certainly try to isolate the system as much as possible from the outside and create a closed system, nature will only allow this in an approximate way. Even small correlations with the environment will lead to an open system, so the ideal unitary gate U, with unitary matrix $\mathcal{U}$, will be implemented as a noisy quantum channel $\mathcal{N}$.
Now we can use the fidelity as a measure of how well $\mathcal{N}$ simulates U. Again starting with a pure input state $\Phi = |\phi\rangle\langle\phi|$, $\mathcal{U}\Phi\mathcal{U}^\dagger$ is still pure so we again use the fidelity Proposition 1.2.10 as our measure:

$$F(\mathcal{U}\,\Phi\,\mathcal{U}^\dagger, \mathcal{N}(\Phi)) = Tr\Big((\mathcal{U}\,\Phi\,\mathcal{U}^\dagger\,\mathcal{N}(\Phi))\Big) = Tr\Big((\Phi\,\mathcal{U}^\dagger\,\mathcal{N}(\Phi)\,\mathcal{U})\Big).$$

Because of cyclic invariance of the trace, we can move the unitary transformation away from the input state and absorb it into the quantum channel. The resulting quantum channel $\mathcal{N}'$ shows how $\mathcal{N}$ deviates from $\mathcal{U}$:

$$F(\mathcal{U}\,\Phi\,\mathcal{U}^\dagger, \mathcal{N}(\Phi)) = F(|\phi\rangle, \mathcal{N}').$$

Note that if $\mathcal{N}$ is unital, so is $\mathcal{N}'$.

3.2. Average Channel Fidelity for multiple Channels

Now, similarly to [11], we begin our investigation by defining the average channel fidelity the average channel fidelity. In Chapter 2, we showed how to average over all states of the according Hilbert space.

Definition 3.2.1 (Average Channel Fidelity)**.** *Let $\mathcal{U}(d)$ be the unitary group acting on a Hilbert space $\mathcal{H} = \mathbb{C}^d$, and $d\mathcal{U}$ the Haar measure on $\mathcal{U}(d)$. Then for a quantum channel $\mathcal{N}$, we define its average channel fidelity as:*

$$F(\mathcal{N}) = \int_{\mathcal{U}(d)} d\mathcal{U}\,\langle 0|\,\mathcal{U}^\dagger\mathcal{N}\left(\mathcal{U}\,|0\rangle\langle 0|\,\mathcal{U}^\dagger\right)\mathcal{U}\,|0\rangle.$$

We are interested in entanglement effects. Since in Definition 3.2.1 we have no restrictions on the structure of the channel, we can naturally calculate the average n-shot channel fidelity from Definition 3.1.2. The overall average will become smaller, simply because we are dealing with a larger Hilbert space, but we are interested in the average per channel, and how this differs from the average for a single channel. This means that we need to regularize here, by taking the n-th root, since there is no logarithm involved.

Definition 3.2.2 (Regularized Average Channel Fidelity)**.** *Consider the same situation as in Definition 3.1.2. The regularized average channel fidelity $F_n(\mathcal{N})$ is defined by:*

$$F_n(\mathcal{N}) = \left(F(\mathcal{N}^{\otimes n})\right)^{\frac{1}{n}}.$$

Now we can calculate these averages using the result of Example 2.3.4 or the general formula from Theorem 2.3.13.

Theorem 3.2.3 (Average Fidelities)**.** *For a quantum channel $\mathcal{N}$ with Kraus operators A_i, its average fidelity is given by:*

$$F(\mathcal{N}) = \frac{d + \sum_i |\operatorname{Tr}(A_i)|^2}{d\,(d+1)}.$$

Furthermore we can state its regularized average channel fidelity,

$$F_n(\mathcal{N}) = \left(\frac{d^n + \left[\sum_i |\operatorname{Tr}(A_i)|^2\right]^n}{d^n\,(d^n+1)}\right)^{\frac{1}{n}},$$

which we can restate as a function of the average channel fidelity,

$$F_n(F(\mathcal{N})) = \left(\frac{1 + [(d+1)F(\mathcal{N}) - 1]^n}{d^n + 1}\right)^{\frac{1}{n}}.$$

Note that the minimal average fidelity for any channel is $\frac{1}{d+1}$. This is achieved for a Pauli channel, Theorem 1.4.3, where the probability that the input state gets through the channel unaffected is 0. The completely depolarizing channel (Definition 1.4.7), however, has a larger average fidelity of $\frac{1}{d}$. The regularized average channel fidelity shows non-trivial n dependence: channel fidelity and the regularized counterpart are only equal for the identity and the completely depolarizing channel — not for a general depolarizing channel (see Proposition 1.4.6). We will discuss this in more detail after the proof.

The following lemma makes the proof more understandable.

Lemma 3.2.4 (Trace Factorization)**.** *Given operators $A^q_{\vec{i}}$ that have a tensor product structure, $A^q_{\vec{i}} = A^q_{i_1} \otimes \cdots \otimes A^q_{i_n}$, using our compact notation from Definition 2.3.12, we have:*

$$(A^1_{\vec{i}}, \ldots, A^q_{\vec{i}})_\sigma = (A^1_{i_1}, \ldots, A^q_{i_1})_\sigma \cdots (A^1_{i_n}, \ldots, A^q_{i_n})_\sigma.$$

Proof. The lemma is just an extension of the trivial statement that for two operators A and B, the trace factorizes their tensor product:

$$\mathrm{Tr}\,(A \otimes B) = \mathrm{Tr}\,(A)\ \mathrm{Tr}\,(B)\,.$$

Since there is no restriction to either of those operators, we can easily do it one tensor factor at a time and get the product result. Furthermore, a cycle is just a product of traces. Such the statement is true for each factor. The factors then again can be brought into the original form and written in the compact form as done in the statement. □

Now we can prove Theorem 3.2.3.

Proof. Let us rewrite $F(\mathcal{N})$ in terms of Example 2.3.4 and use its result or the result for the general q-correlator from Theorem 2.3.13:

$$\begin{aligned} F(\mathcal{N}) &= \int\limits_{\mathcal{U}(d)} d\mathcal{U}\, \langle 0|\mathcal{U}^\dagger \mathcal{N}\left(\mathcal{U}\,|0\rangle\,\langle 0|\,\mathcal{U}^\dagger\right)\mathcal{U}\,|0\rangle \\ &= \sum_i I(A_i, A_i^\dagger) = \sum_i \frac{1}{d\,(d+1)} \sum_{\sigma\in S_2} \left(A_i, A_i^\dagger\right)_\sigma \\ &= \sum_i \frac{\mathrm{Tr}\,(A_i)\,\mathrm{Tr}\left(A_i^\dagger\right) + \mathrm{Tr}\left(A_i A_i^\dagger\right)}{d(d+1)}. \end{aligned}$$

For trace preserving quantum channels, we have $\sum_i \mathrm{Tr}\left(A_i A_i^\dagger\right) = \mathrm{Tr}\,(\mathbb{1}) = d$.
Next we notice that for a quantum channel $\mathcal{N}$ with Kraus operators $\{A_i\}_i$, its n-fold tensor product $\mathcal{N}^{\otimes n}$ has Kraus operators $\{A_{\vec{i}} = A_{i_1} \otimes \cdots \otimes A_{i_n}\}_{i_1,\ldots,i_n}$. The Hilbert space dimension changes accordingly to d^n. We use these simplifications in the calculation of $F_n(\mathcal{N})$:

$$F_n(\mathcal{N}) = \left(\sum_{\vec{i}} I(A_{\vec{i}}, A_{\vec{i}}^\dagger)\right)^{\frac{1}{n}} = \left(\sum_{\vec{i}} \frac{1}{d^n\,(d^n+1)} \sum_{\sigma\in S_2} \left(A_{\vec{i}}, A_{\vec{i}}^\dagger\right)_\sigma\right)^{\frac{1}{n}}.$$

Then we use Lemma 3.2.4:

$$\left(A_{\vec{i}}, A_{\vec{i}}^\dagger\right)_\sigma = \left(A_{i_1}, A_{i_1}^\dagger\right)_\sigma \cdots \left(A_{i_n}, A_{i_n}^\dagger\right)_\sigma,$$

so we can write:

$$\sum_{\vec{i}} \left(A_{\vec{i}}, A_{\vec{i}}^{\dagger} \right)_{\sigma} = \sum_{i_1} \left(A_{i_1}, A_{i_1}^{\dagger} \right)_{\sigma} \cdots \sum_{i_n} \left(A_{i_n}, A_{i_n}^{\dagger} \right)_{\sigma} = \left[\sum_{i} \left(A_i, A_i^{\dagger} \right)_{\sigma} \right]^n .$$

Noticing that all these summations are actually over identical sets, we can get back to the simple i.
After exchanging the two summations we get:

$$F_n(\mathcal{N}) = \left(\frac{1}{d^n \left(d^n + 1 \right)} \sum_{\sigma \in S_2} \left[\sum_{i} \left(A_i, A_i^{\dagger} \right)_{\sigma} \right]^n \right)^{\frac{1}{n}},$$

which is equal to the final form of $F_n(\mathcal{N})$. The expression on the right is then attained by noticing that, for a trace preserving quantum channel $\mathcal{N}$, its n-fold tensor product is also trace preserving.
Finally we transform the equation for $F(\mathcal{N})$,

$$d\left(d+1\right)F(\mathcal{N}) - d = \sum_{i} |\mathrm{Tr}\left(A_i\right)|^2,$$

and put this in the expression for $F_n(\mathcal{N})$:

$$F_n(F(\mathcal{N})) = \left(\frac{d^n + \left[d\left(d+1\right) F(\mathcal{N}) - d \right]^n}{d^n \left(d^n + 1 \right)} \right)^{\frac{1}{n}} .$$

Canceling d^n in all terms leads to the final expression. □

We begin our discussion of these results by stating that they have been calculated before, just the average fidelity by [4] and [19], and all three — the average fidelity, the regularized average fidelity, and the regularized average fidelity as a function of average fidelity — in [11]. However, we have formalized the proof, which will be useful later in the calculation of higher order moments. In the last reference, the results are also discussed in more detail.
We can now motivate the channel fidelity as an interesting property to study. In Figure 3.1, one can see that the average shows a strong entanglement dependence. Only for $F(\mathcal{N}) = \frac{1}{2}$, which is the fidelity for the completely depolarizing channel, and $F(\mathcal{N}) = 1$, which is only achieved for the identity channel, the regularized channel fidelity is equal to the single channel fidelity. For highly disturbed channels — that is, channels with a low channel fidelity — we have an increase in the average per channel, whereas for those channels with high fidelity, the average per channel decreases.
For a single channel fidelity between $\frac{1}{3}$ and $\frac{2}{3}$, the regularized fidelity is constant and equal to $\frac{1}{d}$, independent of the channel.
We saw in Example 2.3.2 that the average overlap of two random unit vectors is also equal to $\frac{1}{d}$. This means that in this region the high tensor powers of every channel, on

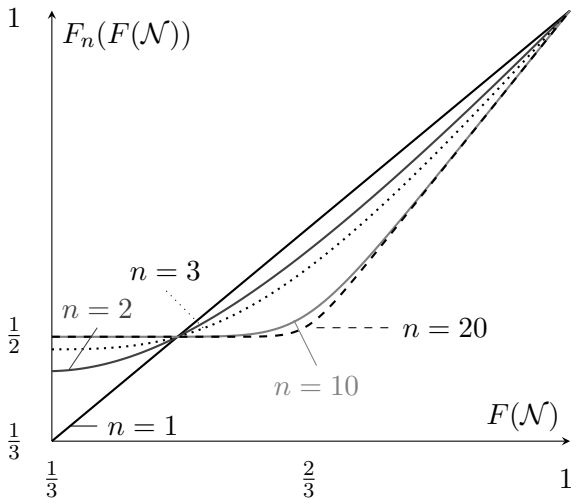


Figure 3.1.: Regularized average channel fidelity as a function of the channel fidelity; $d = 2$

average, completely forget anything about the input state, much like the completely depolarizing channel, introduced in Definition 1.4.7. Still, they are more interesting than the completely depolarizing channel, because they have a non-trivial maximum, which we will discuss in the end of this chapter.

Furthermore, it is somewhat surprising that the average (single-shot) channel fidelity can be worse than the regularized version for highly disturbed channels, even worse than the completely depolarizing channel. We can interpret this as follows: To achieve these low fidelities, a channel cannot be very random in the sense that it will randomize the input state. It has to be deliberately bad, and keep a strong knowledge about the input state. Here we can see that the fidelity does not measure the capability of information transmission. A simple bit flip channel that flips the qubit with 100% probability certainly is a unitary evolution, and as such could easily be corrected, but it has the lowest possible fidelity.

It is most important to realize that we can actually calculate the limit $n \to \infty$! Our observations about the regularized fidelity in Figure 3.1 are not special for $d = 2$. We will see this now, when we formalize this observation further.

Proposition 3.2.5 (Single Letter Formula for F_n)**.** *For the regularized average channel fidelity we can give single letter formulas, see Definition 1.5.6, in different regimes: For a highly disturbed single channel $\mathcal{N}$ with $\frac{1}{d+1} \leq F(\mathcal{N}) \leq \frac{2}{d+1}$, we have:*

$$\lim_{n\to\infty} F_n(F(\mathcal{N})) = \frac{1}{d}.$$

For a weakly disturbed single channel with $\frac{2}{d+1} < F(\mathcal{N}) \leq 1$*, we have:*

$$\lim_{n\to\infty} F_n(F(\mathcal{N})) = \frac{(d+1)F(\mathcal{N}) - 1}{d}. \tag{3.1}$$

Note that the notions of highly and weakly disturbed only really make sense for a low dimensional qudit space. For large qudit spaces we cannot see any entanglement effect; since in the limit $d \to \infty$ only Equation 3.1 applies and shows that $F_\infty = F$.

Proof. Looking at the formula for $F_n(F(\mathcal{N}))$,

$$F_n(F(\mathcal{N})) = \left(\frac{1 + [(d+1)F(\mathcal{N}) - 1]^n}{d^n + 1}\right)^{\frac{1}{n}},$$

we see that the limit $n \to \infty$ is trivial for the denominator, however the numerator is more interesting, especially the second term $P(F(\mathcal{N}))$:

$$P(F(\mathcal{N})) := (d+1)F(\mathcal{N}) - 1.$$

For the highly disturbed case we have

$$0 \leq P(F(\mathcal{N})) < 1,$$

which results in an easy limit for the numerator, which proves the first statement.

In the weakly disturbed case we have

$$1 < P(F(\mathcal{N})),$$

which means it will be the dominating term in the numerator in the limit, giving:

$$\lim_{n\to\infty} F_n(F(\mathcal{N})) \approx \lim_{n\to\infty} \left(\frac{P(F(\mathcal{N}))^n}{d^n}\right)^{\frac{1}{n}} = \frac{P(F(\mathcal{N}))}{d}.$$

□

In order to analyse these results, we will restrict ourselves to a more concrete, but still very general, class of channels, the Pauli channels, from Theorem 1.4.3.

Proposition 3.2.6 (Regularized Average Pauli Channel Fidelity)**.** *The regularized average channel fidelity for Pauli channel* $\mathcal{N}_{\{p_i\}_{i=0}^3}$ *is,*

for $p_0 > \frac{1}{2}$

$$F_\infty(\mathcal{N}_{\{p_i\}_{i=0}^3}) = \lim_{n\to\infty} \left[F_n\left(\mathcal{N}_{\{p_i\}_{i=0}^3}\right)\right]^{\frac{1}{n}} = p_0,$$

and for $p_0 \leq \frac{1}{2}$

$$F_\infty(\mathcal{N}_{\{p_i\}_{i=0}^3}) = \lim_{n\to\infty} \left[F_n\left(\mathcal{N}_{\{p_i\}_{i=0}^3}\right)\right]^{\frac{1}{n}} = \frac{1}{2}.$$

Pauli channels in general still have 3 parameters, which is a complication for plots. However we can investigate the depolarizing channel, see Definition 1.4.5, which is a suitable candidate for several reasons:

- For qubits, it can be written in terms of Pauli matrices with $p_0 = p$ and all other $p_i = \frac{1-p}{3}$, see Proposition 1.4.6, which means $p = \frac{3\lambda+1}{4}$.
- For a single channel, its channel fidelity is constant regardless of the input,

$$F(|\phi\rangle, \mathcal{N}_p) = \frac{1+2p}{3},$$

  which can be seen very easily in Definition 1.4.5.
- For more channels, the fidelity does depend on the input, as we will see later.
- For $p = 1$ it is the identity channel.
- For $p = \frac{1}{4}$ it is called the completely depolarizing channel, which is equivalent to $\lambda = 0$ in Definition 1.4.5.
- It is known to have non-trivial coherent information, [36, 13].

Obviously for the identity and the completely depolarizing channel, the fidelity is always constant, no matter how many channels we use. That means that only for specific values of p will we see entanglement independence, whereas for most p the average will depend on the number of channels.
In Figure 3.2, we see that the depolarizing qubit channel is a suitable exemplary channel, as it very much resembles Figure 3.1, which we have already discussed. Additionally, we plot F_∞ and we can see that F_n approaches F_∞ very quickly. For most p, F_∞ almost perfectly resembles F_n already for $n = 10$ channels. Only around $p = 0.5$ is there a small deviation. If we want to consider the average fidelity for more than 20 channels, it is justified to consider $(F_\infty)^n$ instead of the real average, $(F_n)^n$. Whereas we can certainly see that the regularized average is a function of the number of channels, we cannot yet see that the channel fidelity is no longer constant for each p, independent of the input. We will get back to this point later, when discussing higher moments.
As we are considering tensor products of depolarizing channels, one could have the idea that a tensor product of depolarizing channels resembles the depolarizing channel of the appropriate dimension. More precisely we can ask the question: Is the depolarizing channel with $d = 4$ the same channel as a tensor product of two depolarizing channel with $d = 2$. In Figure 3.3, we plot their average (non-regularized) fidelities. We can see that they are certainly not the same channel.

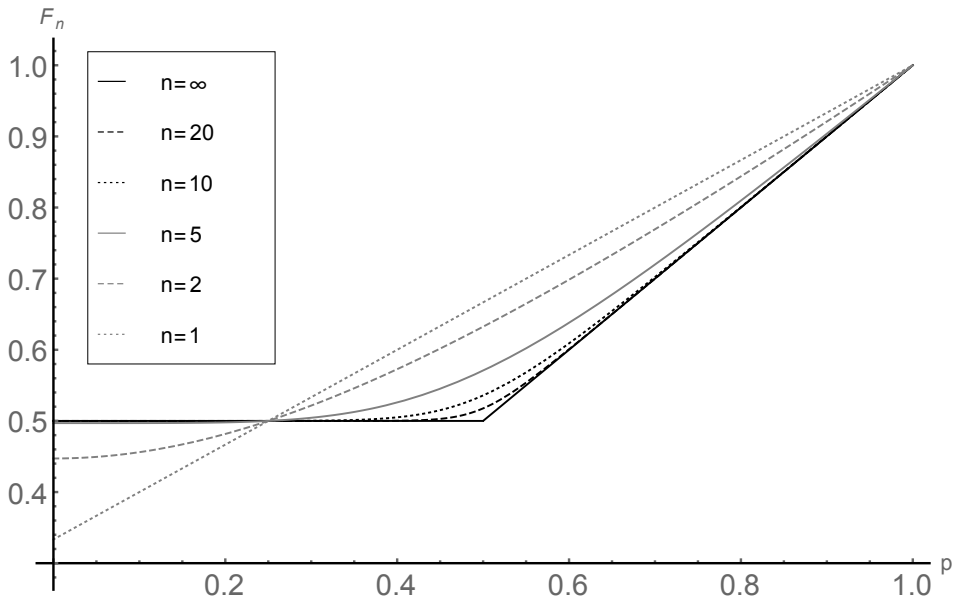


Figure 3.2.: Regularized average channel fidelities for n depolarizing qubit channels

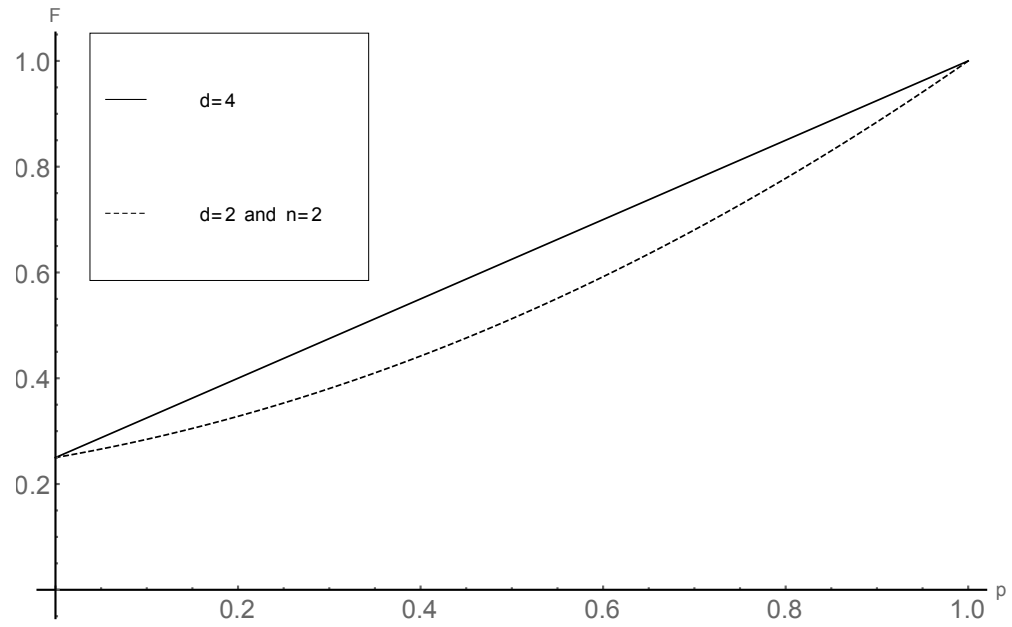


Figure 3.3.: Average channel fidelities for depolarizing channel with $d = 4$, and $\left(\mathcal{N}_\lambda^D\right)^{\otimes 2}$ for $d = 2$

Closing our discussion for now, we can interpret the channel fidelity again as a gate fidelity — see Example 3.1.3. The highly disturbed case — with low gate fidelity — would not be very interesting. A gate that is implemented this badly can be said not to be implemented at all.
However, we can say that for quite accurate gates, an implementation of a large number of the same gate would, on average and per gate, still be implemented very well. A large number of qubit gates, each with accuracy 99%, would be implemented with accuracy per gate of 98.5%.

In this section we could verify that the channel fidelity shows entanglement effects, and yet it is simple enough to allow us to find single letter formulas. Furthermore, we established the depolarizing channel for $d = 2$ as an interesting exemplary channel. We will continue with the investigation of higher order moments.

3.3. Higher Order Moments

We want to learn more about the channel fidelity distribution. We will now introduce and calculate all fidelity moments, first for the single channel and then for the n-channel case.
Motivated by our success in finding and evaluating a single letter formula for the first moment of the channel fidelity, we also seek to find simple formulas for the higher moments.

Definition 3.3.1 (q-th Moment)**.** *For a quantum channel $\mathcal{N}$ with Kraus operators $\{A_i\}_i$, its q-th channel fidelity moment is defined as*

$$\begin{aligned} F^q(\mathcal{N}) &= \int\limits_{U(d)} dU\, F(U|0\rangle, \mathcal{N})^q \\ &= \int\limits_{U(d)} dU \left(\langle 0| U^\dagger \mathcal{N}(U|0\rangle\langle 0| U^\dagger) U|0\rangle \right)^q \\ &= \sum_{\vec{i}} \int\limits_{U(d)} dU \langle 0| U^\dagger A_{i_1} U|0\rangle \langle 0| U^\dagger A_{i_1}^\dagger U|0\rangle \cdots \langle 0| U^\dagger A_{i_q} U|0\rangle \langle 0| U^\dagger A_{i_q}^\dagger U|0\rangle . \end{aligned}$$

Note that for the q-th moment we need q sets of Kraus operators, so our vector valued index here has q components instead of n as before: $\vec{i} = (i_1, \ldots, i_q)$.

The extension to the n-channel case is done by extension to an n-fold tensor product.

Definition 3.3.2 (q-th Moment for n-Channel)**.**

$$\begin{aligned} F_n^q(\mathcal{N}) &= \int\limits_{U(d^n)} dU\, F(U|0\rangle, \mathcal{N}^{\otimes n})^q \\ &= \int\limits_{U(d^n)} dU \left(\langle 0| U^\dagger \mathcal{N}^{\otimes n}(U|0\rangle\langle 0| U^\dagger) U|0\rangle \right)^q . \end{aligned}$$

We refrain from stating this expression in Kraus operators. In order to avoid confusion with the Kraus operators that come from the n channel and the ones that come from the q-th moment. Fortunately, in the main theorem, this confusion will not be an issue.

The aim is now to find an expression for both $F^q(\mathcal{N})$ and $F_n^q(\mathcal{N})$ in terms of their Kraus operators, where we can now really see the advantages of Definition 2.3.12.

Theorem 3.3.3 (q-th Moment)**.** *For a quantum channel $\mathcal{N}$ with Kraus operators $\{A_i\}_i$, the q-th moment of its channel fidelity is:*

$$F^q(\mathcal{N}) = \frac{1}{\prod_{c=0}^{2q-1}(d+c)} \sum_{\sigma \in S_{2q}} \left[\sum_{\vec{i}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger \right)_\sigma \right].$$

Furthermore, the q-th moment of its n-channel fidelity is:

$$F_n^q(\mathcal{N}) = \frac{1}{\prod_{c=0}^{2q-1}(d^n+c)} \sum_{\sigma \in S_{2q}} \left[\sum_{\vec{i}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger \right)_\sigma \right]^n.$$

Proof. Fortunately all the work has already been done in Theorem 2.3.13. The q-th moment is just a $2q$ correlator:

$$\begin{aligned} F^q(\mathcal{N}) &= \sum_{\vec{i}} I(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger) \\ &= \sum_{\vec{i}} \frac{1}{\prod_{c=0}^{2q-1}(d+c)} \sum_{\sigma \in S_{2q}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger \right)_\sigma. \end{aligned}$$

It is obvious that the two sums commute, which gives the desired expression.

Again we obtain the n-channel expression by replacing each A_{i_k} by a tensor product $A_{i_{k_1}} \otimes \cdots \otimes A_{i_{k_n}}$, where we then need to change the summation such that for each i_1 we sum over n channels; we denote this as the summation matrix I. Also we need to change the dimension from d to d^n:

$$F_n^q(\mathcal{N}) = \sum_I \frac{1}{\prod_{c=0}^{2q-1}(d^n+c)} \sum_{\sigma \in S_{2q}} \left(A_{i_{1_1}} \otimes \cdots \otimes A_{i_{1_n}}, \ldots, A_{i_{q_1}}^\dagger \otimes \cdots \otimes A_{i_{qn}}^\dagger \right)_\sigma.$$

The sums still commute:

$$F_n^q(\mathcal{N}) = \frac{1}{\prod_{c=0}^{2q-1}(d^n+c)} \sum_{\sigma \in S_{2q}} \left[\sum_I \left(A_{i_{1_1}} \otimes \cdots \otimes A_{i_{1_n}}, \ldots, A_{i_{q_1}}^\dagger \otimes \cdots \otimes A_{i_{qn}}^\dagger \right)_\sigma \right].$$

Now we use Lemma 3.2.4 and see that the n summations each have their own q summation, are independent and all equal:

$$F_n^q(\mathcal{N}) = \frac{1}{\prod_{c=0}^{2q-1}(d^n+c)} \sum_{\sigma \in S_{2q}} \left[\sum_{\vec{i}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger \right)_\sigma \right]^n.$$

□

The one channel formula has been obtained before by Magesan, Blume-Kohout and Emerson in a different fashion [4]. Whereas they reinvented a simpler version of Collins and Śniady's formula [8], we took the latter result and rigorously simplified it to the appropriate level.

The n channel formula is new. Both results are very similar to Theorem 3.2.3, however there is an important difference: For the average we only had to sum over S_2, which conveniently only has two elements.

One of the trace products resolves easily because we are handling quantum channels: $\sum_i A_i^\dagger A_i = \mathbb{1} \Rightarrow \sum_i \mathrm{Tr}\left(A_i A_i^\dagger\right) = d$. The other trace product, $\left(\mathrm{Tr}\left(A_i\right), \mathrm{Tr}\left(A_i^\dagger\right)\right)_e = \mathrm{Tr}\left(A_i\right)\ \mathrm{Tr}\left(A_i^\dagger\right) = |\mathrm{Tr}\left(A_i\right)|^2$, has no general simplification. However since both cases only depended on this expression we could easily express the n channel average in terms of the single shot average.

We see that in general this is not the case. Already for the second moment a priori we have $4! = 24$ different combinations of traces. One of them can certainly again be eliminated with the quantum channel property leaving us with 23 expressions. Symmetry arguments that we will investigate in the next section will reduce this further, however it is clear already that a 4-cycle will typically give a fundamentally different expression than a pair of transpositions.

Unfortunately, without further assumptions about the quantum channel, we can not derive a general single letter formula for the qth moment. Still more simplification is possible. For Pauli channels, we will even find more single letter formulas.

3.4. Variances for Generic Quantum Channels

In the previous section, we derived general formulas for the moments of the one shot and the n-shot channel fidelity. We will now try to understand the result. First we will argue that the real points of interest are the central moments and not the moments themselves.

Consider the situation in which at least one of the Kraus operators has a trace that is close to its dimension. This is certainly the case for high fidelity channels. In this situation, the q-th moment will be dominated by the permutation (or its corresponding trace product) with the highest dimension (see Definition 2.1.15). Since the dimension is equal to the number of cycles, the trivial permutation has the largest contribution and will dominate every moment. By studying central moments, we can find fluctuations around the average.

We will now illustrate this observation by beginning the investigation with the second moment, which we take from Theorem 3.3.3 by setting $q = 2$.

Definition 3.4.1 (2nd Moments)**.** *For a quantum channel $\mathcal{N}$ with Kraus operators*

$\{A_i\}_i$, *the second moment of its n channel fidelity is given by:*

$$F_n^2(\mathcal{N}) = \frac{1}{\prod_{c=0}^{3}(d^n + c)} \sum_{\sigma \in S_4} \left[\sum_{i_1, i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger \right)_\sigma \right]^n .$$

Note that the A_{i_1} and the A_{i_2} are from the same set of operators, however, the sums are a priori independent.

The square root of the second central moment is usually called variance and is given as follows. Note that here we actually compare the non-regularized moments, but we will want the simple Var_n later for the regularized version and hence have to pay attention to the exponent.

Proposition 3.4.2 (Variances)**.** *For a quantum channel as in Definition 3.4.1, we define the variance of its n channel fidelity by:*

$$(Var_n)^{2n} = F_n^2(\mathcal{N}) - (F_n(\mathcal{N}))^2 .$$

For the following symmetry considerations it is essential to restate the square of the averages in an alternative way.

Proposition 3.4.3 (Alternative Characterization of n Channel Fidelity Averages)**.**

$$\begin{aligned}(F_n(\mathcal{N}))^2 &= \left(\frac{1}{d^n (d^n + 1)} \sum_{\sigma \in S_2} \left[\sum_i \left(A_i, A_i^\dagger \right)_\sigma \right]^n \right)^2 \\ &= \frac{1}{d^{2n} (d^n + 1)^2} \sum_{\sigma \in S_2 \times S_2} \left[\sum_{i_1, i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger \right)_\sigma \right]^n\end{aligned}$$

In this notation, the square of the average looks structurally similar to the second moment. The first simple observation is that the element $e \in S_4$ is certainly also in $S_2 \times S_2$, thus the variance will not directly be dominated by this element.

3.4.1. General Symmetry Observations

We will now take a closer look at the variance for arbitrary channel and we will use symmetry arguments in order to simplify the expression. For now let us ignore prefactors.

Definition 3.4.4 (Central Part)**.** *For a quantum channel as in Definition 3.4.1, we define its central part by:*

$$P_\sigma(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger) = \sum_{i_1, i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger \right)_\sigma .$$

For the central part, it does not make a difference to rename the first set of operators the second and vice versa, $i_1 \leftrightarrow i_2$. Mathematically this means that the central part is invariant under the permutation (13)(24).

Proposition 3.4.5.

$$P_\sigma(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger) = P_\sigma(A_{i_2}, A_{i_2}^\dagger, A_{i_1}, A_{i_1}^\dagger)$$

Proof. In the definition of P,

$$P_\sigma(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger) = \sum_{i_1,i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger\right)_\sigma ,$$

so we can just rename i_1 to i_2 and vice versa:

$$P_\sigma(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger) = \sum_{i_2,i_1} \left(A_{i_2}, A_{i_2}^\dagger, A_{i_1}, A_{i_1}^\dagger\right)_\sigma = P_\sigma(A_{i_2}, A_{i_2}^\dagger, A_{i_1}, A_{i_1}^\dagger).$$

□

The idea is that we can use this symmetry to sum over orbits, Definition 2.1.24, of the isotropy group, Definition 2.1.22, rather than summing over all of S_4.

Definition 3.4.6 (Symmetries of the Central Part)**.** *The isotropy group P_4 of P is the group generated by* (13)(24)*.*

$$P_4 = \{e, (13)(24)\}.$$

In Table 3.2, we decompose S_4 into orbits of P_4. Instead of summing over all elements of S_4, we can now sum over the orbits. The following proposition follows trivially from Theorem 2.1.26.

Proposition 3.4.7. *Let $\{O_i\}_i$ be the orbits of S_4 under P_4 and $\sigma(O_i)$ a representative of orbit O_i. Then we have:*

$$F_n^2(\mathcal{N}) = \frac{1}{\prod_{c=0}^{3}(d^n + c)} \sum_i |O_i| \left[\sum_{i_1,i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger\right)_{\sigma(O_i)}\right]^n .$$

While this gives some simliplification (the sum is now over 16 elements instead of 24) we certainly do not get an expression where we can express F_n^2 as a function of F^2. The full decomposition can be seen in Table 3.2.
In comparison the group $S_2 \times S_2$ is rather simple.

Proposition 3.4.8.

$$S_2 \times S_2 = \{e, (12), (34), (12)(34)\}$$

However we can still decompose this group into orbits of the isotropy group P_4 (Table 3.1).

Proposition 3.4.9. *Let $\{\tilde{O}_i\}_i$ be the orbits of $S_2 \times S_2$ under P_4 and $\sigma(\tilde{O}_i)$ a representative of orbit $\tilde{O}_i$. Then we have:*

$$(F_n(\mathcal{N}))^2 = \frac{1}{d^{2n}\,(d^n+1)^2} \sum_i |\tilde{O}_i| \left[\sum_{i_1,i_2} \left(A_{i_1}, A_{i_1}^\dagger, A_{i_2}, A_{i_2}^\dagger \right)_{\sigma(\tilde{O}_i)} \right]^n$$

Comparing the tables, Table 3.1 and Table 3.2, we see that all three trace products that appear in Table 3.1, all different orbits of $(F_n)^2$, also appear in F_n^2, but not vice versa. For a short-hand notation, it is useful to call the sum of all P_σ that only appear in Table 3.2 R.

Proposition 3.4.10. *A complete expression of the variance is then given by:*

$$\begin{aligned} Var_n &= \frac{1}{d^{2n}\,(d^n+1)^2\,(d^n+2)\,(d^n+3)} \Big(d^n\,(d^n+1) \Big[1\,P_e^n + 2\,P_{(12)}^n + 3\,d^{2n} + R \Big] \\ &\quad - (d^n+2)(d^n+3) \Big[1\,P_e^n + 2\,P_{(12)}^n + 1\,d^{2n} \Big] \Big) \\ &= \frac{1}{d^{2n}\,(d^n+1)^2\,(d^n+2)\,(d^n+3)} \Big((6 - 4\,d^n) \left(P_e + 2\,P_{(12)} \right) + (2\,d^{4n} - 2\,d^{3n} - 6\,d^{2n}) \\ &\quad + (d^{2n} + d^n) R \Big). \end{aligned}$$

We see that the highest orders of P_e and $P_{(12)}$ are cancelled in general. However, without further assumptions, we can clearly not take the limit of infinite channels or give any other channel independent insight.

$P_4\,\sigma$	Repr. σ	$\lvert C_{S_2\times S_2}(\sigma)\rvert$	$\lvert P_4\sigma\rvert$	P_σ
$\tilde{O}_1$	e	1	1	$\sum_{i_1,i_2} \mathrm{Tr}\,(A_{i_1})\,\mathrm{Tr}\left(A_{i_1}^\dagger\right)\mathrm{Tr}\,(A_{i_2})\,\mathrm{Tr}\left(A_{i_2}^\dagger\right)$
$\tilde{O}_2$	(12)	2	2	$d\,\sum_{i_2} \mathrm{Tr}\,(A_{i_2})\,\mathrm{Tr}\left(A_{i_2}^\dagger\right)$
$\tilde{O}_6$	(12)(34)	1	1	d^2

Table 3.1.: The orbits of $S_2 \times S_2$ under P_4.

$P_4\,\sigma$	Repr. σ	$\lvert C_{S_4}(\sigma)\rvert$	$\lvert P_4\sigma\rvert$	P_σ
O_1	e	1	1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1}\right) \operatorname{Tr}\left(A_{i_1}^\dagger\right) \operatorname{Tr}\left(A_{i_2}\right) \operatorname{Tr}\left(A_{i_2}^\dagger\right)$
O_2	(12)	6	2	$d \sum_{i_2} \operatorname{Tr}\left(A_{i_2}\right) \operatorname{Tr}\left(A_{i_2}^\dagger\right)$
O_3	(13)		1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}\right) \operatorname{Tr}\left(A_{i_1}^\dagger\right) \operatorname{Tr}\left(A_{i_2}^\dagger\right)$
O_4	(14)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}^\dagger\right) \operatorname{Tr}\left(A_{i_2}\right) \operatorname{Tr}\left(A_{i_1}^\dagger\right)$
O_5	(24)		1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1}^\dagger A_{i_2}^\dagger\right) \operatorname{Tr}\left(A_{i_1}\right) \operatorname{Tr}\left(A_{i_2}\right)$
O_6	(12)(34)	3	1	d^2
O_7	(13)(24)		1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}\right) \operatorname{Tr}\left(A_{i_1}^\dagger A_{i_2}^\dagger\right)$
O_8	(14)(23)		1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}^\dagger\right) \operatorname{Tr}\left(A_{i_1}^\dagger A_{i_2}\right)$
O_9	(123)	8	2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_1}^\dagger A_{i_2}\right) \operatorname{Tr}\left(A_{i_2}^\dagger\right)$
O_{10}	(124)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_1}^\dagger A_{i_2}^\dagger\right) \operatorname{Tr}\left(A_{i_2}\right)$
O_{11}	(132)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2} A_{i_1}^\dagger\right) \operatorname{Tr}\left(A_{i_2}^\dagger\right)$
O_{12}	(142)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}^\dagger A_{i_1}^\dagger\right) \operatorname{Tr}\left(A_{i_2}\right)$
O_{13}	(1234)	6	1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_1}^\dagger A_{i_2} A_{i_2}^\dagger\right)$
O_{14}	(1243)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_1}^\dagger A_{i_2}^\dagger A_{i_2}\right)$
O_{15}	(1324)		2	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2} A_{i_1}^\dagger A_{i_2}^\dagger\right)$
O_{16}	(1432)		1	$\sum_{i_1,i_2} \operatorname{Tr}\left(A_{i_1} A_{i_2}^\dagger A_{i_2} A_{i_1}^\dagger\right)$

Table 3.2.: The orbits of S_4 under P_4.

3.4.2. Symmetry Observations for Self-Adjoint Channels

We will now restrict ourselves further. The strongest assumption, without choosing concrete operators, is to assume that all Kraus operators are self-adjoint. The restriction to self-adjoint Kraus operators drastically simplifies P.

Proposition 3.4.11 (Central Part for Self-Adjoint Channels)**.** *For a self-adjoint channel $\mathcal{N}$, with self-adjoint Kraus operators $\{A_i\}_i$, its central part is:*

$$P_\sigma(A_{i_1}, A_{i_2}) = \sum_{i_1,i_2} (A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2})_\sigma \,.$$

This expression is now not only invariant under the exchange of the operators blockwise, but also under the exchange within each block.

Definition 3.4.12. *The isotropy group H_4 of P_σ is the group generated by* $(12), (13)(24)$*,*

$$H_4 = \{e, (12), (34), (12)(34), (13)(24), (1324), (1423), (14)(23)\}.$$

Since H_4 is also a subgroup of S_4, and thus has a natural action on S_4 as a set, we can again decompose S_4 into orbits of H_4.

$H_4\,\sigma$	Repr. σ	$\lvert C_{S_4}(\sigma)\rvert$	$\lvert H_4\sigma\rvert$	P_σ
O_1	e	1	1	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}) \operatorname{Tr}(A_{i_1}) \operatorname{Tr}(A_{i_2}) \operatorname{Tr}(A_{i_2})$
O_2	(12)	6	2	$d \sum_{i_2} \operatorname{Tr}(A_{i_2}) \operatorname{Tr}(A_{i_2})$
O_3	(13)		4	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}A_{i_2}) \operatorname{Tr}(A_{i_1}) \operatorname{Tr}(A_{i_2})$
O_4	(12)(34)	3	1	d^2
O_5	(13)(24)		2	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}A_{i_2}) \operatorname{Tr}(A_{i_1}A_{i_2})$
O_6	(123)	8	8	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}A_{i_1}A_{i_2}) \operatorname{Tr}(A_{i_2})$
O_7	(1234)	6	4	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}A_{i_1}A_{i_2}A_{i_2})$
O_8	(1324)		2	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}A_{i_2}A_{i_1}A_{i_2})$

Table 3.3.: The orbits of S_4 under H_4.

Now there are only 8 different orbits, in contrast to the 16 orbits before, which is a great simplification.
The decomposition of $S_2 \times S_2$ under H_4 is actually the same as under P_4.

$H_4\,\sigma$	Repr. σ	$\lvert C_{S_2\times S_2}(\sigma)\rvert$	$\lvert H_4\sigma\rvert$	P_σ
$\tilde{O}_1$	e	1	1	$\sum_{i_1,i_2} \operatorname{Tr}(A_{i_1}) \operatorname{Tr}(A_{i_1}) \operatorname{Tr}(A_{i_2}) \operatorname{Tr}(A_{i_2})$
$\tilde{O}_2$	(12)	2	2	$d \sum_{i_2} \operatorname{Tr}(A_{i_2}) \operatorname{Tr}(A_{i_2})$
$\tilde{O}_4$	(12)(34)	1	1	d^2

Table 3.4.: The orbits of $S_2 \times S_2$ under H_4.

The number of orbits are drastically reduced, however the variance still has the same structure as in Proposition 3.4.10. We will not get more simplification without stronger assumptions for the quantum channel.

3.4.3. Variances for Unital Quantum Channels

In Section 1.4, we introduced the class of unital channels, a special case of self-adjoint channels. We will now investigate the central part, from Definition 3.4.4, for unital qubit channels.

Fact 3.4.13. *As a quick reminder, a unital qubit channel $\mathcal{N}_{\{p_i\}_{i=0}^3}$ can be written as:*

$$\mathcal{N}_{\{p_i\}_{i=0}^3} = \sum_{i=0}^{3} p_i\, \sigma_i\, \rho\, \sigma_i,$$

where σ_0 is the identity, and $\{\sigma_1, \sigma_2, \sigma_3\}$ are the Pauli operators. Then the Kraus operators are $\{A_i = \sqrt{p_i}\sigma_i\}_{i=0}^3$.

For these, the Kraus operators inherit the nice properties of Pauli operators.

Proposition 3.4.14. *Let $\{A_0, \ldots, A_3\}$ be the Kraus operators of a unital qubit channel. Then for $i_1, i_2 = 0, \ldots, 3$:*

$$Tr(A_{i_1}) = \sqrt{p_0}\,\delta_{i_1,0} \quad \textit{and} \quad Tr(A_{i_1}\, A_{i_2}) = p_{i_1}\,\delta_{i_1,i_2}.$$

Furthermore, this also simplifies three cycles:

$$Tr(A_{i_1} A_{i_1} A_{i_2})\; Tr(A_{i_2}) = 2 * \delta_{i_2,0}\,\sqrt{p_0}\; Tr(A_{i_1} A_{i_1} A_{i_2}) = p_{i_1}\, p_0\, 2^2.$$

These properties neatly simplify Table 3.3 and Table 3.4.

$H_4\,\sigma$	Representative σ	$\lvert C_{S_4}(\sigma)\rvert$	$\lvert H_4\sigma\rvert$	P_σ
O_1	e	1	1	$p_0^2\,2^4$
O_2	(12)	6	2	$p_0\,2^3$
O_3	(13)		4	$p_0^2\,2^3$
O_4	(12)(34)	3	1	2^2
O_5	(13)(24)		2	$\sum_i p_i^2\,2^2$
O_6	(123)	8	8	$p_0\,2^2$
O_7	(1234)	6	4	2
O_8	(1324)		2	$(p_1^2 - p_0^2 + (p_2 - p_3)^2 - 2p_1(p_2 + p_3) + 2p_0) * 2 = \alpha_1\, d$

Table 3.5.: The orbits of S_4 under H_4 for Pauli channels.

$H_4\,\sigma$	Repr. σ	$\lvert C_{S_2\times S_2}(\sigma)\rvert$	$\lvert H_4\sigma\rvert$	P_σ
$\tilde{O}_1$	e	1	1	$p_0^2\,2^4$
$\tilde{O}_2$	(12)	2	2	$p_0\,2^3$
$\tilde{O}_4$	(12)(34)	1	1	2^2

Table 3.6.: The orbits of $S_2 \times S_2$ under H_4 for Pauli channels.

If in Table 3.5 we ignore the powers of 2 that come from the dimension $d = 2$, the possible values for P_σ are the same as in Table 3.6, except for the orbits of (13)(24) and the 4-cycles. However, we see that the dimension of (13)(24) is higher than the dimension of the 4-cycles. In the limit this will be the crucial factor.

Proposition 3.4.15 (F_n^2 for Pauli Channels)**.** *Let us denote the sum of all elements of O_7 and O_8 as α. Then the second fidelity moment for n Pauli channel is:*

$$F_n^2(\mathcal{N}_G) = \frac{2^{2n}\left((p_0^n\,2^n+1)^2 + 4\,(p_0^{2n}\,2^n + 2\,p_0^n) + 2\,\left(\sum_i p_i^2\right)^n\right) + \alpha\,2^n}{2^n(2^n+1)(2^n+2)(2^n+3)}.$$

Proposition 3.4.16 ($(F_n)^2$ for Pauli Channels)**.** *The square of the average fidelity for n Pauli channel is:*

$$(F_n(\mathcal{N}_G))^2 = \frac{2^{2n}\,(p_0^n\,2^n+1)^2}{2^{2n}(2^n+1)^2}.$$

In all terms but the term proportional to α, there is a factor 2^{2n}, which we can cancel. In the end, this leads to a very simple and elegant expression for the regularized variance, which is just the n-th root of the n-channel variance (Proposition 3.4.2).

Theorem 3.4.17 (Regularized Variance for Pauli Channel)**.** *The regularized variance for the identity channel and the completely depolarizing channel is* 0*. For all other Pauli channels it is simply:*

$$Var_\infty^2 = \lim_{n\to\infty} \left(Var_n^2\right)^{\frac{1}{n}} = \sum_i \left(\frac{p_i}{2}\right)^2.$$

Proof. For the completely depolarizing channel, every state has fidelity $\frac{1}{2}$. Hence there is no variance. The same is true for the identity channel, only that here the fidelity is 1 regardless of the input.
For all other cases we calculate:

$$\begin{aligned}
\mathrm{Var}_n^2 = &\frac{1}{(2^n+1)^2\,(2^n+2)\,(2^n+3)} \\
&\Bigg[2^n\,(2^n+1)\Bigg((p_0^n\,2^n+1)^2 + 4\,(p_0^{2n}\,2^n + 2\,p_0^n) + 2\left(\sum_i p_i^2\right)^n + \frac{\alpha}{2^n}\Bigg) \\
&- (2^n+2)(2^n+3)(p_0^n\,2^n+1)^2\Bigg] \\
= &\frac{1}{(2^n+1)^2\,(2^n+2)\,(2^n+3)} \\
&\Big[2^{4n}\left(p_0^{2n} - p_0^{2n}\right) + d^{3n}\left(p_0^{2n} + 2\,p_0^n + 4\,p_0^{2n} - 5\,p_0^{2n} - 2\,p_0^n\right) \\
&+ 2^{2n}\left(1 + 2\,p_0^n + 4\,p_0^{2n} + 8\,p_0^n + 2\left(\sum_i p_i^2\right)^n - 1 - 10\,p_0^n - 6\,p_0^{2n}\right) + O(2^n)\Big] \\
= &\frac{2^{2n}\left(2\,\left(\sum_i p_i^2\right)^n - 2\,p_0^{2n}\right) + O(2^n)}{(2^n+1)^2\,(2^n+2)\,(2^n+3)}.
\end{aligned}$$

If $\sum_i p_i^2 > \frac{1}{2}$, we are easily convinced that the terms $O(2^n)$ vanish very quickly in the limit, however the minimum of $\sum_i p_i^2$ is achieved when $p_i = \frac{1}{4}$ for all i, in which case we have

$$\sum_i p_i^2 = \frac{1}{4}.$$

Since $p_i = \frac{1}{4}$ for all i is the completely depolarizing channel, and hence we do not have to consider this situation, we only need to check the next order in 2^n. The calculation is messy and not insightful, and hence we relegate it to Appendix A.

In the end we are left with

$$\lim_{n\to\infty} \left(\mathrm{Var}_n^2\right)^{\frac{1}{n}} = \lim_{n\to\infty} \left[2\left(\sum_i \frac{p_i^2}{2^2}\right)^n - 2\left(\frac{p_0}{2}\right)^{2n}\right]^{\frac{1}{n}}, \tag{3.2}$$

where we notice that $\lim_{n\to\infty} 2^{\frac{1}{n}} = 1$, then we can pull out $\frac{p_0^{2n}}{2^{2n}}$, and get

$$\lim_{n\to\infty} \left(\mathrm{Var}_n^2\right)^{\frac{1}{n}} = \frac{p_0^2}{2^2} \lim_{n\to\infty} \left[\left(1 + \sum_{i\neq 0} \frac{p_i^2}{p_0^2}\right)^n - 1\right]^{\frac{1}{n}} = \sum_i \frac{p_i^2}{2^2},$$

since $\sum_{i\neq 0} \frac{p_i^2}{p_0^2} > 0$ if $p_0 \neq 1$. □

First we should notice that the regularized variance for Pauli channels is astonishingly simple, and its elegance made us wonder why it is that way. We could translate the calculation into graphic structures that explain the expression for Var_∞. We will explore these structures in detail in the next section, since they will allow us to find higher regularized moments.
Since the corrections to Var_∞ are suppressed exponentially, when considering a large number of channel we can replace the real variance Var_n with the variance in the limit, Var_∞. However, if we want to use Var_∞ to approximate the real (unregularized) variance and thus reexponentiate with n, we have to pay attention to the prefactor in Equation 3.2 while the non leading terms are still surpressed exponentially.

$$\left(\mathrm{Var}_n^2\right)^n \approx 2\left(\sum_i \frac{p_i^2}{2^2}\right)^n$$

Let us now interpret what the result means for the channel fidelity distribution for n Pauli channels. For that it is useful to compare the variance to the average from Proposition 3.2.6.

Definition 3.4.18 (Regularized Relative Variance for Pauli Channels)**.** *The regularized variance for a Pauli channel divided by the regularized average channel fidelity is called regularized relative variance for a Pauli channel* $\frac{Var_n}{F_n}$.

Proposition 3.4.19 (Regularized Relative Variance for Pauli Channels in the Limit)**.** *The regularized relative variance in the limit of $n \to \infty$ Pauli channels is defined piecewise:*

- *for* $p_0 = 1$

$$\frac{Var_\infty}{F_\infty} = 0,$$

- *for* $p_0 > \frac{1}{2}$

$$\frac{Var_\infty}{F_\infty} = \frac{\sqrt{1 + \sum_{i\neq 0}\left(\frac{p_i}{p_0}\right)^2}}{2} < 1,$$

- *and for* $p_0 \leq \frac{1}{d}$ *and all* $p_{i\neq 0} < 1$,

$$\frac{Var_\infty}{F_\infty} = \sqrt{\sum_i p_i^2} < 1,$$

- *whereas if any of the other* $p_i = 1$, *then*

$$\frac{Var_\infty}{F_\infty} = 1.$$

Proof. The first expression is obvious: since the variance of the identity channel is zero, so is the relative variance. For the second expression we notice that $\sum_{i\neq 0} p_i^2 \leq (1-p_0)^2$, because the sum of squares is minimal for an equal distribution and maximal if one probability is maximal and all the others are equal 0, and furthermore $\frac{1}{p_0} < 2$ for $p_0 \in]\frac{1}{2}, 1]$. Then we have the chain of inequalities:

$$\frac{\sqrt{1+\sum_{i\neq 0}\left(\frac{p_i}{p_0}\right)^2}}{2} \leq \frac{\sqrt{1+\left(\frac{1}{p_0}-1\right)^2}}{2} < \frac{\sqrt{2}}{2} < 1.$$

For the third and fourth expression we see that $\sum_i p_i^2 = 1$ only if one $p_i = 1$. □

From Proposition 3.4.19, we learn that for basically all Pauli channels, the variance becomes very small compared to the average once we consider enough channels. This means that the channel fidelity distribution is strongly peaked around the mean. Only for those Pauli channels that are a unitary transformation that is not the identity the width of the distribution is the same as the mean.

We will now investigate how the variances approach the elegant approximate formula Theorem 3.4.17. For a more vivid discussion let us again investigate the depolarizing channel (Proposition 1.4.6).

In Figure 3.4, we plot the regularized variance for different numbers of depolarizing channels. It is quite difficult to get good plots for more channels; the high roots can easily pick up more computational noise than data. We can see the beginning of this phenomenon for p close to 1 for 17 channels.

Let us now analyze the plot. First we notice that the variance for $n = 1$ channel does not appear in the plot. As we have mentioned before, the channel fidelity for a single depolarizing channel is independent of the input and thus certainly the variance is 0. Next we see that for more than one channel the variance is not zero. This is sufficient to prove that the channel fidelity is no longer independent of the input, for those tensor products of the depolarizing channel.

Then we can observe that for most p the variance quickly approach the limit of infinite channels, quite similar to Figure 3.1. However, close to the extreme cases,

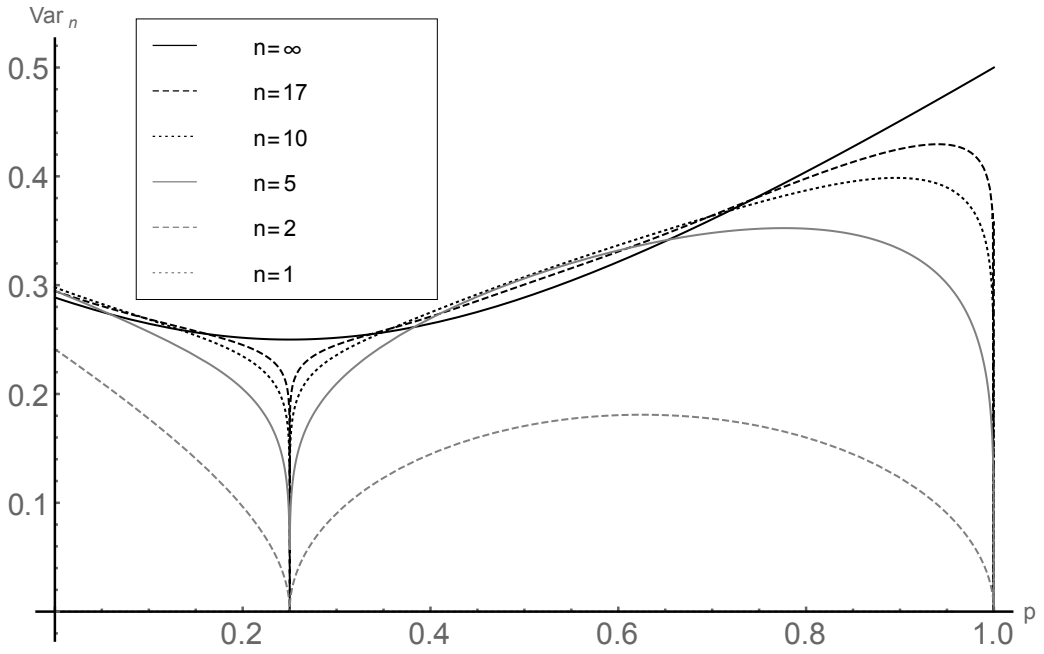


Figure 3.4.: Regularized variances for n depolarizing channels

where the variance is 0, the regularized fidelities deviate noticeable from the limit. This deviation is due to the fact that in the limit, Var_∞ is not a smooth function for $p = \frac{1}{4}$ and $p = 1$. Close to both extreme cases we cannot use the infinite channel formula to approximate the real variance. For $p \in [0, 0.2]$ and $p \in [0.3, 0.8]$, Var_∞ approximates the real variance very well for more than 10 channels.

Between $p = 0$ and $p = 0.5$, the variances look almost symmetric around the completely depolarizing channel. Once $p_0 < 0.5$, where the average is no longer proportional to p_0, it becomes also unimportant for the variance.

In Figure 3.5, we plot the relative variances for different numbers of depolarizing channels. Here we can learn more about the concrete distributions. First of all, we recognize that very similarly to Figure 3.4, for most p, the relative variances quickly approach $\frac{\mathrm{Var}_\infty}{F_\infty}$. Then we see, as predicted in Proposition 3.4.19, that the relative variance is always smaller than 1 and more precisely around $\frac{1}{2}$. Again this is because in the limit of infinite channels, the channel fidelity distribution is peaked very strongly at the mean, which effectively means that the channel fidelity is, for the vast majority of states, independent of the state itself.

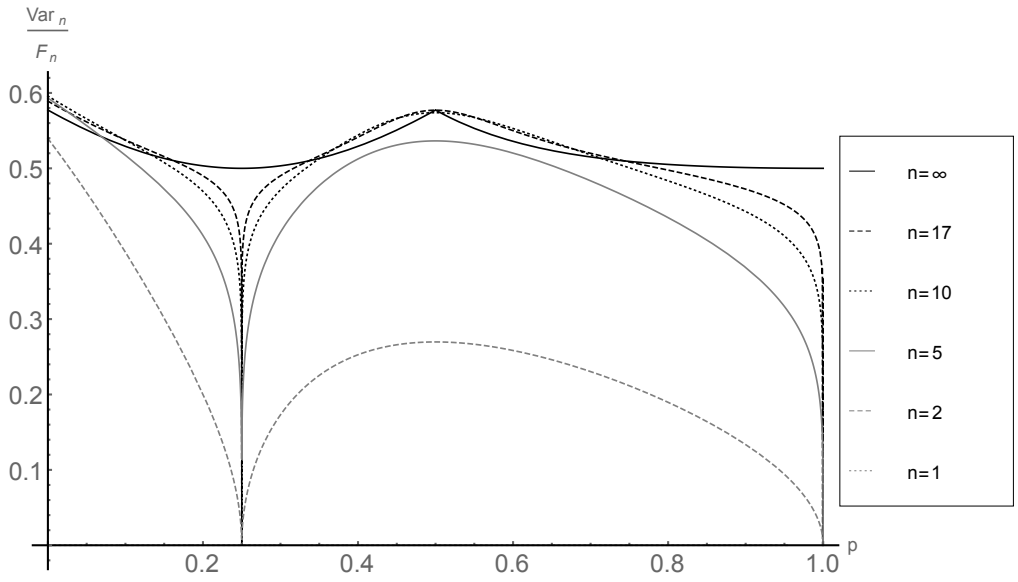


Figure 3.5.: Relative variances for n depolarizing channels

For the channel fidelity, we conclude:

- For $p > \frac{1}{2}$ and large n, we can approximate an n fold qubit depolarizing channel with a depolarizing channel $\mathcal{N}_\lambda^D$ (see Definition 1.4.5) that acts on elements of $\mathcal{L}(\mathcal{H}_1^{\otimes n})$ with $\lambda = \frac{(4p)^n - 1}{4^n - 1} \approx p^n$.

- For $p \leq \frac{1}{2}$ we have to approximate the n fold qubit depolarizing channel with a completely depolarizing channel that acts on elements of $\mathcal{L}(\mathcal{H}_1^{\otimes n})$.

It is fascinating that using a small number of channels can spread out the distribution, but then eventually with more and more channels the distribution becomes singular again. In particular, a residual effect of this transition is that there will be non trivial maximizing states, we will discuss these in the end of this chapter.

Let us get back to discussing plot. If we have a look at the non-regularized relative variances in Figure 3.6, we notice that for small and large p, the distribution is maximally broad for 2 channels, however for p around $\frac{1}{2}$, it is the broadest for 3 channels. The maximum is for sure achieved for a small number of channels and then rapidly approaches 0.

Since the depolarizing channel is somewhat special in the sense that its channel fidelity is independent of its input, and thus its variance is zero for a single channel, in Figure 3.7, we plot the non-regularized relative variance for a bit flip channel.

We spot that there is already a variance for a single channel. It also declines rapidly for most p, again showing that the distribution is strongly peaked already for a small number of channels. However, close to the complete bit flip channel (small p) the relative variance does not decline as quickly, in agreement with our analysis that the large n relative variance for this channel stays constant.

For $p > 0.5$, the broadest distribution is for a single channel, however if we come closer to the complete bitflip channel, $p = 0$, the distribution becomes more spread out for two channels than it was for a single channel. This maximum can certainly be shifted to more channels, but to achieve a maximum for a high channel number we have to go a lot closer, since the relative variance is suppressed exponentially with the number of channels, to the complete bitflip channel — numerically that is not possible.

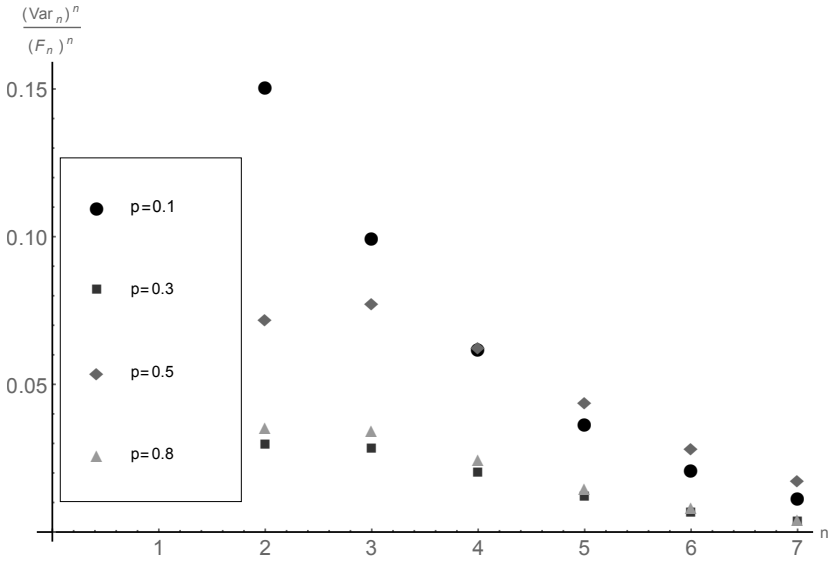


Figure 3.6.: Relative non-regularized variances for depolarizing channel $\mathcal{N}_p^{\otimes n}$

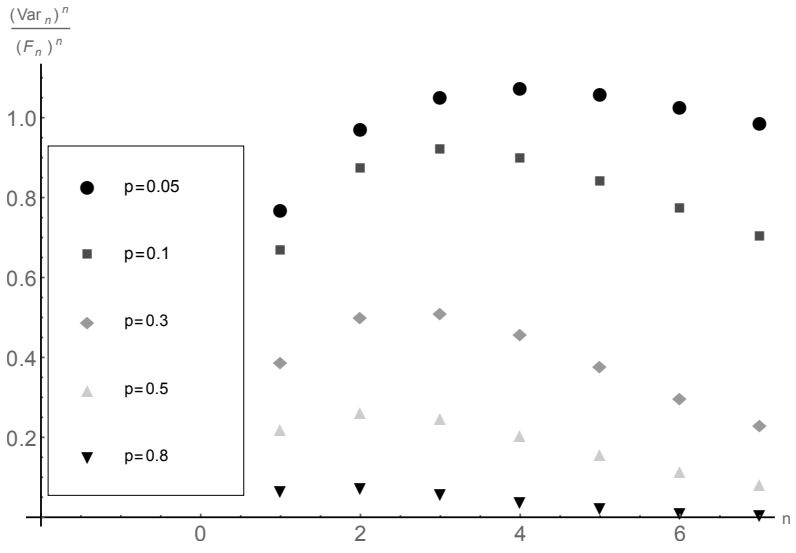


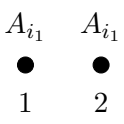
Figure 3.7.: Relative non-regularized variances for different bit flip channel

3.5. Correlations & Diagrams

We seek to find structural reasons for the simplicity of the large n variance. It will be useful to have diagrammatic tools to understand systematically what is happening and which permutations are important.

In Theorem 3.3.3, we noted that as q gets larger we get more Kraus operators, precisely going over to the next higher moment there is an additional pair of operators with the same index. We can think of the Kraus operators with the same index as being part of a box.

Definition 3.5.1 (Box)**.** *A box with a single pair of Kraus operators $(A_{i_1}, A_{i_1})_\sigma$ we represent with two dots.*



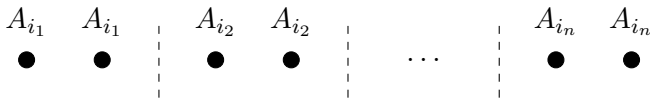
Eventually, we will not name the operators explicitly, it will rather be implicitly clear, which pair of operators is meant.

In the calculation of higher moments, we have expressions like

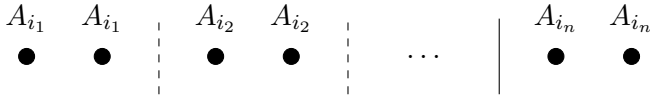
$$(A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2}, \ldots, A_{i_n}, A_{i_n})_\sigma.$$

We will now translate these expressions into boxes. For this we need to consider the two following situations: On the one hand, if $\sigma \in S_{2q}$ there are permutations that mix any operator with any other operator, on the other hand if $\sigma \in S_{2(q-k)} \times S_2^k$ the permutations can only mix within the first $2(q-k)$ operators or within the first $(q-k)$ boxes.

Definition 3.5.2 (Multiple Boxes)**.** *A trace product $(A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2}, \ldots, A_{i_n}, A_{i_n})_\sigma$ with $\sigma \in S_{2n}$ can be represented as n boxes, since σ can mix within all boxes, the boxes are called connected and are visually separated with a dashed line.*

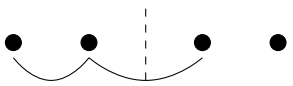


However if $\sigma \in S_{2n-2} \times S_2$, the last box is called not connected and is separated with a solid line.



We will now add permutations and their corresponding trace products to the picture.

Definition 3.5.3 (Permutations in Diagrams)**.** *In an n-box we can inscribe a permutation σ and thus a trace product by splitting σ into its cycles and for each cycle connect the according operators with a curve. Cycles with only one element will not be inscribed.*



The definitions might seem a bit dry. Let us have a look at an example, further motivating the diagrammatic tools.

Example 3.5.4. *In the third moment*

$$F_n^3(\mathcal{N}) = \frac{1}{\prod_{c=0}^{5}(2^n + c)} \sum_{\sigma \in S_6} \left[\sum_{\vec{i}} (A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2}, A_{i_3}, A_{i_3})_\sigma \right]^n$$

the central part will be:

$$P_\sigma = \sum_{\vec{i}} (A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2}, A_{i_3}, A_{i_3})_\sigma$$

We see that we have three pairs of operators and identify them with three boxes, each with two dots, furthermore the sum is over the full S_6 hence all boxes are connected as indicated by the dashed lines.

Now we inscribe the permutation $\sigma = (13)(456)$ into the system.

Remember that the diagram above represents $P_{(13)(456)}$ and that

$$P_{(13)(456)} = \sum_{\vec{i}} Tr(A_{i_1})\ Tr(A_{i_1}\, A_{i_2})\ Tr(A_{i_2}\, A_{i_3}\, A_{i_3})\,.$$

Now that we can interpret the diagrams a priori, we will introduce rules of manipulation, by which diagrams can be reduced.
The most important rule comes from the special property of Pauli Operators; that there is only one with a non-vanishing trace:

- An unconnected point acts as a Kronecker delta on the index of the two operator system it belongs to, also restricting the other operator to be proportional to the identity.

We will say it decouples the box from other boxes and leads to the following rule of manipulation illustrated in the following Example 3.5.5.

- Any line connecting to a decoupled box can be removed, catching a factor $d^{-1} = 2^{-1}$ or $\delta = -1$.

We will understand the notion of coupled and decoupled boxes as we work with the diagrams.

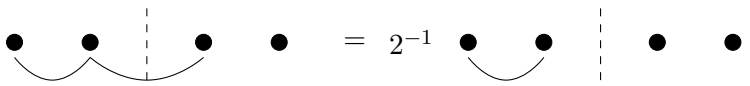
Example 3.5.5 (Rules)**.** *Consider the central part of the second moment for the permutation $\sigma = (123)$,*

$$P_{(123)} = \sum_{\vec{i}} (A_{i_1}, A_{i_1}, A_{i_2}, A_{i_2})_{(123)} = \sum_{i_1} \sum_{i_2} Tr(A_{i_1}\, A_{i_1}\, A_{i_2})\ Tr(A_{i_2})\,.$$

Since one A_{i_2} is in a trace by itself, it collapses the sum over i_2 and also forces the other A_{i_2} to be proportional to the identity. That means that we can effectively remove A_{i_2} from the first trace but have to take into account that now there will a factor d too much and get

$$P_{(123)} = \sum_{i_1} \sum_{i_2} Tr(A_{i_1} A_{i_1}) \, d^{-1} \, Tr(A_{i_1})^2 = d^{-1} P_{(12)}.$$

In the language of diagrams this translates to:



there are no more decoupled boxes to decouple, so we are done.

Because sometimes we will rather talk about the diagrams than about the permutations that are inscribed into them, we will transfer the dimension of a permutation (Definition 2.1.15), explained in Subsection 2.3.6, into our diagrammatic language.

Definition 3.5.6 (Dimension of a Diagram)**.** *The dimension δ of a diagram is simply the dimension of the permutation it represents. It is equal to the number of connected sets.*

Certain diagrams are characteristic for a specific symmetric group and as such for a specific central moment. They are characteristic in the sense that they appear the first time in a specific diagram, where all boxes are connected and cannot be simplified to a diagram with not connected boxes. These concepts will be explained after the following definitions.

Definition 3.5.7 (k-correlation)**.** *A diagram that can not be simplified and still connects k boxes is called a k-correlation.*

We will now re-investigate our calculation for the variance. Let us first have a look at $(F_n)^2$ and thus all the orbits from Table 3.6 in our diagrammatic language.
We can understand $(F_n)^2$ as the sum over all permutations, that do not connect boxes:

We will learn more from this, if we look at it orbit wise.
The orbit $\tilde{O}_1$ is represented by the following diagram, which clearly is a 0-correlation.

Orbit $\tilde{O}_2$ can be written as the diagram, a 1-correlation.

Last we have orbit $\tilde{O}_4$, which has two 1-correlations.

Clearly $(F_n)^2$ has only 1-correlations or connects even less boxes.

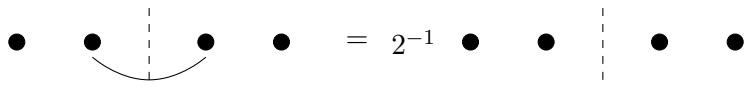
Next we look into the second moment F_n^2 and thus the orbits in Table 3.5 and we can interpret F_n^2 as the sum over all permutations, connecting or not connecting.

$F_n^2 =$ $\sum_{\sigma \in S_4}$

Let us again look at this orbitwise:

The first orbit O_1 is identical to the first orbit above, also the second orbit O_2 has an identical diagram to the second orbit above. The third orbit O_3 has a diagram, that is correlationwise identical to the first diagram.

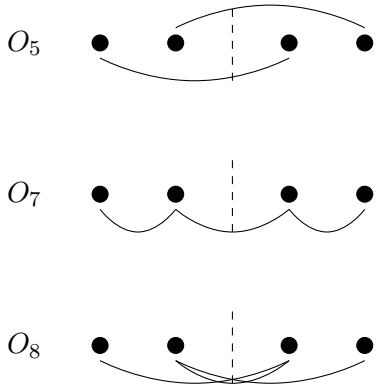
O_3



$= 2^{-1}$

The fourth orbit O_4 again has a diagram identical to $\tilde{O}_4$ and the sixth orbit has been shown to have a 1-correlation only in Example 3.5.5.

However the orbits O_5, O_7 and O_8 give rise to new kinds of diagrams, the 2-correlations.



It is important to note that all of these diagrams cannot be reduced. At first glance one might wonder what is happening in the diagram for O_8, because of the strange crossing. However the nice feature is, that as long as the contribution of the diagram itself is larger than $\frac{1}{2}$, we do not have to understand lower dimensional correlations. Looking at the dimension we see that the dimension of the diagram belonging to O_5 is 2 and the other two both have dimension 1. That means that for $\sum_i p_i \geq \frac{1}{2}$ in the limit of $n \to \infty$ permutations of O_5-type will be dominating the 2-correlations.
The variance is special in the sense that even without the restriction all 0- and 1-correlations vanish up to the relevant order and the regularized variance is dominated by the highest dimensional 2-correlation.

3.6. Central Moments and Diagrams

We have seen that we could give another single letter formula for the variance of the Channel Fidelity. The variance is the second central moment. We will now study the higher central moments in the language of diagrams.

Proposition 3.6.1 (General Formula for Central Moments)**.** *The qth central moment for n channel is given as:*

$$\mu_n^q = \sum_{k=0}^{q} (-1)^k \binom{q}{k} \left(\frac{1}{d^n\,(d^n+1)}\right)^k \frac{1}{\prod_{c=0}^{2(q-k)-1}(d^n+c)} \sum_{\sigma \in S_{2(q-k)} \times S_2^k} \left[\sum_{\vec{i}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger\right)_\sigma\right]^n .$$

Proof. First we use the definition of central moments. Then after inserting the results from Theorem 3.3.3 we see that we can simplify the expression by always summing over a suitable subgroups of the S_{2q}; compare Proposition 3.4.3.

$$\begin{aligned}
\mu_n^q &= \sum_{k=0}^{q} (-1)^k \binom{q}{k} \left(F_n(\mathcal{N})\right)^k F_n^{q-k}(\mathcal{N}) \\
&= \sum_{k=0}^{q} (-1)^k \binom{q}{k} \Bigg(\left(\frac{1}{d^n\,(d^n+1)}\right)^k \sum_{\sigma \in S_2^k} \left[\sum_{i_1,\ldots,i_k} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_k}, A_{i_k}^\dagger\right)_\sigma\right]^n \\
&\quad \times \frac{1}{\prod_{c=0}^{2(q-k)-1}(d^n+c)} \sum_{\sigma \in S_{2(q-k)}} \left[\sum_{i_1,\ldots,i_{q-k}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_{q-k}}, A_{i_{q-k}}^\dagger\right)_\sigma\right]^n \Bigg) \\
&= \sum_{k=0}^{q} (-1)^k \binom{q}{k} \left(\frac{1}{d^n\,(d^n+1)}\right)^k \frac{1}{\prod_{c=0}^{2(q-k)-1}(d^n+c)} \sum_{\sigma \in S_{2(q-k)} \times S_2^k} \left[\sum_{\vec{i}} \left(A_{i_1}, A_{i_1}^\dagger, \ldots, A_{i_q}, A_{i_q}^\dagger\right)_\sigma\right]^n
\end{aligned}$$

□

Without a concrete choice for the Kraus operators, it is difficult to get anything the formula for μ_n^q. However for unital Pauli channel, we could find more interesting structures in higher central moments and we will explain our findings now with the help of diagrams.

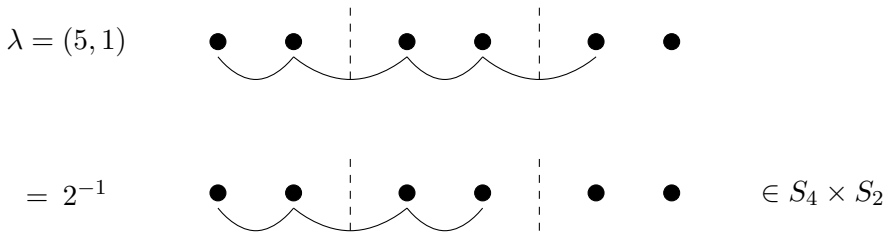
We investigate the third central moment next.

Proposition 3.6.2 (Third central Moment)**.**

$$\begin{aligned}
\mu_n^3 =& \sum_{i=0}^{3}(-1)^k\binom{3}{k}\left(\frac{1}{d^n(d^n+1)}\right)^k \\
&\times\frac{1}{\prod_{c=0}^{5-2k}(d^n+c)}\sum_{\sigma\in S_{2(3-k)}\times S_2^k}\left[\sum_{i_1,i_2,i_3}\left(A_{i_1},A_{i_1},A_{i_2},A_{i_2},A_{i_3},A_{i_3}\right)_\sigma\right]^n \\
=&\frac{1}{\prod_{c=0}^{5}(d^n+c)}\sum_{\sigma\in S_6}\left[\sum_{i_1,i_2,i_3}\left(A_{i_1},A_{i_1},A_{i_2},A_{i_2},A_{i_3},A_{i_3}\right)_\sigma\right]^n \\
&-3\,\frac{1}{d^n\,(d^n+1)}\frac{1}{\prod_{c=0}^{3}(d^n+c)}\sum_{\sigma\in S_4\times S_2}\left[\sum_{i_1,i_2,i_3}\left(A_{i_1},A_{i_1},A_{i_2},A_{i_2},A_{i_3},A_{i_3}\right)_\sigma\right]^n \\
&+3\left(\frac{1}{d^n\,(d^n+1)}\right)^3\sum_{\sigma\in S_2^3}\left[\sum_{i_1,i_2,i_3}\left(A_{i_1},A_{i_1},A_{i_2},A_{i_2},A_{i_3},A_{i_3}\right)_\sigma\right]^n \\
&-\left(\frac{1}{d^n\,(d^n+1)}\right)^3\sum_{\sigma\in S_2^3}\left[\sum_{i_1,i_2,i_3}\left(A_{i_1},A_{i_1},A_{i_2},A_{i_2},A_{i_3},A_{i_3}\right)_\sigma\right]^n
\end{aligned}$$

So far the computations have been for arbitrary quantum channel. Let us now again restrict ourselves to Pauli channel. In the third central moment, we find again the already discussed S_4 and S_2 terms, but now we also find permutations from S_6. Let us now classify these permutations.
First we have elements that couple only two or less boxes. We indicate the cycles by the partition belonging to its conjugation class. We have to be careful though, we have seen before that the symmetry orbits are not identical with the conjugation classes.

- Here it is not necessary to consider the different orbits for 4- respectively 5-cycles, for all 5-cycles one leg will be in an uncorrelated system; thus a 5-cycle will always be reduced to a 4-cycle.



- Here we only consider 4-cycles that are new in the S_6, all of them have the same structure.

$\lambda = (4,1,1)$ 

$= 2^{-2}$  $\in S_2^3$

- This conjugation class, $\lambda = (3,2,1)$, either effectively couples only one system,

$\lambda = (3,2,1)$ 

$= 2^{-2}$  $\in S_2^3$

- or it, again $\lambda = (3,2,1)$, couples two.

$\lambda = (3,2,1)$ 

$= 2^{-1}$  $\in S_4 \times S_2$

- The other 3-cycles are known already and are contained in S_4.

$\lambda = (3,1,1,1)$ 

$= 2^{-2}$  $\in S_2^3$

- Here we have to be careful, there are permutations in this conjugation class, which will lead to 3-correlations – see below, in the last diagram.

$\lambda = (2,2,2)$  $\in S_4 \times S_2$

- The other double transpositions are known.

$\lambda = (2,2,1,1)$ 

$= 2^{-2}$  $\in S_2^3$

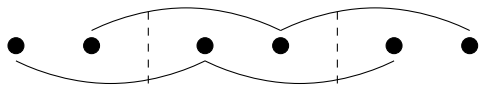
Second there are permutations that connect 3 boxes.

- The 6-cycles with $\delta = 1$. Note that not all 6-cycles belong to the same symmetry orbit, however they all have the same dimension.

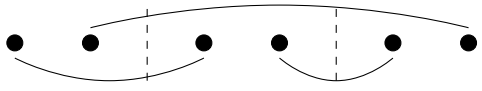
$\lambda = (6)$ 

- There are 2 classes of double 3-cycles, each with dimension $\delta = 2$.

$\lambda = (3,3)$ 

$\lambda = (3,3)$ 

- Last but most important, we have three transpositions connecting three boxes, with dimension $\delta = 3$.

$\lambda = (2,2,2)$ 

After studying all of these diagrams, it seems natural to ask, what has been achieved? We saw first that most of the new permutations in S_6 are equivalent to diagrams we have already found before. This is similar to the situation when we calculated the variance. A careful calculation will show that for the central moment the contributions of these diagrams will vanish up to the order in 2^n where the first 3-correlation appears. We will not show this calculation here, but we have provided Mathematica code in Appendix B, that allows the computation, furthermore we have calculated the central parts for all orbits and put them in tables in Appendix C, such that the results can be recalculated even without a computer.
Studying the diagrams, we found that the 3-correlation consisting of 3 transpositions is the highest dimensional 3-correlation. If its contribution is larger than $\frac{1}{2}$ it will dominate the third central moment and again allow us to find a single letter formula.

Theorem 3.6.3 (Regularized Third Moment in the Limit)**.** *The third regularized moment for the completely depolarizing and the identity channel is clearly* 0*.*
For $\sum_i p_i^3 < \frac{1}{2}$ *its absolute value is upper bounded by* $\frac{1}{2^4}$ *and for* $\sum_i p_i^3 \geq \frac{1}{2}$ *it is*

$$\lim_{n\to\infty} \mu_n^3 = \sum_i \left(\frac{p_i}{2}\right)^3.$$

The single letter formula is very elegant again. However, the condition

$$\sum_i p_i^3 \geq \frac{1}{2}$$

is rather strong, as it restricts the the allowed probability distributions, $\{p_i\}_{i=0}^3$, severely. In contrast do the variance we can no longer ignore the contributions of other trace products. Unfortunately these cannot be simplified in general, this means for the depolarizing channel, for most p, especially small p, we cannot accurately predict the n channel case.
Let us begin the discussion of this result by introducing the skewness of a distribution.

Definition 3.6.4 (Skewness)**.** *The skewness* v *of a distribution is the third central moment* μ_3 *normalized by the variance* σ*,*

$$v = \frac{\mu_3}{(\sigma)^{\frac{3}{2}}}.$$

The regularized skewness of the channel fidelity distribution will be called v_n*.*

The skewness tells us how asymmetric a distribution is, in particular a positive (negative) skewness means that the distribution has more states below (above) the average than we would expect in a gaussian distribution.
Before regularizing the skewness, we have to be careful with the regularization of quantities that can be negative, like the third central moment and the skewness. We can save regularization by only regularizing the absolute value and then multiply the quantity with the correct sign.

Proposition 3.6.5 (Regularized Skewness of Pauli Channel Fidelity Distribution in the limit)**.** *For the identity channel the skewness is obviously* 0 *regardless of the number of channel, if one of the other* $p_i = 1$*, then the skewness is* 1 *in the limit* $n \to \infty$ *channel, for* $\sum_i p_i^3 \geq \frac{1}{2}$ *we have:*

$$v_\infty = \frac{\mu_\infty^3}{Var_\infty^{\frac{3}{2}}} \in]0, 1[.$$

Proof. The first two statements are trivial.

For the third one we can use μ_∞^3 and have

$$0 < \frac{\sum_i p_i^3}{\left(\sum_i p_i^2\right)^{\frac{3}{2}}} < 1,$$

which can be seen using the multinomial theorem. □

Unfortunately we cannot make proper predictions for many $\{p_i\}$. For $\sum_i p_i^3 < \frac{1}{2}$ we can only bound by

$$|v_\infty| < \frac{\frac{1}{2^4}}{\left(\sum_i \left(\frac{p_i}{2}\right)^2\right)^{\frac{3}{2}}} \leq \frac{\frac{1}{2^4}}{\frac{1}{16^{\frac{3}{2}}}} = 2^2,$$

which, once we use it as an estimate for a concrete n, we have $|v_\infty|^n$, scales with n. It is no bound at all. The bound we could give for the third moment is not strong enough to also give a bound to the skewness.
We still do not expect the absolute value of the regularized skewness to be larger than 1 anywhere, that would be a skewness that grows with n, simply because the distribution will be very peaked around the mean and become more peaked with more channel. A negative skewness cannot grow with n, since the average approaches zero, there is simply no space for a few very low fidelity states. A growing positive skewness would then have to be due to a growing number of very high fidelity states, states for which the fidelity does not drop with the number of channel, in the non-regularized distribution. This would be equally surprising, as we will see in the end of this chapter, we have found some high fidelity states in former work, [11]. High fidelity in the non-regularized distribution means orders of magnitude larger than the average, but the fidelity of these states still approaches zero quickly for a large number of channel.
Let us explore our findings now again with the depolarizing channel.
In Figure 3.8 we plot the regularized third central moment for a few numbers of channel, unfortunately 10 channel is the maximal number of channel for which we can produce readable plots. Where the plots of μ_∞^3 and the bound intersect, is the largest p for which we have given the single letter formula. We can somewhat see how the real regularized third moment approaches the limit for a growing number

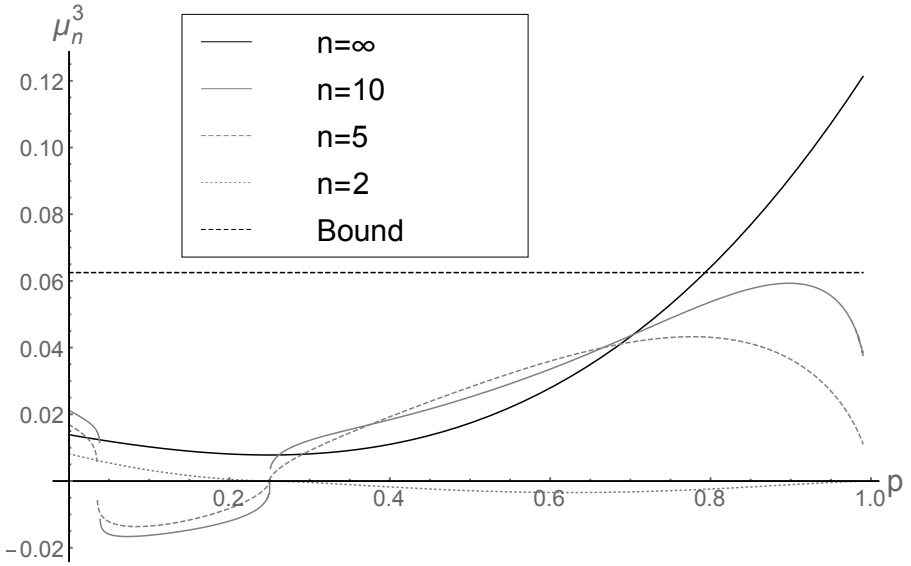


Figure 3.8.: Regularized third central moment for the depolarizing channel

of channel, however the singularity at $p = 1$ still has a strong influence. For smaller p the large n formula does not predict the real third central moment, nonetheless it is still in the same order of magnitude. The bound we gave is not very tight. For two channel the third central moment is very small. For 5 and 10 channel, at the completely depolarizing channel the third central moment drops to and passes through zero, just to come back up for p close to zero.

We can learn more about the small n distributions by looking at the non-regularized skewnessess in Figure 3.9. We see that the skewness is always a bit smaller than 1 but in contrast to the relative variance it does not get drastically smaller, if there is a relevant variance it will be skew. The most remarkable observation is that when passing through the completely depolarizing channel the skewness changes its sign. For 2 depolarizing channel the absolute value of the skewness is constant, it is, however, larger than zero for small p and smaller than 0 for $p > 0.25$. For low probabilities this means that while most states have a fidelity below the average, some states have a significantly higher fidelity and vise versa for $p > 0.25$.

The skewnessess for 8 and 5 channel are not really relevant, their distributions are very peaked already. For 3 channel the absolute value of the skewness is larger than for 2 channel, but the signs are reversed. We can interpret it like this; the average is even for high probabilities fairly low around p_0^3 and most states will be even below this average, however there are still a few states with fairly high probability. These

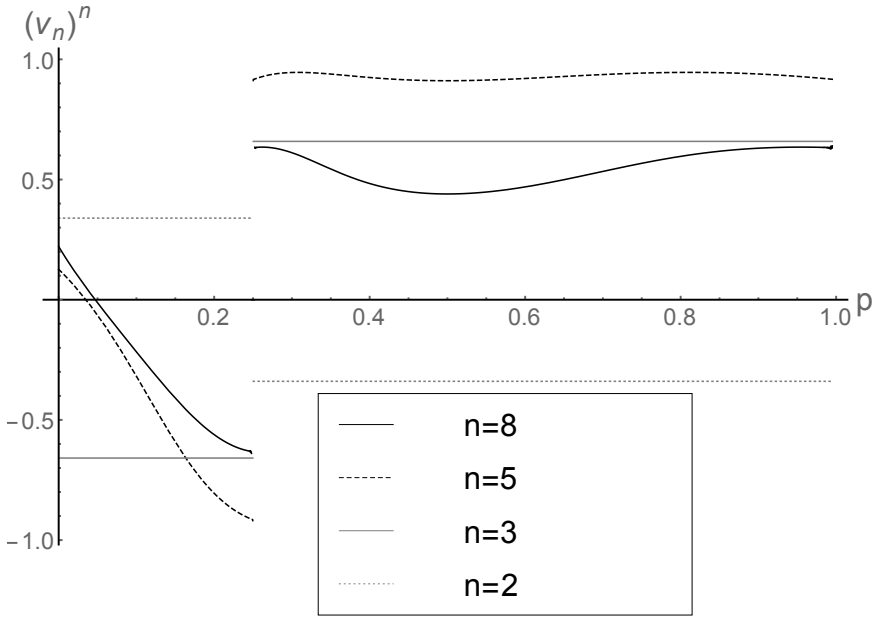


Figure 3.9.: Unregularized Skewnesses

states become much less common for more channel, which is in agreement with our prediction that the regularized skewness will be smaller than one everywhere.

For now we will end our discussion of channel fidelity distributions for a small number of channel. In the end of this chapter we will come back to it and add high fidelity states into the picture we found in former work [11].

Under some restrictions, we again found a very elegant form for the regularized central moment for infinite channel. It is a valid question to ask if and under which restrictions central moments will always be dominated by the q-transpositions, maybe we can find single letter formulas for all central moments. We will address this question in the next section.

3.7. Cumulants in the Limit

Beginning with the 4th central moment we will see a new type of correlation, namely that two or more sets of boxes are correlated independently. We see an example of this type of correlation in Figure 3.10.

Figure 3.10.: Double 2-correlation

This type of correlation cannot vanish, since it cannot appear in any of the other, smaller groups. It is a correlation that first appears in S_8. It is an element of $S_4 \times S_4$, but in the formula for the central moment, Proposition 3.6.1, we can never have a factor S_4, there are only S_2 added.
If we calculate the fourth central moment exactly, using either Mathematica or the presented tables in Appendix C. We find again that all orders of 2^n vanish up to where the first new correlations appear.

Proposition 3.7.1 (Fourth Central Moment)**.**

$$\lim_{n\to\infty} \mu_n^4 = \lim_{n\to\infty} \frac{1}{d^4} \left(48 \left[\sum_i p_i^4 \right]^n + 12 \left[\sum_i p_i^2 \right]^{2n} - 36\, p_0^{4n} - 24 \left[p_0^2 \sum_i p_i^2 \right]^n \right)^{\frac{1}{n}}$$

In contrast to the second and third central moment, where we could give a simple restriction to the p_i and then evaluate the limit for arbitrary p_i, to calculate this limit we need additional information about the probability distribution. A closer look at Proposition 3.7.1 reveals that in the expression for the fourth central moment, we have terms that look very similar to the second central moment; $\sum_i p_i^2$. The fourth cumulant is a function of the second central moment and the fourth central moment; to introduce the concept of cumulants we need to take a small detour.
For any distribution one can define its characteristic function.

Definition 3.7.2 (Characteristic Function)**.** *For a distribution $D(x)$ over a random variable X its characteristic function $\psi_X(t)$ is its inverse Fourier transform:*

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{itx} dD(x).$$

The characteristic function also completely determines the distribution $D(x)$. Moreover the logarithm of the characteristic function is a cumulant generating function.

Fact 3.7.3 (Cumulant Generating Function)**.** *The logarithm of a characteristic function $\psi_X(t)$ belonging to a probability distribution $D(x)$ over the random variable X is a function of the distributions cumulants;*

$$\ln(\psi_X(t)) = \sum_{q=1}^{\infty} \kappa^n \frac{(it)^q}{q!}.$$

In this sense cumulants are characteristic for a probability distribution. The cumulants can be written as functions of central moments.

Fact 3.7.4 (Cumulants)**.** *The first five cumulants κ^i [29] for a probability distribution are given as functions of the central moments μ^i in the following way:*

$$\begin{aligned}\kappa^1 &= \mu^1\\ \kappa^2 &= \mu^2\\ \kappa^3 &= \mu^3\\ \kappa^4 &= \mu^4 - 3\,\mu_2^2\\ \kappa^5 &= \mu^5 - 10\,\mu_3\,\mu_2\end{aligned}$$

We notice that we actually have already calculated the first three cumulants, since they are identical to the central moments. We will now see that the cumulants are actually the objects that have a simple limit.

Proposition 3.7.5 (Regularized Fourth Cumulant)**.** *The regularized fourth cumulant of the channel fidelity distribution for Pauli channel is 0 for the identity and the completely depolarizing channel and if $\sum_i p_i^4 \geq \frac{1}{2}$ it is*

$$\lim_{n\to\infty} \left(\kappa_n^4\right)^{\frac{1}{n}} = \sum_i \left(\frac{p_i}{2}\right)^4.$$

For all other distributions of p_i we can again only give a bound:

$$\kappa_n^4 \leq \frac{1}{2^5}.$$

Proof. The statements for the identity channel and the completely depolarizing channel are trivial. For $\sum_i p_i^4 > \frac{1}{2}$, let us recall the second central moment up to relevant order before taking the limit:

$$\mathrm{Var}_n^2 = \mu_n^2 = \frac{2^{2n}\left(2\,\left(\sum_i p_i^2\right)^n - 2\,p_0^{2n}\right) + O(2^n)}{(2^n+1)^2\,(2^n+2)\,(2^n+3)}.$$

The fourth central moment is.

$$\mu_n^4 = \frac{2^{7n}\left(48\,\left[\sum_i p_i^4\right]^n + 12\,\left[\sum_i p_i^2\right]^{2n} - 36\,p_0^{4n} - 24\,\left[p_0^2\,\sum_i p_i^2\right]^n\right) + O(2^{6n})}{(2^n+7)(2^n+6)(2^n+5)(2^n+4)(2^n+3)(2^n+2)(d^n+1)^4}$$

We are only interested in the highest order of 2^n.

$$\begin{aligned}
\kappa_n^4 &= \mu_n^4 - 3\left(\mu_n^2\right)^2 \\
&= 2^{-4n}\left(48\left[\sum_i p_i^4\right]^n + 12\left[\sum_i p_i^2\right]^{2n} - 36\,p_0^{4n} - 24\left[p_0^2\sum_i p_i^2\right]^n\right. \\
&\quad \left. - \left(12\left(\sum_i p_i^2\right)^{2n} - 24\left(\sum_i p_i^2\right)p_0^{2n} + 12\,p_0^{4n}\right)\right) + O(2^{-5n}) \\
&= 2^{-4n}\left(48\left[\sum_i p_i^4\right]^n - 48\,p_0^{4n}\right) + O(2^{-5n})
\end{aligned}$$

Here it is again easy to take the limit of infinite channel and we obtain the desired result. The bound follows trivially. □

Now that we have this elegant result, let us have a look into the next central moment and the according cumulant in diagrams. We claim that the next central moment will be dominated by the 5-transpositions.
Looking at Figure 3.11 and Figure 3.12 it is easy to see that these are the only new correlations we can create using transpositions only. We see that for the cumulant we have to remove the product of the third and second central moment.
The correlation in Figure 3.12 is exactly that, a product of the second and third central moment.

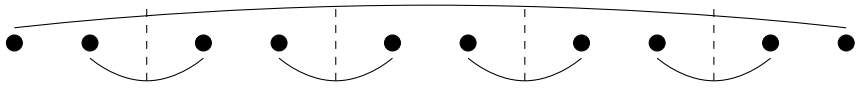


Figure 3.11.: 5-Transposition

Figure 3.12.: 3-Transposition times 2-Transposition

Let us now step up to the general picture. For all cumulants where we could find single letter formulas the dominating correlation had an elegant form.

Definition 3.7.6 (q-transposition)**.** *A permutation that has q disjoint cycles of length 2 is called a q-transposition.*

We now claim that the qth central moment will always be dominated by the q-correlation consisting of q-transpositions. We have seen that a transposition that

only permutes elements within one system cannot contribute to a larger correlation. That means that all correlations of the relevant type have at least to be a product of two or more correlations. In fact it is easy to see that we will find all possible correlations by looking at the partitions of q that do not contain a 1.

Example 3.7.7 (Possible Correlations)**.** *Let be $q = 6$. The possible partitions of 6 that do not contain a 1 are:*

$$\begin{aligned}\lambda_1 &= (6),\\ \lambda_2 &= (4,2),\\ \lambda_3 &= (3,3),\\ \lambda_4 &= (2,2,2).\end{aligned}$$

We expect the sixth cumulant to be a function of μ^6, $\mu^4\,\mu^2$, $\left(\mu^3\right)^2$ and $\left(\mu^2\right)^3$ and indeed it is given as:

$$\kappa^6 = \mu^6 - 15\,\mu^4\,\mu^2 - 10\left(\mu^3\right)^2 + 30\,\left(\mu^2\right)^3.$$

On first glance one might find it surprising that suddenly we have to add the triple 2-correlations. This is due to the fact that the fourth moment already contains correlations of that type and we subtract too many.

We have seen that we can state the fourth cumulant in the limit of infinite channel explicitly. In fact given that our conjecture about q-correlations and the qth central moment is true, we will see that we can take the limit for all cumulants, however there is a trade off. The space of probability distributions that allow the simple limits, becomes smaller and smaller for higher cumulants. Let us formalize the conjecture.

Conjecture 3.7.8 (qth Central Moment)**.** *If $1 > \sum_i p_i^q \geq \frac{1}{2}$ the regularized qth central moment of the of the channel fidelity distribution for Pauli channel in infinite channel limit is dominated by those q-correlations that are created by q-transpositions.*

We did not find a conclusive proof for Conjecture 3.7.8. While it is certainly easy to see, that the q-transpositions that create a q-correlation have in general the highest dimension of any q-correlation, it is more difficult to show that all lower correlations cancel each other out, such that the q-transpositions can dominate the qth moment. It is particularly difficult to count all permutations that relate to a certain correlation. Not only is it difficult to split every conjugation class according to the orbits of the relevant isotropy group, but even if this is done for a concrete q, in any larger group there are inherently new conjugation classes, which have to be classified again.
It would have been interesting to check our conjecture for $q = 5$ or $q = 6$. Unfortunately a calculation of the 5th central moment was already impossible. Our code relies on a Mathematica function called "ToCycles" to convert elements of the symmetric group into cycle form. This function is very slow: the computation time increases

rapidly, while the second moment could be calculated in seconds, the calculation of the third central moment took minutes, of the fourth hours. The calculation for the fifth exceeded the available memory. A manual calculation would have taken too much time as well, there are simply too many elements in S_{10}.

However, with Conjecture 3.7.8 we can find an elegant formula for all cumulants.

Theorem 3.7.9 (qth regularized cumulant in the limit)**.** *Given Conjecture 3.7.8 is true, the regularized qth cumulant of the channel fidelity distribution for Pauli channel is* 0 *for the identity and the completely depolarizing channel and for $\sum_i p_i^q \geq \frac{1}{2}$:*

$$\lim_{n\to\infty} \kappa_n^q = \lim_{n\to\infty} \left(\left[\sum_i \left(\frac{p_i}{2}\right)^q \right]^n - \left[\frac{p_0}{2}\right]^{qn} \right)^{\frac{1}{n}} = \sum_i \frac{p_i^q}{2^q}.$$

For all other distributions of p_i we can only give a bound:

$$\kappa_\infty^q \leq \frac{1}{2^{q+1}}.$$

For the proof we first need to understand the connection between central moments and cumulants. It is usually described either in terms of Faà di Bruno's formula or using the Bell Polynomials [29], [6].

Definition 3.7.10 (Bell Polynomial)**.** *The Bell Polynomials $B_{n,k}(x_1, x_2, \ldots, x_{n-k+1})$ are defined as:*

$$B_{n,k} = \sum_{\sum_i j_i = k,\, \sum_i i\, j_i = n} \frac{n!}{j_1! \cdots j_{n-k+1}!} \left(\frac{x_1}{1!}\right)^{j_1} \cdots \left(\frac{x_{n-k+1}}{(n-k+1)!}\right)^{j_{n-k+1}}.$$

This definition is rather overwhelming. It does, however, a rather simple job. The Bell polynomial $B_{n,k}$ counts partitions of n with length k.

Example 3.7.11 (Bell Polynomial)**.** *Consider the Bell polynomial $B_{3,2}$. There is only one partition of* 3 *with length* 2*, namely* $(2,1)$*. $B_{3,2}$ will give how many different configurations of three elements can lead to this partition and the answer is obviously three.*

$$B_{3,2} = \frac{3!}{1!\,1!} x_1 \frac{x_2}{2} = 3\, x_1 x_2$$

It is a bit more interesting to ask how many partitions of length 2 *does* 4 *have and how often do they appear. Length* 2 *means that at most two of the j_i can be non-zero, however, the restriction $\sum_i i j_i = n$ means that only $j_2 = 2$ ($j_1 = 0$) and $j_1 = j_3 = 1$ ($j_2 = 0$) are valid options.*

$$B_{4,2} = \frac{4!}{2!} \left(\frac{x_2}{2!}\right)^2 + \frac{4!}{1!\,1!} x_1 \frac{x_3}{3!} = 3\, x_2^2 + 4\, x_1 x_3$$

Thus there are three distinct possibilities to form two pairs out of four elements; (12)(34)*,* (13)(24) *and* (14)(23)*, and four ways to pick three;* (1)(234)*,* (2)(134)*,* (3)(124) *and* (4)(123)*.*

Keeping this in mind we state moments as a function of cumulants [29].

Lemma 3.7.12 (Moments as a function Cumulants)**.** *The qth moment* $m^q(\kappa^1, \ldots, \kappa^n)$ *as a function of cumulants* κ^i *is:*

$$m^q(\kappa^1, \ldots, \kappa^q) = \sum_{k=1}^{q} B_{q,k}(\kappa^1, \ldots, \kappa^{q-k+1}).$$

The central moments are obtained by setting $\kappa_1 = 0$*:*

$$\mu^q(\kappa^2, \ldots, \kappa^q) = m^q(0, \kappa^2, \ldots, \kappa^q).$$

With this lemma we can get into the proof of Theorem 3.7.9.

Proof. Given Conjecture 3.7.8 is true, the restriction $\sum_i p_i^q \geq \frac{1}{2}$ guarantees that the q-transpositions will dominate the central moment.
In Lemma 3.7.12 we see that we can express the qth moment as a sum of all cumulants combined such that their indices add up to q. However for the q central moment, we have the extra restriction that any product involving the first cumulant will be equal to 0.
In the language of our diagrams that means that in the expression for central moments, we will only have correlations that at least involve two systems.
A q correlation that is the product of independently correlated boxes, will be reflected as a product of cumulants. This cumulant product, let us call it $c = \beta\kappa_{\alpha_1} \cdots \kappa_{\alpha_m}$, has to obey $\sum_i \alpha_i = q$, since over all it will still correlate q systems.
The coefficient β represents the possibilities to have a $\alpha_1 \cdots \alpha_m$ correlation given q systems, so we can find β using the according Bell polynomial.
This way there is a one to one connection between the qth cumulant and the q-transpositions. □

Even if we assume that the single letter formulas we found for the regularized cumulants of the channel fidelity distribution for Pauli channel are correct, we can still not write down a characteristic function of F_n for any interesting quantum channel.
While we can always find probability distributions such that the cumulant κ_∞^q approximates the real cumulant of the according channel fidelity distribution very well, we cannot find all cumulants for a concrete probability distribution.
We can only find it if one of the probabilities is equal to 1, i.e. if we have a unitary channel.
Concluding, the cumulants are, aside of their elegance, not very useful as they cannot give us more insight into the distribution itself, as they only apply to a smaller and smaller probability space. We still appreciate finding the single letter formulas, as they show it is possible to find properties of a quantity extended to an n-fold Hilbert space.
Furthermore we have learned in Proposition 3.4.19 that in the limit of large channel numbers, the distribution is very peaked, which means that basically all states will

have the same fidelity as the average. We can even approximate the channel as a single depolarizing channel over the n fold Hilbert space. However, in the next section we will review some of our former work, where we managed to find states with much higher channel fidelity. That shows that the approximation loses crucial information.

3.8. Maximizing Channel Fidelities

As we have seen in Theorem 1.5.2 the real challenge is to maximize the coherent information over larger tensor product spaces. In our former work, [11], we tried to do that for channel fidelities.
As reaction to a talk at the Institute for Quantum Computing (Waterloo, Ontario, Canada) that the author gave, we got some external input, which we will present here, [34].
We begin by quickly restating our findings and then discuss the more recent insights, and finally we will discuss our older results in the context of the new results of this chapter and the previous chapter.
The objective is to maximize the channel fidelity,

$$\max_{\phi} \left[F_n(|\phi\rangle, \mathcal{N})\right]^{\frac{1}{n}} =?,$$

by providing a maximizing state.
When looking at the channel fidelity,

$$F(|\phi\rangle, \mathcal{N}) = \langle\phi| \mathcal{N}\left(|\phi\rangle\langle\phi|\right) |\phi\rangle,$$

we see that the operator $\mathcal{N}\left(|\phi\rangle\langle\phi|\right)$ is in the center of attention. It is certainly self-adjoint and as such can be diagonalized, furthermore it is (since a quantum channel is positivity and trace preserving) positive and has trace one. If we were now to find an operator $\mathcal{N}\left(|\phi_0\rangle\langle\phi_0|\right)$ such that its eigenvector with the largest eigenvalue is $|\phi_0\rangle$, then the fidelity of $|\phi_0\rangle$ would be at least $\frac{1}{d}$, $d = \dim \mathcal{H}$, but very likely higher.
For a common operator on a Hilbert space it is well known that, as long as its largest eigenvalue is not degenerate, an iterative application of this operator on itself will quickly produce a projector onto the eigenvector with this largest eigenvalue.
Inspired by this procedure we will now provide an iterative algorithm for finding locally maximal states of $F(|\phi\rangle, \mathcal{N})$. Since $\mathcal{N}\left(|\phi\rangle\langle\phi|\right)$ are not ordinary operators, but depend on the input state, we need to find a name for those $|\phi\rangle$ that are also an eigenvector of $\mathcal{N}\left(|\phi\rangle\langle\phi|\right)$ or rather $\left(\mathcal{N} + \mathcal{N}^\dagger\right)\left(|\phi\rangle\langle\phi|\right)$.
Before we can go into this, we have to understand the adjoint of a quantum channel.

Definition 3.8.1 (Adjoint of a Quantum Channel)**.** *For a quantum channel $\mathcal{N}$ and two operators ρ_1 and ρ_2 on a Hilbert space $\mathcal{H}$, we define its adjoint as usual by conserving the Hilbert Schmidt inner product Definition 1.2.1 in the following way:*

$$(\rho_1, \mathcal{N}(\rho_2))_{HS} = \left(\mathcal{N}^\dagger(\rho_1), \rho_2\right)_{HS}.$$

The adjoint of a quantum channel is certainly still completely positive, as it still has Kraus form, however it is not necessarily trace preserving.

Definition 3.8.2 (Fix Point)**.** *For a quantum channel $\mathcal{N}$ we call $|\phi\rangle$ a fix point of $\mathcal{N}$ if*

$$\left(\mathcal{N}+\mathcal{N}^{\dagger}\right)\left(|\phi\rangle\langle\phi|\right)|\phi\rangle \propto |\phi\rangle\,,$$

which for self-adjoint quantum channel simplifies to

$$\mathcal{N}(\Phi)\,|\phi\rangle \propto |\phi\rangle\,.$$

We will see that critical states of $F(|\phi\rangle\,,\mathcal{N})$ are fix points of $\mathcal{N}$.

Definition 3.8.3 (Critical States)**.** *A state $|\phi\rangle$ is called a critical state of F iff*

$$\frac{\partial}{\partial\,|\chi\rangle}F(|\phi\rangle\,,\mathcal{N})|_{|\chi\rangle=0}=0.$$

Lemma 3.8.4. *If $|\phi\rangle$ is a critical state of $F(|\phi\rangle\,,\mathcal{N})$ then it is a fix point of $\mathcal{N}$.*

Proof. By performing the directional derivative with an arbitrary vector $|\chi\rangle$ orthogonal to $|\phi\rangle$.

$$\begin{aligned}
\frac{\partial}{\partial\,|\chi\rangle}F(|\phi\rangle\,,\mathcal{N})|_{|\chi\rangle=0} &= \frac{d}{dt}\mathrm{Tr}\left(|\phi+t\chi\rangle\langle\phi+t\chi|\;\mathcal{N}(|\phi+t\chi\rangle\langle\phi+t\chi|)\right)\\
&= \mathrm{Tr}\left(|\chi\rangle\langle\phi|\;\mathcal{N}(|\phi\rangle\langle\phi|)\right)+\mathrm{Tr}\left(|\phi\rangle\langle\chi|\;\mathcal{N}(|\phi\rangle\langle\phi|)\right)\\
&+\mathrm{Tr}\left(|\phi\rangle\langle\phi|\;\mathcal{N}(|\chi\rangle\langle\phi|)\right)+\mathrm{Tr}\left(|\phi\rangle\langle\phi|\;\mathcal{N}(|\phi\rangle\langle\chi|)\right)\\
&= \mathrm{Tr}\left(|\chi\rangle\langle\phi|\;\mathcal{N}(\Phi)\right)+\mathrm{Tr}\left(|\phi\rangle\langle\chi|\;\mathcal{N}(\Phi)\right)\\
&+\mathrm{Tr}\left(|\chi\rangle\langle\phi|\;\mathcal{N}^{\dagger}(\Phi)\right)+\mathrm{Tr}\left(|\phi\rangle\langle\chi|\;\mathcal{N}^{\dagger}(\Phi)\right)\\
&= 2\,Re\Big(\langle\chi|\left(\mathcal{N}(\Phi)+\mathcal{N}^{\dagger}(\Phi)\right)|\phi\rangle\Big)
\end{aligned}$$

Since for $|\phi\rangle$ critical the derivative needs to vanish for any choice of $\langle\chi|$, we have the fore mentioned condition. □

In Lemma 3.8.4 we saw that a critical state needs to be a fix point, Definition 3.8.2, and similar to the iteration of an ordinary operator we iterate here, as follows.

Definition 3.8.5 (Fix Point Iteration)**.** *For a quantum channel $\mathcal{N}$ the following procedure is called fix point iteration:*

1. *Begin with a (pseudo) random state $|\phi_0\rangle$.*
2. *Evaluate $\mathcal{N}(\Phi_i)+\mathcal{N}^{\dagger}(\Phi_i)$ and let the resulting operator act on $|\phi_i\rangle$.*
3. *Normalize the resulting vector: $|\phi_{i+1}\rangle \equiv \frac{\left(\mathcal{N}(\Phi_i)+\mathcal{N}^{\dagger}(\Phi_i)\right)|\phi_i\rangle}{||\mathcal{N}(\Phi_i)+\mathcal{N}^{\dagger}(\Phi_i)|\phi_i\rangle||}$.*

4. Go to step 2.

Clearly the "fix points" from Definition 3.8.2 are fix points of the iteration.
For the proof that the iteration actually produces useful results we need more assistance.

Lemma 3.8.6 (Fix Point Basis)**.** *For any fix point ϕ_0 of $\mathcal{N}$ we can represent $\mathcal{N}(|\phi_0\rangle\langle\phi_0|)+\mathcal{N}^\dagger(|\phi_0\rangle\langle\phi_0|)$ as*

$$\mathcal{N}(|\phi_0\rangle\langle\phi_0|)+\mathcal{N}^\dagger(|\phi_0\rangle\langle\phi_0|)=\sum_i \lambda_i |\psi_i\rangle\langle\psi_i|,$$

where one of the ψ_i is equal to ϕ_0 and the other vectors are mutually orthogonal and all $\lambda_i \geq 0$.

Proof. Given that the input to $\mathcal{N}+\mathcal{N}^\dagger$ is hermitian, as is the output by definition, such we can find an orthogonal basis where the operator is diagonal, one of the basis vectors is obviously the fix point, since $\mathcal{N}+\mathcal{N}^\dagger$ acts diagonally on it. Furthermore $\mathcal{N}+\mathcal{N}^\dagger$ is positivity preserving and a rank one projector is a positive operator. □

Corollary 3.8.7. *For a fix point ϕ_0 its channel fidelity is given as*

$$F(\Phi_0,\mathcal{N})=\frac{\lambda_0}{2}.$$

Proof. First we see that we can rewrite the fidelity as a trace and there, understanding the trace as a Hilbert Schmidt inner product, we replace $\mathcal{N}$ with $\frac{1}{2}\left(\mathcal{N}+\mathcal{N}^\dagger\right)$.

$$\begin{aligned}
F(\Phi_0,\mathcal{N}) &= \mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\mathcal{N}(|\phi_0\rangle\langle\phi_0|)\right)\\
&= \frac{1}{2}\mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\mathcal{N}(|\phi_0\rangle\langle\phi_0|)\right)+\frac{1}{2}\mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\mathcal{N}(|\phi_0\rangle\langle\phi_0|)\right)\\
&= \frac{1}{2}\mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\mathcal{N}(|\phi_0\rangle\langle\phi_0|)\right)+\frac{1}{2}\mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\mathcal{N}^\dagger(|\phi_0\rangle\langle\phi_0|)\right)\\
&= \frac{1}{2}\mathrm{Tr}\left(|\phi_0\rangle\langle\phi_0|\left(\mathcal{N}+\mathcal{N}^\dagger\right)(|\phi_0\rangle\langle\phi_0|)\right)=\frac{\lambda_0}{2}
\end{aligned}$$

For the last step we evaluate the trace in the eigenbasis of $\left(\mathcal{N}+\mathcal{N}^\dagger\right)(\Phi_0)$ as in Lemma 3.8.6. □

Lemma 3.8.8. *For any fix point ϕ_0 of $\mathcal{N}$ and an arbitrary orthogonal vector χ we have*

$$\langle\phi_0|\left(\mathcal{N}+\mathcal{N}^\dagger\right)(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|)|\phi_0\rangle=0.$$

Proof. Obviously $\mathcal{N}+\mathcal{N}^\dagger$ is hermitian and as $\left(\mathcal{N}+\mathcal{N}^\dagger\right)(|\phi_0\rangle\langle\phi_0|)$ is symmetric its left and right eigenvectors are the same.

$$\langle\phi_0|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle = \mathrm{Tr}\left(\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle\langle\phi_0|\right)$$
$$= \mathrm{Tr}\left(\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\phi_0|\right)\right)$$
$$= \langle\chi|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\phi_0|\right)|\phi_0\rangle + \langle\phi_0|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\phi_0|\right)|\chi\rangle = 0$$

□

Lemma 3.8.9. *For arbitrary ϕ_0, χ the matrix element $\langle\phi_0|\left[\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)\right]^2|\phi_0\rangle$ is positive.*

Proof. Let $\{\psi_i\}$ be any basis. Then we can insert a $\mathbb{1} = \sum_i |\psi_i\rangle\langle\psi_i|$:

$$\begin{aligned}
&\langle\phi_0|\left[\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)\right]^2|\phi_0\rangle \\
=&\sum_i \langle\phi_0|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\psi_i\rangle\langle\psi_i|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle \\
=&\sum_i |\langle\psi_i|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle|^2 > 0
\end{aligned}$$

□

Corollary 3.8.10. *Especially if we choose χ as the first basis vector and such the basis $\{\chi, \psi_j\}$, we can write*

$$\begin{aligned}
&\langle\phi_0|\left[\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)\right]^2|\phi_0\rangle \\
=&|\langle\chi|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle|^2 + \sum_j |\langle\psi_j|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle|^2
\end{aligned}$$

For simplification we introduce abbreviations $|b|^2$ for the first part and c for the positive sum, whichbe useful for the proof of the following theorem:

$$\begin{aligned}
b &\equiv \langle\chi|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle \\
|b|^2 &= |\langle\chi|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle|^2 \\
c &\equiv \sum_j |\langle\phi_j|\left(\mathcal{N}+\mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi|+|\chi\rangle\langle\phi_0|\right)|\phi_0\rangle|^2.
\end{aligned}$$

Finally we can show that the iteration indeed finds locally maximal states.

Theorem 3.8.11. *If $|\phi\rangle$ is a stable fix point of the iteration, Definition 3.8.5, then it is a local maximum.*

Proof. The idea is to show that a non-maximal point is at best an unstable fix point of the iteration and as such will not be found numerically. This will be done by showing that the iteration transports a state close to a non-maximal state away from it, such the iteration only converges to locally maximal fix points.

Let now be ϕ_0 be a non-maximal fix point with

$$\left(\mathcal{N}(|\phi_0\rangle\langle\phi_0|) + \mathcal{N}^\dagger(|\phi_0\rangle\langle\phi_0|)\right)|\phi_0\rangle = \lambda_0 |\phi_0\rangle,$$

Without loss of generality we can assume that the direction where the fidelity can be increased χ is also an eigenvector of $\mathcal{N}(|\phi_0\rangle\langle\phi_0|) + \mathcal{N}^\dagger(|\phi_0\rangle\langle\phi_0|)$.

We define

$$|\phi_1\rangle \equiv \frac{1}{\sqrt{1+\epsilon^2}}\left(|\phi_0\rangle + \epsilon\,|\chi\rangle\right)$$

and have

$$F(|\phi_0\rangle, \mathcal{N}) < F(|\phi_1\rangle, \mathcal{N}),$$

where the fidelities are with Corollary 3.8.7, and b as in Corollary 3.8.10

$$\begin{aligned} F(|\phi_0\rangle, \mathcal{N}) =& \frac{\lambda_0}{2} \\ F(|\phi_1\rangle, \mathcal{N}) =& \left(\frac{1}{1+\epsilon^2}\right)^2 \left(\frac{\lambda_0}{2} + \epsilon^2(\lambda_\chi + b) + O\left(\epsilon^3\right)\right). \end{aligned}$$

Now the inequality can be simplified with Corollary 3.8.10 to

$$\lambda_0 < \lambda_\chi + b + O(\epsilon). \tag{3.3}$$

Then we define ϕ_2 by iteration

$$\begin{aligned} |\phi_2'\rangle &= \left(\mathcal{N}(\Phi_1) + \mathcal{N}^\dagger(\Phi_1)\right)|\phi_1\rangle \\ |\phi_2\rangle &= \frac{|\phi_2'\rangle}{|\phi_2'|} \end{aligned}$$

and find

$$
\begin{aligned}
|\phi_2'\rangle &= \lambda_0 |\phi_0\rangle + \epsilon \left(\left(\mathcal{N} + \mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi| + |\chi\rangle\langle\phi_0|\right)|\phi_0\rangle + \left(\mathcal{N} + \mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\phi_0|\right)|\chi\rangle \right) \\
&\quad + \epsilon^2 \left(\left(\mathcal{N} + \mathcal{N}^\dagger\right)\left(|\phi_0\rangle\langle\chi| + |\chi\rangle\langle\phi_0|\right)|\chi\rangle + \left(\mathcal{N} + \mathcal{N}^\dagger\right)\left(|\chi\rangle\langle\chi|\right)|\phi\rangle \right) + O(\epsilon^3) \\
|\phi_2'|^2 &= \lambda_0^2 + \epsilon^2 \left(b^2 + 2b\lambda_0 + 2\lambda_\chi\lambda_0 + 2b\lambda_\chi + \lambda_\chi^2 + c\right) + O(\epsilon^3).
\end{aligned}
$$

Where all the first order terms and the last second order term vanish because of Lemma 3.8.8.
Finally we compare $|\langle\phi_0|\phi_1\rangle|^2$ with $|\langle\phi_0|\phi_2\rangle|^2$ and again using Lemma 3.8.8 find the desired:

$$|\langle\phi_0|\phi_1\rangle|^2 > |\langle\phi_0|\phi_2\rangle|^2$$

since

$$
\begin{aligned}
&\frac{1}{1+\epsilon^2} > \frac{\lambda_0^2 + 2\lambda_0\,\epsilon^2(b+\lambda_\chi)}{|\phi_2'|^2} + O(\epsilon^3) \\
\Leftrightarrow\ & |\phi'|^2 > \lambda_0^2 + \epsilon^2\lambda_0^2 + 2\lambda_0\,\epsilon^2(b+\lambda_\chi) + O(\epsilon^3) \\
\Leftrightarrow\ & \lambda_0^2 + \epsilon^2\left(b^2 + 2b\lambda_0 + 2\lambda_\chi\lambda_0 + 2b\lambda_\chi + \lambda_\chi^2 + c\right) > \lambda_0^2 + \epsilon^2\lambda_0^2 + 2\lambda_0\,\epsilon^2(b+\lambda_\chi) + O(\epsilon^3) \\
\Leftrightarrow\ & \lambda_\chi^2 + 2b\lambda_\chi + b^2 + c > \lambda_0^2 + O(\epsilon) \\
\Leftrightarrow\ & (\lambda_\chi + b)^2 + c > \lambda_0^2
\end{aligned}
$$

is obviously true considering Lemma 3.8.9 and inequality 3.3 on the last page within the proof. □

So if the algorithm converges it was successful in finding locally maximal states. The algorithm has a long run time, it becomes impractical already for five qubits. We did extensive calculations with three qubits — possible within minutes — and a few calculations with four qubits, which took about a day. Still we could use it to find distinct local maxima, even with different fidelities, [11]. Because of a huge amount of trials, we assume that we found all maxima.
Unfortunately the algorithm does not increase the fidelity with each step: we entered states with fidelity above the mean and a single iteration lowered the fidelity.

3.8.1. An Improved Algorithm?

Certainly we would have preferred a monotonous iteration. It was suggested [34] to simplify the algorithm and use previous work [33] on directional iterates to construct a monotonous algorithm. Unfortunately the simplification works only for weakly disturbed channel.

We will now introduce the simplified version and show its properties.

Definition 3.8.12 (Tyson)**.** *Simplified fix point iteration for the Channel Fidelity:*

1. *Begin with a (pseudo) random state* $|\phi_0\rangle$.
2. *Define* $|\phi_{i+1}\rangle$ *as one of the eigenvectors of* $\mathcal{N}(\Phi_i) + \mathcal{N}^\dagger(\Phi_i)$ *with largest eigenvalue.*
3. *Go to step 2.*

Testing the algorithm did not give the expected results. To proof that the algorithm is monotonous, Jon Tyson had assumed that we can construct an inner product with $\mathcal{N}$ as one can usually do with ordinary positive definite operators.
It turned out that it is a quite strong restriction for a quantum channel to allow the definition of an inner product. For illustration we will now give a simple example, where a quantum channel does not provide an inner product.

Example 3.8.13 (Counter Example)**.** *Suppose the quantum channel with only one Kraus operator* σ_1*, then define*

$$(A, B)_{\sigma_1} := (A, \sigma_1 B \sigma_1)_{HS},$$

where we easily find operators whose "norm" will be imaginary

$$(\sigma_2, \sigma_2)_{\sigma_1} = Tr(\sigma_2\sigma_1\sigma_2\sigma_1) = -2.$$

This shows clearly that we cannot find an inner product for every quantum channel. Obviously the identity channel would allow it, which then will extend smoothly to weakly disturbed channel. However with the assumption, that the channel will allow the construction of an inner product, we can indeed find a monotonous iteration.

Theorem 3.8.14 (Tyson)**.** *If* $\mathcal{N}$ *allows the definition of an inner product,*

$$(A, B)_{\mathcal{N}} = (A, (\mathcal{N} + \mathcal{N}^\dagger)(B))_{HS},$$

the simplified iteration is monotonous for $\mathcal{N}$*.*

$$F(\Phi_{i+1}, \mathcal{N}) \geq F(\Phi_i, \mathcal{N})$$

Unfortunately the non-trivial (entanglement dependent) high fidelity states are just found for strongly disturbed channel [11].

3.8.2. Final Discussion of Channel Fidelity Distribution

Using the iterative algorithm Definition 3.8.5 we found high fidelity states for Pauli channel, [11]. Since we know that the states we found are locally maximizing states. We assume that these states also achieve the global maximal channel fidelity, more precisely we conjectured the following.

Conjecture 3.8.15. *The maximal states for a Pauli channel fidelity are either:*

$$|\phi_0\rangle = |0\rangle^{\otimes n}$$
$$|\phi_1\rangle = |1\rangle^{\otimes n}$$

or

$$|\phi\rangle_+ = \frac{|+\rangle^{\otimes n} + |-\rangle^{\otimes n}}{\sqrt{2}}$$
$$|\phi\rangle_- = \frac{|+\rangle^{\otimes n} - |-\rangle^{\otimes n}}{\sqrt{2}}$$

with $|+/-\rangle = \frac{|0\rangle+/-|1\rangle}{\sqrt{2}}$.

In particular the maximal regularized fidelity is $\frac{1}{2}$ *for the completely depolarizing channel and otherwise given as:*

$$\begin{aligned} F_{max} &= \max_{|\phi\rangle} \lim_{n\to\infty} \left(F(|\phi\rangle, \mathcal{N}^{\otimes n})\right)^{\frac{1}{n}} \\ &= \max \lim_{n\to\infty} \{p_0 + p_3, \frac{1}{2^{\frac{1}{n}}} \left((p_2+p_3)^n + (p_0+p_1)^n + (p_3-p_2)^n + (p_0-p_1)^n\right)^{\frac{1}{n}}\}. \end{aligned}$$

For the last expression, we assumed without loss of generality that $p_3 \geq p_2 \geq p_1$, *while* p_0 *can be smaller or larger than any of those.*

Now we will compare these presumably maximal states to the average and the variance. For the discussion we want to distinguish between the overall maximal fidelity defined in Conjecture 3.8.15 and the conjectured maximal fidelity for a certain number of channel.

Definition 3.8.16 (Regularized Conjected Maximal Fidelity for n Channel)**.**

$$F_{n,max} = \max\{p_0 + p_3, \frac{1}{2^{\frac{1}{n}}} \left((p_2+p_3)^n + (p_0+p_1)^n + (p_3-p_2)^n + (p_0-p_1)^n\right)^{\frac{1}{n}}\}$$

In Table 3.7, Table 3.8, Table 3.10 and Table 3.9 we have calculated the average fidelity $(F_n)^n$, its variance $(\mathrm{Var}_n)^n$, the large n limit variance Var_∞^n, the conjectured maximal fidelity $F_{n,\max}$ for the actual distribution for 1 to 10 channel, additionally the regularized maximal fidelity and the relative distance.

Definition 3.8.17 (Relative Distance)**.** *As a measure how much the conjectured maximum deviates from a typical state, we define the relative distance as*

$$\Delta_{rel} = \frac{(F_{n,max})^n - (F_n)^n}{(Var_n)^n}$$

In all tables we can see that the predicted maximal fidelity is generally much larger than the average and way out of reach of the variance even for a small number of channel. This means that the high fidelity states can hardly be found by coincidence. As we predicted the variance approaches zero much faster than the average, the distribution becomes strongly peaked already for few channel. The large n formula for the variance predicts the actual variance very well.
Whereas for relatively large p_0, in Table 3.8 and in Table 3.7, the maximal fidelity per channel does not depend on the number of channel and does not change with it, for lower p_0, in Table 3.10 and on Table 3.9, we see that the entangled states do better than the product states and the regularized fidelity increases.
While our work on the channel fidelity distribution has not brought us closer in maximizing quantum information theoretical quantities, we have now certainly more evidence that the states we found before are very good candidates providing a maximal regularized channel fidelity.

n	$(F_n)^n$	$(\mathrm{Var}_n)^n$	Var_∞^n	$F_{n,\max}^n$	$F_{n,\max}$	Δ_{rel}
1	0.8	0	0.51	0.8	0.8	
2	0.59	$0.31 \cdot 10^{-1}$	$1.8 \cdot 10^{-1}$	0.64	0.8	1.53
3	0.42	$0.22 \cdot 10^{-1}$	$0.66 \cdot 10^{-1}$	0.51	0.8	4.28
4	0.28	$0.12 \cdot 10^{-1}$	$0.24 \cdot 10^{-1}$	0.41	0.8	10.77
5	0.19	$0.51 \cdot 10^{-2}$	$0.86 \cdot 10^{-2}$	0.32	0.8	26.33
6	0.13	$0.21 \cdot 10^{-2}$	$0.31 \cdot 10^{-2}$	0.26	0.8	63.52
7	$0.89 \cdot 10^{-1}$	$0.79 \cdot 10^{-3}$	$0.11 \cdot 10^{-3}$	0.21	0.8	$15.15 \cdot 10$
8	$0.61 \cdot 10^{-1}$	$0.30 \cdot 10^{-3}$	$0.40 \cdot 10^{-3}$	0.17	0.8	$35.74 \cdot 10$
9	$0.42 \cdot 10^{-1}$	$0.11 \cdot 10^{-3}$	$0.15 \cdot 10^{-3}$	0.13	0.8	$83.45 \cdot 10$
10	$0.29 \cdot 10^{-1}$	$0.40 \cdot 10^{-4}$	$0.53 \cdot 10^{-4}$	0.11	0.8	$19.31 \cdot 10^2$
25	$0.13 \cdot 10^{-3}$	$0.10 \cdot 10^{-10}$	$0.12 \cdot 10^{-10}$	$0.38 \cdot 10^{-2}$	0.8	$34.87 \cdot 10^7$

Table 3.7.: Average, Variance, Maximal Fidelity, Δ_{rel} and Regularized Maximal Fidelity for n depolarizing channel with $p_0 = 0.7$, $p_1 = p_2 = p_3 = 0.1$

n	$(F_n)^n$	$(\mathrm{Var}_n)^n$	Var_∞^n	$F_{n,\max}^n$	$F_{n,\max}$	Δ_{rel}
1	0.6	0	0.37	0.6	0.6	
2	0.33	$0.21 \cdot 10^{-1}$	$0.99 \cdot 10^{-1}$	0.36	0.6	1.53
3	0.17	$0.11 \cdot 10^{-1}$	$0.26 \cdot 10^{-1}$	0.22	0.6	4.28
4	0.08	$0.43 \cdot 10^{-2}$	$0.69 \cdot 10^{-2}$	0.13	0.6	10.77
5	0.04	$0.14 \cdot 10^{-2}$	$0.18 \cdot 10^{-2}$	$0.78 \cdot 10^{-1}$	0.6	26.40
6	$0.19 \cdot 10^{-1}$	$0.43 \cdot 10^{-3}$	$0.48 \cdot 10^{-3}$	$0.47 \cdot 10^{-1}$	0.6	63.91
7	$0.94 \cdot 10^{-2}$	$0.12 \cdot 10^{-3}$	$0.13 \cdot 10^{-3}$	$0.28 \cdot 10^{-1}$	0.6	$15.32 \cdot 10$
8	$0.45 \cdot 10^{-2}$	$0.34 \cdot 10^{-4}$	$0.34 \cdot 10^{-4}$	$0.17 \cdot 10^{-1}$	0.6	$36.43 \cdot 10$
9	$0.22 \cdot 10^{-2}$	$0.92 \cdot 10^{-5}$	$0.90 \cdot 10^{-5}$	$0.10 \cdot 10^{-1}$	0.6	$85.95 \cdot 10$
10	$0.11 \cdot 10^{-2}$	$0.25 \cdot 10^{-5}$	$0.24 \cdot 10^{-5}$	$0.60 \cdot 10^{-2}$	0.6	$20.14 \cdot 10^2$
25	$0.30 \cdot 10^{-7}$	$0.54 \cdot 10^{-14}$	$0.52 \cdot 10^{-14}$	$0.28 \cdot 10^{-5}$	0.6	$52.07 \cdot 10^7$

Table 3.8.: Average, Variance, Maximal Fidelity, Δ_{rel} and Regularized Maximal Fidelity for n depolarizing channel with $p_0 = 0.4$, $p_1 = p_2 = p_3 = 0.2$

n	$(F_n)^n$	$(\mathrm{Var}_n)^n$	Var_∞^n	$F_{n,\max}^n$	$F_{n,\max}$	Δ_{rel}
1	0.47	$0.79 \cdot 10^{-1}$	0.39	0.6	0.6	1.69
2	0.23	$0.37 \cdot 10^{-1}$	$1.06 \cdot 10^{-1}$	0.36	0.6	3.40
3	0.12	$0.14 \cdot 10^{-1}$	$0.29 \cdot 10^{-1}$	0.22	0.6	6.74
4	$0.60 \cdot 10^{-1}$	$0.49 \cdot 10^{-2}$	$0.80 \cdot 10^{-2}$	$0.79 \cdot 10^{-1}$	0.61	14.01
5	$0.31 \cdot 10^{-1}$	$0.16 \cdot 10^{-2}$	$0.21 \cdot 10^{-2}$	$0.43 \cdot 10^{-1}$	0.62	35.09
6	$0.15 \cdot 1-^{-1}$	$0.47 \cdot 10^{-3}$	$0.60 \cdot 10^{-3}$	$0.25 \cdot 10^{-1}$	0.63	93.68
7	$0.77 \cdot 10^{-2}$	$0.14 \cdot 10^{-3}$	$0.16 \cdot 10^{-3}$	$0.15 \cdot 10^{-1}$	0.64	$24.68 \cdot 10$
8	$0.39 \cdot 10^{-2}$	$0.39 \cdot 10^{-4}$	$0.45 \cdot 10^{-4}$	$0.87 \cdot 1-^{-2}$	0.65	$64.37 \cdot 10$
9	$0.19 \cdot 10^{-2}$	$0.11 \cdot 10^{-4}$	$0.12 \cdot 10^{-4}$	$0.51 \cdot 10-2$	0.65	$16.66 \cdot 10^2$
10	$0.98 \cdot 10^{-3}$	$0.31 \cdot 10^{-5}$	$0.34 \cdot 10^{-5}$	$0.31 \cdot 10^{-2}$	0.65	$42.91 \cdot 10^2$
25	$0.30 \cdot 10^{-7}$	$0.12 \cdot 10^{-13}$	$0.12 \cdot 10^{-13}$	$0.14 \cdot 10^{-5}$	0.68	$54.93 \cdot 10^8$

Table 3.9.: Average, Variance, Maximal Fidelity, Δ_{rel} and Regularized Maximal Fidelity for n Pauli channel with $p_0 = 0.2$, $p_1 = 0.1$, $p_2 = 0.3$ and $p_3 = 0.4$

n	$(F_n)^n$	$(\mathrm{Var}_n)^n$	Var_∞^n	$F_{n,\max}^n$	$F_{n,\max}$	Δ_{rel}
1	0.4	0	$3.74 \cdot 10^{-1}$	0.4	0.4	0
2	0.21	$0.31 \cdot 10^{-1}$	$0.99 \cdot 10^{-1}$	0.28	0.53	2.29
3	0.11	$0.11 \cdot 10^{-1}$	$0.26 \cdot 10^{-1}$	0.14	0.51	2.14
4	$0.59 \cdot 10^{-1}$	$0.37 \cdot 10^{-2}$	$0.69 \cdot 10^{-2}$	$0.78 \cdot 10^{-1}$	0.53	5.29
5	$0.30 \cdot 10^{-1}$	$0.11 \cdot 10^{-2}$	$0.18 \cdot 10^{-2}$	$0.44 \cdot 10^{-1}$	0.54	12.05
6	$0.15 \cdot 10^{-1}$	$0.33 \cdot 10^{-3}$	$0.49 \cdot 10^{-3}$	$0.25 \cdot 10^{-1}$	0.54	30.56
7	$0.77 \cdot 10^{-2}$	$0.93 \cdot 10^{-4}$	$0.13 \cdot 10^{-3}$	$0.15 \cdot 10^{-1}$	0.55	75.81
8	$0.39 \cdot 10^{-2}$	$0.26 \cdot 10^{-4}$	$0.34 \cdot 10^{-4}$	$0.87 \cdot 10^{-2}$	0.55	$18.64 \cdot 10$
9	$0.19 \cdot 10^{-2}$	$0.71 \cdot 10^{-5}$	$0.89 \cdot 10^{-5}$	$0.52 \cdot 10^{-2}$	0.56	$45.07 \cdot 10$
10	$0.98 \cdot 10^{-3}$	$0.20 \cdot 10^{-5}$	$0.24 \cdot 10^{-5}$	$0.31 \cdot 10^{-2}$	0.56	$10.76 \cdot 10^2$
25	$0.30 \cdot 10^{-7}$	$0.50 \cdot 10^{-14}$	$0.52 \cdot 10^{-14}$	$0.14 \cdot 10^{-5}$	0.58	$27.70 \cdot 10^7$

Table 3.10.: Average, Variance, Maximal Fidelity, Δ_{rel} and Regularized Maximal Fidelity for n depolarizing channel with $p_0 = 0.1$, $p_1 = p_2 = p_3 = 0.3$

4. Conclusion

We introduced the quantum capacity of a quantum channel and explained that it is a difficult problem. It involves the maximization of a functional, the regularized coherent information, over large tensor product spaces. Motivated by this problem, we investigated the channel fidelity.

In the main part of this work, we investigated moments of the channel fidelity distribution in the infinite channel limit using Collins' and Śniady's formula, or, more precisely, our simplified version of it. The simplified formula allowed us to obtain a general formula for all moments for any kind of quantum channel.

For Pauli channels, the average and variance could be computed explicitly. The simplified formula has thus proven its usefulness in the study of tensor product spaces. In particular, by expressing the integral over moments of a unitary group as a sum over the corresponding symmetric group, we could identify a concrete class of permutations that dominate the variance in the infinite channel limit. If we directly calculated the integral over the unitary group, we believe it would have been more difficult to find this limit.

In the calculation of higher moments of the Pauli channel fidelity distribution, we found that, under certain restrictions, the moments could also be assigned to a similar class of permutations in the infinite channel limit. Nonetheless, here we can also observe limits to our approach: only if the dimension of a permutation is the dominating parameter for the limit did expressing the integral as a sum over permutations give us an advantage.

Aside from the mathematical results, we could also concretely describe the Pauli channel fidelity distribution for large channel numbers. We found that in this case the distribution is very strongly localized around the mean; in the limit of $n \to \infty$ channels, it becomes a delta distribution. This means that a randomly chosen state will have nearly average channel fidelity. Essentially, for the vast majority of states, a high tensor product of Pauli channels acts as a single depolarizing channel on the corresponding Hilbert space, and as such the typical corresponding channel fidelity does not depend on the input state.

In this context it is remarkable that we could use an the iterative algorithm, presented in the last section, to find locally maximizing states with a channel fidelity that exceeds the mean fidelity by orders of magnitude. We suspect that for a high dimensional tensor product space there are many locally maximizing states that have just remotely higher fidelity than the average states, and thus these states would be found by a standard numerical maximization.

If we suppose that our results about the channel fidelity can be extrapolated to the coherent information, it would mean that most states would have approximately the

same (regularized) coherent information, and that this value is not even close to the maximum. It would be furthermore likely that a numeric maximization would only find states with slightly higher coherent information. Finding a real absolute maximizing state and thus the capacity of a quantum channel remains challenging.

A. Estimates for Variance Calculation

To show that the variance for Pauli channel has the claimed simple form, we need to show that the next order in 2^n is still upper bounded by the simple expression.
The next order in 2^n is

$$O(2^n) = 2^n \left(4\, p_0^n + 2 \left(\sum_i p_i^2 \right)^n + 2 \left(p_1^2 - p_0^2 + (p_2 - p_3)^2 - 2p_1(p_2 + p_3) + 2p_0 \right)^n \right) + O(1).$$

We need to show for every term that in the limit $n \to \infty$ they are smaller than $2^n * \left(\sum_i p_i^2\right)^n$. The prefactors do not scale with n, hence we can drop them.

- For the first summand, we want to show,

  $$\sum_i p_i^2 \geq \frac{p_0}{2}.$$

  Considering that $\sum_{i \neq 0} p_i$ is minimal, if all p_i are the same. We can write,

  $$\sum_i p_i^2 \geq p_0^2 + 3 \left(\frac{1 - p_0}{3} \right)^2,$$

  which leads to the following equation,

  $$p_0^2 + 3 \left(\frac{1 - p_0}{3} \right)^2 - \frac{p_0}{2} = 0,$$

  which has no real solutions. That means since for $p_0 = 1$ the condition is clearly fulfilled it is fulfilled everywhere.

- The second term is trivial.

- For the third we have to show that,

  $$2 \sum_i p_i^2 \geq p_1^2 - p_0^2 + p_2^2 + p_3^2 - 2p_2p_3 - 2p_1p_2 - 2p_1p_3 + 2p_0.$$

It is useful to look at

$$\left(\sum_i p_i\right)^2 = \sum_i p_i^2 + 2p_0(p_1+p_2+p_3) + 2p_1p_2 + 2p_1p_3 + 2p_2p_3$$
$$\Leftrightarrow 1 - \sum_i p_i^2 - 2p_0 + 2p_0^2 = 2p_1p_2 + 2p_1p_3 + 2p_2p_3.$$

Then we get:

$$\begin{aligned} 2\sum_i p_i^2 &\geq p_1^2 - p_0^2 + p_2^2 + p_3^2 - 2p_2p_3 - 2p_1p_2 - 2p_1p_3 + 2p_0 \\ &= \sum_i p_i^2 - 2p_0^2 - (1 - \sum_i p_i^2 - 2p_0 + 2p_0^2) + 2p_0 \\ &= 2\sum_i p_i^2 - 4p_0^2 + 4p_0 - 1 = 2\sum_i p_i^2 - (2p_0-1)^2. \end{aligned}$$

Equality is achieved for $p_0 = \frac{1}{2}$. Equality is not a problem, it just combines the prefactors of the two contributions at this point, however the prefactors are unimportant.

B. Code

We used Mathematica 10 to caculate tables Table 3.3,Table 3.4, Table 3.5,Table 3.6, Table C.1, Table C.2, Table C.3, Table C.4, Table C.5.
We will now show briefly which commands we used and how they work. Mathematica 10 has some easy tools to handle permutation groups. It represents permutations as a list of non-trivial cycles. We can see this, if we use "SymmetricGroup[n]" to generate the symmetric group for n elements and "GroupElements" to show its elements.

Listing B.1: GroupElements

```
In[1]:= GroupElements[SymmetricGroup[2]]
Out[1]= {Cycles[{}], Cycles[{{1, 2}}]}
```

Furthermore it is possible to use the command "PermutationGroup" to generate a permutation group for given permutations. We used this to generate the symmetry groups as discussed in Definition 3.4.6 and Definition 3.4.12.

Listing B.2: Symmetry Group

```
In[2]:= s1 = Cycles[{{1, 2}}];
In[3]:= m1 = Cycles[{{1, 3}, {2, 4}}];
In[4]:= H4 = PermutationGroup[{s1,m1}];
In[5]:= GroupElements[H4]
Out[5]= {Cycles[{}], Cycles[{{3, 4}}], Cycles[{{1, 2}}],
 Cycles[{{1, 2}, {3, 4}}], Cycles[{{1, 3}, {2, 4}}],
 Cycles[{{1, 3, 2, 4}}], Cycles[{{1, 4, 2, 3}}],
 Cycles[{{1, 4}, {2, 3}}]}
```

s1 represents the symmetry that $A_{i_1} = A_{i_1}^\dagger$ for Pauli channel. m1 is due to the fact that we can exchange the operators blockwise. H4 is then the group of all applicable symmetries.
Now we can use orbits of H4 and the command "GroupOrbits" to find all elements that for sure are equivalent with regards to the central part. The command "Length" will give the total number of elements within an orbit.

Listing B.3: Orbits

```
In[6]:= GroupOrbits[H4, {Cycles[{{1, 3, 2, 4}}]}]
Out[6]= {{Cycles[{{1, 3, 2, 4}}], Cycles[{{1, 4, 2, 3}}]}}
In[7]:= Length[GroupOrbits[H4, {Cycles[{{1, 3, 2, 4}}]}][[1]]]
Out[7]= 2
```

For the larger groups it can be a bit tricky to find all different orbits. For example in Table C.3 we have six different orbits for 2-3-permutations. This problem is solved because Mathematica sorts the permutations, such in identical orbits the first permutation is always the same, such that it is sufficient to look at the first element of an orbit to tell them apart.

Listing B.4: Distinguishing Orbits

```
In[8]:= GroupOrbits[H4, {Cycles[{{1, 2, 3, 4}}]}][[1, 1]]
Out[8]= Cycles[{{1, 2, 3, 4}}]
In[9]:= GroupOrbits[H4, {Cycles[{{1, 3, 2, 4}}]}][[1, 1]]
Out[9]= Cycles[{{1, 3, 2, 4}}]
In[10]:= GroupOrbits[H4, {Cycles[{{1, 4, 2, 3}}]}][[1, 1]]
Out[10]= Cycles[{{1, 3, 2, 4}}]
```

The before code can easily be extended to larger groups and we did so for S_6 and S_8, the results can be found in Table C.1, Table C.3 etc..
Now that we know the size of the different orbits, we actually want to calculate their specific central part. For this purpose it is helpful to use an outdated Mathematica package called "Combinatorica". Here one can use the command "ToCycles" to transform a permutation represented as a list into cycles with the small but decisive difference that trivial cycles will still appear. Calling "Combinatorica" Mathematica will warn about compatibility issues, one can ignore this.

Listing B.5: ToCycles

```
<< Combinatorica`
In[11]:= ToCycles[{1, 2, 4, 3}]
Out[11]= {{1}, {2}, {4, 3}}
```

We want Mathematica to calculate for a given list of operators the traces of their products according to a permutation, for example:

$$\sum_{i_1}\sum_{i_2}(A_{i_1},A_{i_1},A_{i_2},A_{i_2})_{(123)}=\sum_{i_1}\sum_{i_2}\mathrm{Tr}\,(A_{i_1}\,A_{i_1}\,A_{i_2})\,\mathrm{Tr}\,(A_{i_2})\,.$$

First we will create a table B, where a list of operators A is organized according to a permutation P. L will be a list representing the summation indices, this will become clear in the end.

Listing B.6: Cycle to List

```
In[12]:= MakeB[P_, A_, L_] :=
 Module[{B = Table[IdentityMatrix[Dimensions[A[[1, 1]]][[1]]],
    {i, 1, Length[P]}]},
   For[
   i1 = 1, i1 < Length[P] + 1, i1++,
```

```
29      For[
30      i2 = 1, i2 < Length[P[[i1]]] + 1, i2++,
31       B[[i1]] = B[[i1]].A[[L[[P[[i1, i2]]]]]]
32       ]
33       ]; B]
```

The module begins by creating a table B of identity matrices, in the respective dimension of the operators in the list A. The table has as many entries as there are separate cycles in the permutation P. Then the first for loop counts through the different cycles and the second counts trough the specific elements in a specific cycle. This way the operators from A are put in the correct order sorted by cycle into the table B. The module outputs the table B.
The next step is relatively easy: We take the trace of every element of B and multiply the traces.

Listing B.7: List to Traceproduct

```
34 In[13]:=MakeB0[B_] :=
35  Module[{B0 = 1},
36   For[
37   i1 = 1, i1 < Length[B] + 1, i1++,
38    B0 = (d/Dimensions[B[[i1]][[1]]]) B0*Tr[B[[i1]]]
39    ]; B0]
```

Again we start by making a trivial list B0, this time it only consists of a single element 1. Then the for loop goes through the list and multiplies the trace of an entry of B into B0. We found it useful to replace dimension factors with the letter d. This way we get outputs that are proportional to powers of d, rather then powers of numbers, where one has to then figure out which power of the dimension it is.
Finally we can evaluate the concrete central part of a permutation. Unfortunately we have not found an easy way to generalize the code.

Listing B.8: Evaluate Central Part

```
40 In[14]:= Be[c_,A_] := FullSimplify[
41   Sum[
42    MakeB0[
43     MakeB[c, A, {k1, k1, k2, k2, k3, k3}]], {k1, 1, Length[A]},
44     {k2, 1, Length[A]}, {k3, 1,Length[A]}]]
```

We see here that Be is made for the third moment as there are three summation indices L = {k1,k1,k2,k2,k3,k3}.
For illustration purposes we will now change Be to the second moment and give a list of operators and then evaluate a concrete permutation.

Listing B.9: Evaluate Concrete Permutation

```
45 In[15]:= Be[c_, A_] := FullSimplify[
46   Sum[
```

```
47      MakeB0[
48       MakeB[c, A, {k1, k1, k2, k2}]], {k1, 1, Length[A]},
49       {k2, 1, Length[A]}, {k3, 1, Length[A]}
50       ]
51       ]
52 In[16]:= A = {Sqrt[p0] PauliMatrix[0], Sqrt[p1] PauliMatrix[1],
53     Sqrt[p2] PauliMatrix[2], Sqrt[p3] PauliMatrix[3]};
54
55 In[17]:= Be[{{1, 2, 3}, {4}}, A]
56 Out[17]= d^2 p0 (p0 + p1 + p2 + p3)
```

Let us now evaluate the third central moment. First we will calculate the sum over all central parts, for the S_4 and the S_6.

Listing B.10: Sum over Central Parts

```
57 In[18]:= S4Ele = Sum[
58 Be[ToCycles[SymmetricGroup[4][[i]]], A], {i, 1, 4!}];
59 In[19]:= S6Ele = Sum[
60 Be[ToCycles[SymmetricGroup[6][[i]]], A], {i, 1, 6!}];
```

In order to automatically evaluate all central parts, we needed to use the ToCycles function on complete groups, not only does the size of permutation groups grow over exponentially, also the ToCycles function seems to be very slow in providing the for our code necessary representation of permutations. Because of these problems we could not evaluate groups bigger than the S_8 and thus only up to the fourth central moment.
Last we calculate the third central moment keeping in mind, but not writing down the common denominator.

Listing B.11: Third Central Moment

```
61 In[20]:= Mom3[d_] := Expand[(d (d + 1))^2 S6Ele -
62   3 (d + 5) (d + 4) d (d + 1) (d^(2) p0 + d) S4Ele +
63   2 (d + 5) (d + 4) (d + 3) (d + 2) (d^(2) p0 + d)^3, d]
```

It is useful to use the function Expand[,d] to get an expression in orders of d. We can now check for specific orders of d.

Listing B.12: Checking Orders of d

```
64 In[21]:=Mom3[d] /. d^k_ /; k != 10 :> 0
65 Out[21] = 0
66 In[22]:=Mom3[d] /. d^k_ /; k != 9 :> 0
67 Out[22] = 3 d^9 p0^2 - 3 d^9 p0^2 (p0 + p1 + p2 + p3)
68 In[23]:=Mom3[d] /. d^k_ /; k != 8 :> 0
69 Out[23] =
70 2 d^7
71 +d^7 (p0 + p1 + p2 + p3)^3
```

```
−3 d^7 (p0 + p1 + p2 + p3)^2
+84 d^7 p0
−84 d^7 p0 (p0 + p1 + p2 + p3)
+219 d^7 p0^2
−219 d^7 p0^2 (p0 + p1 + p2 + p3)
−6 d^7 (p0^2 + p1^2 + p2^2 + p3^2)
+6 d^7 (p0 + p1 + p2 + p3) (p0^2 + p1^2 + p2^2 + p3^2)
+8 d^7 (p0^3 + p1^3 + p2^3 + p3^3)
−8 d^7 p0^3
```

We see that most terms vanish because the probabilities sum to 1 and we are left with essentially the third central moment as presented in Theorem 3.6.3.

C. Tables

In this appendix we provide the complete decomposition of the trace products over S_6 and S_8 under the isotropy groups of the according central parts.

$H_6\,\sigma$	Representative σ	$\lvert C_{S_6}(\sigma)\rvert$	$\lvert H_6\sigma\rvert$	P_σ
O_1	e	1	1	$p_0^3\,2^6$
O_2	(12)	15	3	$p_0^2\,2^5$
O_3	(13)		12	$p_0^3\,2^5$
O_4	(12)(34)	45	3	$p_0\,2^4$
O_5	(13)(24)		6	$p_0\,\sum_i p_i^2\,2^4$
O_6	(12)(35)		12	$p_0^2\,2^4$
O_7	(13)(25)		24	$p_0^3\,2^4$
O_8	(123)	40	24	$p_0^2\,2^4$
O_9	(135)		16	$p_0^3\,2^4$
O_{10}	(1234)	90	12	$p_0\,2^3$
O_{11}	(1324)		6	$p_0\,(p_0^2+p_1^2+(p_2-p_3)^2-2\,p_1\,(p_2+p_3)+2\,p_0\,(p_1+p_2+p_3))2^3 = p_0\,\alpha_1\,2^3$
O_{12}	(1235)		48	$p_0^2\,2^3$
O_{13}	(1325)		24	$p_0^2\,2^3$
O_{14}	(12345)	144	48	$p_0\,2^2$
O_{15}	(13245)		48	$p_0\,(p_0^2+p_1^2+(p_2-p_3)^2-2\,p_1\,(p_2+p_3)+2\,p_0\,(p_1+p_2+p_3))2^2 = p_0\,\alpha_1\,2^2$

Table C.1.: The orbits of S_6 under H_6 for Pauli Channel — first part.

$H_6\,\sigma$	Representative σ	$\lvert C_{S_6}(\sigma)\rvert$	$\lvert H_6\sigma\rvert$	P_σ
O_{16}	(12354)		48	$p_0\,2^2$
O_{17}	(123)(45)	120	48	$p_0^2\,2^3$
O_{18}	(12)(345)		24	$p_0\,2^3$
O_{19}	(13)(245)		48	$p_0\,\sum_i p_i^2\,2^3$
O_{20}	(12)(34)(56)	15	1	2^3
O_{21}	(13)(24)(56)		6	$\sum_i p_i^2\,2^3$
O_{22}	(13)(25)(46)		8	$\sum_i p_i^3\,2^3$
O_{23}	(1234)(56)	90	12	2^2
O_{24}	(1324)(56)		6	$(p_0^2+p_1^2+(p_2-p_3)^2-2\,p_1\,(p_2+p_3)+2\,p_0\,(p_1+p_2+p_3))2^2=\alpha_1\,2^2$
O_{25}	(1235)(46)		48	$\sum_i p_i^2\,2^2$
O_{26}	(1325)(46)		24	$(p_0^3+p_1^3-p_1^2(p_2+p_3)+(p_2-p_3)^2(p_2+p_3)+p_0^2(p_1+p_2+p_3)-p_1(p_2^2+p_3^2)+p_0(p_1^2+p_2^2+p_3^2))\,2^2=\beta_1\,2^2$
O_{27}	(123)(456)	40	24	$p_0\,2^2$
O_{28}	(135)(246)		8	$(p_0^3-6p_1p_2p_3+3p_0(p_1^2+p_2^2+p_3^2))\,2^2=\beta_2\,2^2$
O_{29}	(153)(246)		8	$(p_0^3+6p_1\,p_2\,p_3+3p_0(p_1^2+p_2^2+p_3^2))\,2^2$
O_{30}	(123456)	120	16	2
O_{31}	(123546)		48	$(p_0^2+p_1^2+(p_2-p_3)^2-2p_1(p_2+p_3)+2p_0(p_1+p_2+p_3))\,2=\alpha_1\,2$
O_{32}	(132546)		24	$(p_0^3+p_1^3-p_1^2(p_2+p_3)+(p_2-p_3)^2(p_2+p_3)+3p_0^2(p_1+p_2+p_3)-p_1(p_2^2-6p_2p_3+p_3^2)+p_0(3p_1^2+3p_2^2-2p_2p_3+3p_3^2-2p_1(p_2+p_3)))\,2=\beta_3\,2$
O_{33}	(123564)		24	2
O_{34}	(135246)		8	$(p_0^3+p_1^3+3p_1(p_2-p_3)^2+3p_1^2(p_2+p_3)+(p_2+p_3)^3+3p_0^2(p_1+p_2+p_3)+3p_0(p_1^2+(p_2-p_3)^2-2p_1(p_2+p_3)))\,2=\beta_4\,2$

Table C.2.: The orbits of S_6 under H_6 for Pauli Channel — second part.

$H_8\,\sigma$	Representative σ	$\lvert C_{S_8}(\sigma)\rvert$	$\lvert H_8\sigma\rvert$	P_σ
O_1	e	1	1	$p_0^4\,2^8$
O_2	(12)	28	4	$p_0^3\,2^7$
O_3	(13)		24	$p_0^4\,2^7$
O_4	(12)(34)	210	6	$p_0^2\,2^6$
O_5	(13)(24)		12	$p_0^2\,\sum_i p_i^2\,2^6$
O_6	(12)(35)		48	$p_0^3\,2^6$
O_7	(13)(25)		96	$p_0^4\,2^6$
O_7	(13)(57)		48	$p_0^4\,2^6$
O_8	(123)	112	48	$p_0^3\,2^6$
O_9	(135)		64	$p_0^4\,2^6$
O_{10}	(1234)	420	24	$p_0^2\,2^5$
O_{11}	(1324)		12	$p_0^2\,(p_0^2+p_1^2+(p_2-p_3)^2-2\,p_1\,(p_2+p_3)+2\,p_0\,(p_1+p_2+p_3))2^5 = p_0^2\,\alpha_1\,2^5$
O_{12}	(1235)		192	$p_0^3\,2^5$
O_{13}	(1325)		96	$p_0^3\,2^5$
O_{13}	(1357)		96	$p_0^4\,2^5$
O_{14}	(12345)	1344	192	$p_0^2\,2^4$
O_{15}	(13245)		192	$p_0^2\,(p_0^2+p_1^2+(p_2-p_3)^2-2\,p_1\,(p_2+p_3)+2\,p_0\,(p_1+p_2+p_3))2^4 = p_0^2\,\alpha_1\,2^4$
O_{16}	(12354)		192	$p_0^2\,2^4$
O_{16}	(12357)		384	$p_0^3\,2^4$
O_{16}	(13527)		384	$p_0^3\,2^4$
O_{17}	(123)(45)	1120	192	$p_0^3\,2^5$
O_{18}	(12)(345)		96	$p_0^2\,2^5$
O_{19}	(13)(245)		192	$p_0^2\,\sum_i p_i^2\,2^5$
O_{19}	(12)(357)		64	$p_0^3\,\sum_i p_i^2\,2^5$

Table C.3.: The orbits of S_8 under H_8 for Pauli Channel — first part.

$H_8\,\sigma$	Representative σ	$\lvert C_{S_8}(\sigma)\rvert$	$\lvert H_8\sigma\rvert$	P_σ
O_{19}	(13)(257)		384	$p_0^4\,2^5$
O_{19}	(13)(567)		192	$p_0^3\,2^5$
O_{20}	(12)(34)(56)	420	4	$p_0\,2^5$
O_{22}	(12)(34)(57)		24	$p_0^2\,\sum_i p_i^2\,2^5$
O_{21}	(12)(35)(46)		24	$p_0\,\sum_i p_i^2\,2^5$
O_{21}	(12)(35)(47)		96	$p_0^3\,2^5$
O_{22}	(13)(25)(46)		32	$p_0\,\sum_i p_i^3\,2^5$
O_{22}	(13)(24)(57)		48	$p_0^2\,\sum_i p_i^2\,2^5$
O_{22}	(13)(25)(47)		192	$p_0^4\,2^5$
O_{23}	(12)(3456)	2520	48	$p_0\,2^4$
O_{23}	(12)(3457)		192	$p_0^2\,2^4$
O_{24}	(12)(3546)		24	$p_0(p_0^2+p_1^2+(p_2-p_3)^2-2p_1(p_2+p_3)+2p_0(p_1+p_2+p_3))\,2^4 = p_0\,\alpha_1\,2^4$
O_{23}	(12)(3547)		96	$p_0^2\,2^4$
O_{25}	(13)(2456)		192	$p_0\,\sum_i p_i^2\,2^4$
O_{26}	(13)(2546)		96	$p_0\,(p_0^3+p_1^3-p_1^2(p_2+p_3)+(p_2-p_3)^2(p_2+p_3)+p_0^2(p_1+p_2+p_3)-p_1(p_2^2+p_3^2)+p_0(p_1^2+p_2^2+p_3^2))\,2^4 = p_0\,\beta_4\,2^4$
O_{25}	(13)(2457)		384	$p_0^2\,\sum_i p_i^2\,2^4$
O_{25}	(13)(2547)		192	$p_0^2\sum_i p_i^2\,2^4$
O_{25}	(13)(2567)		384	$p_0^3\,2^4$
O_{25}	(13)(2576)		384	$p_0^3\,2^4$
O_{25}	(13)(2578)		384	$p_0^3\,2^4$
O_{25}	(13)(5678)		96	$p_0^2\,2^4$
O_{25}	(13)(5768)		48	$p_0^2(p_0^2+p_1^2+(p_2-p_3)^2-2p_1(p_2+p_3)+2p_0(p_1+p_2+p_3))\,2^4 = p_0^2\,\alpha_1\,2^4$
O_{27}	(12)(34)(56)(78)	105	1	2^4
O_{27}	(12)(35)(46)(78)		12	$\sum_i p_i^2\,2^4$
O_{27}	(12)(35)(47)(68)		32	$\sum_i p_i^3\,2^4$
O_{27}	(13)(24)(57)(68)		12	$\left(\sum_i p_i^2\right)^2\,2^4$
O_{27}	(13)(25)(47)(68)		48	$\sum_i p_i^4\,2^4$

Table C.4.: The orbits of S_8 under H_8 for Pauli Channel — second part.

$H_8\,\sigma$	Representative σ	$\lvert C_{S_8}(\sigma)\rvert$	$\lvert H_8\sigma\rvert$	P_σ
O_{27}	(12)(34)(567)	1680	48	$p_0\,2^4$
O_{27}	(12)(35)(467)		192	$p_0\,\sum_i p_i^2\,2^4$
O_{27}	(12)(35)(478)		192	$p_0^2\,2^4$
O_{27}	(13)(25)(467)		384	$p_0\,\sum_i p_i^3\,2^4$
O_{27}	(13)(25)(478)		384	$p_0^3\,2^4$
O_{27}	(13)(24)(567)		96	$p_0\,\sum_i p_i^2\,2^4$
O_{27}	(13)(245)(67)		384	$p_0^2\,\sum_i p_i^2\,2^4$
O_{27}	(123)(456)	1120	96	$p_0^2\,2^4$
O_{27}	(123)(457)		384	$p_0^3\,2^4$
O_{29}	(123)(567)		192	$p_0^2\,2^4$
O_{28}	(135)(246)		32	$p_0(p_0^3 - 6p_1p_2p_3 + 3p_0(p_1^2 + p_2^2 + p_3^2))\,2^4 = p_0\,\beta_2\,2^4$
O_{29}	(135)(247)		192	$p_0^2\,\sum_i p_i^2\,2^4$
O_{29}	(135)(264)		32	$p_0(p_0^3 + 6p_1p_2p_3 + 3p_0(p_1^2 + p_2^2 + p_3^2))\,2^4 = p_0\,\beta_3\,2^4$
O_{29}	(135)(267)		192	$p_0^2\,\sum_i p_i^2\,2^4$
O_{27}	(12)(345)(678)	1120	1120	$\propto 2^3$
O_{30}	(123456)	3360	3360	$\propto 2^3$
O_{27}	(1234)(56)(78)	1260	24	$\propto 2^3$
O_{27}	(1234)(567)	3360	24	$\propto 2^3$
O_{27}	(1234)(5678)	1260	24	$\propto 2^2$
O_{27}	(12345)(67)	4032	24	$\propto 2^3$
O_{27}	(12345)(678)	2688	24	$\propto 2^2$
O_{27}	(123456)(78)	3360	24	$\propto 2^2$
O_{27}	(1234567)	5760	24	$\propto 2^2$
O_{27}	(12345678)	5040	24	$\propto 2$

Table C.5.: The orbits of S_8 under H_8 for Pauli Channel — third part.

Bibliography

[1] H. Barnum, M. A. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Phys. Rev.*, A 57, 4153, 1998. 24

[2] I. Bengtsson and K. Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. 2006. 19

[3] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States*. 2006. 8, 20

[4] Easwar Magesan, Robin Blume-Kohout and Joseph Emerson. Gate fidelity fluctuations and quantum process invariants. *Phys. Rev. A*, 84, 012309, 2011. 54, 57, 64

[5] D. J. C. Bures. *Trans. Amer. Math.Sot.*, 135:199, 1969. 8

[6] C. S. Withers andS. Nadarajah. Moments frum cumulants and vice versa. *International Journal of Mathematical Education in Science and Technology*, 40:6:842–845, 2009. 96

[7] M. Choi. *Completely positive linear maps on complex matrices*. 1975. 15

[8] B. Collins and P. Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Commun. Math. Phys.*, 264:773–795, 2006. 27, 44, 64

[9] David P DiVincenzo, Peter W Shor, and John A Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57(2):830, 1998. 22

[10] M.-B. Ruskai, S. Szarek, E. Werner. An analysis of completely-positive trace-preserving maps on 2x2 matrices. *Lin. Alg. Appl.*, 347:159 – 187, 2002. 19

[11] M. Ernst. *Optimization of Quantum Information Theoretical quantities over Tensor product Spaces*. 2010. 53, 54, 57, 89, 91, 99, 104, 105

[12] W. Fulton and J. Harris. *Representation Theory: A First Course*. 1991. 34, 39, 42, 48

[13] G. Smith and J. A. Smolin. Additive extensions of a quantum channel. *IEEE Information Theory*, 54, 368 - 372, 2008. 22, 60

[14] T. S. Cubitt, M.-B. Ruskai, G. Smith. The structure of degradable quantum channels. *J. Math. Phys.*, 49, 102104, 2008. 18, 20

[15] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th ACM Symposium on Theory of Computing*, pages 212–219, 1996. 10

[16] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, H. Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52:3457, 1995. 54

[17] I. Devetak. The private classical and quantum capacity of quantum channels. *IEEE Trans. Inf. Theory*, 51, 2005. 21

[18] N. Johnston and D. W. Kribs. Quantum Gate Fidelity in Terms of Choi Matrices. *Journal of Physics A: Mathematical and Theoretical*, 44:495303, 2011. 54

[19] R. Klesse. A random-coding based proof for the quantum coding theorem. *Open Syst. & Inf. Dyn.*, 15,21, 2008. 21, 57

[20] Christian Krattenthaler. Another involution principle-free bijective proof of Stanley's hook-content formula. *J. Combin. Theory Ser. A*, 88,p 66-92, 1999. 40

[21] K. Kraus. *States, Effects, and Operations: Fundamental Notions of Quantum Theory.* 1987. 15

[22] M. M. Wilde. From classical to quantum shannon theory. *arxiv.org*, 1106.1445v5, 2013. 8

[23] Christian B Mendl and Michael M Wolf. Unital quantum channels–convex structure and revivals of birkhoff's theorem. *Communications in Mathematical Physics*, 289(3):1057–1086, 2009. 19

[24] Ciara Morgan and Andreas Winter. "pretty strong" converse for the quantum capacity of degradable channels. *arxiv.org*, 1301.4927v3, 2013. 18

[25] A. Gilchrist, N. K. Langford, M. A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71, 2005. 5, 7, 8

[26] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* 2000. 1, 8, 16, 17, 20

[27] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* 1949. 24

[28] Peter W Shor and John A Smolin. Quantum error-correcting codes need not completely reveal the error syndrome. *arXiv preprint quant-ph/9604006*, 1996. 22

[29] P. J. Smith. *A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa.* 1995. 93, 96, 97

[30] W. Soergel. Algebra.
http://home.mathematik.uni-freiburg.de/soergel/Skripten/ALGEBRA.pdf, 2011. 28

[31] R. Stanley. *Enumerative Combinatorics*. 1999. 40

[32] S. Sternberg. *Group Theory and Physics*. 1994. 34

[33] J. Tyson. Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates. *Journal of mathematical physics*, 51,092204, 2010. 104

[34] J. Tyson. Personal communication. 2011. 99, 104

[35] A. Uhlmann. The 'transistion probability' in the state space of a *-algebra. *Reports on Mathematical Physics*, 9:273–279, 1976. 8

[36] Y. Ouyang. Improved upper bounds on the quantum capacity of the depolarizing channel with higher dimension amplitude damping channels. *arxiv.org*, 1106.2337v6, 2014. 18, 22, 60

Acknowledgments

First and foremost I want to thank my supervisor Rochus Klesse for the opportunity to do research in this fascinating field. Not only was his door always open when needed but he also allowed and supported my frequent participation in conferences around the world. I would like to thank Simon Trebst for the time he invested in evaluating my work.
This work would not have been possible without the frequent support of my colleagues. I am deeply indebted with all of them:

- Maximilian L. Schaefer for suggesting Collins' and Śniady's paper,
- Daniel J. Wieczorek for teaching representation theory to me and sharing the pain of the world,
- Alexander Alldrige for last minute advice on how to structure my ideas,
- Ricardo Kennedy and Jochen Peschutter for listening to my stories, piloting lessons and a fantastic work environment,
- Roberto Bondesan, Kasper Duivenvoorden, Dominik Ostermayr, Peter Roenne, Abhishek Roy, Thomas Quella, Jan Schmidt, Sebastian Schmittner, and Artur Swiech for making the lunch breaks an interesting part of the day — and cake now and then,
- Martin R. Zirnbauer for his mentorship and for bringing me together with all these fascinating people,
- Wolfgang Palzer for being my private mathematician and most reliable comrade.

Even more I want to thank my family, in particular my uncle Werner Kazmirek for sharing his expertise in physics with me and introducing me to the physics way of thinking at a young age, and my parents Fritz Michael and Petra for encouraging me to follow my passion, listening to my crazy ideas and guiding me so some of them become reality.
Last, but not least, I want to thank my wife Stacey Jeffery for her encouragement, her support, her understanding, her proof reading, her professional advice, and many other things, but most of all for being the love of my life.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit — einschließlich Tabellen, Karten und Abbildungen —, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD Dr. Rochus Klesse betreut worden.

Köln, den 14. Oktober 2014 Moritz Ernst