IRis

Northeastern University

Physics Dissertations

Department of Physics

January 01, 2011

Entropy and energy in quantum networks

David K. Moser
Northeastern University

Recommended Citation

Moser, David K., "Entropy and energy in quantum networks" (2011). *Physics Dissertations.* Paper 29. http://hdl.handle.net/2047/d20002313

This work is available open access, hosted by Northeastern University.

Entropy and Energy in Quantum Networks

Dissertation presented
by

David K. Moser

to
The Departments of
Physics and Mathematics

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy
in Physics and Mathematics

with Specialization in
Quantum Information Theory

Northeastern University
Boston, Massachusetts

December, 2011

Entropy and Energy in Quantum Networks

by

David K. Moser

Abstract of Dissertation

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Physics and Mathematics
in the Graduate School of Northeastern University
December, 2011

Abstract

Quantum channels are functions that map density matrices to density matrices. They can be used to describe a vast array of system environment interactions for quantum mechanical systems. The focus in quantum information theory lies on the study of the information content – the quantum entropy – of input and output system. The property of how much information a channel can convey is called capacity. When two quantum channels are combined they act on a bigger, possibly entangled state in parallel. The effects of entanglement on the capacity of a pair of channels have been widely studied. Still, recent work showed unexpected behavior. Hastings [33] found that the capacities of two channels to transmit classical information does not simply add, but can exceed the sum of the individual quantum capacities, a property called superadditivity. We study Hastings' work and find estimates for how large the dimensions of such channels would have to be. The question of capacity relates to the behavior of quantum channels in extremal cases, when the output entropy is minimized. It is natural to wonder about the "standard" behavior of quantum channels and consider their average output entropy. We have studied this behavior in particular in cases where many relatively simple quantum channels act in parallel. When combining quantum channels in parallel, the parts of the system evolve independently of each other – up to entanglement – and only interact with the environment. Expanding on this idea one can think of a system where the parts not only interact with the environment but also directly with each other, a quantum channel network. Interesting cases of such systems arise naturally in the study of large biological molecules. In this case, information is often transmitted via energy pulses. Thus, the study of entropy and the study of energy overlap. Of particular interest to us are antennas in photo-active systems, they absorb light energy and transports it to a chemical reaction site. Recently, it has been discovered in the FMO molecule [24] that this process evolves coherently, i.e. shows quantum oscillatory behavior. We study such quantum channel networks and how they can be approximated by kinetic networks. When considering the transport efficiency, this helps separating possible coherent effects from incoherent hopping.

Acknowledgments

First and foremost I thank my advisor Chris King for working with me, for his patience in teaching me, his ideas, and many hours of useful discussions that led to this dissertation. I thank Motohisa Fukuda for working with us on the content of Chapter 2. I thank Arthur Jaffe, who introduced me to the field of mathematical physics and without whom I would not have made my way to Northeastern University. Further, I thank professors Alain Karma and Tim Sage for teaching me statistical physics and biophysics and Dashun Wang and Francisco Reynoso for studying this and other subjects with me and discussing problems in many hours of late-night work. Finally, I thank my parents and my sisters for supporting me and my research endeavor through all these years, and my dear friends Adrian, Andrea, James, Mario, Robbie, Thomas and Tong for keeping up my spirits and my physical health.

Publications

1. Chapter 2 is adapted from the paper:

   M. Fukuda, C. King, and D. K. Moser. Comments on Hastings' additivity counterexamples. *Communications in Mathematical Physics*, 296:111–143, 2010.

2. Chapter 3 is adapted from the paper:

   C. King and D. K. Moser. Average output entropy for quantum channels. *Journal of Mathematical Physics*, 52:112202, 2011.

3. Chapter 4 is to be published:

   D. K. Moser. Analysis of energy transfer in quantum networks using kinetic network approximations.

Contents

Abstract **2**

Acknowledgments **4**

Publications **5**

Table of Contents **6**

1 Introduction **8**

1.1 Entropy 9

1.2 Channel capacity and additivity 11

1.3 Counterexamples to additivity 12

1.4 Average output entropy of quantum channels 12

1.5 Photosynthesis 14

1.6 Energy transfer in the Fenna-Matthews-Olson (FMO) complex 15

2 Comments on Hastings' Additivity Counterexamples **16**

2.1 The additivity conjectures 16

2.2 Notation and statement of results 19

2.3 Background on random states and channels 23

2.4 Proof of Theorem 2 26

2.5 Proofs of Lemmas 34

2.6 Discussion 48

2.7 Acknowledgments 49

A Appendix for Chapter 2 **50**

A.1 Derivation of bound for $Z(n, d)$ 50

A.2 Proof of Proposition 14 53

3 Average output entropy for quantum channels **56**

3.1 Introduction 56

3.2 Main result 61

3.3 Proof of lemmas 78

3.4 Other results 86

3.5 Conclusion 87

3.6 Acknowledgments 88

B Appendix for Chapter 3 **89**

B.1 $\mathcal{Q}$ sum diagrams 89

4 Analyzing energy transfer in quantum networks using kinetic network approximations **90**

4.1 Introduction 90

4.2 Kinetic networks 97

4.3 Preliminaries 105

4.4 Bounding relaxation time error 110

4.5 Bounding evolution error 115

4.6 Applications 119

4.7 Resolvent difference bounds 125

4.8 Conclusion 130

4.9 Acknowledgments 130

C Appendix for Chapter 4 **131**

C.1 Three sites 131

C.2 General construction 134

C.3 Calculations for applications 135

Bibliography **139**

Chapter 1

Introduction

A quantum channel offers a very general way to describe the interaction of a quantum system with the environment. First, the input state, or initial state, is paired with the environment, assumed to be in a fixed, initialized state. Then the combined system undergoes a unitary transformation, which can be considered as the time evolution for a given duration, and mixes and entangles the two parts. Finally, the environment is "averaged over", giving the output state. The output state represents all the information that is accessible to direct measurements of the system. In quantum information theory, one compares the *entropy* of input and output, a measure of how much information is needed to describe the state, or conversely, how much information can be stored in the state. Hence, the quantum channel is also a channel of communication.

In this dissertation we employ a large variety of quantum channels. In Chapter 2, we define a randomly chosen quantum channel and its "mirrored version", both acting on very high dimensional input states. We describe how the two channels, used in parallel with entangled input states, can transmit information better than each channel used separately. This property is called superadditivity. In Chapter 3, we study the average output entropy for quantum channels, a measure of the average "noisiness". We achieve explicit results for channels acting on qubits, two-dimensional quantum systems. Considering many copies of these channels in parallel, the results simplify further, and give some nice connections to other properties of the channels. Finally, in Chapter 4, the input and output system is a whole network of quantum systems. These networks are modeled after proteins found in plants and bacteria that transport energy in the form of electronic excitations. They are of primary interest in chemistry and biophysics research, but have recently gained the interest of the quantum computing community, which is interested in the networks quantum coherent properties. For these quantum networks we find a new approximation that helps improve the understanding of the relationship between coherent and non-coherent evolution.

In the following sections we give further introduction and background for the topics covered in Chapters 2-4 .

1.1 Entropy

The concept of entropy was developed in the search to quantify how much of the heat in a heat engine can be converted to useful mechanical energy, or work. The first significant insight is due to Sadi Carnot, who described an idealized version of a heat-work engine, now called the Carnot cycle [14]. This cycle offers an unambiguous way to account for the heat and work that the system exchanges with the environment. The crucial idea contained in the Carnot cycle was extracted by Rudolf Clausius in formulating an infinitesimal law for the change in a quantity he called *entropy* [17]. As long as the system undergoes smooth, reversible changes the law allows to calculate the systems entropy. While the entropy does relate in a crucial way to the work a system can provide, it became a more fundamental object of study by its own.

When considering turbulent, fast changing or dissipative processes, like fiction and temperature relaxation, the assumption of smooth, reversible changes does not apply. One can still find the change in entropy for such processes by considering a reversible process with the same initial and final states. For a closed system – completely isolated from the environment – this assessment shows that the entropy stays constant for reversible processes, while it *always increases* for irreversible processes. The entropy is a measure of how likely a system state is, and for closed systems the states of highest entropy are the most probable states.

A better understanding of this puzzling quantity was gained when scientists considered the microscopic nature of the system, i.e. the speed and locations of the individual particles making up a gas, liquid or solid. Naturally, one does not know all the information about every particle in a system containing billions of them. However, given a few macroscopic variables like temperature, volume, etc. and knowing the microscopic laws, one can *count the number of possible states* a system can take. This way, Ludwig Boltzmann formulated the famous statistical definition of entropy [10]

$$S_B = k \cdot \ln W \tag{1.1}$$

where k is a constant and W is the number of possible microscopic states the system can attain given the (macroscopic) restrictions. With this definition it is obvious that entropy is not only a quantity that points us in the direction of more likely states, but it also represents the amount of *lack of information* we have about the system. One condition on this law is the assumption that the system is in any of the W possible states with equal probability. Especially in quantum mechanics this assumption is often not valid, and a system

takes on a different states $|\psi_i\rangle$ with different probabilities p_i. In this case we can think of an ensemble of N systems where N is a large number. Then the law of large number states that with very high probability $p_i N$ subsystems are in the state $|\psi_i\rangle$. Naturally there are many possibilities of choosing which of the N systems takes on which state, while still having a total of $p_i N$ subsystems in state $|\psi_i\rangle$. All these possibilities are called *typical states* of the ensemble, and they all occur with equal probability

$$P = p_1^{p_1 N} p_2^{p_2 N} \dots$$

Hence, if we consider the ensemble as a whole new system, virtually every state it takes on is a typical state and has equal probability P and we again achieve the condition used in (1.1). Therefore, averaging out over the number of subsystems, we find a generalized form of (1.1) to calculate the entropy

$$\begin{aligned} S_B &= \frac{1}{N} \cdot k \cdot \ln(1/P) \\ &= -k \cdot \sum_i p_i \ln p_i \,. \end{aligned} \tag{1.2}$$

In 1948, Shannon published "a mathematical theory of communication" providing the fundaments of information theory [64]. He studied channels for information transmission by assuming they output the letter i with probability p_i. Comparing to our ensemble of systems this corresponds to thinking of a subsystem as an unknown output of the channel, and of the probabilities for different states as the probabilities of the output being a specific letter. Amazingly, it is just the idea of *typical sequences*, the analogue to the typical states mentioned above, that allowed him to make rigorous statements about how much a sequence of letters can be compressed, how much information can be transmitted through a noisy channel and other questions. Consequently the Shannon entropy

$$H = -\sum_i p_i \ln p_i$$

figures prominently in all of information theory, and is the same as (1.2) up to the the Boltzmann factor k. Informally, H represents the amount of information one gains on average when reading one letter output by the channel, or equivalently, it represents the amount of uncertainty about the next letter that is output by the channel.

In quantum information theory, we consider the quantum analogue of H, the von Neumann entropy, which gives the information content of a quantum ensemble given as a density matrix ρ [54]

$$S(\rho) = -\mathrm{Tr}\, \rho \ln \rho \,. \tag{1.3}$$

This entropy has a generalization, the Renyi entropy $S_r(\rho)$, where r is a positive real. In some cases the Renyi entropy [59] is easier to use for analytical work, and it coincides with the von Neumann entropy when r is one

$$S_1(\rho) = S(\rho)\,.$$

In Chapter 2 we will exclusively work with the von Neumann entropy, and in Chapter 3 we will use the Renyi entropy for $r \geq 1$.

1.2 Channel capacity and additivity

Quantum channels are functions that map density matrices to density matrices. They can be thought of as a wires that transmit quantum systems. In the wire the system may interact with the environment, thus getting entangled or being partially measured by this environment. At the output the state of the environment is considered inaccessible, thus introducing (additional) uncertainty about the state of the system.

For information theorists the quantum channel is the quantum version of a classical data channel. One focal point of research concerns the ability of the channel to transport information – the capacity – when subject to various kinds of environment noise. There are several such capacities useful in different transmission scenarios [66]. The one we are most concerned with is the ability to transmit *classical* information through a quantum channel, the Holevo capacity [37].

When using classical channels in parallel the capacity of the combined channel capacity is obtained by adding the two capacities of the individual channels, this property that is called additivity. In the case of quantum channels additivity of the capacity to transmit classical information has been proven to be equivalent to the additivity of the minimum output entropy [65]. For a quantum channel A the minimum output entropy is defined as

$$S_{\min}(A) = \min_{|\phi\rangle \in \mathcal{H}_A} S\left[A\left(|\phi\rangle\langle\phi|\right)\right]$$

where the minimum is over pure input states $|\phi\rangle\langle\phi|$, and S is the von Neumann entropy (1.3). When considering two channels in parallel we write $A \otimes B$. If A operates on density matrices in $\mathcal{H}_A$ and B operates on density matrices in $\mathcal{H}_B$ then $A \otimes B$ operates on density matrices n $\mathcal{H}_A \otimes \mathcal{H}_B$. Additivity of the minimum output entropy is

$$S_{\min}(A \otimes B) \stackrel{?}{=} S_{\min}(A) + S_{\min}(B)\,.$$

The minimum over pure states in $S_{\min}(A \otimes B)$ includes product states $|\phi_A\rangle \otimes |\phi_B\rangle$ but also entangled states. The entangled states have no classical analogue, and therefore they might yield a lower minimum, breaking

additivity.

1.3 Counterexamples to additivity

One of the earliest counterexamples to an additivity question was found over a decade ago [23], and pertains to the additivity of coherent information transmission. It is a product of the depolarizing channel defined below in (1.4). For the case of classical information transmission through quantum channels the picture looked different for some time, as additivity was proved for several sets of channels. In particular, the capacities of depolarizing channels, of entanglement breaking channels and unital channels were proven to be additive [45]. On the other hand, counterexamples were found in the case of the Renyi entropy S_r for $r > 1$ [35, 69]. But the fact that these examples break down for the most relevant case $r = 1$, where the Renyi entropy equals the von Neumann entropy, left the possibility open that additivity might be true in that case.

However, in 2009, Hastings published a proof showing the existence of counterexamples to additivity [33] for $r = 1$. The proof uses channels of the form

$$A(\rho) = \sum_i p_i U_i^\dagger \rho U_i$$

where U_i is a unitary and $0 \leq p_i \leq 1$, $\sum_i p_i = 1$. Such channels can be chosen randomly, by drawing U_i and p_i from some random distribution. For the right choice of random distribution a nice relationship to random matrix theory arises. Hastings uses some unconventional bounds on this standard construct, and proves that in the limit of large dimension some (in fact most) of the randomly chosen quantum channels are superadditive. In Chapter 2 we make a detailed study of this proof, translating it into rigorous mathematical language. Furthermore, we find some lower bounds on how large the dimension has to be, so that superadditivity might be achieved.

1.4 Average output entropy of quantum channels

In 1.2 we introduced the minimum output entropy of quantum channels $S_{\min}$, and mentioned how it relates to the channel capacity. In the work of Chapter 3 we take a different approach to the study of quantum channels by regarding the *average output entropy* of quantum channels. We start with a pure system – a system that is in a single state and has entropy equal zero –, then we send the system through the channel, where it becomes entangled with the environment. After averaging out the environment the system is mixed

and has positive entropy. By inputting any possible pure state and averaging the output entropy we measure the "average noisiness" of the channel. Explicitly, we try to calculate the value of

$$\bar{S}_r(A) = \mathbb{E}\left[S_r(|\phi\rangle\langle\phi|)\right]$$

where S_r is the Renyi entropy, and where $\mathbb{E}$ is the average over pure input states $|\phi\rangle$.

It is hard to analyze this quantity directly, instead we find a closely related property β_r with

$$\bar{S}_r(A) \geq \beta_r(A)\,,$$

which – using random matrix theory methods – we can write in closed form. The expression we find can easily be extended to quantum channels in parallel $A \otimes B$ and in particular $A^{\otimes n}$. This make some features of the channel more expressed, we take the limit $n \to \infty$ for the expression $\bar{S}_r(A^{\otimes n})$ and $\beta_r(A^{\otimes n})$ and call it $\bar{S_r}^{\,\mathrm{reg}}(A)$ and $\beta_r^{\mathrm{reg}}(A)$ respectively. This approach is similar to statistical physics where one often considers the limit of large systems with infinitely many degrees of freedom.

We make a specific analysis of two groups of channels. First, we consider a subset of entanglement breaking channels (EBC). An entanglement breaking channel paired with any other channel will break the entanglement of input states, i.e. the output will always be a separable density matrix – the density matrix analogue to product states. For some EBC, we can prove that

$$\bar{S_r}^{\,\mathrm{reg}}(A) = \beta_r^{\mathrm{reg}}(A) = \log d$$

i.e. a pure input gets completely mixed with the environment with probability 1.

Second, we consider the depolarization channel

$$\Delta_\lambda(\rho) = \lambda\rho + (1-\lambda)\frac{\mathbb{1}_d}{d} \tag{1.4}$$

where $\mathbb{1}_d$ is the unit matrix in d-dimensional space. One way to think of this channel is, that with probability λ it passes the input to output completely unaffected and with probability $(1-\lambda)$ it replaces the state with the completely mixed state $\mathbb{1}_d/d$. In two dimensions, the states can be considered to be an ensemble of spin-1/2 particles, then the channel Δ_λ simply scales the polarization $\vec{d}$ – the average spin – by a factor of λ, hence its name. We consider this channel for $0 \leq \lambda \leq 1$, which yields a function $\lambda \to S_r^{\mathrm{reg}}(\Delta_\lambda)$. We prove that this function is non-analytic, in fact, it is constant for $\lambda \leq 1/3$ just the region where Δ_λ is entanglement breaking, but for $\lambda > 1/3$ the function decreases and reaches 0 when $\lambda = 1$. This is reminiscent of a

continuous order parameter that is constant for one phase of a system and starts to vary, non-analytically, at the phase transition.

1.5 Photosynthesis

The research presented in Chapter 4 was motivated by quantum networks that arise as parts of biological cells. Here we give a general introduction to these quantum networks, before presenting recent research in the next section.

Plants and some bacteria have the ability to absorb light and convert it to chemical energy. This process called photosynthesis is vital to most lifeforms on earth and consequently has been studied by researchers for centuries. Light from the sun is absorbed by molecules called pigments. Because these molecules are usually only able to absorb part of the visual spectrum, they appear colored. The most prominent are the green chlorophyll, the blue to purple bacteriochlorophyll, and the orange carotene.

In 1932 it was discovered [5] that up to 2400 chlorophyll molecules are required to produce a single oxygen molecule, showing that photosynthesis is not accomplished by single molecules but rather by large, complex structures. This and other experiments are explained by the idea that the large complex captures light as excitations, which then are "funneled" across the complex to a site where the chemical reaction occurs [30]. Today, the pigments and proteins involved in the process are well known, however the geometric structure of some proteins and how proteins are located with respect to each other, is still up to some debate [1]. The antenna structure catching the photon is called chlorosome and consists of a protein scaffolding holding 10-1000 pigments. The pigments absorb a photon and become electronically excited. This excitation, which is usually localized on a few pigments, jumps between pigments via a dipole-dipole interaction, this process is called Förster Resonance Energy Transfer (FRET) [25, 18]. Ultimately, the excitation is transferred to another protein-pigment complex, the reaction center, where the chemical conversion, employing several oxido-reduction reactions, occurs.

The process of transfer does not always succeed. The *quantum efficiency of transfer* is the probability that an absorbed photon reaches the reaction center before it is lost due to other processes. The maximum photsynthetic energy conversion efficiency for plants was calculated to be 6% [72]. This is even less than currently sold solar cells which usually achieve efficiencies between 10-20% and can reach as high as 42.3% [31]. However, the probability for an already absorbed photon to cause a redox reaction, and consequently the quantum efficiency of transfer, is close to 100% [9]. This renders the process of energy transfer in these complexes an interesting object of study not only for researches but also the industry.

1.6 Energy transfer in the Fenna-Matthews-Olson (FMO) complex

The FMO complex consists of three proteins wrapping seven bacteriochlorophyll molecules each. It acts as a mediator between the chlorosomes and the reaction center in green sulfur bacteria [68]. Because of its small size relative to other protein-pigment complexes it provides a simple model system for both experimentalists and theorists. It was the first complex for which the three-dimensional structure was fully determined by X-ray crystallography [52]. This allows the determination of the interaction strength and site energies [1]. In cases of weak interaction Förster derived a rate of transfer that a hybrid of theoretical predictions and measured absorption and emission spectra, this approach has very succesfully applied to describe many experiments. However, it does not capture any coherent, oscillatory population transfer that can occur when the interaction is strong.

However, in 2007, experiments measuring the initial evolution of excitons showed that populations oscillate [55, 24], showing that Försters rate is not the appropriate description in this case. Instead, one needs to consider the system of pigments as a coherent *quantum network*, where the exciton can travel in a wavelike fashion. It was speculated that this coherent evolution might help increase the transfer efficiency, and various approaches of modeling the quantum network were made [42, 43, 57, 22]. Besides chemists and biological physicists, the phenomenon also captured the interest of the quantum information community, because maintaining and utilizing coherence is the central prerequisite for achieving quantum computation.

Because of the complexity of the pigment-protein interaction, one is confronted with the choice between massively simplified models that allow some analytical understanding, and complex models that can only be simulated on computers, giving little guidance to explaining phenomena. Two similar models of the former nature surprisingly showed that the energy transfer efficiency is *increased* by the noisy interaction with the environment [58, 56]. In a sense, these models define a quantum channel that acts on the entire quantum network of pigments. In a subsequent paper, these models of quantum networks were approximated by using *kinetic networks*, giving an analytical insight into parameter values optimal for transfer efficiency [13].

In Chapter 4, we find a new kinetic network that is a much closer approximation to the model of [58]. We then provide a mathematical analysis of the two kinetic networks and the quantum network by finding rigorous bounds for the differences in transfer efficiency. In particular we find the limit where the approximation becomes exact. We further study both approximations in some generic cases and compare the results to simulations, confirming our analytic findings.

Chapter 2

Comments on Hastings' Additivity Counterexamples

Abstract

Hastings [32, 33] recently provided a proof of the existence of channels which violate the additivity conjecture for minimal output entropy. In this paper we present an expanded version of Hastings' proof. In addition to a careful elucidation of the details of the proof, we also present bounds for the minimal dimensions needed to obtain a counterexample.

2.1 The additivity conjectures

The classical capacity of a quantum channel is the maximum rate at which classical information can be reliably transmitted through the channel. This maximum rate is approached asymptotically with multiple channel uses by encoding the classical information in quantum states which can be reliably distinguished by measurements at the output. In general, in order to achieve optimal performance, it is necessary to use measurements which are entangled across the multiple channel outputs. However it was conjectured that product input states are sufficient to achieve the maximal rate of transmission, in other words that there is no benefit in using entangled states to encode the classical information. This conjecture is closely related to other additivity conjectures of quantum information theory, as will be explained below. Recently Hastings [32] disproved all of these additivity conjectures by proving the existence of channels which violate the additivity of minimal output entropy. The purpose of this paper is to present in detail the findings of Hastings' paper, and also to find bounds for the minimal dimensions needed for this type of counterexample.

We begin by formulating the various additivity conjectures. The Holevo capacity of a quantum channel Φ is defined by

$$\chi^*(\Phi) = \sup_{\{p_i,\,\rho_i\}} S\Big(\Phi\big(\sum_i p_i\rho_i\big)\Big) - \sum_i p_i\, S\Big(\Phi(\rho_i)\Big) \tag{2.1}$$

where the supremum runs over ensembles of input states, and where $S(\rho)$ denotes the von Neumann entropy of the state ρ (here and throughout the paper log denotes the natural logarithm) :

$$S(\rho) = -\mathrm{Tr}\,\rho\,\log\rho \tag{2.2}$$

The classical information capacity $C(\Phi)$ is known [37, 63] to equal the following limit:

$$C(\Phi) = \lim_{n\to\infty} \frac{1}{n}\,\chi^*(\Phi^{\otimes n})$$

It has been a longstanding conjecture that the classical information capacity is in fact equal to the Holevo capacity:

$$\boxed{\text{Conjecture 1}} \qquad C(\Phi) = \chi^*(\Phi) \tag{2.3}$$

Conjecture 1 would be implied by additivity of χ^* over tensor products. This led to the following conjecture: for all channels Φ and Ω,

$$\boxed{\text{Conjecture 2}} \qquad \chi^*(\Phi\otimes\Omega) = \chi^*(\Phi) + \chi^*(\Omega) \tag{2.4}$$

Subsequently a third conjecture appeared, namely the additivity of minimum output entropy:

$$\boxed{\text{Conjecture 3}} \qquad S_{\min}(\Phi\otimes\Omega) = S_{\min}(\Phi) + S_{\min}(\Omega) \tag{2.5}$$

where $S_{\min}$ is defined by

$$S_{\min}(\Phi) = \inf_{\rho}\, S(\Phi(\rho)) \tag{2.6}$$

Finally Amosov, Holevo and Werner [4] proposed a generalization of Conjecture 4 with von Neumann entropy replaced by the Renyi entropy: for all $p \geq 1$

$$\boxed{\text{Conjecture 4}} \qquad S_{p,\min}(\Phi\otimes\Omega) = S_{p,\min}(\Phi) + S_{p,\min}(\Omega) \tag{2.7}$$

where $S_{p,\min}$ is the minimal Renyi entropy defined for $p \neq 1$ by

$$S_{p,\min}(\Phi) = \inf_{\rho} S_p(\Phi(\rho)), \qquad S_p(\tau) = \frac{1}{1-p} \log \mathrm{Tr}\, \tau^p \tag{2.8}$$

In 2004 Shor [65] proved the equivalence of several additivity conjectures, including Conjectures 2 and 3 above. In subsequent work [29] it was shown that Conjectures 1 and 2 are equivalent. The conjectures have been proved in several special cases [2, 3, 12, 26, 44, 45, 47, 48], but recently most progress has been made in the search for counterexamples. This started with the Holevo-Werner channel [69] which provided a counterexample to Conjecture 4 with $p > 4.79$, then more recently Winter and Hayden found counterexamples to Conjecture 4 for all $p > 1$ [35], and violations have since been found also for $p = 0$ and p close to zero [21]. Finally in 2008, Hastings [32] produced a family of channels which violate Conjecture 3, namely additivity of minimal output von Neumann entropy, thereby also proving (via [65] and [29]) that Conjectures 1 and 2 are false.

Following Winter's idea, the product channels used by Hastings have the form $\Phi \otimes \overline{\Phi}$ where Φ is a special channel which we call a *random unitary channel*. This means that there are positive numbers $w_1, \ldots, w_d$ with $\sum_i w_i = 1$ and unitary $n \times n$ matrices $U_1, \ldots, U_d$ such that

$$\Phi(\rho) = \sum_{i=1}^{d} w_i\, U_i\, \rho\, U_i^*, \qquad \overline{\Phi}(\rho) = \sum_{i=1}^{d} w_i\, \overline{U_i}\, \rho\, \overline{U_i^*} \tag{2.9}$$

These channels are chosen randomly using a distribution that depends on the two integers n and d, where n is the dimension of the input space and d is the dimension of the environment. Hastings' main result is that for n and d sufficiently large there are random unitary channels which violate Conjecture 3, that is

$$S_{\min}(\Phi \otimes \overline{\Phi}) < S_{\min}(\Phi) + S_{\min}(\overline{\Phi}) \tag{2.10}$$

This result also allows a direct construction of channels which violate Conjectures 1 and 2, as we now show. Using results from the paper [29], the inequality (2.10) implies that the additivity of minimal output entropy does not hold for the product $\Phi' \otimes \Phi'$, where $\Phi' = \Phi \oplus \overline{\Phi}$. In addition, as shown in the paper [27], there is a unital extension of Φ', denoted Φ'', such that the additivity of minimal output entropy does not hold for $\Phi'' \otimes \Phi''$, and such that

$$S_{\min}(\Phi'' \otimes \Phi'') = 2 \log D - \chi^*(\Phi'' \otimes \Phi'').$$

where D is the dimension of the output space for Φ''. Thus Φ'' provides a counterexample for Conjecture 2,

and

$$\lim_{k\to\infty} \frac{1}{2k}\chi^*((\Phi'')^{\otimes 2k}) > \chi^*(\Phi'').$$

Therefore, the classical capacity of Φ'' does not equal its Holevo capacity, and this provides a counterexample for Conjecture 1.

One key ingredient in the proof is the relative sizes of dimensions, namely $n >> d >> 1$, where n is the dimension of the input space, and d is the dimension of the environment. Recall that in the Stinespring representation a channel is viewed as a partial isometry from the input space $\mathcal{H}_{in}$ to the product of output and environment spaces $\mathcal{H}_{out} \otimes \mathcal{H}_{env}$, followed by a partial trace over the environment. The image of $\mathcal{H}_{in}$ under the partial isometry is a subspace of dimension n sitting in the product $\mathcal{H}_{out} \otimes \mathcal{H}_{env}$. Making the environment dimension d much smaller than the input dimension n should guarantee that with high probability this subspace will consist of almost maximally entangled states. For such states the output entropy will be close to the maximal possible value $\log d$, and therefore the minimal entropy of the channel should also (hopefully) be close to $\log d$. At the same time the product channel $\Phi \otimes \overline{\Phi}$ sends the maximally entangled state into an output with one relatively large eigenvalue, and thus one might hope to find a gap between $S_{\min}(\Phi \otimes \overline{\Phi})$ and $S_{\min}(\Phi) + S_{\min}(\overline{\Phi})$. Turning this vague notion into a proof requires considerable insight and ingenuity. In this paper we focus on the technical aspects of Hastings' proof. Some of the estimates and inequalities derived in this paper are new, but all the main ideas and methods are taken from [32].

The paper is organized as follows. In Section 2 we define notation and make a precise statement of Hastings' results. In Section 3 we present some background material on probability distributions for states and channels. In Section 4 we 'walk through' the proof of Hastings' Theorem, stating results where needed and delineating the logic of the argument. In Section 5 we give the proofs of various results needed in Section 4 and elsewhere. Section 6 discusses different aspects of the proof and possible directions for further research. The Appendix contains the derivation of some estimates needed for the proof.

2.2 Notation and statement of results

We will mostly avoid Dirac bra and ket notation, although it will be used in Sections 2.5.1 and 2.5.5.

2.2.1 Notation

Let $\mathcal{M}_n$ denote the algebra of complex $n \times n$ matrices. The identity matrix will be denoted I_n, or just I. The set of *states* in $\mathcal{M}_n$ is defined as

$$\mathcal{S}_n = \{\rho \in \mathcal{M}_n \,:\, \rho = \rho^* \geq 0,\ \mathrm{Tr}\,\rho = 1\}$$

The set of *unit vectors* in $\mathbb{C}^n$ will be denoted

$$\mathcal{V}_n = \{z = (z_1, \ldots, z_n)^T \in \mathbb{C}^n \,:\, z^* z = \sum_{i=1}^{n} |z_i|^2 = 1\}$$

Every unit vector $z \in \mathcal{V}_n$ defines a pure state $\rho = zz^*$ satisfying $\rho^2 = \rho$. The set of unit vectors $\mathcal{V}_n$ is identified with the real $(2n-1)$-dimensional sphere S^{2n-1}, and hence carries a unique uniform probability measure which we denote σ_n.

The set of unitary matrices in $\mathcal{M}_n$ is denoted

$$\mathcal{U}(n) = \{U \in \mathcal{M}_n \,:\, UU^* = I\}$$

We will write H_n for the normalized Haar measure on $\mathcal{U}(n)$.

A *channel* is a completely positive trace-preserving map $\Phi : \mathcal{M}_n \to \mathcal{M}_m$. Recall the definition of random unitary channel (2.9):

$$\Phi(\rho) = \sum_{i=1}^{d} w_i\, U_i\, \rho\, U_i^*$$

The set of all random unitary channels on $\mathcal{M}_n$ with d summands will be denoted $\mathcal{R}_d(n)$. Given a channel $\Phi \in \mathcal{R}_d(n)$ the *complementary* or *conjugate* channel $\Phi^C : \mathcal{M}_n \to \mathcal{M}_d$ is defined by [39, 46]

$$\Phi^C(\rho) = \sum_{i,j=1}^{d} \sqrt{w_i w_j}\ \mathrm{Tr}\,(\rho\, U_j^* U_i)\ |i\rangle\langle j| \tag{2.11}$$

As is well-known, for any input state ρ the output states $\Phi(\rho)$ and $\Phi^C(\rho)$ are related by

$$\Phi(\rho) = \mathrm{Tr}_2\, W\rho W^*, \quad \Phi^C(\rho) = \mathrm{Tr}_1\, W\rho W^* \tag{2.12}$$

Here, $W : \mathbb{C}^n \to \mathbb{C}^{nd}$ is a partial isometry. Also Tr_2 denotes the partial trace over the state space of the environment, and Tr_1 denotes the partial trace over the state space of the system. When $\rho = zz^*$ is a pure

state, the matrices $\Phi(zz^*)$ and $\Phi^C(zz^*)$ are partial traces of the same pure state, and thus have the same non-zero spectrum and the same entropy. Therefore $S_{\min}(\Phi) = S_{\min}(\Phi^C)$. For the purposes of constructing the counterexample it is convenient to work with both Φ and Φ^C. In particular, we are interested in the cases where W consists of rescaled unitary block matrices;

$$W = \begin{pmatrix} \sqrt{w_1}U_1 \\ \vdots \\ \sqrt{w_d}U_d \end{pmatrix}$$

Note that $\sum_i w_i = 1$ as W is a partial isometry. We define a measure on this subset of partial isometries, in Section 2.3.4, as the product of Haar measures and a particular measure on the simplex.

The complex conjugate channel $\overline{\Phi}$ is defined by

$$\overline{\Phi}(\rho) = \sum_{i=1}^{d} w_i \, \overline{U_i} \rho \overline{U_i^*} = \sum_{i=1}^{d} w_i \, \overline{U_i} \rho U_i^T$$

Again note that Φ and $\overline{\Phi}$ have identical minimum output entropies.

2.2.2 The main result

Following the work of Winter and Hayden [35], the counterexample is taken to be a product channel of the form $\Phi \otimes \overline{\Phi}$ where Φ is a random unitary channel. Hastings first proves the following universal upper bound for the minimum output entropy of such a product.

Lemma 1. *For any $\Phi \in \mathcal{R}_d(n)$,*

$$S_{\min}(\Phi \otimes \overline{\Phi}) \leq 2 \log d - \frac{\log d}{d}$$

Lemma 1 will be proved in Section 2.5.1. The counterexample is found by proving the existence of a random unitary channel Φ whose minimum output entropy is greater than one half of this upper bound, that is greater than $\log d - \log d/2d$. For such a channel it will follow that

$$\begin{aligned} S_{\min}(\Phi \otimes \overline{\Phi}) &\leq 2 \log d - \frac{\log d}{d} && (2.13) \\ &< 2 S_{\min}(\Phi) && (2.14) \\ &= S_{\min}(\Phi) + S_{\min}(\overline{\Phi}) && (2.15) \end{aligned}$$

and this will provide the counterexample to Conjecture 3. Hastings [32] proved the existence of such channels

using a combination of probabilistic arguments and estimates involving the distribution of the reduced density matrix of a random pure state. The next Theorem is a precise statement of Hastings' result.

Theorem 2. *There is $h_{\min} < \infty$, such that for all $h > h_{\min}$, all d satisfying $d \log d \geq h$, and all n sufficiently large, there is $\Phi \in \mathcal{R}_d(n)$ satisfying*

$$S_{\min}(\Phi) > \log d - \frac{h}{d} \tag{2.16}$$

By taking d large enough so that $2h_{\min} < \log d$, we deduce from Theorem 2 that there is a channel Φ satisfying

$$S_{\min}(\Phi) > \log d - \frac{\log d}{2d}$$

and this establishes the existence of counterexamples for Conjecture 3. In fact the proof will show that as d, $n \to \infty$, the probability that a randomly chosen channel in $\mathcal{R}_d(n)$ will satisfy the bound (2.16) approaches one.

It would be interesting to determine the set of integers (n, d) for which there are random unitary channels in $\mathcal{R}_d(n)$ violating additivity, and in particular to find the smallest dimensions which allow violations, as well as the size of the largest possible violation. Following this line of reasoning we define

$$\begin{aligned} d_{\min} &= \inf\left\{d \,:\, \exists\, n,\, \exists\, \Phi \in \mathcal{R}_d(n) \,\text{s.t.}\, S_{\min}(\Phi) > \log d - \frac{\log d}{2d}\right\} \\ n_{\min} &= \inf\left\{n \,:\, \exists\, d,\, \exists\, \Phi \in \mathcal{R}_d(n) \,\text{s.t.}\, S_{\min}(\Phi) > \log d - \frac{\log d}{2d}\right\} \\ \Delta S_{\max} &= \sup_{n,d} \sup_{\Phi \in \mathcal{R}_d(n)} \Big(S_{\min}(\Phi) + S_{\min}(\overline{\Phi}) - S_{\min}(\Phi \otimes \overline{\Phi})\Big) \end{aligned} \tag{2.17}$$

The next result gives some bounds on these quantities.

Proposition 3.

$$\begin{aligned} d_{\min} &< 3.9 \times 10^4 \\ n_{\min} &< 7.8 \times 10^{32} \\ \Delta S_{\max} &> 9.5 \times 10^{-6} \end{aligned}$$

Proposition 3 will be proved in Section 2.4.6. The bounds in Proposition 3 are surely not optimal, however they may indicate the delicacy of the non-additivity effect for this class of channels. It would certainly be interesting to tune the estimates in this paper in order to improve the bounds in Proposition 3, or even better to find a different class of channels where the effect is larger.

2.3 Background on random states and channels

As mentioned above, the proof of Theorem 2 relies on probabilistic arguments, involving distributions of pure states and random unitary channels. The next sections explain the distributions which play a role in the proof.

2.3.1 Probability distributions for states

Recall that $\mathcal{V}_n$ is the set of unit vectors in $\mathbb{C}^n$. This set carries a natural uniform measure σ_n, namely the uniform measure on the (real) $(2n-1)$-dimensional sphere. If $\mathbb{C}^{dn} = \mathbb{C}^d \otimes \mathbb{C}^n$ is a product space, then a unit vector $z \in \mathcal{V}_{dn}$ can be written as a $n \times d$ matrix M, with entries

$$M_{ij}(z) = z_{(i-1)d+j}, \quad i = 1, \ldots n, \; j = 1, \ldots, d$$

satisfying $\operatorname{Tr} M^* M = \sum_{ij} |z_{ij}|^2 = 1$. Define the map $G : \mathcal{V}_{dn} \to \mathcal{M}_d$ by

$$G(z) = M(z)^* M(z) \tag{2.18}$$

It follows that $G(z) \geq 0$ and $\operatorname{Tr} G(z) = 1$, and hence the image of G lies in $\mathcal{S}_d$ (the set of d-dimensional states). Since z is a random vector (with distribution σ_{dn}) it follows that $G(z)$ is a $\mathcal{S}_d$-valued random variable, or more simply a random state. Its distribution has been studied in many other contexts (see for example [34]) and it plays a key role in the proof here.

2.3.2 Probability distributions on the simplex Δ_d

Let Δ_d denote the simplex of d-dimensional probability distributions:

$$\Delta_d = \{(x_1, \ldots, x_d) \subset \mathbb{R}^d \, : \, x_i \geq 0, \; \sum_{i=1}^{d} x_i = 1\} \tag{2.19}$$

We define below three different probability distributions on Δ_d. One is the uniform measure inherited from $\mathbb{R}^d$, and the others are defined by the diagonal entries and the eigenvalues of $G(z)$ where z is a random unit vector in $\mathcal{V}_{dn}$.

$\boxed{\text{Uniform distribution}}$ The simplex Δ_d carries a natural measure inherited from Lebesgue measure on

$\mathbb{R}^d$: this is conveniently written as

$$\delta\Big(\sum_{i=1}^d w_i - 1\Big) dw_1 \dots dw_d = \delta\Big(\sum_{i=1}^d w_i - 1\Big)\,[dw] \tag{2.20}$$

where $\delta(\cdot)$ is the Dirac δ-function. Integrals with respect to this measure can be evaluated by introducing local coordinates on $\mathbb{R}^d$ in a neighborhood of Δ_d. In particular the volume of Δ_d with respect to the measure (2.20) can be computed:

$$\int_{\Delta_d} \delta\Big(\sum_{i=1}^d w_i - 1\Big)\,[dw] = \frac{1}{(d-1)!} \tag{2.21}$$

$\boxed{\text{Diagonal distribution } \nu_{d,n}}$ Let $z \in \mathcal{V}_{dn}$ be a random unit vector in $\mathbb{C}^n \otimes \mathbb{C}^d$. The joint distribution of the diagonal entries $(G_{11}(z), \dots, G_{dd}(z))$ will be denoted $\nu_{d,n}$. It is possible to find an explicit formula for the density of $\nu_{d,n}$, however we will not need it in this paper. It is sufficient to note that a collection of d random variables $Y_1, \dots, Y_d$ have the joint distribution $\nu_{d,n}$ if and only if they can be written as

$$Y_j = \sum_{i=1}^n |z_{ij}|^2, \quad j = 1, \dots, d \tag{2.22}$$

where $\{z_{ij}\}$ are the components of a uniform random vector on the unit sphere in $\mathbb{C}^n \otimes \mathbb{C}^d$. We come back to this problem in Section 2.5.3.

$\boxed{\text{Eigenvalue distribution } \mu_{d,n}}$ As noted above the eigenvalues of $G(z)$ are non-negative and sum to one. [1] However the eigenvalues are not ordered and so define a map not into Δ_d but rather into the quotient Δ_d/Σ_d where Σ_d is the symmetric group. Thus when $z \in \mathcal{V}_{dn}$ is a random vector the eigenvalues of $G(z)$ are Δ_d/Σ_d-valued random variables. However it is convenient to use a joint density for the eigenvalue distribution on Δ_d, with the understanding that it should be evaluated only on events which are invariant under Σ_d. This density is known explicitly [50, 73]: for any event $A \subset \Delta_d$

$$\mu_{d,n}(A) = Z(n,d)^{-1} \int_A \prod_{1 \le i < j \le d} (w_i - w_j)^2 \prod_{i=1}^d w_i^{n-d}\, \delta\Big(\sum_{i=1}^d w_i - 1\Big)\,[dw] \tag{2.23}$$

where $Z(n,d)$ is a normalization factor. The distribution $\mu_{d,n}$ plays an essential role in the proof of Theorem 2. Explicit expressions for $Z(n,d)$ are known [73]. In Appendix A we derive the following bound: for n sufficiently large,

$$Z(n,d)^{-1} \le n^{d^2}\, d^{d\,(n-d)} \tag{2.24}$$

[1] $G(z)$ gives the complex Wishart matrix when $z \in \mathbb{C}^n \otimes \mathbb{C}^d$ with each entry z_{ij} having IID complex normal distribution. The eigenvalue distribution was shown to be proportional to $\prod_{1 \le i < j \le d} (w_i - w_j)^2 \prod_{i=1}^d w_i^{n-d} [dw]$, for example, in [11].

2.3.3 Estimates for $\mu_{d,n}$

Define the function

$$\boxed{F(x) = -\log x + x - 1} \tag{2.25}$$

Lemma 4. *For all d, for n sufficiently large, and for any event $A \subset \Delta_d$,*

$$\mu_{d,n}(A) \leq \int_A \exp\left[d^2 \log n - (n-d)\sum_{i=1}^{d} F(dw_i)\right] \delta\left(\sum_{i=1}^{d} w_i - 1\right)[dw] \tag{2.26}$$

This Lemma will be proved in Section 2.5.2. Using (2.21) we immediately get the following bound.

Corollary 5. *For all d, for n sufficiently large, and for any event $A \subset \Delta_d$,*

$$\mu_{d,n}(A) \leq \exp\left[d^2 \log n - \log(d-1)! - (n-d)\inf_{w\in A}\sum_{i=1}^{d} F(dw_i)\right] \tag{2.27}$$

Note that $F(x)$ is convex, and also $F(1) = F'(1) = 0$. The Taylor expansion around 1 gives

$$F(1 + d\delta w) = \frac{1}{2}d^2(\delta w)^2 + R \tag{2.28}$$

where the remainder is

$$R = -\frac{1}{3}(1 + d\delta)^{-3}(d\delta w)^3. \tag{2.29}$$

and δ is some value between 0 and δw. Note that $-1/d < \delta w < (d-1)/d$ as $0 \leq w \leq 1$. Also, $R > 0$ if $\delta w < 0$. When $\delta w > 0$

$$0 > R > -\frac{1}{3}d^3(\delta w)^3. \tag{2.30}$$

Recall that $F(x) \geq 0$, so we have the bound $F(dw_i) \geq 0$ for all i. Thus feeding (2.28) into Corollary 5 gives the following estimate, which will be used in Section 2.5.4.

Corollary 6. *For all d, for n sufficiently large, and for any $i = 1, \ldots, d$,*

$$\mu_{d,n}\left\{w : \left|w_i - \frac{1}{d}\right| \geq t\right\} \leq \exp\left[d^2 \log n - \log(d-1)! - \frac{n-d}{2}d^2t^2 + \frac{n-d}{3}d^3t^3\right] \tag{2.31}$$

2.3.4 Probability distribution for random unitary channels

A random unitary channel (2.9) is determined by the coefficients w_i and the unitary matrices U_i. Thus the set of random unitary channels $\mathcal{R}_d(n)$ is naturally identified with $\Delta_d \times \mathcal{U}(n)^d$. Recall the distribution $\nu_{d,n}$

defined in Section 2.3.2 for the diagonal entries of $G(z)$, and the Haar measure H_n defined on $\mathcal{U}(n)$. We define the following product probability measure on $\mathcal{R}_d(n)$:

$$\mathcal{P}_{d,n} = \nu_{d,n} \times H_n \times \cdots \times H_n \tag{2.32}$$

where $H_n \times \cdots \times H_n$ is the d-fold product Haar measure on $\mathcal{U}(n)^d$. Using the measure $\mathcal{P}_{d,n}$ on $\mathcal{R}_d(n)$ means that the unitaries U_i are selected randomly and independently, while the coefficients w_j have the joint distribution $\nu_{d,n}$, and thus can be written in the form (2.22) where $\{z_{ij}\}$ $(i = 1, \ldots, n;\ j = 1, \ldots, d)$ are the components of a random unit vector in $\mathcal{V}_{nd}$.

Recall the definition (2.11) of the conjugate channel. Define the map

$$H : \mathcal{R}_d(n) \times \mathcal{V}_n \to \mathcal{M}_d, \quad (\Phi, z) \mapsto \Phi^C(zz^*) \tag{2.33}$$

Recall the definition (2.18) of the map $G : \mathcal{V}_{dn} \to \mathcal{M}_d$. The following relation between the distributions $\mathcal{P}_{d,n}$, σ_n and σ_{dn} is crucial to the proof. Given a measurable map $f : X \to Y$ between measure spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ (where $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras on X, Y respectively), and given a measure μ on $(X, \mathcal{A})$, we define the push-forward measure $f^*(\mu)$ on $(Y, \mathcal{B})$ by $f^*(\mu)(B) = \mu(f^{-1}(B))$ for all $B \in \mathcal{B}$.

Lemma 7.

$$H^*(\mathcal{P}_{d,n} \times \sigma_n) = G^*(\sigma_{dn})$$

Lemma 7 will be proved in Section 2.5.3. It implies that if Φ is chosen randomly according to the measure $\mathcal{P}_{d,n}$ and z is chosen randomly and uniformly in $\mathcal{V}_n$, then the eigenvalues of the matrix $\Phi^C(zz^*)$ will have the distribution $\mu_{d,n}$.

2.4 Proof of Theorem 2

The main idea of the proof is to isolate some properties of random unitary channels which are typical for large values n and d. These properties will then be used to prove that large minimum output entropy is also typical for random unitary channels when n, d are large.

Recall that the environment dimension d will be chosen to be much smaller than the input dimension n. As the identity (2.12) shows, selecting a channel in $\mathcal{R}_d(n)$ corresponds to selecting a subspace of dimension n in the product space $\mathbb{C}^n \otimes \mathbb{C}^d$. The structure of random bipartite subspaces was analyzed in the paper [34], and it was shown that in some circumstances most states in a randomly selected subspace will be close

to maximally entangled. In such a situation the reduced density matrix of a randomly selected output state $\Phi^C(zz^*)$ will be close to the maximally mixed state I/d, and hence its entropy will be close to $\log d$. Although this observation plays an essential role in Hastings' proof, the methods used in [34] do not directly yield the bounds needed.

2.4.1 Definition of the typical channel

A channel Φ will be called typical if Φ^C maps at least one half of input states into a small ball centered at the maximally mixed output state I/d. The size of the small ball in question involves a numerical parameter b and is defined as follows:

$$B_d(n) = \left\{ \rho \in \mathcal{S}_d \,:\, \left\| \rho - \frac{1}{d} I \right\|_\infty \leq b \sqrt{\frac{\log n}{n}} \right\}$$

Definition 8. A random unitary channel Φ is called *typical* if with probability at least $1/2$ a randomly chosen input state is mapped by Φ^C into the set $B_d(n)$. The set of typical channels is denoted T:

$$T = \left\{ \Phi \,:\, \sigma_n\Big(z \,:\, \Phi^C(zz^*) \in B_d(n) \Big) \geq 1/2 \right\}$$

As the next result shows, for large n most channels are typical.

Lemma 9. *For all $b > \sqrt{3}$, $d \geq 2$ and $0 < \alpha < b^2/3 - 1$, taking n sufficiently large,*

$$\mathcal{P}_{d,n}(T^c) \leq \frac{2\,d}{(d-1)!} \, \exp[-\alpha\, d^2 \, \log n] \tag{2.34}$$

Thus if $b > \sqrt{3}$, then as $n \to \infty$ with high probability a randomly chosen channel will lie in the set T. In particular $\mathcal{P}_{d,n}(T^c) < 1$ for n sufficiently large. The number α can be chosen to satisfy

$$\alpha = \frac{b^2(n-d)}{3n} - 1 \tag{2.35}$$

The dimension n must be large enough so that the right side of (2.35) is positive, and also so that $n \geq 4b^2d^2 \log n$ (this is a technical condition needed in the proof, see Section 2.5.4).

The second property of a typical channel Φ is the existence of a 'tube' of output states surrounding $\Phi^C(zz^*)$ for every input state $z \in \mathcal{V}_n$. This property is used to eliminate the possibility of isolated output states with low entropy: if for some z the output entropy $S(\Phi^C(zz^*))$ is small, then there is a nonzero fraction of input states whose outputs also have low entropy. In order to define the tube we first construct a line segment $Y(\rho)$ pointing from a general state ρ toward the maximally mixed state I/d. The length of

the segment depends on a parameter γ, which satisfies $0 < \gamma < 1$:

$$Y(\rho) = \left\{ r\rho + (1-r)\frac{1}{d}I \,:\, \gamma \leq r \leq 1 \right\} \tag{2.36}$$

The tube at ρ is defined to be the set of states which lie within a small distance of the set $Y(\rho)$, and thus form a thickened line segment pointing from ρ toward the maximally mixed state. The definition of 'small' here depends on the size of the ball $B_d(n)$, and also on another numerical parameter t.

Definition 10. Let $\rho \in \mathcal{S}_d$, then the Tube *at* ρ is defined as

$$\text{Tube}\,(\rho) = \left\{ \theta \in \mathcal{S}_d \,:\, dist(\theta, Y(\rho)) \leq t\sqrt{\frac{d\log n}{n}} \right\}, \quad dist(\theta, Y(\rho)) = \inf_{\tau \in Y(\rho)} \|\theta - \tau\|_\infty \tag{2.37}$$

The next result shows that for a channel Φ in the typical set T, and for any state $\rho = \Phi^C(zz^*)$ in the image of Φ^C, there is a uniform lower bound for the probability that a randomly chosen state belongs to the tube at ρ. As explained before, this means that an output state $\Phi^C(zz^*)$ cannot be too isolated from the other output states.

Lemma 11. *For all $d \geq 3$ there is $\beta > 0$ such that for n sufficiently large, for all $t \geq b + 4$, and for all $\Phi \in T$ and $\rho \in \text{Im}(\Phi^C)$,*

$$\sigma_n\Big(z \,:\, \Phi^C(zz^*) \in \text{Tube}\,(\rho)\Big) \geq \beta\left(1-\gamma\right)^{n-1}$$

Lemma 11 will be proved in Section 2.5.5. The number β is given by the following expression:

$$\beta = \frac{1}{2} - (d^2+2)\left(1 - \frac{d\log d}{n}\right)^{n-1} \tag{2.38}$$

It can be easily seen that for all $d \geq 3$ the right side of (2.38) is positive for n sufficiently large.

2.4.2 Definition of the low-entropy events E

Define the set of channels whose minimum output entropy does not satisfy our requirements for a violation:

$$C_{d,n} = \left\{ \Phi \in \mathcal{R}_d(n) \,:\, S_{\min}(\Phi) \leq \log d - \frac{h}{d} \right\} \tag{2.39}$$

The goal is to show that for d, n and h sufficiently large we have $\mathcal{P}_{d,n}(C_{d,n}) < 1$, implying that $\mathcal{P}_{d,n}(C^c_{d,n}) > 0$, and thus that there exist random unitary channels with $S_{\min}(\Phi) > \log d - h/d$. The proof will hold for all h, d sufficiently large, and thus by taking $\log d \geq 2h$ this will provide a counterexample to additivity. The

method is to find useful upper and lower bounds for the probability of a particular event E in $\mathcal{R}_d(n) \times \mathcal{V}_n$. The event E is chosen to contain all the pairs (Φ, z) where $\Phi^C(zz^*)$ lies in a tube connected to a state of low entropy. This set of tubes is defined by

$$J = \bigcup_{\rho} \left\{ \text{Tube}\,(\rho) \,:\, S(\rho) \leq \log d - \frac{h}{d} \right\} \tag{2.40}$$

Then the main event of interest for us is the following subset of $\mathcal{R}_d(n) \times \mathcal{V}_n$:

$$E = \{(\Phi, z) \,:\, \Phi^C(zz^*) \in J\} = H^{-1}(J)$$

where H is the map defined in (2.33). The proof will proceed by proving upper and lower bounds for the probability of E, that is $(\mathcal{P}_{d,n} \times \sigma_n)(E)$. These bounds will hold for any $0 < \gamma < 1$; the parameter γ will be 'tuned' at the end in order to derive an estimate for the minimal size $h_{\min}$ needed for the counterexample. As noted the construction works for any values of the parameters b, t satisfying $b > \sqrt{3}$ and $t \geq b + 4$. The sizes of b and t do not play a crucial role, and they can be set to the values $b = 2$ and $t = 6$ without changing anything in the proof.

2.4.3 The upper bound for $Prob(E)$

Note that by Lemma 7,

$$\begin{aligned} (\mathcal{P}_{d,n} \times \sigma_n)(E) &= (\mathcal{P}_{d,n} \times \sigma_n)(H^{-1}(J)) \\ &= H^*(\mathcal{P}_{d,n} \times \sigma_n)(J) = G^*(\sigma_{dn})(J) \end{aligned} \tag{2.41}$$

Let ρ be a fixed state in the set of tubes J. Then by definition there is a state $\tau \in \mathcal{S}_d$ with low entropy such that ρ lies in the tube at τ. Thus for some r satisfying $\gamma \leq r \leq 1$

$$\left\| \rho - \left(r\tau + (1-r)\frac{1}{d} I \right) \right\|_\infty \leq t\sqrt{\frac{d \log n}{n}}, \quad S(\tau) \leq \log d - \frac{h}{d} \tag{2.42}$$

Letting q_i, p_i denote the eigenvalues of ρ, τ respectively, it follows that

$$q_i = rp_i + (1-r)\frac{1}{d} + \epsilon_i, \quad i = 1, \ldots, d \tag{2.43}$$

where p_i, ϵ_i satisfy

$$-\sum_i p_i \log p_i \leq \log d - \frac{h}{d}, \qquad \sum_{i=1}^{d} \epsilon_i = 0 \tag{2.44}$$

Weyl's perturbation theorem [8] and (2.42) imply that

$$|\epsilon_i| \leq t\sqrt{\frac{d \log n}{n}} \tag{2.45}$$

The entropy condition (2.44) can be written as

$$\sum_{i=1}^{d} p_i d \, \log(p_i d) = \sum_{i=1}^{d} \Big(p_i d \, \log(p_i d) - p_i d + 1 \Big) \geq h \tag{2.46}$$

Define the function

$$\boxed{f(x) = x \, \log x - x + 1} \tag{2.47}$$

Lemma 12.

$$\sup_{x \geq 0, \; \gamma \leq r \leq 1} \frac{f(x)}{f(rx + 1 - r)} = \frac{f(0)}{f(1-\gamma)} = \frac{1}{f(1-\gamma)} \tag{2.48}$$

Lemma 12 will be proved in Section 2.5.6. Recall (2.43) and define

$$z_i = q_i - \epsilon_i = r p_i + (1-r)\frac{1}{d} \tag{2.49}$$

Then Lemma 12 implies that for each $i = 1, \ldots, d$,

$$p_i d \, \log(p_i d) - p_i d + 1 = f(p_i d) \leq \frac{1}{f(1-\gamma)} \, f(z_i d) \tag{2.50}$$

Therefore from (2.46) it follows that

$$\sum_{i=1}^{d} \Big(z_i d \, \log(z_i d) - z_i d + 1 \Big) = \sum_{i=1}^{d} f(z_i d) \geq h f(1-\gamma) \tag{2.51}$$

We will use a standard bound for the difference between the entropies of z_i and q_i in terms of the l_1-norm of their difference [20]:

$$\Big| -\sum z_i \log z_i + \sum_i q_i \log q_i \Big| \leq \epsilon_m \, (\log d + \log \frac{1}{\epsilon_m}) \tag{2.52}$$

where

$$\epsilon_m = \sum_{i=1}^{d} |z_i - q_i| = \sum_{i=1}^{d} |\epsilon_i| \leq t \, d \sqrt{\frac{d \log n}{n}}$$

Define

$$\eta = d\,\epsilon_m\,(\log d + \log \frac{1}{\epsilon_m})$$

Note that for all d and t, $\epsilon_m \to 0$ as $n \to \infty$, and hence also $\eta \to 0$ as $n \to \infty$. From (2.52) and (2.51) we deduce

$$\sum_{i=1}^{d} f(q_i d) \geq hf(1-\gamma) - \eta \tag{2.53}$$

To summarize what we have shown so far: if $\rho \in J$ has eigenvalues $(q_1, \ldots, q_d)$ then (2.53) holds. Thus we may upper bound the probability (2.41) by the probability of the state ρ satisfying the inequality (2.53). Since this event depends only on the eigenvalues of ρ, we obtain

$$G^*(\sigma_{dn})(J) \leq \mu_{d,n}\Big\{q \,:\, \sum_{i=1}^{d} f(q_i d) > hf(1-\gamma) - \eta\Big\} \tag{2.54}$$

This probability is estimated using the bound (2.27): given a positive number $x \leq d \log d$, define

$$M_d(x) = \inf_{q \in \Delta_d} \Big\{ \sum_{i=1}^{d} F(q_i d) \,:\, \sum_{i=1}^{d} f(q_i d) \geq x \Big\} \tag{2.55}$$

where $F(x) = -\log x + x - 1$ as defined in (2.25). Then from (2.54) and (2.27) we deduce

$$(\mathcal{P}_{d,n} \times \sigma_n)(E) = G^*(\sigma_{dn})(J) \leq \exp\Big[d^2 \log n - \log(d-1)! - (n-d) M_d\Big(hf(1-\gamma) - \eta\Big)\Big] \tag{2.56}$$

The next Lemma gives a lower bound for $M_d(x)$ which is not optimal but is sufficient for our purposes.

Lemma 13. *The function $M_d(x)$ is increasing. Suppose that $2e^2 \leq x \leq d \log d$. Then*

$$M_d(x) \geq \log(x-1) - \log(2e^2 - 1) \tag{2.57}$$

Lemma 13 will be proved in Section 2.5.7. Applying (2.57) to (2.56) gives

$$(\mathcal{P}_{d,n} \times \sigma_n)(E) \leq \exp\Big[d^2 \log n - \log(d-1)! - (n-d) \log\Big(\frac{hf(1-\gamma) - \eta - 1}{2e^2 - 1}\Big)\Big] \tag{2.58}$$

where h is assumed to satisfy the bounds

$$2e^2 \leq hf(1-\gamma) - \eta \leq d \log d \tag{2.59}$$

2.4.4 The lower bound for $Prob(E)$

First we write

$$\begin{aligned}(\mathcal{P}_{d,n} \times \sigma_n)(E) &= \mathbb{E}_\Phi[\sigma_n(z : \Phi^C(zz^*) \in J)] \\ &\geq \mathbb{E}_\Phi[1_{C_{d,n}\cap T}\, \sigma_n(z : \Phi^C(zz^*) \in J)]\end{aligned}$$

where $\mathbb{E}_\Phi$ denotes expectation over $\mathcal{R}_d(n)$ with respect to the measure $\mathcal{P}_{d,n}$, and $1_{C_{d,n}\cap T}$ is the characteristic function of the event $C_{d,n} \cap T$. Given that $\Phi \in C_{d,n}$ there is a state $v \in \mathbb{C}^n$ such that

$$S(\Phi^C(vv^*)) \leq \log d - \frac{h}{d}$$

Since Tube $(\Phi^C(vv^*)) \subset J$ it follows that

$$(\mathcal{P}_{d,n} \times \sigma_n)(E) \geq \mathbb{E}_\Phi[1_{C_{d,n}\cap T}\, \sigma_n(z : \Phi^C(zz^*) \in \text{Tube}\,(\Phi^C(vv^*)))] \tag{2.60}$$

Applying Lemma 11 to (2.60) gives

$$\begin{aligned}(\mathcal{P}_{d,n} \times \sigma_n)(E) &\geq \beta\,(1-\gamma)^{n-1}\, \mathbb{E}_\Phi[1_{C_{d,n}\cap T}] && (2.61)\\ &= \beta\,(1-\gamma)^{n-1}\, \mathcal{P}_{d,n}(C_{d,n} \cap T) && (2.62)\\ &\geq \beta\,(1-\gamma)^{n-1}\, (\mathcal{P}_{d,n}(C_{d,n}) - \mathcal{P}_{d,n}(T^c)) && (2.63)\end{aligned}$$

2.4.5 Combining the bounds for $Prob(E)$ and finishing the proof

Putting together the upper and lower bounds for $(\mathcal{P}_{d,n} \times \sigma_n)(E)$ and using Lemma 9 produces the following bound: for all $d \geq 3$, for all $b > \sqrt{3}$ and $t \geq b + 4$, for all $0 < \gamma < 1$, for h, d satisfying (2.59), and for n sufficiently large

$$\begin{aligned}\mathcal{P}_{d,n}(C_{d,n}) &\leq \mathcal{P}_{d,n}(T^c) + \frac{1}{\beta}\left(\frac{1}{1-\gamma}\right)^{n-1} (\mathcal{P}_{d,n} \times \sigma_n)(E) \\ &\leq \frac{2\,d}{(d-1)!}\, \exp[-\alpha d^2\, \log n] \\ &+ \frac{1}{\beta}\left(\frac{1}{1-\gamma}\right)^{n-1} \exp[d^2 \log n - \log(d-1)! - (n-d)\log \tilde{h}] \\ &= \frac{2\,d}{(d-1)!}\, \exp[-\alpha d^2\, \log n] \\ &+ \frac{1-\gamma}{\beta(d-1)!}\, \exp[d^2 \log n + d \log \tilde{h} - n \log(1-\gamma)\tilde{h}] && (2.64)\end{aligned}$$

where $\tilde{h} = (hf(1-\gamma) - \eta - 1)/(2e^2 - 1)$. Define

$$h_{\min} = \frac{2e^2 - \gamma}{(1-\gamma)f(1-\gamma)} \tag{2.65}$$

(note that $h_{\min}$ satisfies the lower bound in (2.59)). As $n \to \infty$ the parameter η approaches zero, and therefore for $h > h_{\min}$ the second term on the right side of (2.64) is controlled by the factor

$$\exp\left[-n \log \frac{(1-\gamma)(hf(1-\gamma)-1)}{2e^2-1}\right] = \left(\frac{hf(1-\gamma)-1}{h_{\min}f(1-\gamma)-1}\right)^{-n} \tag{2.66}$$

The first factor on the right side of (2.64) approaches zero as $n \to \infty$, therefore (2.66) implies that for $h > h_{\min}$,

$$\mathcal{P}_{d,n}(C_{d,n}) \to 0 \quad \text{as } n \to \infty$$

$\boxed{\text{Summary and conclusion}}$ We have shown that for any $0 < \gamma < 1$, for $h > h_{\min}$ as defined in (2.65), for any $b > \sqrt{3}$ and $t \geq b + 4$, for any $d \geq 3$ satisfying $d \log d > hf(1-\gamma)$ (this comes from the second inequality in (2.59)), there is $N < \infty$ such that for all $n \geq N$ we have $\mathcal{P}_{d,n}(C_{d,n}) < 1$. In this case we also have $\mathcal{P}_{d,n}(C_{d,n}^c) > 0$, and thus a guarantee that the set $C_{d,n}^c$ is non-empty. Referring to (2.39), this means that there exists a random unitary channel Φ such that

$$S_{\min}(\Phi) > \log d - \frac{h}{d}$$

2.4.6 Optimizing the bounds for $Prob(E)$ and the proof of Proposition 3

First consider the value $h_{\min}$ defined in (2.65). Varying γ shows that the right side achieves its minimum value at $\gamma = 0.72$. In order to achieve a counterexample we need $\log d \geq 2h$, so this implies the existence of counterexamples for all $d \geq d_0$ with

$$d_0 = \exp[2h_{\min} + 1] \simeq \exp[276] \tag{2.67}$$

In order to get a better estimate of $d_{\min}$, we return to the bound (2.56) and look for the smallest value of d satisfying

$$M_d\Big(\frac{f(1-\gamma)}{2} \log d\Big) + \log(1-\gamma) > 0 \tag{2.68}$$

For n sufficiently large this will yield a counterexample. This is a straightforward numerical problem: for

each γ we find the smallest d so that

$$-\log(1-\gamma) < \inf_{z>1}\{-\log z - (d-1)\log\frac{d-z}{d-1} : \\ z\log z + (d-z)\log\frac{d-z}{d-1} = \frac{f(1-\gamma)}{2}\log d\} \tag{2.69}$$

and then minimize over γ. The solution occurs at $\gamma = 0.762$ and yields $d_0 = 38578$. This also proves the first statement in Proposition 3.

For the second statement we estimate the smallest value of n which yields $\mathcal{P}_{d,n}(C_{d,n}) < 1$. Using the values $b = 2$, $t = 6$, $\gamma = 0.762$, and with $d = 50,000$, crude numerical estimates show that we can achieve this with $n = d^7$. This proves the second statement in Proposition 3.

For the third statement, we note from Lemma 1 that for any random unitary channel Φ

$$\Delta S_{\max} \geq 2S_{\min}(\Phi) - 2\log d + \frac{\log d}{d}$$

Thus for every $\Phi \in C^c_{d,n}$ we have

$$\Delta S_{\max} \geq \frac{\log d - 2h}{d} \tag{2.70}$$

For a fixed value h, the right side of (2.70) achieves its maximum value when $d = [\exp(2h+1)]$, and this maximum value is $1/d$. Numerical calculation shows that we can achieve $M_d(f(1-\gamma)h) + \log(1-\gamma) > 0$ using the values $\gamma = 0.762$, $h = \log(38590)/2$ and $d = [\exp(2h+1)]$, and then $1/d$ yields the lower bound for $\Delta S_{\max}$ stated in Proposition 3.

2.5 Proofs of Lemmas

2.5.1 Proof of Lemma 1

First, note that for any unit vectors $\{|\psi_k\rangle\}$ and probability distribution $\{p_k\}$,

$$S\left(\sum_k p_k|\psi_k\rangle\langle\psi_k|\right) \leq -\sum_k p_k \log p_k.$$

Let $|\hat{\psi}\rangle$ be the maximally entangled state. Then

$$(\Phi \otimes \overline{\Phi})(|\hat{\psi}\rangle\langle\hat{\psi}|) = \sum_{i,j=1}^{d} w_i w_j \; U_i \otimes \overline{U_j} |\hat{\psi}\rangle\langle\hat{\psi}| U_i^* \otimes U_j^T \tag{2.71}$$

$$= \left(\sum_{i=1}^{d} w_i^2\right) |\hat{\psi}\rangle\langle\hat{\psi}| + \sum_{i\neq j} w_i w_j \; U_i \otimes \overline{U_j} |\hat{\psi}\rangle\langle\hat{\psi}| U_i^* \otimes U_j^T \tag{2.72}$$

where we used the identity $U_i \otimes \overline{U_i} |\hat{\psi}\rangle\langle\hat{\psi}| U_i^* \otimes U_i^T = |\hat{\psi}\rangle\langle\hat{\psi}|$ for all i. Hence,

$$S\left((\Phi \otimes \overline{\Phi})(|\hat{\psi}\rangle\langle\hat{\psi}|)\right) \leq -\left(\sum_{i=1}^{d} w_i^2\right) \log\left(\sum_{i=1}^{d} w_i^2\right) - \sum_{i\neq j} w_i w_j \log(w_i w_j). \tag{2.73}$$

Write $p = \sum_{i=1}^{d} w_i^2$ and then $\sum_{i\neq j} w_i w_j = 1 - p$. Hence

$$S\left((\Phi \otimes \overline{\Phi})(|\hat{\psi}\rangle\langle\hat{\psi}|)\right) \leq -p \log p + \sup\left\{ -\sum_{k=1}^{d^2-d} v_k \log v_k \; : \; v_k \geq 0, \; \sum_{k=1}^{d^2-d} v_k = 1 - p \right\} \tag{2.74}$$

The supremum on the right side of (2.74) is achieved with $v_k = (1-p)/(d^2 - d)$ for all k, hence

$$S\left((\Phi \otimes \overline{\Phi})(|\hat{\psi}\rangle\langle\hat{\psi}|)\right) \leq h(p) = -p \log p - (1-p) \log\left(\frac{1-p}{d^2 - d}\right),$$

where $1/d \leq p \leq 1$. However,

$$h'(p) = -\log p + \log(1-p) - \log(d^2 - d) = \log\left(\left(\frac{1}{p} - 1\right) \frac{1}{d(d-1)}\right) \leq \log\left(\frac{1}{d}\right) < 0$$

for all d. This implies that the above upper bound $h(p)$ is maximized when $p = 1/d$ and the maximum is

$$\frac{1}{d} \log d + \left(1 - \frac{1}{d}\right) \log(d^2) = 2 \log d - \frac{1}{d} \log d.$$

2.5.2 Proof of Lemma 4

By dropping the terms $(w_i - w_j)^2$ in (2.23) we get

$$\mu_{d,n}(A) \leq Z(n,d)^{-1} \int_A \prod_{i=1}^{d} w_i^{n-d} \, \delta\left(\sum_{i=1}^{d} w_i - 1\right) [dw] \tag{2.75}$$

Applying (2.24) to (2.75) leads to

$$\mu_{d,n}(A) \leq \int_A \exp[d^2 \log n + (n-d) \sum_{i=1}^{d} \log(dw_i)]\, \delta\Big(\sum_{i=1}^{d} w_i - 1 \Big)\, [dw] \tag{2.76}$$

Noting that

$$\sum_{i=1}^{d} F(dw_i) = -\sum_{i=1}^{d} \log(dw_i)$$

the result follows.

2.5.3 Proof of Lemma 7

For any event $A \subset \mathcal{M}_d$,

$$\begin{aligned} H^*(\mathcal{P}_{d,n} \times \sigma_n)(A) &= (\mathcal{P}_{d,n} \times \sigma_n)(H^{-1}(A)) \\ &= \int d\sigma_n(z)\, \mathcal{P}_{d,n}\Big\{ \Phi \,:\, \Phi^C(zz^*) \in A \Big\} \end{aligned} \tag{2.77}$$

A random unitary channel Φ is determined by the coefficients $\{w_1, \ldots, w_d\}$ and the unitary matrices $\{U_1, \ldots, U_d\}$. Given a unitary matrix V define the transformation $T_V : \mathcal{R}_d(n) \to \mathcal{R}_d(n)$ by

$$T_V \,:\, \{w_1, \ldots, w_d; U_1, \ldots, U_d\} \to \{w_1, \ldots, w_d; U_1 V, \ldots, U_d V\} \tag{2.78}$$

The measure $\mathcal{P}_{d,n} = \nu_{d,n} \times H_n \times \cdots \times H_n$ contains the product of d independent copies of Haar measure H_n on the group $\mathcal{U}_n$. Since Haar measure is invariant under group multiplication, for any event $C \subset \mathcal{R}_d(n)$ we have

$$\begin{aligned} \mathcal{P}_{d,n}(C) &= \int_C d\nu_{d,n}(w_1, \ldots, w_d)\, dH_n(U_1) \cdots dH_n(U_n) \\ &= \int_{T_V(C)} d\nu_{d,n}(w_1, \ldots, w_d)\, dH_n(U_1) \cdots dH_n(U_n) \\ &= \mathcal{P}_{d,n}(T_V(C)) \end{aligned} \tag{2.79}$$

Thus in particular for any $z \in \mathcal{V}_n$,

$$\mathcal{P}_{d,n}\Big\{ \Phi \,:\, \Phi^C(zz^*) \in A \Big\} = \mathcal{P}_{d,n}\Big\{ \Phi \,:\, \Phi^C(Vz(Vz)^*) \in A \Big\} \tag{2.80}$$

Since $\mathcal{U}_n$ acts transitively on $\mathcal{V}_n$, (2.80) shows that the probability is independent of z. Hence from (2.77) we obtain that for any fixed $z_0 \in \mathcal{V}_n$,

$$H^*(\mathcal{P}_{d,n} \times \sigma_n)(A) = \mathcal{P}_{d,n}\Big\{\Phi \,:\, \Phi^C(z_0 z_0^*) \in A\Big\} \tag{2.81}$$

For a given channel Φ the $d \times d$ matrix $\Phi^C(z_0 z_0^*)$ can be written in terms of a $n \times d$ matrix $K(\Phi)$ as follows:

$$\Phi^C(z_0 z_0^*) = K(\Phi)^* K(\Phi), \quad K(\Phi) = \begin{pmatrix} \sqrt{w_1} v_1 & \cdots & \sqrt{w_d} v_d \end{pmatrix} \tag{2.82}$$

where for $i = 1, \ldots, d$

$$v_i = \overline{U_i z_0} \tag{2.83}$$

Thus (2.81) can be written as

$$H^*(\mathcal{P}_{d,n} \times \sigma_n)(A) = \mathcal{P}_{d,n}\Big\{\Phi \,:\, K(\Phi)^* K(\Phi) \in A\Big\} \tag{2.84}$$

Recall from (2.18) that $G(z) = M(z)^* M(z)$ where $z \in \mathcal{V}_{nd}$ and where the $n \times d$ matrix $M(z)$ has entries

$$M_{ij}(z) = z_{(i-1)d+j}, \quad i = 1, \ldots n, \; j = 1, \ldots, d \tag{2.85}$$

It follows that for any event $A \subset \mathcal{M}_d$,

$$G^*(\sigma_{dn})(A) = \sigma_{dn}\Big\{z \,:\, M(z)^* M(z) \in A\Big\} \tag{2.86}$$

We wish to prove that $H^*(\mathcal{P}_{d,n} \times \sigma_n)(A) = G^*(\sigma_{dn})(A)$ for every event $A \subset \mathcal{M}_d$. Comparing (2.86) and (2.84), it is sufficient to show that the $n \times d$ matrices $K(\Phi)$ and $M(z)$ have the same distribution. We will do this by showing that the columns of $K(\Phi)$ and $M(z)$ have the same joint distribution. Before showing the result, we need the following observation on a normally distributed vector.

Let $Z_1, \ldots, Z_m$ be IID complex valued normal random variables with mean zero and variance two. Let $R = \sqrt{\sum_{i=1}^m |Z_i|^2}$ and define the vector

$$\xi = \frac{1}{R} \begin{pmatrix} Z_1 \\ \vdots \\ Z_m \end{pmatrix} \tag{2.87}$$

Then R, ξ are independent, ξ is a random pure state in $\mathcal{V}_m$, and R has density

$$P(r) \propto e^{-r^2/2}\, r^{2m-1} \tag{2.88}$$

This result may be easily seen by transforming the joint density for $Z_1, \ldots, Z_m$ to polar coordinates:

$$(2\pi)^{-m} \prod_{i=1}^{m} e^{-|z_i|^2/2}\, d^2 z_i = \pi^{-m} e^{-r^2/2}\, r^{2m-1}\, dr\, d\Omega \tag{2.89}$$

where $d\Omega$ is the uniform measure on S^{2m-1}.

We look at $M(z)$ first. Let $\{Z_{ij}\}$ $(i = 1, \ldots, n,\ j = 1, \ldots, d)$ be a collection of IID complex valued normal random variables with mean zero and variance two, arranged into a $n \times d$ matrix Z as in (2.85). Applying the previous observation to Z and also to each column of Z yields

$$Z = R\, M = \begin{pmatrix} R_1\, \xi_1 & \cdots & R_d\, \xi_d \end{pmatrix} \tag{2.90}$$

Here $\{\xi_1, \ldots, \xi_d\}$ are IID random unit vectors in $\mathcal{V}_n$, and M is a random unit vector in $\mathcal{V}_{nd}$. The vectors $\{\xi_1, \ldots, \xi_d\}$ are independent of the numbers $R_1, \ldots, R_d$. Also $R^2 = R_1^2 + \cdots + R_d^2$, hence $\{\xi_1, \ldots, \xi_d\}$ are also independent of R. Dividing by R, $M(z)$, a random unit vector in $\mathcal{V}_{nd}$, can be reconstructed as

$$M(z) = \begin{pmatrix} \sqrt{Y_1}\, \xi_1 & \cdots & \sqrt{Y_d}\, \xi_d \end{pmatrix}, \quad Y_i = \frac{R_i^2}{R^2} \tag{2.91}$$

Note that $Y_1, \ldots, Y_d$ have the joint distribution $\nu_{d,n}$, and are independent of $\{\xi_1, \ldots, \xi_d\}$.

Next, turning into $K(\Phi)$, recall that

$$K(\Phi) = \begin{pmatrix} \sqrt{w_1} v_1 & \cdots & \sqrt{w_d} v_d \end{pmatrix}$$

where $v_i = \overline{U_i z_0}$. Since the unitaries U_i are independently and uniformly selected (this is part of the definition of the measure $\mathcal{P}_{d,n}$), it follows that the vectors $\{v_i\}$ are IID random unit vectors in $\mathcal{V}_n$. Furthermore the coefficients $\{w_i\}$ have the joint distribution $\nu_{d,n}$. This verifies our claim.

2.5.4 Proof of Lemma 9

Define the following subset of $\mathcal{R}_d(n) \times \mathcal{V}_n$:

$$K = \{(\Phi, z) \,:\, \Phi^C(zz^*) \notin B_d(n)\} = H^{-1}(B_d(n)^c) \tag{2.92}$$

where the map H was defined in (2.33). Then

$$\begin{aligned}(\mathcal{P}_{d,n} \times \sigma_n)(K) &= \mathbb{E}_\Phi[\sigma_n(z : \Phi^C(zz^*) \notin B_d(n))] \\ &\geq \mathbb{E}_\Phi[1_{T^c}\, \sigma_n(z : \Phi^C(zz^*) \notin B_d(n))] \end{aligned} \tag{2.93}$$

where $\mathbb{E}_\Phi$ denotes expectation over $\mathcal{R}_d(n)$ with respect to the measure $\mathcal{P}_{d,n}$, and 1_{T^c} is the characteristic function of the event T^c. Note that if $\Phi \in T^c$ then $\sigma_n(z : \Phi^C(zz^*) \notin B_d(n)) \geq 1/2$, hence

$$(\mathcal{P}_{d,n} \times \sigma_n)(K) \geq \frac{1}{2}\,\mathbb{E}_\Phi[1_{T^c}] = \frac{1}{2}\,\mathcal{P}_{d,n}(T^c) \tag{2.94}$$

Furthermore from Lemma 7 it follows that

$$\begin{aligned}(\mathcal{P}_{d,n} \times \sigma_n)(K) &= (\mathcal{P}_{d,n} \times \sigma_n)(H^{-1}(B_d(n)^c)) \\ &= H^*(\mathcal{P}_{d,n} \times \sigma_n)(B_d(n)^c) \\ &= G^*(\sigma_{dn})(B_d(n)^c) \end{aligned} \tag{2.95}$$

Combining these bounds shows that

$$\begin{aligned}\mathcal{P}_{d,n}(T^c) &\leq 2\,G^*(\sigma_{dn})(B_d(n)^c) \\ &= 2\,\mu_{d,n}\left\{(q_1,\ldots,q_d) : \left|q_i - \frac{1}{d}\right| > b\sqrt{\frac{\log n}{n}} \text{ some } i = 1,\ldots,d\right\} \\ &= 2\mu_{d,n}\Big(\bigcup_{i=1}^{d} L_i\Big) \end{aligned} \tag{2.96}$$

where the events L_i are defined by $L_i = \{(q_1,\ldots,q_d) : |q_i - 1/d| > b\sqrt{\log n/n}\}$. Thus we have

$$\mathcal{P}_{d,n}(T^c) \leq 2\sum_{i=1}^{d} \mu_{d,n}(L_i) = 2\,d\,\mu_{d,n}(L_i) \tag{2.97}$$

We use the bound (2.31) of Corollary 6 with $t = b\sqrt{\log n/n}$ to estimate $\mu_{d,n}(L_i)$. In addition we assume that n is large enough so that

$$d\,t = d\,b\sqrt{\frac{\log n}{n}} \leq \frac{1}{2}$$

and hence

$$\frac{n-d}{2}d^2t^2 - \frac{n-d}{3}d^3t^3 \geq \frac{n-d}{3}d^2t^2$$

Thus (2.96) gives

$$\begin{aligned}\mathcal{P}_{d,n}(T^c) &\leq 2\,d\,\exp\Big[d^2\log n - \log(d-1)! - \frac{n-d}{3}d^2b^2\frac{\log n}{n}\Big] \\ &= \frac{2\,d}{(d-1)!}\,\exp\Big[-d^2\log n\Big(\frac{b^2(n-d)}{3n}-1\Big)\Big] \end{aligned} \tag{2.98}$$

2.5.5 Proof of Lemma 11

This result relies on several properties of random states. We will switch to Dirac bra and ket notation throughout this Section, as it lends itself well to the arguments used in the proof. To set up the notation, let $|\psi\rangle$ be a fixed state in $\mathcal{V}_n$, and let $|\theta\rangle$ be a random pure state in $\mathcal{V}_n$, with probability distribution σ_n. Without loss of generality we assume that a basis is chosen so that $|\psi\rangle = (1,0,\ldots,0)^T$. We write $x = \langle\psi|\theta\rangle$, and let $|\phi\rangle$ be the state orthogonal to $|\psi\rangle$ such that

$$|\theta\rangle = x\,|\psi\rangle + \sqrt{1-|x|^2}\,|\phi\rangle \tag{2.99}$$

Thus $|\phi\rangle$ is also a random state, defined by its relation to the uniformly random state $|\theta\rangle$ in (2.99). The following results are proved in Appendix B.

Proposition 14. *x and $|\phi\rangle$ are independent. $|\phi\rangle$ is a random vector in $\mathcal{V}_{n-1}$ with distribution σ_{n-1}. For all $0 \leq t \leq 1$*

$$\sigma_n\{|\theta\rangle \,:\, |\,\langle\psi|\theta\rangle\,| = |x| > t\} = (1-t^2)^{n-1} \tag{2.100}$$

Proposition 14 implies that as $n \to \infty$ the overlap $x = \langle\psi|\theta\rangle$ becomes concentrated around zero. In other words, with high probability a randomly chosen state will be almost orthogonal to any fixed state. As a consequence, from (2.99) it follows that $|\phi\rangle$ will be almost equal to $|\theta\rangle$. This statement is made precise by noting that

$$\||\theta\rangle - |\phi\rangle\|_2 \leq \sqrt{2}|\,\langle\psi|\theta\rangle\,|$$

Then (2.100) immediately implies that

$$\sigma_n(|\theta\rangle \,:\, \||\theta\rangle - |\phi\rangle\|_2 > t) \leq \left(1-\frac{t^2}{2}\right)^{n-1} \tag{2.101}$$

The second property relies on the particular form of the random unitary channel, or more precisely on the form of the complementary channel Φ^C. Roughly, this property says that for any fixed random unitary channel Φ and random state $|\theta\rangle$, with high probability the norm of the matrix $\Phi^C(|\theta\rangle\langle\psi|)$ is small, and

approaches zero as $n \to \infty$. We will prove the following bound: for any $\Phi \in \mathcal{R}_d(n)$, and for all $0 \le t \le 1$,

$$\sigma_n(|\theta\rangle \,:\, \|\Phi^C(|\theta\rangle\langle\psi|)\|_2 > t) \le d^2\,(1-t^2)^{n-1} \tag{2.102}$$

As a first step toward deriving (2.102), note that for any states $|u\rangle$ and $|v\rangle$,

$$\|\Phi^C(|u\rangle\langle v|)\|_2 = \left(\sum_{k,l=1}^{d} w_k w_l |\langle v|U_l^* U_k|u\rangle|^2\right)^{\frac{1}{2}} \le \max_{k,l} |\langle v|U_l^* U_k|u\rangle|, \tag{2.103}$$

In particular this implies that

$$\|\Phi^C(|u\rangle\langle v|)\|_2 \le \||u\rangle\|_2\,\||v\rangle\|_2 \tag{2.104}$$

To derive (2.102) we apply (2.103) with $u = \theta$ and $v = \psi$ and deduce that

$$\begin{aligned} \sigma_n(|\theta\rangle \,:\, \|\Phi^C(|\theta\rangle\langle\psi|)\|_2 > t) &\le \sigma_n(|\theta\rangle \,:\, \max_{k,l} |\langle\psi|U_l^* U_k|\theta\rangle| > t) \\ &\le d^2\,\sigma_n(|\theta\rangle \,:\, |\langle\psi|U_l^* U_k|\theta\rangle| > t) \\ &= d^2\,(1-t^2)^{n-1} \end{aligned} \tag{2.105}$$

where the last equality follows from (2.100).

With these ingredients in place the proof of Lemma 11 can proceed. By assumption Φ is a random unitary channel belonging to the typical set T, and $\rho = \Phi^C(|\psi\rangle\langle\psi|)$ is some state in $\mathrm{Im}(\Phi^C)$. Let $|\theta\rangle$ be a random input state, then as in (2.99) we write

$$|\theta\rangle = x\,|\psi\rangle + \sqrt{1-|x|^2}\,|\phi\rangle$$

It follows that

$$|\theta\rangle\langle\theta| = |x|^2\,|\psi\rangle\langle\psi| + (1-|x|^2)\,|\phi\rangle\langle\phi| + \sqrt{1-|x|^2}\,(x\,|\phi\rangle\langle\phi| + \overline{x}\,|\phi\rangle\langle\psi|) \tag{2.106}$$

Write $r = |x|^2$, then (2.106) yields

$$\begin{aligned} \Phi^C(|\theta\rangle\langle\theta|) \;&-\left(r\Phi^C(|\psi\rangle\langle\psi|) + (1-r)\frac{1}{d}I\right) \\ &= (1-r)\left(\Phi^C(|\phi\rangle\langle\phi|) - \frac{1}{d}I\right) + \sqrt{r(1-r)}\Phi^C\left(e^{i\xi}\,|\psi\rangle\langle\phi| + e^{-i\xi}\,|\phi\rangle\langle\psi|\right) \end{aligned} \tag{2.107}$$

where ξ is the phase of x. Since $r \leq 1$ this implies

$$\left\| \Phi^C(|\theta\rangle\langle\theta|) - \left(r\Phi^C(|\psi\rangle\langle\psi|) + (1-r)\frac{1}{d}I \right) \right\|_\infty \leq \left\| \Phi^C(|\phi\rangle\langle\phi|) - \frac{1}{d}I \right\|_\infty + \left\| \Phi^C(|\psi\rangle\langle\phi|) \right\|_\infty \tag{2.108}$$

Referring to the definition (2.37) of Tube (ρ), recall that $\Phi^C(|\theta\rangle\langle\theta|)$ belongs to Tube (ρ) if and only if for some r satisfying $\gamma \leq r \leq 1$,

$$\left\| \Phi^C(|\theta\rangle\langle\theta|) - \left(r\Phi^C(|\psi\rangle\langle\psi|) + (1-r)\frac{1}{d}I \right) \right\|_\infty \leq t\sqrt{\frac{d\log n}{n}} \tag{2.109}$$

(the set $Y(\rho)$ defined in (2.36) is closed so the infimum in (2.37) is achieved). Define the following three events in $\mathcal{V}_n$:

$$A_1 = \{|\theta\rangle \,:\, r = |\langle\psi|\theta\rangle|^2 \geq \gamma\} \tag{2.110}$$

$$A_2 = \left\{ |\theta\rangle \,:\, \left\| \Phi^C(|\phi\rangle\langle\phi|) - \frac{1}{d}I \right\|_\infty \leq 2\sqrt{2}\sqrt{\frac{d\log d}{n}} + b\sqrt{\frac{d\log n}{n}} \right\} \tag{2.111}$$

$$A_3 = \left\{ |\theta\rangle \,:\, \left\| \Phi^C(|\psi\rangle\langle\phi|) \right\|_\infty \leq (1+\sqrt{2})\sqrt{\frac{d\log d}{n}} \right\} \tag{2.112}$$

Assume that $d^2 \leq n$ and then since $t \geq b+4$ it follows from (2.108) and (2.109) that

$$A_1 \cap A_2 \cap A_3 \subset \{|\theta\rangle \,:\, \Phi^C(|\theta\rangle\langle\theta|) \in \text{Tube}\,(\rho)\}$$

Note that $n \geq d^2$ Furthermore by Proposition 14, A_1 is independent of A_2 and A_3, hence

$$\sigma_n(|\theta\rangle : \Phi^C(|\theta\rangle\langle\theta|) \in \text{Tube}\,(\rho)) \geq \sigma_n(A_1 \cap A_2 \cap A_3) = \sigma_n(A_1)\,\sigma_n(A_2 \cap A_3) \tag{2.113}$$

Proposition 14 immediately yields

$$\sigma_n(A_1) = (1-\gamma)^{n-1}$$

From (2.113) this gives

$$\sigma_n(|\theta\rangle : \Phi^C(|\theta\rangle\langle\theta|) \in \text{Tube}\,(\rho)) \geq (1-\gamma)^{n-1}\,(1 - \sigma_n(A_2^c) - \sigma_n(A_3^c)) \tag{2.114}$$

In order to bound $\sigma_n(A_3^c)$ we first use (2.104) to deduce

$$\|\Phi^C(|\psi\rangle\langle\phi|)\|_\infty \leq \|\Phi^C(|\psi\rangle\langle\phi|)\|_2 \leq \|\Phi^C(|\psi\rangle\langle\theta|)\|_2 + \||\theta\rangle - |\phi\rangle\|_2 \tag{2.115}$$

Thus

$$\begin{aligned}
\sigma_n(A_3^c) &= \sigma_n\left\{|\theta\rangle \,:\, \left\|\Phi^C(|\psi\rangle\langle\phi|)\right\|_\infty > (1+\sqrt{2})\sqrt{\frac{d\log d}{n}}\right\} \\
&\le \sigma_n\left\{|\theta\rangle \,:\, \|\Phi^C(|\psi\rangle\langle\theta|)\|_2 + \||\theta\rangle - |\phi\rangle\|_2 > (1+\sqrt{2})\sqrt{\frac{d\log d}{n}}\right\} \\
&\le \sigma_n\left\{|\theta\rangle \,:\, \|\Phi^C(|\psi\rangle\langle\theta|)\|_2 > \sqrt{\frac{d\log d}{n}}\right\} + \sigma_n\left\{|\theta\rangle \,:\, \||\theta\rangle - |\phi\rangle\|_2 > \sqrt{2}\sqrt{\frac{d\log d}{n}}\right\} \\
&\le (d^2+1)\left(1 - \frac{d\log d}{n}\right)^{n-1}
\end{aligned} \tag{2.116}$$

where the last inequality follows from (2.105) and (2.101).

Turning now to $\sigma_n(A_2^c)$, note first that

$$\begin{aligned}
\left\|\Phi^C(|\phi\rangle\langle\phi| - \frac{1}{d}I\right\|_\infty &\le \left\|\Phi^C(|\phi\rangle\langle\phi| - \Phi^C(|\theta\rangle\langle\theta|)\right\|_\infty + \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty \\
&\le \left\|\Phi^C(|\phi\rangle\langle\phi| - \Phi^C(|\theta\rangle\langle\theta|)\right\|_2 + \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty \\
&\le 2\,\||\theta\rangle - |\phi\rangle\|_2 + \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty
\end{aligned} \tag{2.117}$$

where we used (2.104) for the last inequality. As in (2.116) this gives

$$\begin{aligned}
\sigma_n(A_2^c) &= \sigma_n\left\{|\theta\rangle \,:\, \left\|\Phi^C(|\phi\rangle\langle\phi| - \frac{1}{d}I\right\|_2 > 2\sqrt{2}\sqrt{\frac{d\log d}{n}} + b\sqrt{\frac{d\log n}{n}}\right\} \\
&\le \sigma_n\left\{|\theta\rangle \,:\, 2\,\||\theta\rangle - |\phi\rangle\|_2 > 2\sqrt{2}\sqrt{\frac{d\log d}{n}}\right\} \\
&\qquad + \sigma_n\left\{|\theta\rangle \,:\, \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty > b\sqrt{\frac{d\log n}{n}}\right\} \\
&\le \left(1 - \frac{d\log d}{n}\right)^{n-1} + \sigma_n\left\{|\theta\rangle \,:\, \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty > b\sqrt{\frac{d\log n}{n}}\right\}
\end{aligned} \tag{2.118}$$

where we used (2.101) for the last inequality. By assumption $\Phi \in T$, and therefore there is a set of input states L with $\sigma_n(L) \ge 1/2$ such that

$$|\theta\rangle \in L \Rightarrow \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty \le b\sqrt{\frac{d\log n}{n}} \tag{2.119}$$

Thus

$$\sigma_n\left\{|\theta\rangle \,:\, \left\|\Phi^C(|\theta\rangle\langle\theta| - \frac{1}{d}I\right\|_\infty > b\sqrt{\frac{d\log n}{n}}\right\} \le \sigma_n(L^c) \le \frac{1}{2} \tag{2.120}$$

Putting together the bounds (2.114), (2.116), (2.118) and (2.120) we get

$$\begin{aligned}\sigma_n(\Phi^C(|\theta\rangle\langle\theta|) \in \mathrm{Tube}\,(\rho)) &\geq (1-\gamma)^{n-1}\left(1-\sigma_n(A_2^c)-\sigma_n(A_3^c)\right)\\ &\geq (1-\gamma)^{n-1}\left(1-\left(1-\frac{d\log d}{n}\right)^{n-1}-\frac{1}{2}-(d^2+1)\left(1-\frac{d\log d}{n}\right)^{n-1}\right)\\ &= (1-\gamma)^{n-1}\left(\frac{1}{2}-(d^2+2)\left(1-\frac{d\log d}{n}\right)^{n-1}\right) \qquad (2.121)\end{aligned}$$

This completes the proof, with

$$\beta = \left(\frac{1}{2}-(d^2+2)\left(1-\frac{d\log d}{n}\right)^{n-1}\right)$$

2.5.6 Proof of Lemma 12

It is clear that $f(rx+1-r)$ is monotone increasing in r, and therefore

$$\sup_{x\geq 0}\sup_{\gamma\leq r\leq 1}\frac{f(x)}{f(rx+1-r)} = \sup_{x\geq 0}\frac{f(x)}{f(\gamma x+1-\gamma)} \qquad (2.122)$$

The function $f(x)\,f(\gamma x+1-\gamma)^{-1}$ is analytic and decreasing at $x=1$ for $\gamma<1$. Thus either the supremum in (2.122) is achieved at $x=0$ or else there is a critical point of the function $f(x)\,f(\gamma x+1-\gamma)^{-1}$ in the interval $(0,\infty)$. In order to rule out the second possibility, we introduce a Lagrange multiplier and define the function

$$h(x,y,\beta) = \log f(x) - \log f(y) - \beta(\gamma x+1-\gamma-y) \qquad (2.123)$$

To find the critical points of h we solve

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial y} = \frac{\partial h}{\partial \beta} = 0$$

Solving for β leads to

$$\frac{f'(x)}{f(x)} = \gamma\,\frac{f'(y)}{f(y)}$$

Since $y-1=\gamma(x-1)$ this is equivalent to

$$\frac{(x-1)\log x}{x\log x - x + 1} = \frac{(y-1)\log y}{y\log y - y + 1} \qquad (2.124)$$

Direct computation shows that

$$\frac{d}{dx}\left(\frac{(x-1)\log x}{x\log x - x + 1}\right) = f(x)^{-2}\left((\log x)^2 - \frac{(x-1)^2}{x}\right)$$
$$= f(x)^{-2}(\log x)^2\left(1 - \left(\frac{x^{1/2} - x^{-1/2}}{\log x}\right)^2\right) \tag{2.125}$$

Furthermore, the function $x^{1/2} - x^{-1/2} - \log x$ is monotone increasing for all $x > 0$, and thus $x^{1/2} - x^{-1/2} > \log x$ for $x > 1$. Thus for $x \geq 1$ the derivative (2.125) is negative, and therefore (2.124) has no solution with $x > 1$. Similarly $x^{1/2} - x^{-1/2} < \log x$ for $0 < x < 1$, and hence again (2.125) is negative for $0 < x < 1$. So there are no solutions of (2.124) except $x = y = 1$. Therefore (2.123) has no critical points except $x = y = 1$, and thus the function $f(x)f(\gamma x + 1 - \gamma)^{-1}$ achieves its supremum at $x = 0$.

2.5.7 Proof of Lemma 13

Suppose first that $0 < h < d\log d$. Recall the definition

$$M_d(h) = \inf_{q\in\Delta_d}\left\{\sum_{i=1}^d F(q_i d) \,:\, \sum_{i=1}^d f(q_i d) \geq h\right\} \tag{2.126}$$

where $F(x) = -\log x + x - 1$ and $f(x) = x\log x - x + 1$. Letting $x_i = q_i d$ we have

$$M_d(h) = \inf_{x_i\geq 0}\left\{\sum_{i=1}^d F(x_i) \,:\, \sum_{i=1}^d f(x_i) \geq h, \qquad \sum_{i=1}^d x_i = d\right\} \tag{2.127}$$

The gradient of the function $\sum_{i=1}^d F(x_i)$ is zero only at $x_1 = \cdots = x_d = 1$, hence since $h > 0$ there are no critical points of $\sum_{i=1}^d F(x_i)$ in the region $\sum_{i=1}^d f(x_i) \geq h$. Thus the infimum in (2.127) is achieved at the boundary where $\sum_{i=1}^d f(x_i) = h$, and so

$$M_d(h) = \inf_{x_i\geq 0}\left\{\sum_{i=1}^d F(x_i) \,:\, \sum_{i=1}^d f(x_i) = h, \qquad \sum_{i=1}^d x_i = d\right\} \tag{2.128}$$

We introduce Lagrange multipliers and define

$$H(x_i, \alpha, \beta) = \sum_{i=1}^d F(x_i) - \alpha\Big(\sum_{i=1}^d f(x_i) - h\Big) - \beta\Big(\sum_{i=1}^d x_i - d\Big) \tag{2.129}$$

The critical equations for H are

$$\frac{\partial H}{\partial x_i} = 1 - \frac{1}{x_i} - \alpha\log x_i - \beta = 0$$

The constraints can be used to eliminate β and obtain

$$\Big(1+\alpha\frac{h}{d}\Big)x_i - 1 = \alpha\, x_i \log x_i, \quad i = 1,\dots,d \tag{2.130}$$

If $\alpha \le 0$ the equations (2.130) have the unique solution $x_i = 1$ for all $i = 1,\dots,d$. However this does not satisfy the constraint $\sum_{i=1}^{d} f(x_i) = h$ for $h > 0$. Thus $\alpha > 0$, in which case there are positive numbers w and z satisfying

$$0 < w < 1 < z < d$$

such that the solutions of (2.130) are

$$x_1 = \cdots = x_k = w, \quad x_{k+1} = \cdots = x_d = z$$

for some $1 \le k \le d-1$. The constraint conditions imply that

$$kw + (d-k)z = d, \quad kw\log w + (d-k)z\log z = h$$

Thus (2.128) can be reformulated as

$$\begin{aligned} M_d(h) = \inf_{0<w<z,\ 1\le k\le d-1} \{-k\log w - (d-k)\log z \quad : \quad & kw\log w + (d-k)z\log z = h, \\ & kw + (d-k)z = d\} \end{aligned} \tag{2.131}$$

We claim that $-k\log w - (d-k)\log z$, subject to the constraints $kw\log w + (d-k)z\log z = h$ and $kw + (d-k)z = d$, is a decreasing function of k. In order to show this, we divide (2.131) by d and write $k = td$, and consider the function

$$Q(w,z,t) = -t\log w - (1-t)\log z \tag{2.132}$$

along with the constraints

$$tw\log w + (1-t)z\log z = \frac{h}{d}, \qquad tw + (1-t)z = 1 \tag{2.133}$$

The constraints (2.133) allow w, z to be defined locally as functions of t. This follows from the implicit function theorem since the Jacobian is $t(1-t)\log(w/z)$; we must have $w < z$ and hence the Jacobian is

nonzero. Solving these constraint equations for the derivatives gives

$$\frac{dw}{dt} = \frac{-w + z - w\log z + w\log w}{t\log(z/w)} \tag{2.134}$$

$$\frac{dz}{dt} = \frac{w - z - z\log w + z\log z}{(1-t)\log(z/w)} \tag{2.135}$$

Returning to (2.132) we can now compute its derivative with respect to t:

$$\begin{aligned}\frac{dQ}{dt} &= -\log w + \log z - \frac{t}{w}\frac{dw}{dt} - \frac{1-t}{z}\frac{dz}{dt} \\ &= -\frac{1}{\log(z/w)}\left[\frac{(z-w)^2}{zw} - \Big(\log(z/w)\Big)^2\right]\end{aligned} \tag{2.136}$$

Note that $2\log u \le u - 1/u$ for all $u \ge 1$, hence

$$\log\left(\frac{z}{w}\right) \le \sqrt{\frac{z}{w}} - \sqrt{\frac{w}{z}}$$

and therefore the right side of (2.136) is negative. Thus Q is a decreasing function of t, and hence the infimum in (2.131) is achieved at the largest possible value of k, namely $k = d-1$. This leads to

$$M_d(h) = \inf_{z>1}\{-\log z - (d-1)\log\frac{d-z}{d-1} \;:\; z\log z + (d-z)\log\frac{d-z}{d-1} = h\} \tag{2.137}$$

The function $z\log z + (d-z)\log\frac{d-z}{d-1}$ is monotone increasing, reaching its maximum value $d\log d$ at $z = d$. Thus for any $0 < h < d\log d$ there is a unique value $z(d,h)$ satisfying the constraint condition in (2.137). Its derivative is

$$\frac{\partial z}{\partial h} = \frac{d-z}{d\log z - h} \ge 0$$

Furthermore the function $g(z) = -\log z - (d-1)\log\frac{d-z}{d-1}$ is also monotone increasing for $1 < z < d$, with derivative $g'(z) = d(z-1)/z(d-z)$. Thus

$$\begin{aligned}M_d(h) - M_d(0) &= g(z(d,h)) - g(z(d,0)) \\ &= \int_0^h g'(z(d,h))\,\frac{\partial z}{\partial h}\,dh \\ &= \int_0^h \frac{d(1-z^{-1})}{d\log z - h}\,dh \\ &\ge \int_0^h \frac{1}{z}\,\frac{d(z-1)}{d\log z - h}\,dh \\ &\ge \int_0^h \frac{1}{z}\,dh\end{aligned} \tag{2.138}$$

The constraint condition in (2.137) implies that

$$h \leq z \log z \leq h + (d-z) \log \frac{d-1}{d-z} \leq h + z - 1 \tag{2.139}$$

Thus

$$z \leq \frac{h-1}{\log z - 1} \tag{2.140}$$

If $h \geq 2e^2$ then the first inequality in (2.139) implies that $z(d,h) \geq e^2$, and therefore $\log z \geq 2$. From (2.140) it follows that

$$h \geq 2e^2 \Rightarrow z(d,h) \leq h-1$$

Thus from (2.138) we deduce that for $h \geq 2e^2$

$$M_d(h) - M_d(0) \geq \int_{2e^2}^{h} \frac{dh}{z} \geq \int_{2e^2}^{h} \frac{dh}{h-1} = \log(h-1) - \log(2e^2 - 1)$$

Since $z(d,0) = 1$ and $g(1) = 0$ it follows that $M_d(0) = 0$, and hence (2.57) holds.

Finally, to show that $M_d(x)$ is increasing, note that

$$\frac{dM_d}{dx} = \Big[-\frac{1}{z} + \frac{d-1}{d-z} \Big] \frac{\partial z}{\partial x} \tag{2.141}$$

where z solves the constraint equation

$$z \log z + (d-z) \log \left(\frac{d-z}{d-1} \right) = x \tag{2.142}$$

Differentiating (2.142) gives

$$\frac{\partial z}{\partial x} = \left(\log \big(\frac{z(d-1)}{d-z} \big) \right)^{-1} > 0$$

since $z > 1$. Also $\Big[-\frac{1}{z} + \frac{d-1}{d-z} \Big] > 0$ hence (2.141) shows that M_d is increasing.

2.6 Discussion

Hastings' Theorem finally settles the question of additivity of Holevo capacity for quantum channels, as well as additivity of minimal output entropy and entanglement of formation. In this paper we have explored in detail the proof of Hastings' result, and we have provided some estimates for the minimal dimensions necessary in order to find a violation of additivity. The violation of additivity seems to be a small effect for

this class of models, requiring delicate and explicit estimates for the proof. It is an open question whether there are random unitary channels with large violations of additivity. Hastings' Theorem is non-constructive, and it would be extremely interesting to find explicit channels which demonstrate the effect. Presumably non-additivity of Holevo capacity is generic, and there may be other classes of channels where the effect is larger.

Having established non-additivity of Holevo capacity, one is led to the question of finding useful bounds for the channel capacity $C(\Phi)$. One may even hope to find a compact 'single-letter' formula for $C(\Phi)$, though that possibility seems remote. It is likely that the methods introduced by Hastings will prove to be useful in addressing these questions.

2.7 Acknowledgments

M. F. thanks B. Nachtergaele, A. Pizzo and A. Soshnikov for numerous discussions, M. Hastings for answering questions, A. Holevo for useful comments on an early draft of this paper and R. Siegmund-Schultze for sending his related slides. C. K. thanks P. Gacs, A. Harrow, T. Kemp, M. B. Ruskai, P. Shor and B. Zeng for useful conversations. This collaboration began at the March 2009 workshop "Entropy and the Quantum" at the University of Arizona, and the authors are grateful to the organizers of the workshop.

Appendix A

Appendix for Chapter 2

A.1 Derivation of bound for $Z(n,d)$

We consider the following integral.

$$Z = \int \delta\left(1 - \sum_{i=1}^{d} p_i\right) \prod_{1\le i<j\le d} (p_i - p_j)^2 \prod_{k=1}^{d} p_k^{n-d} dp_k \tag{A.1}$$

$$= \frac{1}{(dn-1)!} \int e^{-r} r^{dn-1} dr \int \delta\left(1 - \sum_{i=1}^{d} p_i\right) \prod_{1\le i<j\le d} (p_i - p_j)^2 \prod_{k=1}^{d} p_k^{n-d} dp_k \tag{A.2}$$

Consider the following change of variables.

$$q_1 = rp_1 \tag{A.3}$$

$$\vdots \tag{A.4}$$

$$q_{d-1} = rp_{d-1} \tag{A.5}$$

$$q_d = r(1 - p_1 - \ldots - p_{d-1}) \tag{A.6}$$

The Jacobian is

$$\frac{\partial(q_1,\ldots,q_d)}{\partial(p_1,\ldots,p_{d-1},r)} = \begin{vmatrix} r & \ldots & 0 & p_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & r & p_{d-1} \\ -r & \ldots & -r & 1 - p_1 - \ldots - p_{d-1} \end{vmatrix} = \begin{vmatrix} r & \ldots & 0 & p_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & r & p_{d-1} \\ 0 & \ldots & 0 & 1 \end{vmatrix} = r^{d-1} \tag{A.7}$$

After the change of variables we have

$$Z = \frac{1}{(dn-1)!} \int \prod_{1 \le i < j \le d} (q_i - q_j)^2 \prod_{k=1}^{d} e^{-q_k} q_k^{n-d} dq_k. \tag{A.8}$$

However,

$$\prod_{1 \le i < j \le d} (q_i - q_j)^2 = \begin{vmatrix} 1 & \dots & 1 \\ q_1 & \dots & q_d \\ \vdots & \ddots & \vdots \\ q_1^{d-1} & \dots & q_d^{d-1} \end{vmatrix}^2 = \begin{vmatrix} 1 & \dots & 1 \\ p_1(q_1) & \dots & p_1(q_d) \\ \vdots & \ddots & \vdots \\ p_{d-1}(q_1) & \dots & p_{d-1}(q_d) \end{vmatrix}^2. \tag{A.9}$$

Here, p_k is any monic polynomial [1] of degree k. So, set $p_k = (-1)^k k! L_k^{n-d}$, where L_k^{n-d} is the Laguerre polynomial. Then, we have

$$\int p_k(x) p_l(x)\, e^{-x} x^{n-d} dx = \delta_{kl} \Gamma(n-d+k+1)\Gamma(k+1) \tag{A.10}$$

Hence

$$Z = \frac{1}{(dn-1)!} \int \left(\sum_{\sigma} \mathrm{sign}(\sigma) \prod_{i=1}^{d} p_{\sigma(i)}(q_i) \right)^2 \prod_{k=1}^{d} e^{-q_k} q_k^{n-d} dq_k \tag{A.11}$$

$$= \frac{1}{(dn-1)!} \sum_{\sigma} \prod_{k=1}^{d} \int \left(p_{\sigma(k)}(q_k) \right)^2 e^{-q_k} q_k^{n-d} dq_k \tag{A.12}$$

$$= \frac{\Gamma(d+1)}{\Gamma(dn)} \prod_{k=0}^{d-1} \Gamma(n-d+k+1)\Gamma(k+1) \tag{A.13}$$

$$= \frac{1}{\Gamma(dn)} \prod_{k=1}^{d} \Gamma(n-d+k)\Gamma(k+1). \tag{A.14}$$

Here, σ are "permutations"; $\sigma : \{1, \dots, d\} \to \{0, \dots, d-1\}$.

To evaluate this quantity we use the following fact: $\Gamma(s)$ is approximated by

$$\exp\left\{ s \log s - s - \frac{1}{2} \log s + \log \sqrt{2\pi} + \log \left(1 + O\left(\frac{1}{|s|^{\frac{1}{2}}} \right) \right) \right\} \tag{A.15}$$

as $s \to +\infty$. Then, we have

$$\exp\{(s-1)\log s - s\} < (A.15) < \exp\{s \log s - s\}.$$

[1] A monic polynomial is a polynomial whose highest degree term has coefficient 1. These polynomials are inserted in (A.9) by using properties of the determinant.

Note that the above upper bound is true only for large enough s but in our case it is not a problem. By using these bounds we get a lower bound for (A.14).

First, $\log(1/\Gamma(dn)))$ is lower bounded by

$$-(dn)\log(dn) + dn = -dn\log d - dn\log n + dn. \tag{A.16}$$

Secondly, $\log(\prod_{k=1}^{d}\Gamma(n-d+k))$ is lower bounded by

$$\sum_{k=1}^{d}(n-d+k-1)\log(n-d+k) - \sum_{k=1}^{d}(n-d+k) \tag{A.17}$$

$$= n^2\sum_{k=1}^{d}\frac{1}{n}\frac{n-d+k-1}{n}\log\left(\frac{n-d+k}{n}\right) + \sum_{k=1}^{d}(n-d+k-1)\log n - \sum_{k=1}^{d}(n-d+k). \tag{A.18}$$

The first sum in(A.18) is approximately lower bounded by

$$n^2 \times \frac{d}{n} \times \frac{n-d}{n}\log\left(\frac{n-d}{n}\right) = d(n-d)(\log(n-d) - \log n) \approx -d^2,$$

as $n \to \infty$. Also, the remaining part in (A.18) is

$$\left(dn - \frac{1}{2}d^2 - \frac{1}{2}d\right)\log n - \left(dn - \frac{1}{2}d^2 + \frac{1}{2}d\right). \tag{A.19}$$

Thirdly, $\log(\prod_{k=1}^{d}\Gamma(k+1))$ is lower bounded by

$$\sum_{k=1}^{d} k\log(k+1) - \sum_{k=1}^{d}(k+1) \tag{A.20}$$

$$= d^2\sum_{k=1}^{d}\frac{1}{d}\frac{k}{d}\log\left(\frac{k+1}{d}\right) + \sum_{k=1}^{d} k\log d - \sum_{k=1}^{d}(k+1). \tag{A.21}$$

Again, the first term in (A.21) is approximately lower bounded by $-d^2$. Also, the remaining part in (A.21) is

$$\left(\frac{1}{2}d^2 + \frac{1}{2}d\right)\log d - \left(\frac{1}{2}d^2 + \frac{3}{2}d\right). \tag{A.22}$$

As a whole, we know that the inside of exp in (A.14) is lower bounded by

$$
\begin{aligned}
&(A.16) + (A.19) + (A.22) - 2d^2 && \text{(A.23)}\\
&= \left[-\frac{1}{2}d^2 - \frac{1}{2}d\right]\log n + \left[-dn + \frac{1}{2}d^2 + \frac{1}{2}d\right]\log d - 2d^2 - 2d && \text{(A.24)}\\
&= -d^2 \log n + (d^2 - dn)\log d + \frac{1}{2}d(d-1)(\log n - \log d - 4) && \text{(A.25)}\\
&\geq -d^2 \log n + (d^2 - dn)\log d && \text{(A.26)}
\end{aligned}
$$

if $(\log n - \log d) \geq 4$. Therefore, we get an upper bound for the normalization constant:

$$Z(n,d)^{-1} \leq n^{d^2} d^{d(n-d)}$$

in this case.

A.2 Proof of Proposition 14

Let $Z_1, \ldots, Z_n$ be IID complex Gaussian random vectors with mean zero and variance two. Apply the result (2.87) with $m = n-1$ to deduce that

$$(Z_2, \ldots, Z_n)^T = \rho\,|\phi\rangle$$

where $|\phi\rangle$ is a random unit vector in $\mathbb{C}^{n-1} \hookrightarrow \mathbb{C}^n$, independent of $\rho = (|Z_2|^2 + \cdots + |Z_n|^2)^{1/2}$. Then apply (2.87) with $m = n$ to deduce

$$(Z_1, \ldots, Z_n)^T = R\,|\theta\rangle$$

where $|\theta\rangle$ is a random unit vector in $\mathbb{C}^n$, independent of $R = (|Z_1|^2 + \cdots + |Z_n|^2)^{1/2}$. Let

$$x = \frac{Z_1}{R} = \frac{Z_1}{\sqrt{|Z_1|^2 + \rho^2}} \tag{A.27}$$

and recall that $|\psi\rangle = (1, 0, \ldots, 0)^T$. Then

$$|\theta\rangle = \frac{1}{R}(Z_1, \ldots, Z_n)^T = x|\psi\rangle + \frac{\rho}{R}|\phi\rangle = x|\psi\rangle + \sqrt{1-|x|^2}\,|\phi\rangle$$

This proves the first part Proposition 14 since $|\phi\rangle$ is independent of Z_1 and ρ, and hence is independent of x.

To prove the second part of Proposition 14 we use the representation (A.27) to derive the distribution

of $|x|^2$. We write $Z_j = X_{2j-1} + iX_{2j}$ where $\{X_j\}$ are IID real normal random variables with mean zero and variance one, so from (A.27) it follows that

$$|x|^2 = \frac{X_1^2 + X_2^2}{X_1^2 + \ldots + X_{2n}^2}.$$

Note that $X_1^2 + \ldots + X_k^2$ has the following Chi-square probability distribution:

$$f_k(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

Hence, set

$$X = X_1^2 + X_2^2 \tag{A.28}$$

$$Y = X_3^2 + \ldots + X_{2n}^2 \tag{A.29}$$

and then X and Y are independent and have the following probability distributions

$$f_X(x) = \frac{1}{2} e^{-\frac{x}{2}} \tag{A.30}$$

$$f_Y(y) = \frac{1}{2^{n-1}\Gamma(n-1)} y^{n-2} e^{-\frac{y}{2}} \tag{A.31}$$

However,

$$\frac{X}{X+Y} \leq t \Leftrightarrow X(1-t) \leq tY$$

implies that the cumulative function of $\frac{X}{X+Y}$ is

$$\int_0^\infty \left(\int_0^{\frac{t}{1-t}y} f_X(x)\, dx \right) f_Y(y)\, dy = \int_0^\infty \left(1 - e^{-\frac{ty}{2(1-t)}}\right) f_Y(y)\, dy \tag{A.32}$$

$$= 1 - \int_0^\infty e^{-\frac{ty}{2(1-t)}} f_Y(y)\, dy \tag{A.33}$$

$$= 1 - \frac{1}{2^{n-1}(n-2)!} \int_0^\infty y^{n-2} e^{-\frac{y}{2(1-t)}}\, dy \tag{A.34}$$

Here,

$$\int_0^\infty y^{n-2} e^{-\frac{y}{2(1-t)}}\, dy = (2(1-t))^{(n-2)} (n-2)! \int_0^\infty e^{-\frac{y}{2(1-t)}}\, dy \tag{A.35}$$

$$= 2^{n-1}(1-t)^{n-1}(n-2)! \tag{A.36}$$

Therefore

$$F_{\frac{X}{X+Y}}(t) = 1 - (1-t)^{n-1}$$

Chapter 3

Average output entropy for quantum channels

Abstract

We study the regularized average Renyi output entropy $\overline{S}_r^{\text{reg}}$ of quantum channels. This quantity gives information about the average noisiness of the channel output arising from a typical, highly entangled input state in the limit of infinite dimensions. We find a closed expression for β_r^{reg}, a quantity which we conjecture to be equal to $\overline{S}_r^{\text{reg}}$. We find an explicit form for β_r^{reg} for some entanglement-breaking channels, and also for the qubit depolarizing channel Δ_λ as a function of the parameter λ. We prove equality of the two quantities in some cases, in particular we conclude that for Δ_λ both are non-analytic functions of the variable λ.

3.1 Introduction

The noisiness of a quantum channel is closely related to its ability to transfer information, and is reflected in the values of the various channel capacities. Much work has been done on understanding the capacities, for example [66] provides a recent survey. These capacities are sometimes difficult to analyze directly, for example the Holevo capacity is computed using multiple output states. Accordingly other more mathematically tractable quantities have been used to measure the amount of noise introduced by the channel. One example is the minimal output Renyi entropy [4] of a channel $\mathcal{A}$, defined for $r \geq 1$:

$$S_{r,\min}(\mathcal{A}) = \min_{|\phi\rangle} \frac{1}{1-r} \log \operatorname{Tr} \left(\mathcal{A}(|\phi\rangle\langle\phi|)\right)^r$$

At $r = 1$ this yields the minimal output von Neumann entropy, which has a close connection to the classical capacity of the channel [51]. The entropies for $r > 1$ also provide useful properties of the channel, and in some cases are easier to analyze and compute. The famous additivity conjecture concerns the regularized version of this quantity, which is defined as

$$S^{\text{reg}}_{r,\min}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} S_{r,\min}(\mathcal{A}^{\otimes n})$$

(the existence of the limit is an easy consequence of the sub-additivity bound $\mathcal{S}_{r,\min}(\mathcal{A}\otimes\mathcal{B}) \leq \mathcal{S}_{r,\min}(\mathcal{A}) + \mathcal{S}_{r,\min}(\mathcal{B})$). While the inequality $S^{\text{reg}}_{r,\min}(\mathcal{A}) \leq S_{r,\min}(\mathcal{A})$ is always true, it was an open question for several years whether equality holds. It is now known that equality does not hold in general [35, 33], so this raises the interesting question of determining $S^{\text{reg}}_{r,\min}(\mathcal{A})$. Except for those channels where additivity does hold, the value of this regularized quantity is unknown. For channels with non-additive Holevo capacity the classical capacity is also defined by such a regularized quantity, so it is an important problem to find new ways to calculate these regularized limits. In a sense we follow a strategy opposite to the random channel methods used to disprove the additivity conjecture; our high-dimensional channels are products of fixed channels, and thus are constructed explicitly.

In the hopes of finding some new insights into these regularized channel properties we consider a related quantity which also measures the noisiness of the channel, namely the average output Renyi entropy. This measures the entropy of the channel output for typical input states, rather than the smallest value which is used to compute the minimal output entropy. For a finite-dimensional channel this is defined for all $r \geq 1$ by

$$\overline{S}_r(\mathcal{A}) = \mathbb{E}\left[\frac{1}{1-r}\log \text{Tr}\,\left(\mathcal{A}(|\phi\rangle\langle\phi|)\right)^r\right]$$

where the expectation is computed using the uniform probability measure on the set of input pure states. We also consider the related quantity

$$\beta_r(\mathcal{A}) = \frac{1}{1-r}\log \mathbb{E}\left[\text{Tr}\,\left(\mathcal{A}(|\phi\rangle\langle\phi|)\right)^r\right]$$

Note that by Jensen's inequality

$$\overline{S}_r(\mathcal{A}) \geq \beta_r(\mathcal{A})\,. \tag{3.1}$$

These quantities can be computed (at least numerically) for any given channel. In operational terms, they describe the long-run average output Renyi entropy of the channel for a sequence of random pure input states. Loosely speaking, they measure the average noisiness of an output state from the channel.

As with the minimal Renyi entropy, we also consider the regularized versions of these quantities. However unlike the minimal Renyi entropy, the existence of these regularized limits is not obvious, so we define them conservatively using the lim inf:

$$\overline{S}_r^{\mathrm{reg}}(\mathcal{A}) = \liminf_{n\to\infty} \frac{1}{n}\,\overline{S}_r(\mathcal{A}^{\otimes n}) \tag{3.2}$$
$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \liminf_{n\to\infty} \frac{1}{n}\,\beta_r(\mathcal{A}^{\otimes n})$$

We conjecture that the two quantities in (3.2) are equal, however we do not yet have a proof of this for a general channel. For the specific channels we look at in more detail the quantity $\beta_r^{\mathrm{reg}}(\mathcal{A})$ is given by the expression above with lim inf replaced by lim. But from the closed expression for $\beta_r^{\mathrm{reg}}(\mathcal{A})$ presented below there is a possibility for more complex limiting behavior.

For one special class of channels we can compute a simple formula for $\beta_r^{\mathrm{reg}}(\mathcal{A})$ for integer values of r. These are a subset of the entanglement breaking (E-B) channels [60, 40], which can be written in the form $\mathcal{A}(\rho) = \sum_k \sigma_k \mathrm{Tr}\,(X_k\rho)$ – for some states σ_k and with X_k a POVM – meeting the additional condition $\mathrm{Tr}\,(\prod_{i=1}^r \sigma_{k_i}) \geq 0$. For $r = 2$ this includes all the E-B channels. For higher r a particular class of E-B channels that fulfill the condition are the QC channels as defined by Holevo [38], where $\sigma_k = |k\rangle\langle k|$ are pure states formed by an orthonormal basis. For unital E-B channels which satisfy the condition on the σ_k we can prove even more, namely that $\overline{S}_r^{\mathrm{reg}}(\mathcal{A})$=$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \log d$.

We consider only finite-dimensional channels with equal input and output dimensions, and we define the dimension of the channel to be this common value. The identity matrix is denoted by $\mathbb{1}$.

Theorem 15 (proof in 3.2.5)**.**

(a) Let $r \geq 2$ be an integer, and $\mathcal{A}(\rho) = \sum \sigma_k \mathrm{Tr}\,(X_k\rho)$ an entanglement breaking channel satisfying the condition $\mathrm{Tr}\,(\prod_{i=1}^r \sigma_{k_i}) \geq 0$ for all choices of $\{k_i\}$. Then

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n}\beta_r(\mathcal{A}^{\otimes n}) = \frac{1}{1-r}\log \mathrm{Tr}\,(\mathcal{A}(\mathbb{1}/d)^r)\,.$$

(b) Let $\mathcal{A}$ be a d-dimensional unital entanglement-breaking channel, $\mathcal{A}(\mathbb{1}) = \mathbb{1}$, satisfying the condition $\mathrm{Tr}\,(\prod_{i=1}^m \sigma_{k_i})$ for all integer $m \geq 2$. Then for all real $r \geq 1$

$$\overline{S}_r^{\mathrm{reg}}(\mathcal{A}) = \beta_r^{\mathrm{reg}}(\mathcal{A}) = \log d\,.$$

We also derive an explicit expression for $\beta_r^{\mathrm{reg}}(\mathcal{A})$ in the general case. The statement of this result requires

some additional notation. First recall the definition of the Choi-Jamiolkowski representation [16] of a channel, namely

$$Choi(\mathcal{A}) = \sum_{x,y} \mathcal{A}(|x\rangle\langle y|) \otimes |x\rangle\langle y|$$

where $|x\rangle$ and $|y\rangle$ are orthonormal bases of pure input states. Also let $Sym(r)$ denote the symmetric group on r letters. Then every element $\alpha \in Sym(r)$ defines a permutation operator on $(\mathbb{C}^d)^{\otimes r}$ by $\mathcal{R}(\alpha)(v_1 \otimes \cdots \otimes v_r) = v_{\alpha(1)} \otimes \cdots \otimes v_{\alpha(r)}$.

Definition 16. Let $\mathcal{A}$ be a channel. For all $\alpha \in Sym(r)$ define

$$\mathcal{Q}_{\mathcal{A},\,r}(\alpha) = \mathrm{Tr}\left[Choi(\mathcal{A})^{\otimes r}(\mathcal{R}(123\ldots r) \otimes \mathcal{R}(\alpha))\right] . \tag{3.3}$$

Furthermore, for some channels β_r^{reg} is a simple limit. One such class are the entrywise positive maps defined studied in [47].

Definition 17. A channel $\mathcal{A}$ is called *entrywise positive* if there exist a bases for input and output space such that $\langle s|\mathcal{A}(|x\rangle\langle y|)|t\rangle \geq 0$ for all x, y, s, t.

It is clear that this definition is equivalent to $Choi(\mathcal{A})$ being entrywise positive.

Theorem 18 (proof in 3.2.4)**.**

(a) Let $r \geq 2$ be an integer and let $\mathcal{Q}_{\max} = \max_{\alpha \in Sym(r)} |\mathcal{Q}_{\mathcal{A},\,r}(\alpha)|$. Then

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \frac{r \log d - \log \mathcal{Q}_{\max}}{r-1} .$$

(b) Let $r \geq 2$ be an integer. If the maximum $\mathcal{Q}_{\max}$ is attained for a unique α then the $\liminf$ in β_r^{reg} can be replaced with a regular limit:

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \beta_r(\mathcal{A}^{\otimes n}) .$$

(c) If the channel $\mathcal{A}$ is entrywise positive then for all integer $r \geq 2$

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \beta_r(\mathcal{A}^{\otimes n}) .$$

The evaluation of $\mathcal{Q}_{\max}$ seems to be a difficult problem in general for large values of r. However for two

special permutations the quantity $\mathcal{Q}_{\mathcal{A},\,r}(\alpha)$ can be evaluated easily, namely the identity permutation and the full cycle:

$$\mathcal{Q}_{\mathcal{A},\,r}(id) \quad = \quad \mathrm{Tr}\,\mathcal{A}(\mathbb{1})^r, \qquad \mathcal{Q}_{\mathcal{A},\,r}(123\ldots r) = \mathrm{Tr}\,(Choi(\mathcal{A})^r)\,.$$

Thus for $r = 2$ the result can be stated more explicitly as follows.

Corollary 19. *Let $\mathcal{A}$ be a d-dimensional channel, then*

$$\beta_2^{\mathrm{reg}}(\mathcal{A}) = 2\log d - \log\max[\mathrm{Tr}\,\mathcal{A}(\mathbb{1})^2, \mathrm{Tr}\,(Choi(\mathcal{A})^2)]$$

To make further progress we now focus on one of the simplest cases, namely the qubit depolarizing channel Δ_λ [61], where we are able to prove a number of additional results. In particular using concentration of measure arguments [49, 53] we compute the regularized quantity $\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda)$ – which we call regularized output entropy in the remainder of the paper – for integer values of r, and for a range of values of λ. One interesting consequence is that this quantity is a non-analytic function of the depolarizing parameter λ. Recall the definition of this channel:

$$\Delta_\lambda(\rho) = \lambda\rho + \frac{1-\lambda}{2}\mathbb{1}$$

The channel is completely positive for $-1/3 \leq \lambda \leq 1$ and is entanglement breaking for $-1/3 \leq \lambda \leq 1/3$.

Theorem 20 (proof in 3.2.6.2)**.**

(a) For all $r \in \mathbb{N}$, $r \geq 2$, and $\lambda \in [0,1]$,

$$\beta_r^{\mathrm{reg}}(\Delta_\lambda) = \lim_{n\to\infty}\frac{1}{n}\,\beta_r(\Delta_\lambda^{\otimes n}) = \min\left\{1, \frac{2r - \log\left[(1+3\lambda)^r + 3(1-\lambda)^r\right]}{r-1}\right\}$$

in particular

$$\beta_2^{\mathrm{reg}}(\Delta_\lambda) = \begin{cases} 1 & \lambda \leq 1/\sqrt{3} \\ 2 - \log(1+3\lambda^2) & \lambda > 1/\sqrt{3} \end{cases}$$

$$\beta_\infty^{\mathrm{reg}}(\Delta_\lambda) = \begin{cases} 1 & \lambda \leq 1/3 \\ 2 - \log(1+3\lambda) & \lambda > 1/3 \end{cases}$$

where $\beta_\infty^{\mathrm{reg}}(\Delta_\lambda) = \lim_{r\to\infty} \beta_r^{\mathrm{reg}}(\Delta_\lambda)$.

(b) For all $r \in \mathbb{N}$, $r \geq 2$, *and* $\lambda \in J_r$,

$$\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda) = \beta_r^{\mathrm{reg}}(\Delta_\lambda)\,,$$

where $J_r = [0, c_r] \cup [d_r, 1] \subset [0, 1]$ *for some* $0 < c_r < d_r < 1$ *(see Table 3.1 in 3.2.6.2).*

From the explicit form given in (a) it is clear that $\beta_r^{\mathrm{reg}}(\Delta_\lambda)$ does not have a continuous first derivative for some $\lambda_r \in [1/3, 1/\sqrt{3}]$, in particular it is non-analytic. Because $\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda)$ is defined as the lim inf of a series upper bounded by 1 it is well defined for all $0 \leq \lambda \leq 1$. Furthermore we know from (b) that it is equal to 1 for some range $\lambda \in [0, c_r]$ but at $\lambda = 1$ its value is 0. Therefore $\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda)$ has least one non-analytic point somewhere in the range $\lambda \in [c_r, d_r]$.

It is tempting to associate this non-analyticity with a transition between distinct phases of the model, but the operational meaning of this is unclear at the moment. We conjecture that the quantities $\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda)$ and $\beta_r^{\mathrm{reg}}(\Delta_\lambda)$ are in fact equal for all λ, and for all $r \geq 1$. It is noteworthy that the channel Δ_λ is entanglement-breaking at and below the value $\lambda = 1/3$.

Counterexamples to the additivity conjecture have been found so far by using randomization techniques [34, 28, 6, 33]. This has led to an understanding of the behavior of a typical high-dimensional channel, at least insofar as it affects the minimal output Renyi entropy, by proving the generic existence of channels all of whose output states have high entropy. Here we look from a different point of view, by considering the properties of a typical output state for a product of many copies of a channel. Open questions remain, for example the amount of entanglement in a typical output state. We note that the questions addressed here have a different flavor from arguments based on locality, since here the system is fully entangled across all copies.

3.2 Main result

3.2.1 Notation

We work with a general channel $\mathcal{A}$ and its tensor product $\mathcal{C} = \mathcal{A}^{\otimes n}$. The dimension of $\mathcal{A}$ is

$$d = \dim \mathcal{A}$$

To achieve the results for $\overline{S}_r^{\text{reg}}$ and β_r^{reg} in Theorems 15, 18 and 20 we first find a closed expression for the average moments for integer $r \geq 2$

$$M_r(\mathcal{C}) = \mathbb{E}\left[\text{Tr}\left(\mathcal{C}(|\phi\rangle\langle\phi|)^r\right)\right] , \tag{3.4}$$

where averaging is over random pure input states $|\phi\rangle = U|0\rangle$ with U distributed according to the Haar measure on $SU(d^n)$. We then use the relation

$$\beta_r^{\text{reg}}(\mathcal{A}) = \liminf_{n\to\infty} \frac{1}{n(1-r)} \log M_r(\mathcal{A}^{\otimes n})$$

3.2.2 Evaluating trace moments

We rewrite the trace moment (3.4) by inserting four complete sums that run over the entire input space

$$\begin{aligned}
&\mathbb{E}\left[\text{Tr}\left(\mathcal{C}(|\phi\rangle\langle\phi|)^r\right)\right] \\
&= \mathbb{E}\left[\text{Tr}\left(\sum_{a,b,x,y} |a\rangle\langle a|\mathcal{C}\left(|x\rangle\langle x|\phi\rangle\langle\phi|y\rangle\langle y|\right)|b\rangle\langle b|\right)^r\right] \\
&= \mathbb{E}\left[\sum_{\substack{\{a_i,x_i,y_i\}\\ i=1\ldots r}} \prod_{i=1}^{r} \mathcal{C}_{a_i x_i y_i a_{i+1}} \langle x_i|\phi\rangle\langle\phi|y_i\rangle\right] \\
&= \sum_{\substack{\{a_i,x_i,y_i\}\\ i=1\ldots r}} \prod_{i=1}^{r} \mathcal{C}_{a_i x_i y_i a_{i+1}} \mathbb{E}\left[\prod_{j=1}^{r} \langle x_j|\phi\rangle\langle\phi|y_j\rangle\right]
\end{aligned} \tag{3.5}$$

with the identification $a_{r+1} \equiv a_1$, and with

$$\mathcal{C}_{axyb} = \langle a|\mathcal{C}(|x\rangle\langle y|)|b\rangle$$

the matrix elements of the channel. The expectation value in (3.5)

$$\mathbb{E}\left[\prod_{j=1}^{r} \langle x_j|\phi\rangle\langle\phi|y_j\rangle\right] = \mathbb{E}\left[\prod_{j=1}^{r} \langle x_j|U|0\rangle\langle 0|U^*|y_j\rangle\right] \tag{3.6}$$

of products of matrix elements of unitaries distributed according to the Haar measure may be calculated using Weingarten calculus [67]. The Weingarten function $\text{Wg} : \mathbb{N} \times Sym(r) \to \mathbb{R}$ maps pairs of dimension k

and elements of the symmetric group $Sym(r)$ into the reals. The general expression is

$$\begin{aligned}
&\mathbb{E}\left[U_{i_1 j_1}\dots U_{i_r j_r}\overline{U_{i'_1 j'_1}}\dots\overline{U_{i'_r j'_r}}\right] \\
&= \sum_{\alpha,\beta\in Sym(r)} \delta_{i_{\alpha(1)}i'_1}\dots\delta_{i_{\alpha(r)}i'_r}\delta_{j_{\beta(1)}j'_1}\dots\delta_{j_{\beta(r)}j'_r}\mathrm{Wg}\left(k,\beta^{-1}\alpha\right) .
\end{aligned}$$

Thus (3.6) simplifies to

$$\mathbb{E}\Big[\prod_{j=1}^{r}\langle x_j|U|0\rangle\langle 0|U^*|y_j\rangle\Big] = \sum_{\alpha,\beta\in Sym(r)} \delta_{x_{\alpha(1)}y_1}\dots\delta_{x_{\alpha(r)}y_r}\mathrm{Wg}\left(k,\beta\alpha^{-1}\right) \tag{3.7}$$

$$= \sum_{\alpha\in Sym(r)} \delta_{x_{\alpha(1)}y_1}\dots\delta_{x_{\alpha(r)}y_r}C_{k,\,r} \tag{3.8}$$

where k is the input dimension for $\mathcal{C}$, and in the last step the Weingarten function is summed over all permutations $\gamma = \beta\alpha^{-1}$. This sum can be evaluated explicitly, as was shown for example in [19]:

$$C_{k,\,r} = \sum_{\gamma\in Sym(r)} \mathrm{Wg}\left(k,\gamma\right) = \prod_{j=0}^{r-1}\frac{1}{k+j} . \tag{3.9}$$

We plug the evaluated expectation value (3.8) in our original expression (3.5) and get

$$\begin{aligned}
&\sum_{\substack{\{a_i,x_i,y_i\}\\ i=1\dots r}}\left(\prod_{i=1}^{r}\mathcal{C}_{a_i x_i y_i a_{i+1}}\cdot\sum_{\alpha\in Sym(r)}\prod_{j=1}^{r}\delta_{x_{\alpha(j)}y_j}\cdot C_{k,\,r}\right) \\
&\qquad = C_{k,\,r}\sum_{\substack{\{a_i,x_i\}\\ i=1\dots r}}\sum_{\alpha\in Sym(r)}\prod_{i=1}^{r}\mathcal{C}_{a_i x_i x_{\alpha(i)} a_{i+1}} .
\end{aligned} \tag{3.10}$$

Now define

$$\mathcal{Q}_{\mathcal{C},\,r}(\alpha) = \sum_{\substack{\{a_i,x_i\}\\ i=1\dots r}}\prod_{i=1}^{r}\mathcal{C}_{a_i x_i x_{\alpha(i)} a_{i+1}} = \sum_{\substack{\{x_i\}\\ i=1\dots r}}\mathrm{Tr}\prod_{i=1}^{r}\mathcal{C}_{x_i x_{\alpha(i)}} , \tag{3.11}$$

where the matrices $\mathcal{C}_{xy}$ have entries $(\mathcal{C}_{xy})_{ab} = \mathcal{C}_{axyb}$ or more simply $\mathcal{C}_{xy} = \mathcal{C}(|x\rangle\langle y|)$. If it is clear from context we may omit one or both subscripts $\mathcal{Q}(\alpha) = \mathcal{Q}_{\mathcal{C}}(\alpha) = \mathcal{Q}_{\mathcal{C},\,r}(\alpha)$. These are the terms we will analyze to a great length in the rest of the work.

The $\mathcal{Q}$ defined in this way is identical to the one from Definition 16. This can be seen from the following calculation

$$
\begin{aligned}
&\mathrm{Tr}\left(Choi(\mathcal{C})^{\otimes r}(\mathcal{R}(123\ldots r)\otimes\mathcal{R}(\alpha))\right)\\
=&\sum_{\{a,x\}}\langle a_1,x_1|\otimes\cdots\otimes\langle a_r,x_r|Choi(\mathcal{C})^{\otimes r}|a_2,x_{\alpha(1)}\rangle\otimes\cdots\otimes|a_1,x_{\alpha(r)}\rangle\\
=&\sum_{\{a,x\}}\langle a_1|\mathcal{C}(|x_1\rangle\langle x_{\alpha(1)}|)|a_2\rangle\langle a_2|\mathcal{C}(|x_2\rangle\langle x_{\alpha(2)}|)|a_3\rangle\cdots\\
=&\sum_{\substack{\{a_i,x_i\}\\ i=1\ldots r}}\prod_{i=1}^{r}\mathcal{C}_{a_ix_ix_{\alpha(i)}a_{i+1}}\\
=&\mathcal{Q}_{\mathcal{C},\,r}(\alpha)\,.
\end{aligned}
$$

In terms of the $\mathcal{Q}_{\mathcal{C}}$ we get our final working expression for the average moments

$$
M_r(\mathcal{C})=C_{k,\,r}\sum_{\alpha\in Sym(r)}\mathcal{Q}_{\mathcal{C}}(\alpha)\,.
$$

In general $\mathcal{Q}_{\mathcal{C}}(\alpha)$ could have complex values. However in some cases it can be shown to be real. In particular for the depolarizing channel it follows directly from $\mathcal{C}_{axyb}\geq 0$ that $\mathcal{Q}_{\mathcal{C}}(\alpha)$ is positive.

3.2.3 Product channels

In the case where $\mathcal{C}$ is a tensor product $\mathcal{C}=\mathcal{D}\otimes\mathcal{E}$ we work in the product base $|x\rangle=|x'x''\rangle$. Now the tensor and channel application are interchangeable

$$
\begin{aligned}
\mathcal{D}\otimes\mathcal{E}_{xy}&=\mathcal{D}\otimes\mathcal{E}(|x'x''\rangle\langle y'y''|)\\
&=\mathcal{D}(|x'\rangle\langle y'|)\otimes\mathcal{E}(|x''\rangle\langle y''|)\\
&=\mathcal{D}_{x'y'}\otimes\mathcal{E}_{x''y''}\,.
\end{aligned}
$$

And therefore, the $\mathcal{Q}_{\mathcal{D}\otimes\mathcal{E}}(\alpha)$ factors

$$\begin{aligned}
\mathcal{Q}_{\mathcal{D}\otimes\mathcal{E}}(\alpha) &= \sum_{\substack{\{x_i\}\\ i=1\ldots r}} \operatorname{Tr} \prod_{i=1}^{r} \mathcal{D}\otimes\mathcal{E}_{x_i x_{\alpha(i)}} \\
&= \sum_{\substack{\{x'_i, x''_i\}\\ i=1\ldots r}} \operatorname{Tr} \prod_{i=1}^{r} \mathcal{D}_{x'_i x'_{\alpha(i)}} \otimes \mathcal{E}_{x''_i x''_{\alpha(i)}} \\
&= \sum_{\substack{\{x'_i\}\\ i=1\ldots r}} \operatorname{Tr} \prod_{i=1}^{r} \mathcal{D}_{x'_i x'_{\alpha(i)}} \sum_{\substack{\{x''_i\}\\ i=1\ldots r}} \operatorname{Tr} \prod_{i=1}^{r} \mathcal{E}_{x''_i x''_{\alpha(i)}} \\
&= \mathcal{Q}_{\mathcal{D}}(\alpha)\mathcal{Q}_{\mathcal{E}}(\alpha)\,. \qquad (3.12)
\end{aligned}$$

3.2.4 Average moments of $\mathcal{A}^{\otimes n}$

If we set $\mathcal{C} = \mathcal{A}^{\otimes n}$ the average moment factors as above, the dimension is $k = d^n$ and according to (3.12) we get

$$M_r(\mathcal{A}^{\otimes n}) = C_{d^n,\,r} \sum_{\alpha\in Sym(r)} \mathcal{Q}_{\mathcal{A},\,r}(\alpha)^n\,. \qquad (3.13)$$

When the meaning is clear from the context we suppress the index in $\mathcal{Q}_{\mathcal{A}}$. The limiting behavior of this sum is relatively simple and determines the quantity $\beta_r^{\mathrm{reg}}(\mathcal{A})$ as described in the following theorem.

Theorem 18. *(a) Let $r \geq 2$ be an integer and let $\mathcal{Q}_{\max} = \max_{\alpha\in Sym(r)} |\mathcal{Q}_{\mathcal{A},\,r}(\alpha)|$. Then*

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \frac{r\log d - \log \mathcal{Q}_{\max}}{r-1}\,.$$

(b) Let $r \geq 2$ be an integer. If the maximum $\mathcal{Q}_{\max}$ is attained for a unique α then the $\liminf$ *in β_r^{reg} can be replaced with a regular limit:*

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n}\beta_r(\mathcal{A}^{\otimes n})\,.$$

(c) If the channel $\mathcal{A}$ is entrywise positive then for all integer $r \geq 2$

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n}\beta_r(\mathcal{A}^{\otimes n})\,.$$

Proof. (a) From the definition of (3.2) and (3.13) we have

$$\begin{aligned}\beta_r^{\mathrm{reg}}(\mathcal{A}) &= \liminf_{n\to\infty} \frac{1}{n(1-r)} \log M_r(\mathcal{A}^{\otimes n}) \\ &= \liminf_{n\to\infty} \frac{1}{n(1-r)} \log\left[C_{d^n,\,r} \sum_\alpha \mathcal{Q}(\alpha)^n\right]\end{aligned}$$

Note that

$$\begin{aligned}C_{d^n,\,r} \sum_\alpha \mathcal{Q}(\alpha)^n &= \left|C_{d^n,\,r} \sum_\alpha \mathcal{Q}(\alpha)^n\right| \\ &\le C_{d^n,\,r} \sum_\alpha |\mathcal{Q}(\alpha)|^n \\ &\le d^{-nr} \sum_\alpha \mathcal{Q}_{\max}^n \\ &= r!\, d^{-nr} \mathcal{Q}_{\max}^n\end{aligned}$$

Since $M_r(\mathcal{A}^{\otimes n}) \le 1$ we have

$$\begin{aligned}\beta_r^{\mathrm{reg}}(\mathcal{A}) &= \frac{1}{r-1} \liminf_{n\to\infty} \frac{1}{n} \log\left[C_{d^n,\,r} \sum_\alpha \mathcal{Q}(\alpha)^n\right]^{-1} \\ &\ge \frac{1}{r-1} \liminf_{n\to\infty} \frac{1}{n} \log\left(r!\, d^{-nr} \mathcal{Q}_{\max}^n\right)^{-1} \\ &= \frac{1}{r-1} \log\left(d^r \mathcal{Q}_{\max}^{-1}\right) \\ &= \frac{r\log d - \log \mathcal{Q}_{\max}}{r-1}\,.\end{aligned}$$

In order to prove equality, we will use the existence of a subsequence $\{n_j\}$ such that

$$\lim_{j\to\infty} \frac{1}{n_j(1-r)} \log\left[C_{d^{n_j},\,r} \sum_\alpha \mathcal{Q}(\alpha)^{n_j}\right] \le \frac{r\log d - \log \mathcal{Q}_{\max}}{r-1}\,. \tag{3.14}$$

To this end, let $\{\alpha_1, \ldots, \alpha_N\}$ be the maximizers satisfying $|\mathcal{Q}(\alpha_i)| = \mathcal{Q}_{\max}$, so that $\mathcal{Q}(\alpha_i) = \mathcal{Q}_{\max} e^{2\pi i \gamma_i}$ for some $\gamma_i \in [0,1)$. It is a basic result from simultaneous Diophantine approximations (using the Dirichlet box principle) that for any $\epsilon > 0$ there is an increasing sequence of positive integers n_j such that $\max_i\{n_j\gamma_i\} < \epsilon$ for all j, where $\{x\}$ denotes the distance to the closest integer. Choose $\epsilon = 1/6$, then we have

$$\begin{aligned}\left|\sum_{i=1}^N \mathcal{Q}(\alpha_i)^{n_j}\right| &= \mathcal{Q}_{\max}^{n_j} \left|\sum_{i=1}^N e^{2\pi i n_j \gamma_i}\right| \\ &\ge \mathcal{Q}_{\max}^{n_j} \cdot \frac{N}{2}\,.\end{aligned} \tag{3.15}$$

Furthermore, there is $\Theta < 1$ such that

$$|\mathcal{Q}(\alpha')| \leq \mathcal{Q}_{\max} \cdot \Theta$$

for all α' which are not maximizers. Thus

$$\begin{aligned} C_{d^{n_j},r} \sum_{\alpha} \mathcal{Q}(\alpha)^{n_j} &= C_{d^{n_j},r} \left| \sum_{\alpha} \mathcal{Q}(\alpha)^{n_j} \right| \\ &\geq C_{d^{n_j},r} \mathcal{Q}_{\max}^{n_j} \cdot \frac{N}{2} - C_{d^{n_j},r} \left| \sum_{\alpha'} \mathcal{Q}(\alpha')^{n_j} \right| \\ &\geq C_{d^{n_j},r} \mathcal{Q}_{\max}^{n_j} \left(\frac{N}{2} - r!\, \Theta^{n_j} \right) \end{aligned}$$

Since $\Theta < 1$, for j sufficiently large we have $\frac{N}{2} - r!\Theta^{n_j} \geq \frac{1}{3}$, thus for j sufficiently large

$$\begin{aligned} &\frac{1}{n_j(1-r)} \left[\log C_{d^{n_j},r} \sum_{\alpha} \mathcal{Q}(\alpha)^{n_j} \right] \\ &\qquad \leq \frac{1}{r-1} \frac{1}{n_j} \log \left(3 \mathcal{Q}_{\max}^{-n_j} C_{d^{n_j},r}^{-1} \right) . \end{aligned}$$

Using $\lim_{n\to\infty} (d^{nr} C_{d^n,r}) = 1$, the result follows immediately.

(b) Let $\alpha_0 \in Sym(r)$ be the unique permutation satisfying $|\mathcal{Q}(\alpha_0)| = \mathcal{Q}_{\max}$, so that $N = 1$ in the notation of (a). Then (3.15) is replaced by the equality

$$|\mathcal{Q}(\alpha_0)^n| = \mathcal{Q}_{\max}^n$$

which holds for every n. Thus the inequality (3.14) is true for every n, hence the upper and lower bound yield the existence of the limit.

(c) If $\mathcal{A}$ is entrywise positive then every term $\mathcal{Q}(\alpha)$ is also positive. Thus the inequality (3.15) is replaced by

$$\left| \sum_{i=1}^{N} \mathcal{Q}(\alpha_i)^n \right| = N \mathcal{Q}_{\max}^n$$

and this holds for every n. Thus again (3.14) is true for every n, and the result follows. □

3.2.5 Entanglement breaking channels

After dealing with basic facts about β_r^{reg} we turn our attention to the special case of entanglement breaking channels. In this case calculating the relevant $\mathcal{Q}$-terms and studying their properties is particularly easy. We restate and then prove Theorem 15 from the introduction.

Theorem 15. *(a) Let $r \geq 2$ be an integer, and $\mathcal{A}(\rho) = \sum \sigma_k \mathrm{Tr}\,(X_k \rho)$ an entanglement breaking channel satisfying the condition $\mathrm{Tr}\,(\prod_{i=1}^r \sigma_{k_i}) \geq 0$ for all choices of $\{k_i\}$. Then*

$$\beta_r^{\mathrm{reg}}(\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \beta_r(\mathcal{A}^{\otimes n}) = \frac{1}{1-r} \log \mathrm{Tr}\,(\mathcal{A}(\mathbb{1}/d)^r)\,.$$

(b) Let $\mathcal{A}$ be a d-dimensional unital entanglement-breaking channel, $\mathcal{A}(\mathbb{1}) = \mathbb{1}$, satisfying the condition $\mathrm{Tr}\,(\prod_{i=1}^m \sigma_{k_i})$ for all integer $m \geq 2$. Then for all real $r \geq 1$

$$\overline{S}_r^{\mathrm{reg}}(\mathcal{A}) = \beta_r^{\mathrm{reg}}(\mathcal{A}) = \log d\,.$$

Proof. (a) The result follows immediately from Lemma 21 and Theorem 18.

(b) For a unital entanglement breaking channel $\beta_r^{\mathrm{reg}}(\mathcal{A})$ attains its maximal value $\log d$ which is at the same time the maximal possible value of $\overline{S}_r^{\mathrm{reg}}(\mathcal{A})$. So, with inequality (3.1) it follows that $\overline{S}_r^{\mathrm{reg}}(\mathcal{A}) = \beta_r^{\mathrm{reg}}(\mathcal{A}) = \log d$ for all integer $r \geq 2$. In the following we extend this result to all real $r \geq 1$.

Consider the derivatives of the function $f_n(r) = \frac{1}{n} \ln \mathbb{E}\left[\mathrm{Tr}\,(\mathcal{A}^{\otimes n}(|\phi\rangle\langle\phi|)^r)\right] / \ln 2$ with respect to r, and set $\rho = \mathcal{A}^{\otimes n}(|\phi\rangle\langle\phi|)$

$$\begin{aligned} f_n'(r) &= \frac{1}{n} \mathbb{E}\left[\mathrm{Tr}\,\rho^r\right]^{-1} \mathbb{E}\left[\mathrm{Tr}\,(\rho^r \ln \rho)\right] / \ln 2 < 0 \\ f_n''(r) &= \frac{1}{n} \mathbb{E}\left[\mathrm{Tr}\,\rho^r\right]^{-1} \mathbb{E}\left[\mathrm{Tr}\left(\rho^r (\ln \rho)^2\right)\right] / \ln 2 - \frac{1}{n} \mathbb{E}\left[\mathrm{Tr}\,\rho^r\right]^{-2} \mathbb{E}\left[\mathrm{Tr}\,(\rho^r \ln \rho)\right]^2 / \ln 2 > 0 \end{aligned} \tag{3.16}$$

To prove the second inequality we use two applications of the Cauchy-Schwarz inequality. First $|\mathrm{Tr}\,(AB)| \leq \mathrm{Tr}\left(A^2\right)^{1/2} \mathrm{Tr}\left(B^2\right)^{1/2}$ with $A = \rho^{r/2}$ and $B = \rho^{r/2} \ln \rho$, and then $\mathbb{E}\left[XY\right]^2 \leq \mathbb{E}\left[X^2\right] \mathbb{E}\left[Y^2\right]$ to deduce

$$\mathbb{E}\left[\mathrm{Tr}\,(\rho^r \ln \rho)\right]^2 \leq \mathbb{E}\left[(\mathrm{Tr}\left(\rho^r (\ln \rho)^2\right))^{1/2} (\mathrm{Tr}\,\rho^r)^{1/2}\right]^2 \leq \mathbb{E}\left[\mathrm{Tr}\left(\rho^r (\ln \rho)^2\right)\right] \mathbb{E}\left[\mathrm{Tr}\,\rho^r\right]$$

Therefore, the function $f_n(r)$ is convex in r for $r > 0$. We also know that $f_n(r) \geq (1-r) \log d$ for any real $r \geq 1$ and $\lim_{n\to\infty} f_n(r) = (1-r) \log d$ for any integer $r \geq 1$. Therefore, for any integer $r_0 \geq 1$ and $r \in [r_0, r_0+1]$ we have upper and lower bounds

$$f_n(r_0)(r_0 + 1 - r) + f_n(r_0+1)(r - r_0) \geq f_n(r) \geq (1-r) \log d\,. \tag{3.17}$$

Thus we have $\lim_{n\to\infty} f_n(r) = (1-r) \log d$ for all real $r \geq 1$. Dividing by $1-r$ gives $\lim_{n\to\infty} \frac{1}{n} S_r(\mathcal{A}^{\otimes n}) = \log d$ the desired equality for all $r > 1$.

Finally, for $r = 1$ the Renyi entropy is defined as the Neumann entropy, which equals the limit $\frac{1}{n} S_1(\mathcal{A}^{\otimes n}) =$

$\lim_{r\to 1}\frac{f_n(r)}{1-r} = -f_n'(1)$. Again, using the bound (3.17) with $r_0 = 1$ (and noting that $f_n(1) = 0$) we get $-f_n(2)\,(1-r) \geq f_n(r) \geq (1-r)\log d$ and so

$$-f_n(2) \leq \frac{f_n(r)}{1-r} \leq \log d$$

which implies $-f_n(2) \leq \frac{1}{n}S_1(\mathcal{A}^{\otimes n}) \leq \log d$. But as $\lim_{n\to\infty} -f_n(2) = \log d$ both sides of the bound become equal and we have $\lim_{n\to\infty}\frac{1}{n}S_1(\mathcal{A}^{\otimes n}) = \log d$ as well. □

Lemma 21. *Let $\mathcal{A}$ be an entanglement breaking channel with*

$$\mathrm{Tr}\left(\prod_{i=1}^{r}\sigma_{k_i}\right) \geq 0 \tag{3.18}$$

for all choices of $\{k_i\}$, then $\alpha = \mathrm{id}$ is the unique maximum of $Q_{\mathcal{A},\,r}$ so $\mathcal{Q}_{\max} = \mathcal{Q}_{\mathcal{A},\,r}(\mathrm{id}) = \mathrm{Tr}\,(\mathcal{A}(\mathbb{1})^r)$.

Proof. Entanglement breaking channels are of the form $\mathcal{A}(\rho) = \sum \sigma_k \mathrm{Tr}\,(X_k\rho)$ where the σ_k are density matrices and the X_k constitute a POVM. Therefore we calculate

$$\begin{aligned}
|\mathcal{Q}(\alpha)| &= \Big|\sum_{\substack{\{x_i\}\\ i=1\ldots r}} \mathrm{Tr}\prod_{i=1}^{r}\mathcal{A}_{x_i x_{\alpha(i)}}\Big| \\
&= \Big|\sum_{\substack{\{x_i\}\\ i=1\ldots r}} \mathrm{Tr}\left(\prod_{i=1}^{r}\sum_k \sigma_k \mathrm{Tr}\left(X_k|x_i\rangle\langle x_{\alpha(i)}|\right)\right)\Big| \\
&= \Big|\sum_{\substack{\{x_i,k_i\}\\ i=1\ldots r}} \mathrm{Tr}\left(\prod_{i=1}^{r}\sigma_{k_i}\right)\prod_{i=1}^{r}\mathrm{Tr}\left(X_{k_i}|x_i\rangle\langle x_{\alpha(i)}|\right)\Big| \\
&\leq \sum_{\substack{\{k_i\}\\ i=1\ldots r}} \mathrm{Tr}\left(\prod_{i=1}^{r}\sigma_{k_i}\right)\Big|\sum_{\substack{\{x_i\}\\ i=1\ldots r}}\prod_{i=1}^{r}\langle x_{\alpha(i)}|X_{k_i}|x_i\rangle\Big|
\end{aligned} \tag{3.19}$$

where condition (3.18) is used in the last step. The term on the right side of the last line can be rewritten as a product of traces

$$\begin{aligned}
\sum_{\substack{\{x_i\}\\ i=1\ldots r}}\prod_{i=1}^{r}\langle x_{\alpha(i)}|X_{k_i}|x_i\rangle &= \sum_{\substack{\{x_i\}\\ i=1\ldots r}} \ldots \langle x_{\alpha^2(1)}|X_{k_{\alpha(1)}}|x_{\alpha(1)}\rangle\langle x_{\alpha(1)}|X_{k_1}|x_1\rangle \\
&= \prod_{\gamma\in\alpha}\mathrm{Tr}\prod_{i\in\gamma^{-1}} X_{k_i}
\end{aligned}$$

where $\gamma \in \alpha$ are the sub-cycles of α and $\prod_{i\in\gamma^{-1}}$ is a product over the numbers in γ^{-1}.

Now consider any set of operators $Y_j \geq 0$, using the spectral decomposition $Y_j = \sum_k \lambda_{k,\,j}|k\rangle_j\langle k|_j$ we

have

$$\left|\operatorname{Tr}\prod_{j=1}^{m} Y_j\right| = \left|\sum_{\substack{\{k_j\}\\ j=1\ldots m}} \lambda_{k_1,1}\ldots\lambda_{k_m,m}\operatorname{Tr}\left(|k_1\rangle_1\langle k_1|_1\ldots|k_m\rangle_m\langle k_m|_m\right)\right| \tag{3.20}$$

$$\leq \sum_{\substack{\{k_j\}\\ j=1\ldots m}} \lambda_{k_1,1}\ldots\lambda_{k_m,m} = \prod_{j=1}^{m}\operatorname{Tr} Y_j\,.$$

Equality holds only in the following cases:

- If $m = 1$.
- If any of the Y_j's equals zero.
- If all the Y_j are a multiple of a one-dimensional projection.

To see that there are no other possibilities consider the case where $m \geq 2$ and all Y_j are rank one but they don't have the same eigenvectors. Now the sum in (3.20) contains the overlap of the eigenvectors which is smaller than one in absolute value because some eigenvectors are not the same. Therefore there is no equality. Finally, consider the case where $m \geq 2$, where all the Y_j have at least rank one and where there exists a j_0 such that Y_{j_0} has rank two or higher. On the RHS of (3.20), choose the k_j such that all λ_{k_j} are non-zero. For k_{j_0} there are two or more choices and for one of those choices the trace term has to be smaller than one in absolute value. Therefore equality cannot hold in this case.

Returning to the X_j we see that if any $X_j = 0$ we can drop it from our POVM without changing the channel $\mathcal{A}$. Also if say $X_1 = qX_2$ are multiples of each other then we can combine them to $\tilde{X}_1 = X_1 + X_2$ and $\tilde{\sigma}_1 = (\sigma_1 + q\sigma_2)/(1+q)$ again without changing $\mathcal{A}$. Therefore we can assume no X_j equals zero and no two X_j are multiples of each other, then equality is only possible if $m = 1$. It follows that

$$\left|\prod_{\gamma\in\alpha}\operatorname{Tr}\prod_{i\in\gamma^{-1}} X_{k_i}\right| \leq \prod_{i=1}^{r}\operatorname{Tr} X_{k_i}\,,$$

can only be equality if all cycles in α have length one, i.e. $\alpha = \mathrm{id}$.

Combining the last inequality with (3.19) we get

$$|\mathcal{Q}(\alpha)| \leq \sum_{\substack{\{k_i\}\\ i=1\ldots r}} \operatorname{Tr}\left(\prod_{i=1}^{r}\sigma_{k_i}\right)\prod_{i=1}^{r}\operatorname{Tr} X_{k_i} = \mathcal{Q}(\mathrm{id})$$

with equality if and only if $\alpha = \mathrm{id}$. And finally,

$$\begin{aligned}
\mathcal{Q}(\mathrm{id}) &= \sum_{\substack{\{k_i\}\\ i=1\dots r}} \mathrm{Tr}\left(\prod_{i=1}^{r} \sigma_{k_i}\right) \prod_{i=1}^{r} \mathrm{Tr}\, X_{k_i} \\
&= \sum_{\substack{\{k_i\}\\ i=1\dots r}} \mathrm{Tr}\left(\prod_{i=1}^{r} \sigma_{k_i} \mathrm{Tr}\, X_{k_i}\right) \\
&= \mathrm{Tr}\left(\prod_{i=1}^{r} \sum_{k} \sigma_k \mathrm{Tr}\, X_k\right) \\
&= \mathrm{Tr}\left(\mathcal{A}(\mathbb{1})^r\right) .
\end{aligned}$$

□

3.2.6 The qubit depolarizing channel

3.2.6.1 Evaluating $\mathcal{Q}_{\Delta_\lambda(\alpha)}$

In this section we calculate $\mathcal{Q}_{\Delta_\lambda}(\alpha)$ for $\alpha = \mathrm{id}$ and $\alpha = (1 \dots r)$. As we prove in Lemma 27 in 3.3.1 one of these two terms is always maximal, $\mathcal{Q}_{\max} = \max\{\mathcal{Q}_{\Delta_\lambda}(\mathrm{id}), \mathcal{Q}_{\Delta_\lambda}((1\dots r))\}$, so they are of particular interest.

We have

$$\mathcal{Q}_{\Delta_\lambda}(\mathrm{id}) = \mathrm{Tr}\, \Delta_\lambda(\mathbb{1})^r = \mathrm{Tr}\, \mathbb{1} = 2$$

To evaluate the second $\mathcal{Q}$-term we consider a slightly more general channel $\mathcal{A}$ with Choi-Jamiolkowski representation

$$Choi(\Delta_\lambda) = \begin{pmatrix} \mu & 0 & 0 & \lambda \\ 0 & \nu & \kappa & 0 \\ 0 & \kappa & \nu & 0 \\ \lambda & 0 & 0 & \mu \end{pmatrix} .$$

This matrix has diagonal block form with blocks

$$\begin{pmatrix} \mu & \lambda \\ \lambda & \mu \end{pmatrix}, \begin{pmatrix} \nu & \kappa \\ \kappa & \nu \end{pmatrix} .$$

We need to raise the matrix to the r-th power, which gives

$$\begin{pmatrix} \mu & \lambda \\ \lambda & \mu \end{pmatrix}^r = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} (\mu+\lambda)^r & 0 \\ 0 & (\mu-\lambda)^r \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

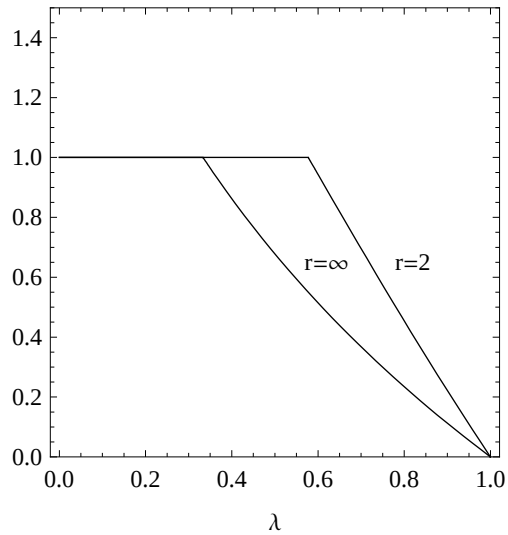


Figure 3.1: S_2 and S_∞.

r	c_r	d_r
2	.577	.732
3	.5	.835
4	.458	.878
10	.381	.953
100	.338	.995

Table 3.1: Range parameters of Theorem 20.

and similarly for the second matrix. Therefore we get

$$\mathcal{Q}_{\mathcal{A}}((1\dots r)) = \mathrm{Tr}\left(Choi^r_{\Delta_\lambda}\right) = (\mu+\lambda)^r + (\mu-\lambda)^r + (\nu+\kappa)^r + (\nu-\kappa)^r\,. \tag{3.21}$$

For $\mathcal{A} = \Delta_\lambda$ we have $\mu = \frac{1+\lambda}{2}$, $\nu = \frac{1-\lambda}{2}$ and $\kappa = 0$, so

$$\mathcal{Q}_{\Delta_\lambda}((1\dots r)) = \left(\frac{1+3\lambda}{2}\right)^r + 3\left(\frac{1-\lambda}{2}\right)^r\,.$$

In 3.4.2 we compute the $\mathcal{Q}((1\dots r))$ for any dimension $d \geq 2$.

More generally, whenever α is a product of cycles of consecutive numbers the sum factors as shown in Lemma 24, e.g. $\mathcal{Q}((123)(45)) = \frac{1}{d}\mathcal{Q}((123))\mathcal{Q}((45))$.

3.2.6.2 Regularized output entropy of $\Delta_\lambda^{\otimes n}$

We are now ready to prove the main theorem.

Theorem 20. *(a) For all $r \in \mathbb{N}$, $r \geq 2$, and $\lambda \in [0,1]$,*

$$\beta_r^{\mathrm{reg}}(\Delta_\lambda) = \lim_{n\to\infty} \frac{1}{n}\beta_r(\Delta_\lambda^{\otimes n}) = \min\left\{1, \frac{2r - \log\left[(1+3\lambda)^r + 3(1-\lambda)^r\right]}{r-1}\right\} \tag{3.22}$$

in particular

$$\beta_2^{\mathrm{reg}}(\Delta_\lambda) = \begin{cases} 1 & \lambda \leq 1/\sqrt{3} \\ 2 - \log(1+3\lambda^2) & \lambda > 1/\sqrt{3} \end{cases},$$

$$\beta_\infty^{\mathrm{reg}}(\Delta_\lambda) = \begin{cases} 1 & \lambda \leq 1/3 \\ 2 - \log(1+3\lambda) & \lambda > 1/3 \end{cases},$$

where $\beta_\infty^{\mathrm{reg}}(\Delta_\lambda) = \lim_{r\to\infty} \beta_r^{\mathrm{reg}}(\Delta_\lambda)$.

(b) For all $r \in \mathbb{N}$, $r \geq 2$, and $\lambda \in J_r$,

$$\overline{S}_r^{\mathrm{reg}}(\Delta_\lambda) = \beta_r^{\mathrm{reg}}(\Delta_\lambda) \tag{3.23}$$

where $J_r = [0, c_r] \cup [d_r, 1] \subset [0,1]$ for some $0 < c_r < d_r < 1$ (see Table 3.1).

The regularized output entropy for $r = 2$ and the lower bound of the same for $r = \infty$ are plotted in Figure 3.1. We expect that (3.23) holds for all λ. The regularized output entropy is maximal when $2 \geq \left(\frac{1+3\lambda}{2}\right)^r + 3\left(\frac{1-\lambda}{2}\right)^r$ which holds for all r when $\lambda \leq 1/3$. It's interesting to note that $\lambda \leq 1/3$ is also the condition for Δ_λ to be entanglement breaking.

Proof. (a) From 3.4.2 we know that $\mathcal{Q}(\mathrm{id}) = 2$ and $\mathcal{Q}((1\ldots r)) = \left(\frac{1+3\lambda}{2}\right)^r + 3\left(\frac{1-\lambda}{2}\right)^r$. In Lemma 27 in 3.3.1 we prove that one of these two $\mathcal{Q}$-terms yields the maximal value, i.e. for any $\lambda \in [0,1]$

$$\mathcal{Q}_{\max} = \max\left\{2, \left(\frac{1+3\lambda}{2}\right)^r + 3\left(\frac{1-\lambda}{2}\right)^r\right\}.$$

Additionally, Δ_λ is entrywise positive for $\lambda \in [0,1]$. Therefore we can apply parts (a) and (c) of Theorem 18 and (3.22) follows immediately.

In the case $r = 2$ we have

$$\begin{aligned} \frac{2\cdot 2 - \log\left[(1+3\lambda)^2 + 3(1-\lambda)^2\right]}{2-1} &= 4 - \log(4+12\lambda^2) \\ &= 2 - \log(1+3\lambda^2), \end{aligned}$$

and in the case $r = \infty$ we have

$$\begin{aligned}\lim_{r\to\infty} \beta_r^{\mathrm{reg}}(\Delta_\lambda) &= \lim_{r\to\infty} \frac{2r - \log\left[(1+3\lambda)^r + 3(1-\lambda)^r\right]}{r-1} \\ &= \lim_{r\to\infty} \frac{2r - r\,\log(1+3\lambda)}{r-1} \\ &= 2 - \log(1+3\lambda)\,.\end{aligned}$$

(b) Define the functions

$$\begin{aligned} f(|\phi\rangle) &= \mathrm{Tr}\left(\Delta_\lambda^{\otimes n}(|\phi\rangle\langle\phi|)^r\right) \\ g(x) &= \frac{1}{n(1-r)}\log x\,, \end{aligned}$$

and the two series (notice the n dependency in the definition of f)

$$\begin{aligned} a_n &= g(\mathbb{E}f)\,, \\ b_n &= \mathbb{E}g\circ f\,. \end{aligned}$$

The only difference between a_n and b_n is the position at which the averaging over pure inputs $|\phi\rangle$ takes places. As a result, $a_n = \frac{1}{n}\beta_r(\Delta_\lambda^{\otimes n})$ contains the average moments that have been considered in previous sections, and $b_n = \frac{1}{n}S_r(\Delta_\lambda^{\otimes n})$ is the Renyi output entropy per systems. By Jensen's inequality and because the maximal output entropy is $\log 2 = 1$ we have the bounds $a_n \le b_n \le 1$ for any λ. Our goal is to prove that the two series have the same limit for $\lambda \in [0, c_r] \cup [d_r, 1]$.

From part (a) we know

$$\lim_{n\to\infty} a_n = \frac{r - \log \mathcal{Q}_{\max}}{r-1}\,.$$

First, choose $c(r)$ such that $2 \ge \left(\frac{1+3\lambda}{2}\right)^r + 3\left(\frac{1-\lambda}{2}\right)^r$ for all $\lambda \in [0, c]$. Then we have $\mathcal{Q}_{\max} = 2$, and so

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \overline{S}_r^{\mathrm{reg}}(\Delta_\lambda) = 1\,.$$

Now only (3.23) remains to be proved for the range $\lambda \in [d, 1]$ with d still to be determined.

In Proposition 32 in 3.3.2 we prove that the Lipschitz constant η of the map f is bounded by $\eta \le \sqrt{2}r\kappa^n$ for $\kappa = \lambda + \frac{1-\lambda}{\sqrt{2}}$. From Levy's Lemma (according to Lemma III.1 in [34]) we know that the values of f

concentrate around their average

$$\Pr(|f(|\phi\rangle) - \mathbb{E}f| > \alpha_n) \leq 4\exp\left(-C(k+1)\frac{\alpha_n^2}{\eta^2}\right) =: \epsilon_n \tag{3.24}$$

with $k = 2 \cdot 2^n - 1$ the (real) dimension of the sphere of input states, $C = (9\pi^3 \ln 2)^{-1}$, and we choose the deviation

$$\alpha_n = \frac{1}{2}N\left(\frac{\mathcal{Q}_{\max}}{2^r}\right)^n \approx \frac{1}{2}\mathbb{E}f$$

where $N \geq 1$ is the multiplicity of the maximum of $\mathcal{Q}(\alpha)$, that is, α_n is half of the dominant term of $\mathbb{E}f$ (see prove of Theorem 18). Now $\alpha_n \to 0$ (apart from the special case $\lambda = 1$) and $\mathbb{E}f - \alpha_n > 0$ which is required in a later step. To ensure concentration for large n the exponent $(k+1)\frac{\alpha_n^2}{\eta^2}$ needs to become large. Taking the $2n$-th root we get

$$\begin{aligned}\left((k+1)\frac{\alpha_n^2}{\eta^2}\right)^{1/2n} &= \left(2 \cdot \sqrt{2}^{2n}\frac{\frac{N}{4}\left(\frac{\mathcal{Q}_{\max}}{2^r}\right)^{2n}}{2r^2\kappa^{2n}}\right)^{1/2n} \\ &\approx \sqrt{2}\frac{\frac{\mathcal{Q}_{\max}}{2^r}}{\kappa}\end{aligned}$$

where the approximation becomes equality in the large n limit. If this term is larger than 1 we have the required divergence, this gives the following inequality

$$\frac{\mathcal{Q}_{\max}}{2^r} > \frac{\kappa}{\sqrt{2}}\,. \tag{3.25}$$

This condition gives the lower bound d_r for the λ values for which (3.23) holds. Both sides of the inequality are plotted in Figure 3.2. For some values of r the range parameters can be found in Table 3.1.

To transform (3.24) into a statement about the range of g – and whence about a_n – we need a condition that implies $|f(|\phi\rangle) - \mathbb{E}f| > \alpha_n$. We set $\alpha_n' = \frac{1}{n(r-1)}\frac{\alpha_n}{\mathbb{E}f - \alpha_n} > |g(\mathbb{E}f) - g(\mathbb{E}f - \alpha_n)| > |g(\mathbb{E}f) - g(\mathbb{E}f + \alpha_n)|$ where the second inequality follows from convexity of g. Then for any x we have

$$|g(\mathbb{E}f) - g(x)| > \alpha_n' \Rightarrow |\mathbb{E}f - x| > \alpha_n\,.$$

Therefore, the values of $g \circ f$ concentrate around $a_n = g(\mathbb{E}f)$

$$\Pr(|g \circ f(|\phi\rangle) - a_n| > \alpha_n') \leq \epsilon_n\,.$$

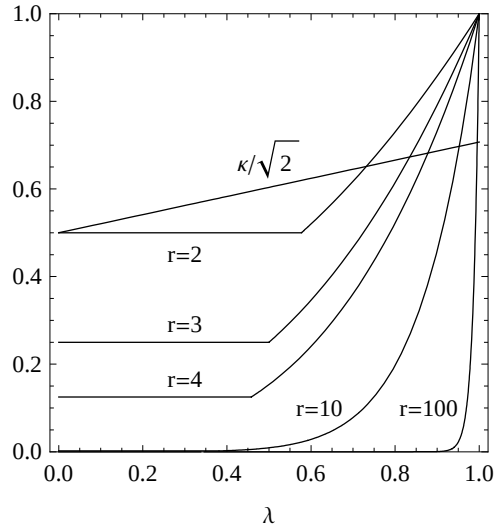


Figure 3.2: Plots of both sides of (3.25) for some r values to determine the range of validity.

For α'_n we calculate

$$\begin{aligned}\lim_{n\to\infty} \alpha'_n &= \lim_{n\to\infty} \frac{1}{n(r-1)} \frac{\alpha_n}{\mathbb{E}f - \alpha_n} \\ &= \lim_{n\to\infty} \frac{1}{n(r-1)} \frac{1/2}{1/2} \\ &= 0\,.\end{aligned}$$

To find an upper bound on b_n assume all $|\phi\rangle \in (g \circ f)^{-1}[a_n - \alpha'_n, a_n + \alpha'_n] = B$ map to the maximal value $a_n + \alpha'_n$ and all $|\phi\rangle \in B^c$ map to the maximal value 1. This gives a possible range for the average

$$b_n \in [a_n, (1-\epsilon_n)(a_n + \alpha'_n) + \epsilon_n \cdot 1]\,. \tag{3.26}$$

For $\lambda \in [d, 1]$ both ϵ_n and α'_n tend to 0 for large n and with (3.26) and because we know the limit of a_n we have

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \frac{r - \log \mathcal{Q}_{\max}}{r-1}\,.$$

□

3.2.6.3 Output entropy of random sequences

Considering sequences of random pure input states $|\phi_n\rangle$ with increasing dimension 2^n we have the following statement.

Proposition 22. *Let $|\phi_n\rangle \in \mathbb{C}^{2^n}$ be a sequence of random pure states and*

$$c_n = \frac{1}{n} S_r(\Delta_\lambda^{\otimes n}(|\phi_n\rangle\langle\phi_n|))$$

the sequence of output Renyi entropies per system. Then

$$c_n \overset{\text{a.s.}}{\to} \frac{r - \log \mathcal{Q}_{\max}}{r-1}$$

if λ is restricted as in Theorem 20.

Proof. Define $f_n(|\phi\rangle) = \frac{1}{n} S_r(\Delta_\lambda^{\otimes n}(|\phi\rangle\langle\phi|))$, the limit value $c = \frac{r - \log \mathcal{Q}_{\max}}{r-1} = \lim_{n\to\infty} \mathbb{E} f_n$, and $\delta_n = |\mathbb{E} f_n - c|$. Then for any $\epsilon > 0$

$$\begin{aligned} \Pr(|f_n(|\phi\rangle) - c| > \epsilon) &\leq \Pr(|f_n(|\phi\rangle) - \mathbb{E} f_n| > \epsilon - \delta_n) \\ &\leq 4 \exp\left(-C(k+1)\frac{(\epsilon - \delta_n)^2}{\eta^2}\right), \end{aligned} \tag{3.27}$$

with C, $k = 2 \cdot 2^n - 1$, and $\eta \leq \sqrt{2} r \kappa^n$ with $0 < \kappa < 1$ as in the proof of Theorem 20. If $\lambda \in J_r$ then $\lim \delta_n = 0$. When n becomes large then k becomes large, η becomes small, and $\epsilon - \delta_n$ is close to $\epsilon > 0$. Therefore the probabilities in (3.27) become small. Let N be such that for $n > N$ we have $\delta_n < \epsilon/2$. Then we have the bound

$$\begin{aligned} \sum_{n=1}^{\infty} \Pr(|f_n(|\phi_n\rangle) - c| > \epsilon) &\leq \sum_{n=1}^{\infty} 4 \exp\left(-C(k+1)\frac{(\epsilon - \delta_n)^2}{\eta^2}\right) \\ &\leq N + \sum_{n>N} 4 \exp\left(-C(k+1)\frac{(\epsilon/2)^2}{\eta^2}\right) \\ &\leq N + \sum_{n>N} 4 \exp\left(-\tilde{C}(2/\kappa^2)^n\right) \\ &< \infty, \end{aligned}$$

with $\tilde{C} = C \cdot 2 \cdot \frac{(\epsilon/2)^2}{(\sqrt{2}r)^2}$ independent of n. The Proposition follows by the lemma of Borel-Cantelli. □

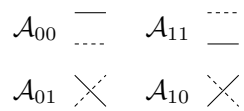


Figure 3.3: The diagrams for the $\mathcal{A}_{xy}$ matrices.

$\mathcal{A}_{00}\mathcal{A}_{01}\mathcal{A}_{10}\mathcal{A}_{11}\mathcal{A}_{10}$

Figure 3.4: A product of $\mathcal{A}_{xy}$ matrices and the corresponding diagram.

3.3 Proof of lemmas

3.3.1 Maximal $\mathcal{Q}$ for Δ_λ

Before proving Proposition 27 which is required for the proof of Theorem 20 we introduce some new notation and we present four lemmas. We only prove our Lemmas in two dimensions, but similar results will hold for higher dimensions.

For the most part we consider a channel $\mathcal{A}$ slightly more general than the depolarization channel with

$$\begin{aligned} \mathcal{A}_{00}&=\begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix} & \mathcal{A}_{11}&=\begin{pmatrix} \nu & 0 \\ 0 & \mu \end{pmatrix} \\ \mathcal{A}_{10}&=\begin{pmatrix} 0 & \kappa \\ \lambda & 0 \end{pmatrix} & \mathcal{A}_{01}&=\begin{pmatrix} 0 & \lambda \\ \kappa & 0 \end{pmatrix} \end{aligned},$$

where $\kappa, \lambda, \mu, \nu \in \mathbb{R}_0^+$ and $\mu \geq \nu$, $\lambda \geq \kappa$. We call this channel the *two-rail channel*. Remember that in the case $\mathcal{A} = \Delta_\lambda$ we have $\mu = \frac{1+\lambda}{2}$, $\nu = \frac{1-\lambda}{2}$ and $\kappa = 0$. Because all these matrix entries are positive, so are the $\mathcal{Q}$-terms, and therefore the largest positive $\mathcal{Q}$-term will yield $\mathcal{Q}_{\max}$.

We think of these matrices as the diagrams in Figure 3.3. We refer to the lines as rails and to their vertical position as their track (starting with track 0)

A product looks like the diagram in Figure 3.4, notice that we read from right to left, the same way that matrix multiplication applies. Consider a vector multiplying with this product, the diagram can be thought of as presenting two rails along which the two entries of the vector pass through to the left. On the way, continuous lines multiply the entries with factors μ or λ, dashed lines with a factor ν or κ. Because the products are inside a trace, they will only contribute, if the rail starting at the top on the right, ends on the top at the left (giving the $\langle 0| \ldots |0\rangle$ contribution), and the same for the rail starting at the bottom (giving

the $\langle 1|\ldots|1\rangle$ contribution). Using this, we can compare contributions to $\mathcal{Q}(\alpha)$ for different α.

We rewrite

$$\mathcal{Q}(\alpha) = \sum_{\delta\in\mathcal{D}(\alpha)} \mathrm{Tr}\,(\delta)$$

where $\mathcal{D}(\alpha)$ is the set of 2^r diagrams corresponding to α and we identify the diagram δ with the corresponding matrix.

Furthermore, for a diagram δ we define δ^1 to be the unchanged diagram and δ^{-1} to be the horizontally reflected diagram. If we take $\mathcal{D}_0(\alpha)$ to be the set of all diagrams in $\mathcal{D}(\alpha)$ that start high (at track 0) then obviously $\mathcal{D}(\alpha) = \mathcal{D}_0(\alpha) \cup \mathcal{D}_0^{-1}(\alpha)$ provides a convenient splitting of the sum in $\mathcal{Q}(\alpha)$.

We say an α is *non-overlapping* if all its cycles permute consecutive numbers, e.g. $\alpha = (123)(45)(678)$. For such α we write $\alpha = \alpha_1 \ldots \alpha_s$, where all the α_i are the cycles. We define a product for diagrams by simply concatenating them. With this product we get $\mathcal{D}(\alpha) = \mathcal{D}(\alpha_1)\ldots\mathcal{D}(\alpha_s)$.

Lemma 23. *Consider a diagram $\delta \in \mathcal{D}((1\ldots r))$ of a two-rail channel. Then we have*

$$\delta + \delta^{-1} \propto \mathbb{1}\,.$$

Proof. In any diagram $\delta \in \mathcal{D}((1\ldots r))$ one rail starts and ends at track 0 and the other starts and ends at track 1. That means the corresponding matrix is diagonal. Reflecting the diagram simply means exchanging the two diagonal entries, and if we sum $\delta + \delta^{-1}$ then the diagonal entries both have the same sum, i.e. it is proportional to unity. □

Lemma 24. *Let $\alpha = \alpha_1 \ldots \alpha_s \in Sym(r)$ be a non-overlapping permutation consisting of s cycles. Then the $\mathcal{Q}(\alpha)$ of a two-rail channel factors like*

$$\mathcal{Q}(\alpha) = \frac{1}{2^{s-1}}\mathcal{Q}(\alpha_1)\ldots\mathcal{Q}(\alpha_s)\,.$$

Proof. First we use the fact, that the diagrams for a non-overlapping α can be split between cycles, i.e. the sum splits into sums over separate diagrams,

$$\begin{aligned}
\mathcal{Q}(\alpha) &= \sum_{\delta\in\mathcal{D}(\alpha)} \mathrm{Tr}\,(\delta) \\
&= \sum_{\delta_1\in\mathcal{D}(\alpha_1)} \cdots \sum_{\delta_s\in\mathcal{D}(\alpha_s)} \mathrm{Tr}\,(\delta_1\ldots\delta_s) \\
&= \mathrm{Tr}\left(\sum_{\delta_1\in\mathcal{D}(\alpha_1)} \delta_1 \cdots \sum_{\delta_s\in\mathcal{D}(\alpha_s)} \delta_s\right).
\end{aligned}$$

Using the splitting $\mathcal{D}(\alpha) = \mathcal{D}_0(\alpha) \cup \mathcal{D}_0^{-1}(\alpha)$ and Lemma 23 we have

$$\sum_{\delta \in \mathcal{D}(\alpha_i)} \delta = \sum_{\delta \in \mathcal{D}_0(\alpha_i)} \delta + \delta^{-1} \propto \mathbb{1}$$

for all s cycles α_i. Therefore, we can split the single trace into s separate traces

$$\begin{aligned}\mathcal{Q}(\alpha) &= \frac{1}{2^{s-1}} \sum_{\delta_1 \in \mathcal{D}(\alpha_1)} \mathrm{Tr}\,(\delta_1) \cdots \sum_{\delta_1 \in \mathcal{D}(\alpha_s)} \mathrm{Tr}\,(\delta_s) \\ &= \frac{1}{2^{s-1}} \mathcal{Q}(\alpha_1) \ldots \mathcal{Q}(\alpha_s)\,.\end{aligned}$$

□

Lemma 25. *For $\mathcal{A} = \Delta_\lambda$ the depolarizing channel in two dimensions, $\mathcal{Q}$ restricted to non-overlapping permutations $\alpha \in Sym(r)$ is either maximal when $\alpha = (1 \ldots r)$ or when $\alpha = \mathrm{id}$.*

Proof. Remember from 3.2.6.1 that $\mathcal{Q}((1 \ldots r)) = (\mu + \lambda)^r + 3\nu^r$. It is convenient to introduce

$$f(x) = (\mu + \lambda)^x + 3\nu^x\,,$$

as a function with range $\mathbb{R}$. For a pure $\alpha = \alpha_1 \ldots \alpha_s$ according to Lemma 24 we get

$$\mathcal{Q}(\alpha) = \frac{1}{2^{s-1}} \prod_{i=1}^{s} f(|\alpha_i|)\,. \tag{3.28}$$

with $|\alpha_i|$ denoting the length of a cycle.

In the following we keep the number s of cycles and the total length r of the permutation invariant. Now, if we increase the length of one cycle and decrease the length of another, we are only changing two factors in the product, $f(x)f(k-x)$, where x is the length of the first cycle and k the (invariant) sum of the lengths of both cycles. We rewrite

$$\begin{aligned}f(x)f(k-x) &= ((\mu+\lambda)^x + 3\nu^x)((\mu+\lambda)^{k-x} + 3\nu^{k-x}) \\ &= (\mu+\lambda)^k + 9\nu^k + 3\nu^k \left(\frac{\mu+\lambda}{\nu}\right)^x + 3(\mu+\lambda)^k \left(\frac{\nu}{\mu+\lambda}\right)^x\,.\end{aligned}$$

Because of $\mu + \lambda, \nu \geq 0$ the function $f(x)f(k-x)$ is convex in x. Therefore it is maximal at the boundaries, i.e. when one cycle has the minimal length of 1. If we repeat this procedure $s-1$ times we end up with one large cycle of length $t = r - s + 1$ while all other cycles are of length 1. In every step we increase $\mathcal{Q}$, so we

get the bound

$$\mathcal{Q}(\alpha) \leq \mathcal{Q}((1)(2)\dots(s-1)(s\dots r)) = \frac{1}{2^{s-1}} f(1)^{s-1} f(t) = f(t) \tag{3.29}$$

for non-overlapping permutations α consisting of s cycles.

Now compare the upper bounds given by (3.29) for permutations of the same total length r but different number of cycles s. This is the same as varying t . Because f is convex we get a maximal upper bound if t is minimal or maximal. The minimal value $t = 1$ is achieved when $s = r$ and all the cycles are of length 1. Then (3.29) becomes an equality - there are no steps necessary in the maximization procedure - and we have $\mathcal{Q}(\mathrm{id}) = f(1) = 2$. The maximal value $t = r$ is achieved when α is simply one large cycle. Again (3.29) becomes equality, and $\mathcal{Q}(\alpha) = f(r)$. One of these upper bounds is the highest upper bound possible in (3.29), and because they are achieved by $\mathcal{Q}(\mathrm{id})$ and $\mathcal{Q}((1\dots r))$ we know that one of these $\mathcal{Q}(\alpha)$ is maximal over $\alpha \in Sym(r)$. □

Lemma 26. *Let $\mathcal{A}$ be a two-rail channel, and $[\beta]$ the conjugacy class of a permutation. Then $\mathcal{Q}$ restricted to $[\beta]$ is maximal on non-overlapping members $\alpha \in [\beta]$.*

Proof. Let $\alpha = \alpha_1 \dots \alpha_s$ be a non-overlapping member of the class and $\beta = \gamma\alpha\gamma^{-1}$ be any other member of the class. First, remember

$$\mathcal{Q}(\alpha) = \sum_{\substack{\{x_i\} \\ i=1\dots r}} \mathrm{Tr}\left(\prod_{i=1}^{r} \mathcal{A}_{x_i x_{\alpha(i)}}\right) = \sum_{\delta_1 \in \mathcal{D}(\alpha_1)} \cdots \sum_{\delta_s \in \mathcal{D}(\alpha_s)} \mathrm{Tr}\left(\delta_1 \dots \delta_s\right) . \tag{3.30}$$

With the first way of writing $\mathcal{Q}(\alpha)$ in mind we define a 1-1-mapping between terms in the sum of $\mathcal{Q}(\alpha)$ and $\mathcal{Q}(\beta)$ via a mapping of indices $x_i \to x_{\gamma^{-1}(i)}$ (or $x_{\gamma(i)} \to x_i$). Then the products of matrices are mapped like

$$\begin{aligned} \mathcal{A}_{x_1 x_{\alpha(1)}} \dots \mathcal{A}_{x_r x_{\alpha(r)}} &\to \mathcal{A}_{x_{\gamma(1)} x_{\gamma(\beta(1))}} \dots \mathcal{A}_{x_{\gamma(r)} x_{\gamma(\beta(r))}} \\ &= \mathcal{A}_{x_{\gamma(1)} x_{\alpha(\gamma(1))}} \dots \mathcal{A}_{x_{\gamma(r)} x_{\alpha(\gamma(r))}} . \end{aligned}$$

In terms of diagrams this corresponds to permuting the "tiles" (crossings or straight pieces) $\mathcal{A}_{x_i x_{\alpha(i)}}$ according to the permutation γ. Some examples are shown in Appendix B.1. In particular, this mapping does not change the number of any kind of tile $\mathcal{A}_{00}$, $\mathcal{A}_{01}$, $\mathcal{A}_{10}$ or $\mathcal{A}_{11}$.

Now, consider the second way of writing $\mathcal{Q}(\alpha)$ in (3.30) and split the sums over subdiagrams $\mathcal{D}(\alpha_i)$ according to $\mathcal{D}(\alpha_i) = \mathcal{D}_0(\alpha_i) \cup \mathcal{D}_0^{-1}(\alpha_i)$

$$\mathcal{Q}(\alpha) = \sum_{\delta_1 \in \mathcal{D}_0(\alpha_1)} \cdots \sum_{\delta_s \in \mathcal{D}_0(\alpha_s)} \sum_{\substack{\{t_i = \pm 1\} \\ i=1\dots s}} \mathrm{Tr}\left(\delta_1^{t_1} \dots \delta_s^{t_s}\right) .$$

We consider a subsum

$$\sum_{\substack{\{t_i=\pm 1\}\\ i=1\dots s}} \mathrm{Tr}\left(\delta_1^{t_1}\dots\delta_s^{t_s}\right)$$

for fixed subdiagrams $\delta_i \in \mathcal{D}_0(\alpha_i)$. In the following we will prove that the contribution of this subsum to $\mathcal{Q}(\alpha)$ is larger or equal to the contribution of the subsum of the corresponding diagrams to $\mathcal{Q}(\beta)$. From this it immediately follows that $\mathcal{Q}(\alpha) \geq \mathcal{Q}(\beta)$.

It follows a general proof of the inequality between corresponding subsums. For illustration one subsum is evaluated in full detail with diagrams in Appendix B.1.

For the non-overlapping permutation α in the subdiagrams δ_i all the straight lines and dashed lines are aligned, i.e. we have a weak and a strong rail. Let m_i be the number of crossings and n_i the number of straight pieces in δ_i. The strong rail in subdiagram δ_i contributes a factor $\lambda^{m_i}\mu^{n_i}$ and the weak rail contributes the factor $\kappa^{m_i}\nu^{n_i}$. Summing over reflections the subsum equals

$$2\prod_{i=1}^{s} \lambda^{m_i}\mu^{n_i} + \kappa^{m_i}\nu^{n_i}\,.$$

On the other hand, for the possibly overlapping β some of the tiles are permuted and for one particular subdiagram, not all the strong rail pieces might be on the same rail. Let $\tilde{m}_i$ be the number of crossings that are thus misaligned and $\tilde{n}_i$ the number of straight pieces that are misaligned. The two rails in subdiagram δ_i now contribute the factors $\lambda^{m_i-\tilde{m}_i}\kappa^{\tilde{m}_i}\mu^{n_i-\tilde{n}_i}\nu^{\tilde{n}_i}$ and $\lambda^{\tilde{m}_i}\kappa^{m_i-\tilde{m}_i}\mu^{\tilde{n}_i}\nu^{n_i-\tilde{n}_i}$. Summing over reflections the subsum adding to $\mathcal{Q}(\beta)$ equals

$$2\prod_{i=1}^{s} \lambda^{m_i-\tilde{m}_i}\kappa^{\tilde{m}_i}\mu^{n_i-\tilde{n}_i}\nu^{\tilde{n}_i} + \lambda^{\tilde{m}_i}\kappa^{m_i-\tilde{m}_i}\mu^{\tilde{n}_i}\nu^{n_i-\tilde{n}_i}\,.$$

Because $\lambda \geq \kappa$ and $\mu \geq \nu$ it follows that $\lambda^{m_i}\mu^{n_i} + \kappa^{m_i}\nu^{n_i} \geq \lambda^{m_i-\tilde{m}_i}\kappa^{\tilde{m}_i}\mu^{n_i-\tilde{n}_i}\nu^{\tilde{n}_i} + \lambda^{\tilde{m}_i}\kappa^{m_i-\tilde{m}_i}\mu^{\tilde{n}_i}\nu^{n_i-\tilde{n}_i}$ (the strong and weak rail dominate the two mixed rails), and hence we have the desired inequality between corresponding contributions to $\mathcal{Q}(\alpha)$ and $\mathcal{Q}(\beta)$. □

Proposition 27. *For $\mathcal{A} = \Delta_\lambda$ the depolarizing channel in two dimensions, $\mathcal{Q}(\alpha)$ is either maximal when $\alpha = (1\dots r)$ or when $\alpha = \mathrm{id}$.*

Proof. First, consider permutations that consist of non-overlapping α, Lemma 25 proves that either $\alpha = \mathrm{id}$ or $\alpha = (1\dots r)$ yields the maximum $\mathcal{Q}(\alpha)$ amongst these permutations. Every conjugacy class has a non-overlapping representant, and Lemma 26 states that these have maximal $\mathcal{Q}$-value. Therefore either $\alpha = \mathrm{id}$ or $\alpha = (1\dots r)$ yield the maximal $\mathcal{Q}$-value amongst all the permutations $\alpha \in Symm(r)$. □

3.3.2 Bound on Lipschitz constant

We derive an upper bound on the Lipschitz constant of the function $f : S^{2^{n+1}-1} \to \mathbb{R}$

$$f(|\phi\rangle) = \mathrm{Tr}\left(\Delta_\lambda^{\otimes n}(|\phi\rangle\langle\phi|)^r\right)$$

with respect to the Euclidean norm on $S^{2^{n+1}-1} \subset \mathbb{R}^{2^{n+1}}$.

We divide the function into four steps and prove bounds on the Lipschitz constants for each step. The splitting is $f = d \circ c \circ b \circ a$ with

$$\begin{aligned}
a &: |\phi\rangle \to |\phi\rangle\langle\phi| = \rho \\
b &: \rho \to \Delta_\lambda^{\otimes n}(\rho) = \rho' \\
c &: \rho' \to \text{eigenvalues of } \rho' = \vec{v} \\
d &: \vec{v} \to \sum_i v_i^r \,.
\end{aligned}$$

Let $\mathcal{M}_m$ be the space of complex $m \times m$ matrices containing the set of states $\mathcal{S}_m = \{\rho \in \mathcal{M}_m : \rho = \rho^*, \mathrm{Tr}\,\rho = 1\}$.

Lemma 28. *Let $a(|\phi\rangle) = |\phi\rangle\langle\phi|$ a map $\mathbb{C}^m \to \mathcal{M}_m$ where $\langle\phi|\phi\rangle = 1$. Then the Lipschitz constant of a with respect to the Euclidean norm in the domain and the Frobenius norm $\|M\|_2 = \mathrm{Tr}\,(MM^*)^{1/2}$ in the range is upper bounded by $\sqrt{2}$.*

Proof. For $\langle\phi|\phi\rangle = \langle\psi|\psi\rangle = 1$ set $c = \langle\phi|\psi\rangle$. Now

$$\||\phi\rangle - |\psi\rangle\|_2^2 = \langle\phi|\phi\rangle - \langle\phi|\psi\rangle - \langle\psi|\phi\rangle + \langle\psi|\psi\rangle = 2 - 2\Re(c)\,,$$

and

$$\||\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|\|_2^2 = \langle\phi|\phi\rangle^2 - 2|\,\langle\phi|\psi\rangle\,|^2 + \langle\psi|\psi\rangle^2 = 2 - 2|c|^2\,.$$

Because of $\Re(c) \leq |c|$ and the inequality derived as follows

$$\begin{aligned}
(1 - |c|)^2 &\geq 0 \\
2 - 2|c| &\geq 1 - |c|^2 \\
2(2 - 2|c|) &\geq 2 - 2|c|^2\,,
\end{aligned}$$

we arrive at $2(2 - 2\Re(c)) \geq 2 - 2|c|^2$ or $\sqrt{2}\||\phi\rangle - |\psi\rangle\|_2 \geq \||\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|\|_2$. □

Lemma 29. *Let $b(\rho) = \Delta_\lambda^{\otimes n}(\rho)$ a map $\mathcal{S}_{2^n} \to \mathcal{S}_{2^n}$, then the Lipschitz constant of b with respect to the Frobenius norm in domain and range is upper bounded by κ^n, where $\kappa = \lambda + \frac{1-\lambda}{\sqrt{2}}$ so $0 < \kappa < 1$ for $0 < \lambda < 1$.*

Proof. It is useful to use the notation

$$\Delta_\lambda^{\otimes n} = \sum_{J \subset \mathbb{Z}_n} \lambda^{n-|J|}(1-\lambda)^{|J|} \mathrm{Tr}_J \otimes \left(\frac{\mathbb{1}}{2}\right)^{\otimes |J|}$$

where Tr_J is the partial trace over the systems with indices in J. Notice that we use a loose notation of the tensor product as the systems that are partially traced out and replaced by the totally mixed states are not necessarily all on the right side of the tensor product.

We will bound the operator norm $\|\mathcal{A}\|_{op} := \sup_{\rho \in S} \frac{\|\mathcal{A}(\rho)\|_2}{\|\rho\|_2}$ where the supremum is over the set $S = \{|\phi\rangle\langle\phi| - |\psi\rangle\langle\psi| \,|\, |\phi\rangle, |\psi\rangle \in \mathcal{S}_{2^n-1}\}$. Because of

$$\|\mathcal{A}(\rho) - \mathcal{A}(\tau)\|_2 = \|\mathcal{A}(\rho - \tau)\|_2 \leq \|\mathcal{A}\|_{op} \|\rho - \tau\|_2$$

for a linear map $\mathcal{A}$ bounding $\|\mathcal{A}\|_{op}$ immediately gives a bound of the Lipschitz constant as well. Now

$$\begin{aligned}
\|\Delta_\lambda^{\otimes n}\|_{op} &\leq \sum_J \lambda^{n-|J|}(1-\lambda)^{|J|} \left\| \mathrm{Tr}_J \otimes \left(\frac{\mathbb{1}}{2}\right)^{\otimes |J|} \right\|_{op} \\
&\leq \sum_J \lambda^{n-|J|} \left(\frac{1-\lambda}{\sqrt{2}}\right)^{|J|} \\
&= \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k} \left(\frac{1-\lambda}{\sqrt{2}}\right)^k \\
&= \left(\lambda + \frac{1-\lambda}{\sqrt{2}}\right)^n = \kappa^n ,
\end{aligned}$$

where we used the bound

$$\begin{aligned}
\left\| \mathrm{Tr}_{\{1\ldots k\}} \otimes \left(\frac{\mathbb{1}}{2}\right)^{\otimes k} \right\|_{op}^2 &= \sup_{\rho \in S} \frac{\mathrm{Tr}\left(\mathrm{Tr}_{\{1\ldots k\}} \rho \otimes \left(\frac{\mathbb{1}}{2}\right)^{\otimes k}\right)^2}{\mathrm{Tr}\,\rho^2} \\
&= \mathrm{Tr}\left(\left(\frac{\mathbb{1}}{2}\right)^{\otimes k}\right)^2 \sup_{\rho \in S} \frac{\mathrm{Tr}\left(\mathrm{Tr}_{\{1\ldots k\}} \rho\right)^2}{\mathrm{Tr}\,\rho^2} \\
&= \left(\frac{1}{2}\right)^k .
\end{aligned} \tag{3.31}$$

The supremum in (3.31) was evaluated as follows. First, consider that $\rho = |\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|$ can be written as $\rho = \alpha|0\rangle\langle 0| - \alpha|1\rangle\langle 1|$ with $0 \leq \alpha \leq 1$, where $|0\rangle$ and $|1\rangle$ are orthonormal states. Then the supremum runs

over all possible orientations of $|0\rangle$ and $|1\rangle$

$$\begin{aligned}\sup_{\rho\in S}\frac{\operatorname{Tr}\left(\operatorname{Tr}_{\{1\ldots k\}}\rho\right)^2}{\operatorname{Tr}\rho^2} &= \sup_{|0\rangle,|1\rangle}\frac{\operatorname{Tr}\left(\operatorname{Tr}_{\{1\ldots k\}}(\alpha|0\rangle\langle 0|-\alpha|1\rangle\langle 1|)\right)^2}{\operatorname{Tr}(\alpha|0\rangle\langle 0|-\alpha|1\rangle\langle 1|)^2}\\ &= \sup_{|0\rangle,|1\rangle}\frac{\operatorname{Tr}\left(\operatorname{Tr}_{\{1\ldots k\}}(|0\rangle\langle 0|-|1\rangle\langle 1|)\right)^2}{\operatorname{Tr}(|0\rangle\langle 0|-|1\rangle\langle 1|)^2}\\ &= \sup_{|0\rangle,|1\rangle}\frac{\operatorname{Tr}\left(\operatorname{Tr}_{\{1\ldots k\}}(|0\rangle\langle 0|-|1\rangle\langle 1|)\right)^2}{2}\,.\end{aligned}$$

Now assume $\rho_0=\operatorname{Tr}_{\{1\ldots k\}}|0\rangle\langle 0|$ and $\rho_1=\operatorname{Tr}_{\{1\ldots k\}}|1\rangle\langle 1|$ are arbitrary density matrices. Then

$$\begin{aligned}\sup_{|0\rangle,|1\rangle}\frac{\operatorname{Tr}\left(\operatorname{Tr}_{\{1\ldots k\}}(|0\rangle\langle 0|-|1\rangle\langle 1|)\right)^2}{2} &= \sup_{\rho_0,\rho_1}\frac{\operatorname{Tr}(\rho_0-\rho_1)^2}{2}\\ &= \sup_{\rho_0,\rho_1}\frac{\operatorname{Tr}\left(\rho_0^2+\rho_1^2-2\rho_0\rho_1\right)}{2}=1\,.\end{aligned}$$

The last equality follows from the fact, that $\operatorname{Tr}\rho_{0,1}^2\leq 1$ and $\operatorname{Tr}\rho_0\rho_1\geq 0$. The suprema are achieved when $|\phi\rangle=|00\rangle$ and $|\psi\rangle=|11\rangle$ so that $\rho_0=|0\rangle\langle 0|$ and $\rho_1=|1\rangle\langle 1|$. □

Remark 30. Let $c:\mathcal{S}_m\to\mathbb{R}^m$ be the map that sends density matrices to their eigenvalues, ordered high to low. The fact that the Lipschitz constant of c is upper bounded by 1 is equivalent to the Hoffman-Wielandt inequality [36]

$$\|\rho-\tau\|_2\geq\|c(\rho)-c(\tau)\|_2\,.$$

Lemma 31. *Let $d(\vec{v})=\sum_i v_i^r$ a map from $\{\vec{w}\in\mathbb{R}_+^m|\sum_i w_i=1\}$ to $\mathbb{R}$. Then the Lipschitz constant of d is upper bounded by r.*

Proof. We have $\frac{\partial d}{\partial v_j}=rv_j^{r-1}$ so

$$\sup_{\vec{v}\in\operatorname{Dom}d}|\vec{\nabla}d|^2=\sup\sum r^2v_i^{2(r-1)}\leq r^2\,.$$

By integration we get

$$|d(\vec{v})-d(\vec{w})|\leq r\|\vec{v}-\vec{w}\|_2\,.$$

□

Proposition 32. *The Lipschitz constant η of $\operatorname{Tr}\left(\Delta_\lambda^{\otimes n}(|\phi\rangle\langle\phi|)^r\right)$, with respect to the Euclidean norm in the domain, is upper bounded by $\sqrt{2}r\kappa^n$, with κ as in Lemma 29.*

Proof. From $\mathrm{Tr}\left(\Delta_\lambda^{\otimes n}(|\phi\rangle\langle\phi|)^r\right) = d \circ c \circ b \circ a(|\phi\rangle)$ with a, b, c and d as defined in Lemmas/Remark 28-31. Thus, we simply combine the upper bounds of the lemmas and get the bound $\sqrt{2} \cdot \kappa^n \cdot 1 \cdot r$. □

3.4 Other results

3.4.1 $\mathcal{Q}((1 \dots r))$ for a more general qubit channel

For a channel $\mathcal{A}$ that maps that scales the Bloch sphere like

$$\vec{n} \to \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \vec{n}\,,$$

the Choi-Jamiolkowski representation is

$$Choi(\Delta_\lambda) = \begin{pmatrix} \mu & 0 & 0 & \lambda \\ 0 & \nu & \kappa & 0 \\ 0 & \kappa & \nu & 0 \\ \lambda & 0 & 0 & \mu \end{pmatrix},$$

i.e. this is a two-rail channel with $\mu = \frac{1+\lambda_3}{2}$, $\nu = \frac{1-\lambda_3}{2}$, $\lambda = \frac{\lambda_1+\lambda_2}{2}$ and $\kappa = \frac{\lambda_1-\lambda_2}{2}$. Therefore we have

$$\mathcal{Q}(\mathrm{id}) = \mathrm{Tr}\,\mathcal{A}(\mathbb{1})^r = \mathrm{Tr}\,\mathbb{1} = 2$$

and according to (3.21)

$$\begin{aligned} \mathcal{Q}((1 \dots r)) &= (\mu+\lambda)^r + (\mu-\lambda)^r + (\nu+\kappa)^r + (\nu-\kappa)^r \\ &= \left(\frac{1+\lambda_1+\lambda_2+\lambda_3}{2}\right)^r + \left(\frac{1-\lambda_1-\lambda_2+\lambda_3}{2}\right)^r \\ &+ \left(\frac{1+\lambda_1-\lambda_2-\lambda_3}{2}\right)^r + \left(\frac{1-\lambda_1+\lambda_2-\lambda_3}{2}\right)^r. \end{aligned}$$

Assuming $\mathcal{Q}_{\max} = \max\{2, \mathcal{Q}((1 \dots r))\}$ the output is maximally mixed if $|\lambda_1| + |\lambda_2| + |\lambda_3| \leq 1$. But this is exactly the condition for $\mathcal{A}$ to be entanglement breaking, according to Theorem 3 in [60].

3.4.2 $\mathcal{Q}((1\dots r))$ of Δ_λ for any dimension d

For the d-dimensional depolarizing channel we have $\mathcal{Q}(\mathrm{id}) = \mathrm{Tr}\,\mathbb{1} = d$ as usual. Furthermore, the Choi-Jamiolkowsi representation consists of two blocks, one d-dimensional block with diagonal entries $\mu = \frac{1+\lambda}{d}$ and off-diagonal entries λ, and another block that is $\nu = \frac{1-\lambda}{d}$ times identity on the other $d^2 - d$ dimensions. For example for $d = 3$ the representation looks like

$$Choi(\Delta_\lambda) = \begin{pmatrix} \mu & & & & \lambda & & & & \lambda \\ & \nu & & & & & & & \\ & & \nu & & & & & & \\ & & & \nu & & & & & \\ \lambda & & & & \mu & & & & \lambda \\ & & & & & \nu & & & \\ & & & & & & \nu & & \\ & & & & & & & \nu & \\ \lambda & & & & \lambda & & & & \mu \end{pmatrix}.$$

The eigenvalues of the first block are $\mu - \lambda$ with $(d-1)$-multiplicity and a single eigenvalue $\mu + (d-1)\lambda$. Therefore,

$$\begin{aligned} \mathcal{Q}((1\dots r)) &= \mathrm{Tr}\, Choi(\Delta_\lambda)^r \\ &= (\mu + (d-1)\lambda)^r + (d-1)(\mu - \lambda)^r + (d^2 - d)\nu^r \\ &= \left(\frac{1 + (d^2-1)\lambda}{d}\right)^r + (d^2 - 1)\left(\frac{1-\lambda}{d}\right)^r. \end{aligned}$$

This result agrees with the result for $d = 2$ found in 3.2.6.1. The critical value is $\lambda = \frac{1}{d+1}$, below this value the average output is maximally mixed.

3.5 Conclusion

We found a general explicit form for $\beta_r(\mathcal{A})$ depending on the function $\mathcal{Q}_\mathcal{A} : Sym(r) \to \mathbb{R}$. In the limit $n \to \infty$ the maximal term $\mathcal{Q}_{\max}$ is dominant in $\beta_r^{\mathrm{reg}}(\mathcal{A})$ and therefore the only relevant term. However, finding $\mathcal{Q}_{\max}$ is not easy in general. In the case of the qubit depolarizing channel we proved that $\mathcal{Q}_{\max} = \max\{2, \mathcal{Q}_{\Delta_\lambda}((1\dots r))\}$ and also that $\overline{\beta}_r(\Delta_\lambda) = \overline{S}_r^{\mathrm{reg}}(\Delta_\lambda)$ for some λ. For all $r \in \mathbb{N}$ and $\lambda \leq 1/3$ the regularized output entropy becomes 1. Because the typical high-dimensional random state is highly

entangled our result for $\lambda \leq 1/3$ agrees with the notion that the tensor product of an entanglement breaking channel "chops up" these highly entangled states and produces maximally mixed states with almost certain probability.

For future work it would be interesting to also study channels other than the depolarizing channel, especially channels that are not known to be additive. Also, it might be of interest to study quantities similar to $\overline{S}_r^{\text{reg}}$ with the same general procedure, for example $\mathbb{E}\left[\operatorname{Tr} \mathcal{A} \otimes \mathbb{1}(|\phi\rangle\langle\phi|)^r\right]$. It would be insightful to gain a better understanding of the typical output and input states by finding explicit examples that conform with the average output.

3.6 Acknowledgments

We thank the reviewer for important corrections and improvements which helped the quality and presentation of the paper.

Appendix B

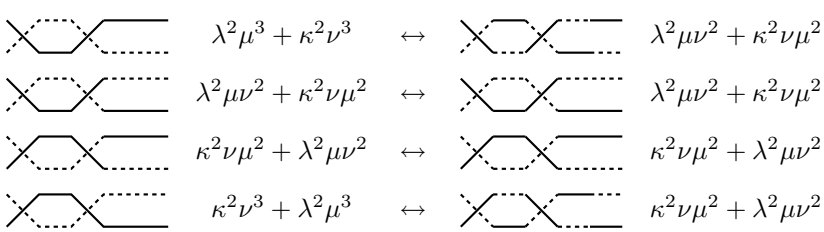
Appendix for Chapter 3

B.1 $\mathcal{Q}$ sum diagrams

We calculate an example of the subsums appearing in the proof of Lemma 26. Let $\alpha = (123)(45)$, $\beta = (143)(25)$ and $\gamma = (24)$. With this choice the correspondence of diagrams consists of switching tiles 2 and 4. The choice of indices

x_1	x_2	x_3	x_4	x_5
0	1	1	0	0

gives the diagrams and totals shown in Table B.1. The contribution to $\mathcal{Q}(\alpha)$ dominates in both factors as expected because α is non-overlapping.



$\lambda^2\mu^3 + \kappa^2\nu^3$	$\leftrightarrow$		$\lambda^2\mu\nu^2 + \kappa^2\nu\mu^2$
$\lambda^2\mu\nu^2 + \kappa^2\nu\mu^2$	$\leftrightarrow$		$\lambda^2\mu\nu^2 + \kappa^2\nu\mu^2$
$\kappa^2\nu\mu^2 + \lambda^2\mu\nu^2$	$\leftrightarrow$		$\kappa^2\nu\mu^2 + \lambda^2\mu\nu^2$
$\kappa^2\nu^3 + \lambda^2\mu^3$	$\leftrightarrow$		$\kappa^2\nu\mu^2 + \lambda^2\mu\nu^2$
Total for $\mathcal{Q}(\alpha)$		Total for $\mathcal{Q}(\beta)$	
$2(\lambda^2\mu + \kappa^2\nu)(\mu^2 + \nu^2)$		$2(\lambda^2\nu + \kappa^2\mu)(\mu\nu + \nu\mu)$	

Table B.1: Contributions to the sums $\mathcal{Q}(\alpha)$ and $\mathcal{Q}(\beta)$.

Chapter 4

Analyzing energy transfer in quantum networks using kinetic network approximations

Abstract

Coherent energy transfer in pigment-protein complexes has been studied by mapping the quantum network to a kinetic network. This gives an analytic way to find parameter values of optimal efficiency. In the case of the Fenna-Matthews-Olson (FMO) complex, the comparison of quantum and kinetic network evolution shows that dephasing-assisted energy transfer is driven by two-site coherent interaction, and not system-wide coherence. Using the Schur complement, we find a new kinetic network that gives a closer approximation to the quantum network. For both kinetic networks we study the system relaxation time, which gives a good estimate of the transfer time for strong trapping. We make mathematically rigorous estimates of the relaxation time when comparing kinetic and quantum network. Numerical simulations comparing the coherent model and the two kinetic network models, confirm our bounds, and show that the relative error of the new kinetic network approximation is several orders of magnitude smaller.

4.1 Introduction

Since coherent exciton transfer in the Fenna-Matthews-Olson complex (FMO) has been observed [24, 55, 70], extensive experimental and theoretical research has been dedicated to studying coherent resonant transfer

[18] and the coherent pigment-protein interaction[62, 42]. In particular, numerical solutions of simple models have shown that dephasing – the destruction of the coherences – at an intermediate rate helps to increase the energy transfer efficiency [56, 58]. This has been called dephasing- or environment-assisted energy transfer.

The models are based on a set of assumptions. First, only a single excitation is present, it is located at any of the seven pigments. The pigment excitation energy, and the pigment dipole-dipole interaction [15, 1] then lead to an oscillatory evolution of the system. And second, the site-environment interactions are assumed to be purely Markovian without any temporal or spatial correlations. The environment interactions are dephasing, recombination and trapping. Dephasing destroys the site coherences without destroying the exciton itself, it is an elastic environment interaction. Phonon recombination or photon re-emission lead to loss of the exciton to the environment. Trapping is the transfer of the exciton to the reaction center, where the electronic energy is converted to chemical energy, it occurs at pigment 3. The *transfer efficiency* is the probability that an exciton starting at site 1 or site 6 reaches the reaction center. Because the trapping rate at site 3 is very strong, the deciding factor for a good transfer efficiency is the speed at which the exciton spreads through the system, the *exciton relaxation time*.

Conceptually, the reason that an intermediate dephasing rate leads to a maximal transfer efficiency is as follows. For low dephasing, the system undergoes Rabi oscillations which, because of high site energy gaps, are only weakly mixing the site populations, therefore the transfer efficiency is low. For strong dephasing, the quantum Zeno effect occurs as the system is constantly measured in the site basis, trapping the exciton at one site and again preventing efficient transfer. Only for some intermediate region, dephasing and coupling can work together in the most constructive fashion, to allow optimal energy transfer.

To study population transfer channels and conditions for optimal transfer, a mapping to kinetic networks has been proposed [13, 41]. A kinetic network is a system where the exciton jumps incoherently between sites according to some fixed rates, i.e. a continuous-time Markov process. In its simplest version this approximation only takes into account the coherent interaction between pairs of sites to derive the transfer rate between them. If the two sites interact with strength V, have an energy separation E, and both sites experience dephasing at rate γ and population loss at rate κ then the rate is

$$\mu = \frac{2\left|V\right|^2(\gamma+\kappa)}{(\gamma+\kappa)^2+E^2}\,. \tag{4.1}$$

This rate is maximized for the intermediate dephasing rate $\gamma = E - \kappa$ so the phenomenon of dephasing-assisted transfer is maintained in this approximation. For a system with n sites, these rates constitute the

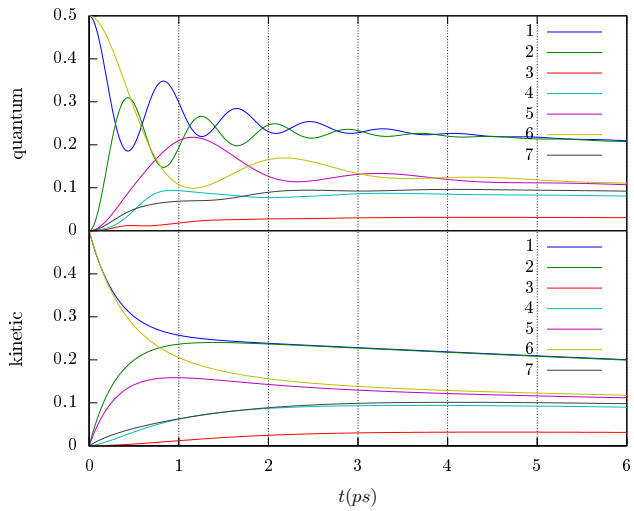


Figure 4.1: The initial evolution of quantum network and kinetic network N_0 for $\gamma = 50cm^{-1}$

off-diagonals of a rate $n \times n$-matrix N_0, and the system populations evolve according to

$$\dot{\vec{p}} = N_0 \vec{p}$$

where $\vec{p} \in \mathbb{R}^n$ is the time-dependent population vector. As the simulation for the FMO complex in Figure 4.1 shows, this simple network is surprisingly accurate in approximating the coherent network dynamics once the oscillations die down. Figure 4.2 displays the transfer efficiency for different γ, the dephasing-assisted regime clearly shows as a peak around $\gamma \approx 170cm^{-1}$. At the peak the coherent model is well approximated by N_0, therefore dephasing-assisted energy transfer can be explained by the relatively simple coherent dynamic between pairs of sites that enters the rate μ and the influence of system-wide coherence is small.

In this paper, we generalize the procedure of finding a kinetic network approximation in a mathematically appealing way using block matrices. We find a kinetic network matrix N that follows the coherent evolution much closer, it is over three orders of magnitude more precise than the network N_0 as shown in Figure 4.2. It can be expanded as

$$N = \sum_{k=0}^{\infty} N_k$$

where N_0 is the approximation described above, and the N_k are rate corrections due to coherent interactions via k intermediate sites. The expansion terms become smaller for increasing k proportional to the k-th power of a small parameter. By stopping the expansion at a finite k kinetic networks approximation of varying accuracy can be formed allowing the study of coherent interaction at different "scales" or number of involved sites. We restrict our further investigation to the dominant contribution N_0 and the entire sum N.

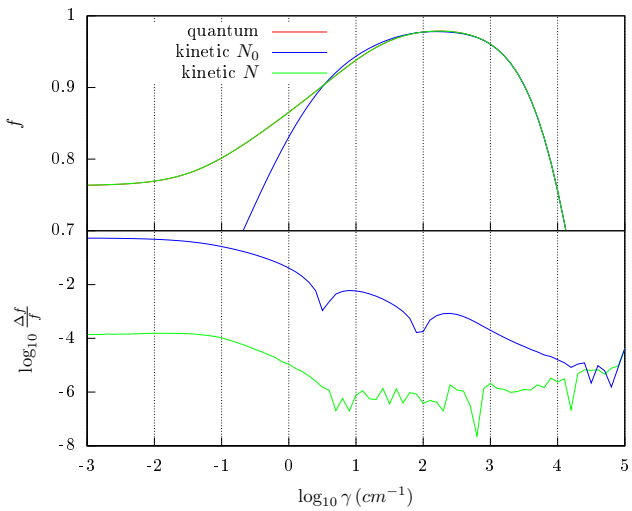


Figure 4.2: The efficiency of energy transfer in the FMO monomer. Model parameters are described in 4.6.3

For these two networks we study the limit in which their approximation to the quantum network becomes exact. We reduce the system complexity by ignoring exciton re-emission, recombination and trapping, such that dephasing is the only environment influence. For this system we bound the error in exciton relaxation time. In the case where the trapping rate at the site connected to the reaction center is very high, the rate of exciton decay due to re-emission or recombination competes with the rate of system relaxation. Thus finding the asymptotics of the difference in relaxation time between quantum and kinetic network is equivalent to finding the asymptotics of the difference in transfer efficiency.

4.1.1 Master equation

We consider the same quantum mechanical system studied in [58] with n sites carrying a single excitation which is equivalent to a system with n states/levels. The site energies are $E_k \in \mathbb{R}$ so the energy operator is

$$H = \sum_{k=1}^{n} E_k |k\rangle\langle k| \,.$$

The site k couples to site l with interaction strength $V_{kl} \in \mathbb{C}$ so the interaction operator is

$$V = \sum_{k \neq l} V_{kl} |k\rangle\langle l|$$

where $V_{kl} = \overline{V}_{lk}$. Site trapping, re-emission and recombination can be incorporated by an anti-hermitian operator A. Let κ_k be the combined rate of exciton loss at site k due to these effects, then A is defined as

$$A = \frac{-i}{2} \sum_{k=1}^{n} \kappa_k |k\rangle\langle k| \,.$$

Finally, every site is also under the influence of dephasing at rate $\gamma_k \geq 0$ incorporated in the Lindbladian superoperator

$$\mathcal{L}(\rho) = \sum_{k=1}^{n} L_k \rho L_k^\dagger - \frac{1}{2}\{\rho, L_k^\dagger L_k\}$$

with $L_k = \sqrt{\gamma_k}|i\rangle\langle i|$. Setting $\hbar = 1$, the single exciton manifold of the quantum network is described by the master equation

$$\dot{\rho} = -i[H + V, \rho] - i\{A, \rho\} + \mathcal{L}(\rho) \tag{4.2}$$

where square and curly brackets represent commutator and anti-commutator respectively.

For now we set $A = 0$, ignoring exciton depleting processes as explained above. We will mention how to include them in the kinetic network approximations later on. Our approximation becomes exact in the limit where the energy difference between sites is large, the dephasing is large and the interactions are small. To be specific, we introduce scaling parameters Γ and Θ and consider the limit $\Theta\Gamma^{-1} \to 0$. Energies and dephasing scale like Γ and interactions scale like Θ

$$E_k \propto \Gamma \,,$$

$$\gamma_k \propto \Gamma \,,$$

$$V_{kl} \propto \Theta \,.$$

With these assumptions the master equation turns into

$$\dot{\rho} = -i[\Gamma H + \Theta V, \rho] + \Gamma \mathcal{L}(\rho) \,. \tag{4.3}$$

Because this equation is linear in ρ it can be converted into vector form

$$\dot{\vec{\rho}} = M\vec{\rho}$$

where $\vec{\rho} \in \mathbb{R}^{n^2}$ is the density matrix in vector form and M is a real $n^2 \times n^2$-matrix. Two procedures to find M are presented in 4.2.1 and 4.2.7.

4.1.2 Transfer time, relaxation time and quantum efficiency

Let τ be the exciton relaxation time and $\mu = 1/\tau$ the rate of relaxation. We give rigorous definitions of these terms in the next section. Further, let κ_{T} be the trapping rate at which the exciton is captured by the reaction center and let κ_{L} be the rate at which the exciton is lost due to recombination and re-emission. Because the trapping rate is very large $\kappa_{\mathrm{T}} \gg \mu$ the rate of successful exciton transfer is approximately equal to the relaxation rate

$$\mu \approx (\mu^{-1} + \kappa_{\mathrm{T}}^{-1})^{-1} .$$

And hence the transfer efficiency is

$$f \approx \frac{\mu}{\mu + \kappa_{\mathrm{L}}} .$$

Furthermore, if the τ calculated for a kinetic network shows a small deviation of $\Delta\tau$ from the quantum network, then the deviation of kinetic network and quantum network efficiency is

$$\Delta f \approx f^2 \kappa_{\mathrm{L}} \Delta\tau .$$

Hence the asymptotic behavior we derive for $\Delta\tau$ translates directly to the asymptotic behavior for Δf.

4.1.3 Bounds

In [13], the kinetic network N_0 is derived by arguing that for strong dephasing γ the coherences decay fast, and can therefore be assumed to be stationary. However, upon light absorption the exciton is usually assumed to be localized on a few pigments only , and Figure 4.1 shows that the system initially shows a complex dynamic before the coherences decay. Our analysis takes into account any initial state ρ_0 with zero coherences, and any initial dynamic, by directly bounding the operator norms of differences between time evolution operators e^{Mt}, e^{Nt} and $e^{N_0 t}$.

To give a precise version of our bounds we have to make definitions of the quantities we analyze. We use the Euclidean norm $\|\vec{p}\|_2 = \sqrt{\sum_{i=1}^{n} p_i^2}$ to compare population vectors.

Definition 33.

1. The map $T : \mathbb{R}^{n^2} \to \mathbb{R}^n$ is the restriction of density vectors $\vec{\rho}$ to population vectors $\vec{p}$, and consequently $T^\dagger$ gives the embedding of population vector space in density vector space. In particular, if the first n components of $\vec{\rho}$ represent the site populations, then $T = (\mathbb{1}_n, 0_{n\times(n^2-n)})$.

2. The maximum relaxation time is

$$\tau = \max_{\vec{p}_0} \left\| \int_0^\infty e^{Nt}\vec{p}_0 - \frac{1}{n}\, dt \right\|_2$$

and the corresponding minimal relaxation rate is

$$\mu = \frac{1}{\tau}$$

3. The maximum deviation of relaxation time between the quantum network M and the kinetic network N is

$$\Delta\tau = \max_{\vec{p}_0} \left\| \int_0^\infty \left(Te^{Mt}T^\dagger - e^{Nt}\right)\vec{p}_0\, dt \right\|_2 .$$

4. Define τ_0, μ_0 and $\Delta\tau_0$ in the same way, replacing N with N_0.

For our bounds we always assume that our networks are *connected*, that means any two sits can exchange populations, at least via some intermediate sites (exact definition in 4.3.2) and that the site interactions are real. A system can always be decomposed into connected components, so the result generalizes easily. The second restriction makes our proofs simpler, but it is clear that the generalization to complex interactions could be treated in a similar manner to yield the same bounds.

Our first results shows how fast the relaxation time of the two kinetic networks N_0 and N approximate that of the quantum network M as $\Theta\Gamma^{-1}$ gets small.

Theorem 34. *Assume the network is connected, the site interactions are real $V_{kl} \in \mathbb{R}$ and that every site experiences dephasing $\gamma_k > 0$. Then V_{kl}, E_k and γ_k determine scaling invariant constants k_1 and k_2, and for $\Theta\Gamma^{-1}$ small enough we have the following bounds:*

1. *The relative difference of relaxation time between quantum evolution M and kinetic evolution N_0 is bounded by*

$$\Delta\tau_{0,\,\mathrm{rel}} = \Delta\tau_0/\tau_0 \leq k_1\Theta\Gamma^{-1} .$$

2. *The relative difference of relaxation time between quantum evolution M and kinetic evolution N is bounded by*

$$\Delta\tau_{\mathrm{rel}} = \Delta\tau/\tau \leq k_2\Theta^2\Gamma^{-2} .$$

This Theorem follows from Theorem 37 and Corollary 39 in Section 4.4.

We also find the following exponential bounds on the time dependence.

Theorem 35. *Assume the network is connected, the site interactions are real $V_{kl} \in \mathbb{R}$ and that every site experiences dephasing $\gamma_k > 0$. The V_{kl}, E_k and γ_k determine scaling invariant constants k_3, k_4 and k_5. Let $\vec{p}_0$ be any initial population distribution with total population 1. If $\Theta\Gamma^{-1}$ is small enough we have the following bounds:*

1. *For all times $t \geq 0$*

$$\left\| Te^{Mt}T^\dagger \vec{p}_0 - e^{N_0 t}\vec{p}_0 \right\|_2 \leq k_3 e^{-\mu_0 t/2} \cdot \Theta\Gamma^{-1} .$$

2. *For all times $t \geq 0$*

$$\left\| Te^{Mt}T^\dagger \vec{p}_0 - e^{Nt}\vec{p}_0 \right\|_2 \leq k_4 e^{-\mu t/2} \cdot \Theta^2\Gamma^{-2}(1 + k_5 \log \Theta\Gamma^{-1}) .$$

This Theorem follows from Theorem 40 and Corollary 42 in Section 4.5. We expect that more sophisticated methods might yield without the $\Theta^2\Gamma^{-2} \log \Theta\Gamma^{-1}$ term.

4.2 Kinetic networks

In this section we show how kinetic networks emerge naturally out of the study of the resolvent $(z - M)^{-1}$. First, we split M into block matrices separating populations and coherences. Then the population matrix block of the resolvent can be extracted and a simplification yields the network N. To separate the contributions of N into parts that scale like Θ and parts that scale like Γ we expand N in powers of $\Theta\Gamma^{-1}$, giving the series

$$N = \sum_{k=0}^{\infty} N_k$$

with the leading order contribution being N_0. Because M is a $n^2 \times n^2$ matrix we do not describe it in full detail here. However, in Appendix (C.1) we follow the procedure of deriving N_0 and N described below, giving the full expressions in the case $n = 3$.

4.2.1 Converting the master equation

We rewrite the master equation (4.3), skipping the scaling factors Θ and Γ, it is easy to reintroduce them at a later point

$$\dot{\rho} = -i[H + V, \rho] + \mathcal{L}(\rho) . \tag{4.4}$$

Our first goal is to convert this into the differential equation

$$\dot{\vec{\rho}} = M\vec{\rho}$$

for density "vectors" $\vec{\rho} \in \mathbb{R}^{n^2}$. Notice that because $\rho = \rho^\dagger$ the space of density matrix has n^2 real dimensions, so we are not using any information when mapping ρ to $\vec{\rho}$.

We use the following conversion:

1. The first n entries of the density vector are the populations – the real diagonal entries of ρ.

2. For the entries $n+1$ to n^2 we alternate between real and imaginary parts of the coherences – the off-diagonal entries of ρ – starting with entry ρ_{kl} where $k = 1$ and $l = 2$ continuing by increasing l until $l = n$, then moving to the entry ρ_{23}. We multiply all these entries with $\sqrt{2}$, a normalization factor useful to achieve simpler expressions later on.

In terms of index equations this is:

1. For $k = 1 \ldots n$

$$\vec{\rho}_k = \rho_{kk}\,.$$

2. For $k, l \in \{1, \ldots, n\}$ with $k < l$

$$\begin{aligned}\vec{\rho}_{n+2n(k-1)-k(k+1)+2l-1} &= \sqrt{2}\,\mathrm{Re}\,\rho_{kl}\\ \vec{\rho}_{n+2n(k-1)-k(k+1)+2l} &= \sqrt{2}\,\mathrm{Im}\,\rho_{kl}\,.\end{aligned}$$

Other mappings will yield the same kinetic networks, as long as they allow for an easy separation of population and coherence space.

While somewhat tedious, it is now relatively straightforward to find the matrix M such that

$$\dot{\vec{\rho}} = M\vec{\rho}.$$

To find the rows $k = 1 \ldots n$ we write out the diagonal components of the RHS of (4.4), and to find rows $k = n+1, \ldots, n^2$ we write out the off-diagonals of the RHS of (4.4). We follow this procedure explicitly for the case $n = 3$ in Appendix C.1. From there it is obvious how the procedure generalizes to larger n. Here we will only present the final form.

4.2.2 The coherent evolution matrix M

For simple notation and to simply extract the kinetic networks we split up the density vector space $\mathbb{R}^{n^2}$. Let $P = \mathbb{R}^n$ be the space of populations and let $C = \mathbb{R}^{n^2-n}$ be the space of coherences. We can then write density vectors as $\vec{\rho} = \begin{pmatrix} \vec{p} \\ \vec{c} \end{pmatrix}$ with $\vec{p} \in P$ and $\vec{c} \in C$. With this splitting the matrix M describing the quantum network looks like

$$M = \begin{pmatrix} 0 & -a^\dagger \\ a & b \end{pmatrix}$$

where $a : P \to C$ and $b : C \to C$ are real matrices (so $a^\dagger = a^\top$, but we'll keep the more general notation for later). Notice that the populations do not affect each other directly, but only via the coherences.

Matrix a describes how populations couple to coherences, its entries are real and imaginary parts of V_{kl}, naturally, site k will only couple to coherences kl with $l \neq k$, thus of the $(n^2 - n)$ entries in the k-th column of a only $2(n-1)$ are nonzero. Matrix b describes how coherences couple to other coherences, if considered as a block matrix with 2×2-blocks the diagonal block for the coherence between site k and site l is of the form

$$\begin{pmatrix} -\gamma_{kl} & -E_{kl} \\ E_{kl} & -\gamma_{kl} \end{pmatrix} \tag{4.5}$$

where

$$\gamma_{kl} = \frac{1}{2}(\gamma_k + \gamma_l)$$
$$E_{kl} = E_k - E_l \,.$$

The off-diagonal blocks consist of real and imaginary parts of V_{kl}.

From the form of M, when ignoring the off-diagonal blocks of b, we see that the site k couples to the site l via the coupling strength V_{kl}, then some mixture of γ_{kl} and E_{kl} and then again via the coupling strength V_{kl}. This reminds us of the rates of the form $\mu = \frac{2V^2\gamma}{\gamma^2+E^2}$ described in (4.1) that make up the matrix N_0. We will make this intuition precise is the next subsections.

4.2.3 Extracting the kinetic network N

To extract kinetic networks from M we consider its resolvent $(z-M)^{-1}$. Remember that for any holomorphic function f we have

$$f(M) = \frac{1}{2\pi i} \oint f(z)(z-M)^{-1}\, dz \,.$$

Therefore, if one can bound the resolvent appropriately, one can also bound the evolution operator e^{Mt} and other related quantities. Because we only care about approximating the population dynamics we restrict our view to the population block of the resolvent of M. The Banchiewicz formula [7] gives the inverse of a 2×2-block matrix. The first block of the inverse – in our case the population block – is called the Schur complement, and due to its basic nature has many applications in applied mathematics, statistics and physics [71]. Here we use it to "pull" the coherence dynamic back into population space. Only writing the Schur complement and skipping the other blocks of the resolvent we have

$$(z - M)^{-1} = \begin{pmatrix} (z - a^\dagger (b - z)^{-1} a)^{-1} & \cdot \\ \cdot & \cdot \end{pmatrix} .$$

Remember the operator T, the restriction to population space. With our choice of density vector basis it has the form

$$T = \begin{pmatrix} \mathbb{1}_n & 0_{n \times (n^2 - n)} \end{pmatrix} .$$

The difference of evolution for initial conditions $\vec{\rho}_0 = \begin{pmatrix} \vec{p}_0 \\ 0 \end{pmatrix} = T^\dagger \vec{p}_0$ (zero coherences) between quantum network M and kinetic network N is thus

$$\left(T e^{Mt} T^\dagger - e^{Nt} \right) \vec{p}_0 = \frac{1}{2\pi i} \oint e^{zt} \left((z - a^\dagger (b - z)^{-1} a)^{-1} - (z - N)^\dagger \right) dz \, .$$

For a good approximation we require

$$(z - a^\dagger (b - z)^{-1} a)^{-1} \approx (z - N)^\dagger \, . \tag{4.6}$$

At this point it is a small step to drop the second z on the LHS, in which case the formula becomes equality if we set

$$N = a^\dagger b^{-1} a \, .$$

To see intuitively that this approximation is good, consider the following. Matrix b contains terms proportional to Γ on its diagonal and terms proportional to Θ on its off-diagonal, matrix a is proportional to Θ, therefore

$$N \propto \Theta^2 \Gamma^{-1}$$

when $\Theta\Gamma^{-1}$ becomes small. For values of z that are smaller than eigenvalues of b the approximation (4.6) is good because $(b-z)^{-1} \approx b^{-1}$, for values of z larger than eigenvalues of b is good, because then z is much larger than the eigenvalues of N, and so both sides of (4.6) are approximately z^{-1}. This basic insight is what drives our bounds in Section 4.7.

4.2.4 Expanding N

As mentioned in 4.2.2, b consists of 2×2-blocks proportional to Γ on the diagonal and 2×2-blocks proportional to Θ on the off-diagonal. We separate this contributions defining

$$b = b_0 + \nu$$

where $b_0 \propto \Gamma$ and $\nu \propto \Theta$ is the block-diagonal and block-off-diagonal of b respectively. If $\Theta\Gamma^{-1}$ is small enough and if b_0 is invertible we can expand

$$b^{-1} = \sum_{k=0}^{\infty} b_0^{-1} \left(-\nu b_0^{-1}\right)^k .$$

This leads to the expansion

$$N = a^\dagger b^{-1} a = \sum_{k=0}^{\infty} N_k$$

with

$$N_k = a^\dagger b_0^{-1} \left(-\nu b_0^{-1}\right)^k a \,. \tag{4.7}$$

When using explicit forms of a, b_0 and ν one can see that the rates in N_k consist of corrections due to interactions via k intermediates. Roughly speaking, every of the $(k+1)$ sites along the chain contributes a factor of Θ, every of the k coherences (links) contributes a factor of Γ^{-1}, thus N_k scales like $\Theta^{k+1}\Gamma^{-k}$.

4.2.5 The network N_0

We now present the explicit form of

$$N_0 = a^\dagger b_0^{-1} a$$

the dominant contribution to N. We only show the crucial parts of the calculations that should make clear how to get the result for general n.

Notice that, from 4.2.4 and (4.5), it follows that b_0 is a $(n^2-n)\times(n^2-n)$ matrix with the only nonzero

entries being 2×2 blocks

$$\begin{pmatrix} -\gamma_{kl} & -E_{kl} \\ E_{kl} & -\gamma_{kl} \end{pmatrix}$$

along the diagonal. With the unitary transformation

$$U_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$$

we can diagonalize these 2×2 blocks. Hence, the entire matrix b_0 can be diagonalized by applying the transformation

$$U = \mathbb{1}_{(n^2-n)/2} \otimes U_0 \tag{4.8}$$

and

$$\tilde{b}_0 = U^\dagger b_0 U = \mathrm{diag}(\alpha_{12}, \bar{\alpha}_{12}, \alpha_{13}, \bar{\alpha}_{13}, \ldots, \bar{\alpha}_{n-1,n}) \tag{4.9}$$

with

$$\alpha_{kl} = -\gamma_{kl} + iE_{kl}$$

where diag denotes a diagonal matrix with given diagonal entries. In fact, U also helps to simplify a, consider the case $n = 3$

$$\tilde{a} = U^\dagger a = \begin{pmatrix} \overline{V}_{12} & -\overline{V}_{12} & \\ V_{12} & -V_{12} & \\ \overline{V}_{13} & & -\overline{V}_{13} \\ V_{13} & & -V_{13} \\ & \overline{V}_{23} & -\overline{V}_{23} \\ & V_{23} & -V_{23} \end{pmatrix}, \tag{4.10}$$

and the same happens for $\tilde{\nu} = U^\dagger \nu U$ (derivation in Appendix C.1). Notice that both $\tilde{b}_0$ and $\tilde{a}$ are complex matrices, still, we can use the transformed matrices $\tilde{a}$, $\tilde{b}_0$, and $\tilde{\nu}$ when finding explicit expressions for the real matrices N_k, because U cancels out. For example

$$\begin{aligned} N_0 &= a^\dagger b_0^{-1} a \\ &= a^\dagger U U^\dagger b_0^{-1} U U^\dagger a\,. \\ &= (U^\dagger a)^\dagger (U^\dagger b_0^{-1} U)^{-1} (U^\dagger a) \\ &= \tilde{a}^\dagger \tilde{b}_0^{-1} \tilde{a}\,. \end{aligned}$$

In the case $n = 3$ we get

$$N_0 = \begin{pmatrix} -\mu_{12} - \mu_{13} & \mu_{12} & \mu_{13} \\ \mu_{12} & -\mu_{12} - \mu_{23} & \mu_{23} \\ \mu_{13} & \mu_{23} & -\mu_{13} - \mu_{23} \end{pmatrix} \tag{4.11}$$

with

$$\mu_{kl} = \frac{2\,|V_{kl}|^2\,\gamma_{kl}}{\gamma_{kl}^2 + E_{kl}^2}\,.$$

The following simplified calculation illustrates how the rates μ_{kl} result from the matrix multiplication $\tilde{a}^\dagger \tilde{b}_0^{-1} \tilde{a}$

$$\begin{aligned} \begin{pmatrix} \bar{V} \\ V \end{pmatrix}^\dagger \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \bar{\alpha}^{-1} \end{pmatrix} \begin{pmatrix} -\bar{V} \\ -V \end{pmatrix} &= -V\bar{V}(\alpha^{-1} + \bar{\alpha}^{-1}) \\ &= \frac{2\,|V|^2\,\gamma}{\gamma^2 + E^2}\,. \end{aligned}$$

More generally for any n we have

$$(N_0)_{kl} = \mu_{kl} \tag{4.12}$$

for $i \neq j$ and

$$(N_0)_{kk} = -\sum_{l \neq k} \mu_{kl} \tag{4.13}$$

This is just the network described in [13] and the introduction.

4.2.6 Including re-emission, recombination and trapping

The population decreasing effects of re-emission, recombination and trapping can all be described by the rates κ_k of the diagonal anti-hermitian operator

$$A = \frac{-i}{2} \sum_{k=1}^{n} \kappa_k |k\rangle\langle k|$$

included in our general master equation (4.2). The contribution to the rate of change $\dot{\rho}$ is easily calculated

$$-i\{A, \rho\} = -\sum_{k,l} \frac{1}{2}(\kappa_k + \kappa_l)|i\rangle\langle j|\,,$$

and M becomes

$$M = \begin{pmatrix} c_1 & -a^\dagger \\ a & b + c_2 \end{pmatrix}$$

with the new contributions

$$c_1 = -\mathrm{diag}(\kappa_1, \kappa_2, \ldots, \kappa_n)$$

and

$$c_2 = -\mathrm{diag}(\kappa_{12}, \kappa_{12}, \kappa_{13}, \kappa_{13}, \ldots, \kappa_{n-1,n})$$

with $\kappa_{kl} = \frac{1}{2}(\kappa_k + \kappa_l)$ the rate that decreases the coherence of sites k and l. With this the networks become

$$\begin{aligned} N_0 &= a^\dagger (b_0 + c_2)^{-1} a + c_1 \\ N &= a^\dagger (b + c_2)^{-1} a + c_1 \\ N_k &= a^\dagger (b_0 + c_2)^{-1} \left(-\nu (b_0 + c_2)^{-1}\right)^k a \end{aligned} \tag{4.14}$$

which also hold with the replacements $a \to \tilde{a}$, $b \to \tilde{b}$ and $\nu \to \tilde{\nu}$, while leaving c_1 and c_2 unchanged.

The rates in N_0 can again be calculated directly

$$(N_0)_{kl} = \mu_{kl} = \frac{2\,|V_{kl}|^2\,(\gamma_{kl} + \kappa_{kl})}{(\gamma_{kl} + \kappa_{kl})^2 + E_{kl}^2}$$

for $k \neq l$ and

$$(N_0)_{kk} = -\kappa_k - \sum_{l \neq k} \mu_{kl}\,.$$

4.2.7 Numerical simulations

According to the last two subsections, network N_0 is easy to calculate directly, while network N and any k-site contribution N_k can be formed from the general definition of $\tilde{a}$, $\tilde{b}_0$ and $\tilde{\nu}$ (see Appendix (C.2)) which can be somewhat tedious. However, there is another approach related to numerical simulations. When running numerical calculations to simulate a complex master equation (4.2) on a software like Octave or Matlab, the need to convert the equation to the form

$$\dot{\vec{\rho}} = M\vec{\rho}$$

with a real M arises in any case. This can be done as we describe it in 4.2.1, or more easily – because we have the help of a computer – by defining an orthonormal density space basis. For example set

$$\begin{aligned}
\sigma_k &= |k\rangle\langle k| \\
\xi_{kl} &= (|k\rangle\langle l| + |l\rangle\langle k|)/\sqrt{2} \\
\eta_{kl} &= (-i|k\rangle\langle l| + i|l\rangle\langle k|)/\sqrt{2}
\end{aligned}$$

for $k < l$. Then matrix M can be formed by applying the master equation to those vectors and finding their coordinates. The following gives the population space block of M

$$M_{kl} = \mathrm{Tr}\left(\sigma_k^\dagger \mathcal{M}(\sigma_l)\right)$$

where $\mathcal{M}$ is the superoperator formed by the RHS of the master equation (4.2). Once the entire real matrix is found it is cut into population and coherence blocks

$$M = \begin{pmatrix} m_{PP} & m_{PC} \\ m_{CP} & m_{CC} \end{pmatrix}$$

and a generalized kinetic network of the same form as N is calculated as

$$N = m_{PC} m_{CC}^{-1} m_{CP} + m_{PP}\,.$$

Hence, if one has already calculated M in order to simulate a quantum network, it only takes a few steps to find the kinetic network approximation N.

4.3 Preliminaries

In this section we give some definitions and conditions. The conditions allow us to infer basic facts about the spectra of the operators N_0, N and M, which are required for all our bounds in Sections (4.4), (4.5) and (4.7).

4.3.1 Norm

Because for our bounds of the relaxation time we remove all population decreasing effects all evolutions M, N_0, and N leave the total population invariant. Therefore we split up the space of populations P. Set

$$\vec{e} = (1, 1, \ldots, 1)^\dagger / n \in P$$

the equal population vector. As we will prove in Proposition 36, as long as the network meets certain conditions, both quantum and kinetic evolutions will tend to $\vec{e}$ for any initial condition with total population 1. Consequently, we are only interested in the properties of our matrices in the space of population inequalities

$$I = \vec{e}^\top = \left\{ \vec{v} \middle| \sum_k v_k = 0 \right\} .$$

This is reflected in the norm we use, defines as follows. For $A : X_1 \to X_2$ where X_1 and X_2 are equal to I or C we define the operator norm as

$$\|A\| = \sup_{v \in X_1} \frac{\|Av\|_2}{\|v\|_2}$$

where $\|\cdot\|_2$ is the Euclidean norm. Hence, from now on, we think of our matrix blocks as

$$\begin{aligned} a :& I \to C \\ b :& C \to C \\ a^\dagger :& C \to I . \end{aligned}$$

Note that $\|a\|$ is the same if we maximize over I or P because $a\vec{e} = 0$, and that $\|a\| = \left\|a^\dagger\right\|$. Also, according to Proposition 36 $N_0 < 0$ on I and therefore N_0^{-1} is well-defined. The same holds for N. Define

$$\mu = \left\|N^{-1}\right\|^{-1}$$

to be the eigenvalue closest to 0 in N in I, define μ_0 the same way for N_0.

4.3.2 Conditions

For all our following bounds we have a set of conditions.

- First, we require that the network is *connected*, in the sense that any two sites k and l are coupled, at least via some intermediates, i.e. for some integer $p \geq 0$ there are sites m_j, $j = 1 \ldots p$ such that the

product

$$V_{km_1}V_{m_1m_2}\dots V_{m_{p-1}m_p}V_{k_pj}$$

is nonzero. This condition ensures that all sites can exchange population and the evolution ultimately converges to $\vec{e}$.

- Second, we require all the site dephasing rates to be strictly positive, $\gamma_k > 0$. This condition is essential for our approximation, as the coherences need to decay for the evolution M to become non-oscillatory. Notice that the limit $\Theta\Gamma^{-1} \to 0$ does not require that the dephasing rates get larger, but they will be much larger than the magnitude of eigenvalues of N or N_0, the population decay rates, because $\Gamma \gg \Theta^2\Gamma^{-1}$.

- Finally, we require that the V_{kl} are *real*. This ensures that N is symmetric and has a real spectrum (see Proposition 36), which allows simpler bounds in our proofs. While N_0 is always symmetric, we first compare the evolutions of M and N, and then the evolutions of N and N_0. Therefore we require this condition for both N_0 and N. We are confident that our methods would extend to the case of complex V_{kl}, but for the sake of clarity we restrict ourselves to the simpler case.

4.3.3 Inverse bounds

Or proofs consist mainly of using the following two bounds on the inverse on different parts of resolvents.

First, consider the Taylor series of the inverse close to 1, which for real numbers x gives

$$|(1+x)^{-1} - 1| \leq 2|x|$$

for $|x| \leq 1/2$. This is readily translated to a bound for operators

$$\left\|(A+B)^{-1} - A^{-1}\right\| \leq 2\left\|A^{-1}\right\|^2 \|B\| \tag{4.15}$$

for $\|B\| \leq \frac{1}{2}\left\|A^{-1}\right\|^{-1}$.

Second, if $A < -c < 0$ is a negative definite, self-adjoint, finite dimensional operator and $z \in \mathbb{C}$ with $\operatorname{Re} z \geq 0$ then

$$\left\|(z-A)^{-1}\right\| \leq c^{-1} \tag{4.16}$$

and

$$\left\|(z-A)^{-1}\right\| \leq |z|^{-1} \ . \tag{4.17}$$

these two bounds follow from the fact

$$\begin{aligned}\left\|(z-A)^{-1}\right\| &= \max\left\{|\lambda| \middle| \lambda \in \operatorname{Spec}(z-A)^{-1}\right\} \\ &= \max\left\{\left|(z-\lambda)^{-1}\right| \middle| \lambda \in \operatorname{Spec} A\right\} .\end{aligned}$$

4.3.4 Spectral properties

The following Proposition gives some basic facts about the spectra of the kinetic networks N and N_0. We will use these properties for the proofs of our bounds.

Proposition 36. *The matrices N_0 and N as defined in 4.2.5 and 4.2.3 have the following properties*

1. *N_0 is real and symmetric.*
2. *N is real.*
3. *If the interactions V_{kl} are real then N is symmetric.*
4. *$N_0\vec{e} = N\vec{e} = 0$*
5. *If $\gamma_k > 0$ and the network is connected network then $N_0 < 0$ on I.*
6. *For $\Theta\Gamma^{-1}$ small enough, $N_0 < -\mu/2$ and $N < -\mu_0/2$ on I.*

Proof. 1. These properties follow directly from the form in (4.12) and (4.13).

2. N is real because it is a product of a, b^{-1} and $a^\dagger$ which are also real.
3. If V_{kl} is real then $\tilde{a}$ (see (4.10))is real, so

$$N = \tilde{a}^\top \tilde{b}^{-1} \tilde{a}$$

Furthermore, $\tilde{b}^\top = \tilde{b}$ (see Appendix C.1) therefore

$$\begin{aligned}N^\top &= \tilde{a}^\top \left(\tilde{b}^{-1}\right)^\top \tilde{a} \\ &= \tilde{a}^\top \left(\tilde{b}^\top\right)^{-1} \tilde{a} \\ &= N .\end{aligned}$$

4. From (4.10) it is not hard to understand how $\tilde{a}$ looks for any n. One sees that the two rows for the coherence between sites k and l have exactly two non-zero entries, the first has V_{kl} and $-V_{kl}$, and the second has $\bar{V}_{kl}$ and $-\bar{V}_{kl}$. Therefore $\tilde{a}\vec{e} = 0$ and so $N_0\vec{e} = N\vec{e} = 0$.

5. For $\vec{v} \in I$ we have $\sum_k v_k = 0$. Now

$$\begin{aligned}\vec{v}^\dagger N_0 \vec{v} &= -\sum_{k<l} \mu_{kl}(v_k - v_l)^2 \\ &\leq 0\end{aligned}$$

The condition for equality is as follows. Because $\gamma_k > 0$ we have

$$V_{kl} \neq 0 \iff \mu_{kl} \neq 0\,.$$

Hence, because the network is connected, we have $v_k = v_l$ for all k and l and with $\sum_k v_k = 0$ it follows that $\vec{v} = 0$, thus $N_0 < 0$ on I.

6. Because $\mu_0 = \left\|N_0^{-1}\right\|^{-1}$ and $N_0 < 0$ we have $N_0 \leq -\mu_0$ on I. Note that $\mu \propto \Theta^2\Gamma^{-1}$ can grow as $\Theta\Gamma^{-1}$ gets small, so $\mu - \mu_0$ can grow in absolute value, however, as we now show the spectra of N and N_0 approach each other relative to their "size"

$$\mu - \mu_0 \ll \mu_0$$

We bound the distance of N and N_0 with the inverse bound. For $\Theta\Gamma^{-1}$ small enough we have $\|\nu\| \leq \frac{1}{2}\left\|b_0^{-1}\right\|^{-1}$ and we can apply (4.15) on

$$\left\|(b_0+\nu)^{-1} - b_0^{-1}\right\| \leq 2\left\|b_0^{-1}\right\|^2 \|\nu\|\,.$$

Now

$$\begin{aligned}\|N - N_0\| &= \left\|a^\dagger\left(b^{-1} - b_0^{-1}\right)a\right\| \\ &= \|a\|^2\, 2\left\|b_0^{-1}\right\|^2 \|\nu\|\,.\end{aligned}$$

So, the distance of N and N_0 is proportional to $\Theta^2\Gamma^{-2}\Theta$, and the eigenvalues in N_0 and N – in particular μ_0 and μ – are proportional to $\Theta^2\Gamma^{-1}$. Comparing the two gives

$$\Theta^2\Gamma^{-2}\Theta \ll \Theta^2\Gamma^{-1}$$

because $\Theta\Gamma^{-1} \ll 1$. That means the eigenvalues are approaching each other relative to their magnitude, in

particular N becomes negative definite like N_0, and

$$\frac{|\mu - \mu_0|}{\mu_0} \to 0\,.$$

Now it immediately follows that

$$N < -\mu < \mu_0/2$$
$$N_0 < -\mu_0 < \mu/2\,.$$

□

4.4 Bounding relaxation time error

We now give an explicit definition of relaxation time and the norms we use to control it. Then we derive bounds first comparing the quantum network M to the kinetic network N, and then comparing the kinetic networks N and N_0.

As a simple check of sanity consider the following. If we scale $\Gamma \propto s$ and $\Theta \propto s$ then also $M, N \propto s$ and time scales inversely $\Delta\tau, \tau \propto s^{-1}$. Therefore the relative error $\Delta\tau_{\text{rel}} = \Delta\tau/\tau$ stays unchanged and we expect bounds in terms of positive powers of $\Theta\Gamma^{-1}$. Our two bounds show exactly this behavior. The approximation of N to M is proportional to $\Theta^2\Gamma^{-2}$, while the approximation of N_0 to N is proportional to $\Theta\Gamma^{-1}$, combining the two approximations it follows that the approximation of N_0 to M is also proportional $\Theta\Gamma^{-1}$.

Note that all the results in this Section require the conditions in 4.3.2.

4.4.1 Relaxation time

By Proposition (36), the eigenvalues of N and N_0 on I are all negative for $\Theta\Gamma^{-1}$ small enough, so for any initial distribution $\vec{p}_0 \in I$

$$e^{Nt}\vec{p}_0 \to 0$$
$$e^{N_0 t}\vec{p}_0 \to 0$$

for large t. We can integrate

$$\int_0^\infty e^{Nt}\,dt = N^{-1}$$

and applying the operator norm maximizes the relaxation time for the kinetic network N over all possible population inequalities $\vec{p}_0 \in I$, set

$$\begin{aligned}\tau = \mu^{-1} &= \left\|N^{-1}\right\| = \left\|(a^\dagger b^{-1} a)^{-1}\right\| \\ &= \left\|\int_0^\infty e^{Nt}\,dt\right\|\end{aligned}$$

and in the same way we define $\tau_0 = \mu_0^{-1}$ for the network N_0.

We define the error in relaxation time as the relaxation time difference maximized over I

$$\Delta\tau = \left\|\int_0^\infty Te^{Mt}T^\dagger - e^{Nt}\,dt\right\| .$$

Hence, bounding $\Delta\tau$ means controlling the *worst possible* error in relaxation time when approximating M by N. The relative error is

$$\Delta\tau_{\text{rel}} = \Delta\tau/\tau$$

notice that we compare the worst possible relaxation time error to the longest possible relaxation time, those two do not necessarily occur for the same initial condition. We define $\Delta\tau_0$ and $\Delta\tau_{0,\,\text{rel}}$ in the same way, comparing N and N_0.

4.4.2 Resolvent difference

Converting the operator for the relaxation time error we get

$$\begin{aligned}\int_0^\infty Te^{Mt}T^\dagger - e^{Nt}\,dt &= TM^{-1}T^\dagger - N^{-1} \\ &= \frac{1}{2\pi i}\oint \frac{1}{z}\left(T\frac{1}{z-M}T^\dagger - \frac{1}{z-N}\right)dz \\ &= \frac{1}{2\pi i}\oint \frac{1}{z}\left(\frac{1}{z - a^\dagger(b-z)^{-1}a} - \frac{1}{z - a^\dagger b^{-1}a}\right)dz\,,\end{aligned}$$

where the complex integration follows a contour surrounding both $\operatorname{Spec} M$ and $\operatorname{Spec} N$. Define $S(z)$ to be the difference of the two resolvents

$$S(z) = \frac{1}{z - a^\dagger (b-z)^{-1} a} - \frac{1}{z - a^\dagger b^{-1} a}\,.$$

We now seek a bound on

$$\left\| \int_0^\infty T e^{Mt} T^\dagger - e^{Nt}\, dt \right\| = \left\| \frac{1}{2\pi i} \oint \frac{1}{z} S(z)\, dz \right\|\,. \tag{4.18}$$

4.4.3 Comparing the relaxation time of M and N

When bounding second order terms with the inverse bound we encounter

$$\kappa = \|a\|^2 \left\|b^{-1}\right\|^2$$

and κ_0 for the corresponding terms with b_0 instead of b. Notice the scaling behavior $\mu, \mu_0 \propto \Theta^2 \Gamma^{-1}$ and $\kappa, \kappa_0 \propto \Theta^2 \Gamma^{-2}$.

We will change the contour integration in (4.18) to be along the imaginary axis $z = iy$ for $y \in \mathbb{R}$. We prove the somewhat technical bounds on $S(iy)$ in Lemma 43 in Section 4.7.

Theorem 37. *If $\Theta\Gamma^{-1}$ is small enough then*

$$\Delta\tau = \left\| \int_0^\infty T e^{Mt} T^\dagger - e^{Nt}\, dt \right\| \leq \frac{4}{\pi} \kappa \mu^{-1} (1+\beta)$$

where $\beta > 0$ is the scaling independent constant from Lemma 43. This gives a bound on the relative error

$$\Delta\tau_{\text{rel}} = \Delta\tau/\tau \leq \frac{4}{\pi}\kappa(1+\beta) = k_2 \Theta^2 \Gamma^{-2}$$

where k_2 is scaling invariant.

Proof. We set the integration contour in (4.18) to be along the complex axis $z = iy$ for $y \in \mathbb{R}$ with y going from $-R$ to $+R$. We close the contour to the left in the half plane of negative real parts along a circle of radius R. According to Lemma 43, $S(z)$ has no poles with $\operatorname{Re} z \geq 0$ and so all poles lie within this contour for R large enough and $\Theta\Gamma^{-1}$ small enough. As R tends to infinity the integrand behaves like $\frac{1}{z^3}$ so the half-circle does not contribute to the integral. We can therefore change to complex integral to an integral in y over all of $\mathbb{R}$

$$\left\| \int_0^\infty T e^{Mt} T^\dagger - e^{Nt}\, dt \right\| = \left\| \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1}{iy} S(iy)\, dy \right\|\,.$$

Now split up the integral into two regions $|y| \leq \mu$ and $|y| \geq \mu$ and then use the corresponding bounds from Lemma 43. Choose $\Theta\Gamma^{-1}$ small enough so that $\mu < \alpha$ and use part 1 of the Lemma to bound

$$\begin{aligned}\left\| \int_{-\mu}^{\mu} \frac{1}{iy} S(iy)\, dy \right\| &\leq \int_{-\mu}^{\mu} \frac{1}{|y|} \cdot 4\kappa\mu^{-2}|y|\, dy \\ &\leq 8\kappa\mu^{-1}\,,\end{aligned}$$

and use part 2 of the Lemma to bound

$$\begin{aligned}\left\| \int_{\mu}^{\infty} \frac{1}{iy} S(iy)\, dy \right\| &\leq \int_{\mu}^{\infty} \frac{1}{|y|} \cdot 4\beta\kappa|y|^{-1}\, dy \\ &\leq 4\beta\kappa\mu^{-1}\,.\end{aligned}$$

Adding the two bounds gives the result

$$\begin{aligned}\left\| \int_0^{\infty} Te^{Mt}T^{\dagger} - e^{Nt}\, dt \right\| &\leq \left\| \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{1}{iy} S(iy)\, dy \right\| \\ &\leq \frac{1}{2\pi} 8\kappa\mu^{-1}(1+\beta)\,.\end{aligned}$$

□

4.4.4 Comparing the relaxation time of N and N_0

Theorem 38. *If $\Theta\Gamma^{-1}$ is small enough then*

$$\Delta\tau_1 = \left\| \int_0^{\infty} e^{Nt} - e^{N_0 t}\, dt \right\| \leq 4\kappa\mu^{-2} \|\nu\|$$

where μ and κ can also be replaced by μ_0 and κ_0. This gives a bound on the relative error

$$\Delta\tau_{1,\,\mathrm{rel}} = \Delta\tau_1/\tau \leq 4\kappa\mu^{-1} \|\nu\| = k_2'\Theta\Gamma^{-1}$$

where k_2' is scaling invariant.

Proof. In this case we don't need to bound the resolvent, instead we can evaluate the integral

$$\int_0^{\infty} e^{Nt} - e^{N_0 t}\, dt = N^{-1} - N_0^{-1}\,.$$

We use the inverse bound (4.15) twice. First, because $\|\nu\| \leq \frac{1}{2} \left\| b^{-1} \right\|^{-1}$ as long as $\Theta\Gamma^{-1}$ is small enough, we

can apply the bound on

$$\left\|(b-\nu)^{-1}-b^{-1}\right\| \leq 2\left\|b^{-1}\right\|^2 \|\nu\| \,. \tag{4.19}$$

Now, apply the bound again with $A = N$ and $B = N_0 - N$. The condition for B is

$$\begin{aligned}
\|B\| &\leq \|a\|^2 \left\|(b-\nu)^{-1}-b^{-1}\right\| \\
&\leq \|a\|^2 \, 2\left\|b^{-1}\right\|^2 \|\nu\| \\
&= 2\kappa \|\nu\| \\
&\leq 2\left\|A^{-1}\right\|^{-1} = 2\mu
\end{aligned}$$

where we used (4.19) in the second step. The last inequality is again achieved for $\Theta\Gamma^{-1}$ small enough because the two sides scale like

$$\Theta^2\Gamma^{-2}\Theta \leq \Theta^2\Gamma^{-1} \,.$$

Now it follows that

$$\left\|N^{-1}-N_0^{-1}\right\| \leq 2\left\|A^{-1}\right\|^2 \|B\| = 4\kappa\mu^{-2}\|\nu\|$$

as claimed. By switching the role of b and b_0 we receive the corresponding bound with κ_0 and μ_0. □

As a corollary we receive a bound on the relaxation time difference between the fully quantum mechanical evolution of M and the simple kinetic network evolution of N_0.

Corollary 39. *If $\Theta\Gamma^{-1}$ is small enough then for some scaling independent constant k_1*

$$\Delta\tau_{0,\,\mathrm{rel}} \leq k_1\Theta\Gamma^{-1}$$

Proof. According to Proposition 36 we have $|\mu-\mu_0|/\mu_0 \to 0$, and therefore there is a c such that

$$c \geq \tau/\tau_0$$

for $\Theta\Gamma^{-1}$ small enough. Then with Theorems 37 and 38 we have

$$\begin{aligned}\Delta\tau_{0,\,\mathrm{rel}} &= \Delta\tau_0/\tau_0 \\ &\leq (\Delta\tau + \Delta\tau_1)/\tau_0 \\ &\leq c(\Delta\tau + \Delta\tau_1)/\tau \\ &\leq c\left(\Delta\tau_{\mathrm{rel}} + \Delta\tau_{1,\,\mathrm{rel}}\right) \\ &\leq k_1\Theta\Gamma^{-1}\,.\end{aligned}$$

□

4.5 Bounding evolution error

In this chapter we bound the difference of time evolution operators for M, N and N_0. Our error bounds looks as follows

$$\left\|e^{Mt} - e^{Nt}\right\| \leq e^{-\mu t/2} \cdot X$$

where X is proportional to $\Theta^2\Gamma^{-2}$ up to a logarithmic term, and proportional to $\Theta\Gamma^{-1}$ if N is replaced with N_0. The logarithmic term appears due to intermediate times. It seems the integral over time performed in the last chapter seems to have conveniently guided us around that logarithm. As for the time dependence, using a shifting integration contour might give a bound like $e^{-\mu t}\mu t$, but a better control of the spectrum would be necessary to shift the contour close to $-\mu$ for long times.

As in the last chapter in 4.4.2, we write the evolution difference as a complex integral before we prove bounds

$$\left\|Te^{Mt}T^\dagger - e^{Nt}\right\| = \left\|\frac{1}{2\pi i}\oint e^{zt}S(z)dt\right\|\,. \tag{4.20}$$

Note that all the results in this Section again require the conditions in 4.3.2.

4.5.1 Comparing the evolution of M and N

We will change the contour integration in (4.20) to be parallel to the imaginary axis $z = iy - \mu/2$ for $y \in \mathbb{R}$. With this choice the exponential in the integral yields exponential decay at rate $\mu/2$. Again we give the technical bounds on $S(iy - \mu/2)$ in Lemma 44 in Section 4.7.

Theorem 40. *If $\Theta\Gamma^{-1}$ is small enough then for all $t \geq 0$ we have*

$$\left\|Te^{Mt}T^{\dagger} - e^{Nt}\right\| \leq e^{-\mu t} \cdot k_4 \Theta^2 \Gamma^{-2} \left(1 + k_5 \ln \Theta^{-1}\Gamma\right)$$

where k_4 and k_5 are a scaling independent constants.

Proof. We set the integration contour in (4.20) to parallel to the complex axis $z = iy - \mu/2$ for $y \in \mathbb{R}$ with y going from $-R$ to $+R$. We close the contour to the left in the half plane of negative real parts along a circle or radius R. According to Lemmas 43 and 44, $S(z)$ is bounded for $\operatorname{Re} z \geq -\mu/2$ and hence has no poles. Therefore all the poles lie within the contour for R large enough and $\Theta\Gamma^{-1}$ small enough. As R tends to infinity the integrand behaves like $\frac{1}{z^2} e^{\operatorname{Re} z}$ so the half-circle does not contribute to the integral. We can therefore change to complex integral to an integral in y over all of $\mathbb{R}$

$$\left\|Te^{Mt}T^{\dagger} - e^{Nt}\right\| = \left\| \frac{1}{2\pi} \int_{\mathbb{R}} e^{(iy-\mu/2)t} S(iy - \mu/2)\, dy \right\| .$$

Now split up the integral into three regions with $|y|$ in the intervals $[0, \mu]$, $[\mu, \hat{\alpha}]$ and $[\hat{\alpha}, +\infty)$ and then use the bounds from Lemma 44. Choose $\Theta\Gamma^{-1}$ small enough so that $\mu < \hat{\alpha}$ and use part 1 of the Lemma to bound

$$\begin{aligned} \left\| \int_0^{\mu/2} e^{(iy-\mu/2)t} S(iy - \mu/2)\, dy \right\| &\leq e^{-\mu t/2} \int_0^{\mu/2} 16\kappa\mu^{-2} |iy - \mu/2|\, dy \\ &\leq e^{-\mu t/2} 16\kappa \end{aligned}$$

and

$$\begin{aligned} \left\| \int_{\mu/2}^{\hat{\alpha}} e^{(iy-\mu/2)t} S(iy - \mu/2)\, dy \right\| &\leq e^{-\mu t/2} \int_{\mu/2}^{\hat{\alpha}} 4\kappa \cdot |y|^{-2} |iy - \mu/2|\, dy \\ &\leq e^{-\mu t/2} \int_{\mu/2}^{\hat{\alpha}} 4\kappa \cdot 2y^{-1}\, dy \\ &= e^{-\mu t/2} 8\kappa \ln(2\hat{\alpha}/\mu) , \end{aligned}$$

and we use part 2 of the Lemma to bound

$$\begin{aligned} \left\| \int_{\hat{\alpha}}^{\infty} e^{(iy-\mu/2)t} S(iy - \mu/2)\, dy \right\| &\leq e^{-\mu t/2} \int_{\hat{\alpha}}^{\infty} 4|y|^{-2} \|a\|^2 (b_{\min} - \mu/2)^{-1}\, dy \\ &\leq e^{-\mu t/2} 4\hat{\alpha}^{-1} \|a\|^2 (b_{\min} - \mu/2)^{-1} . \end{aligned}$$

Adding the three bounds gives the result

$$\left\|Te^{Mt}T^{\dagger} - e^{Nt}\right\| \leq e^{-\mu t/2} 4 \left(4\kappa + 2\kappa \ln(2\hat{\alpha}/\mu) + \hat{\alpha}^{-1} \|a\|^2 (b_{\min} - \mu/2)^{-1}\right) .$$

The middle term of the parenthesis has the worst scaling behavior

$$2\kappa \ln(2\hat{\alpha}/\mu) \propto \Theta^2 \Gamma^{-2} \ln \Theta^{-1}\Gamma$$

while the other two terms scale like $\Theta^2\Gamma^{-2}$. Therefore there are some scaling independent constants k_4 and k_5 such that

$$\left\|Te^{Mt}T^{\dagger} - e^{Nt}\right\| \leq e^{-\mu t/2} \cdot k_4 \Theta^2 \Gamma^{-2} \left(1 + k_5 \ln \Theta^{-1}\Gamma\right) .$$

□

4.5.2 Comparing the evolution of N and N_0

Theorem 41. *If $\Theta\Gamma^{-1}$ is small enough then for all $t \geq 0$ we have*

$$\left\|e^{Nt} - e^{N_0 t}\right\| \leq e^{-\mu t/2} \cdot k_4' \Theta\Gamma^{-1}$$

where k_4' is scaling independent, and where μ and κ can also be replaced by μ_0 and κ_0.

Proof. We are bounding the integral

$$\frac{1}{2\pi i} \oint e^{zt} \tilde{S}(z) dz$$

with resolvent difference

$$\tilde{S}(z) = \frac{1}{z - N} - \frac{1}{z - N_0} .$$

We use the same contour as in Proposition 40, $z = iy - \mu/2$. According to Proposition 36, all poles of $\tilde{S}(z)$ lie within this contour when $\Theta\Gamma^{-1}$ is small enough and R is large enough. Because of the e^{zt} factor and $T(z)$ tending to zero, the integral over the half-circle tends to 0 as R becomes large.

We bound $\tilde{S}(z)$ in much the same way that we bounded $S(z)$ in Lemma 44, however, the procedure is more straightforward. Set

$$X = (b - \nu)^{-1} - b^{-1} ,$$

because $\|\nu\| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$ we can use the inverse bound (4.15)

$$\|X\| \leq 2\left\|b^{-1}\right\|^2 \|\nu\| \ .$$

Now rewrite

$$\tilde{S}(z) = (z - a^\dagger b^{-1} a)^{-1} + (z - a^\dagger b^{-1} a - a^\dagger X a)^{-1} \ .$$

For any z with $\operatorname{Re} z = -\mu/2$ and for $\Theta\Gamma^{-1}$ small enough we have

$$\begin{aligned}\left\|a^\dagger X a\right\| &\leq 2\|a\|^2 \left\|b^{-1}\right\|^2 \|\nu\| \\ &\leq \frac{1}{2}\left\|z - a^\dagger b^{-1} a\right\|\end{aligned}$$

and so we can apply (4.15) again

$$\begin{aligned}\left\|\tilde{S}(z)\right\| &\leq 2\left\|(z - a^\dagger b a)^{-1}\right\|^2 \left\|a^\dagger X a\right\| \\ &\leq 4\left\|(z - a^\dagger b a)^{-1}\right\|^2 \kappa \|\nu\| \ .\end{aligned}$$

Now we apply inverse bounds (4.16) and (4.17) to receive the bounds

$$\left\|\tilde{S}(z)\right\| \leq 16\mu^{-2}\kappa \|\nu\| \tag{4.21}$$

$$\left\|\tilde{S}(z)\right\| \leq 4|z|^{-2} \kappa \|\nu\| \tag{4.22}$$

as long as $\operatorname{Re} z = -\mu/2$.

To estimate the integral

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{(iy-\mu/2)t} T(iy - \mu/2)\, dy$$

we split it into the two regions $|y| \leq \mu/2$ and $|y| > \mu/2$. Use (4.21) to bound

$$\left\|\int_{-\mu/2}^{\mu/2} e^{(iy-\mu/2)t} \tilde{S}(iy - \mu/2)\, dy\right\| \leq e^{-\mu t/2} \cdot \mu \cdot 16\mu^{-2}\kappa \|\nu\| \ ,$$

and use (4.22) to bound

$$2\left\|\int_{\mu/2}^{\infty} e^{(iy-\mu/2)t} \tilde{S}(iy - \mu/2)\, dy\right\| \leq 2e^{-\mu t/2} \cdot 8\mu^{-1}\kappa \|\nu\| \ .$$

Adding the two bounds gives

$$\begin{aligned}\left\|e^{Nt} - e^{N_0 t}\right\| &\leq 32 e^{-\mu t/2} \mu^{-1} \kappa \left\|\nu\right\| \\ &\leq e^{-\mu t/2} \cdot k_6 \Theta \Gamma^{-1} .\end{aligned}$$

where k_6 is scaling independent. The whole proof works just as well when exchanging μ with μ_0, κ with κ_0 giving a similar bound. □

As a corollary we receive a bound on the decay time difference between the fully quantum mechanical evolution of M and the simple kinetic network evolution of N_0.

Corollary 42. *If $\Theta\Gamma^{-1}$ is small enough then for all $t \geq 0$ we have*

$$\left\|Te^{Mt}T^{\dagger} - e^{N_0 t}\right\| \leq e^{-\mu t/2} \cdot k_3 \Theta \Gamma^{-1}$$

where k_3 is a scaling independent constant.

Proof. The bound follows from Theorems 40 and 41 and the fact that

$$\Theta^2 \Gamma^{-2} \ln \Theta^{-1} \Gamma \leq \Theta \Gamma^{-1}$$

for $\Theta\Gamma^{-1} \leq 1$. □

4.6 Applications

The rate of direct population exchange

$$\mu_{kl} = \frac{2\left|V_{kl}\right|^2 \gamma_{kl}}{\gamma_{kl}^2 + E_{kl}^2}$$

determines the strength of the link between sites k and l for the network N_0. Because of our condition that $\gamma_k > 0$, the network topology is fully determined by the V_{kl}, but the strength of the links is also affected by γ_k and E_k.

As applications, we consider two idealized networks. The first is a highly connected network where all sites are linked, the second is a circular chain where where only nearest neighbors are linked. We numerically calculate the relaxation times for the networks M, N_0 and N and compare the relative errors. Then we compare these networks to randomized networks with the same network topology. We also discuss the dimension dependence of our bounds from Sections 4.4 and again compare it to numerical simulations. All

the simulations agree with our bounds, but they show much room for improvement when considering large dimensions.

Finally, we discuss the FMO-complex and our model for which the results were shown in Figure 4.1 and 4.2.

For clarity of notation we recall that $\Delta\tau$, $\Delta\tau_0$ and $\Delta\tau_1$ are relaxation time differences between the network pairs $M-N$, $M-N_0$ and $N-N_0$ respectively. This only makes the discussion more precise, while generally $\Delta\tau_0$ and $\Delta\tau_1$ show the same dimension and scaling behavior, with small corrections to constants.

4.6.1 Highly connected network

Consider a highly connected network

$$V_{kl} = \Theta$$
$$E_k = 0$$
$$\gamma_k = \Gamma\,.$$

In Figure 4.3 we made a plot for the computed relative relaxation time differences $\Delta\tau_{\mathrm{rel}}$ and $\Delta\tau_{0,\,\mathrm{rel}}$ for different $\Theta\Gamma^{-1}$ with the initial state localized at site 1. Both axes plot logarithms, hence a straight line with slope n represents a $(\Theta\Gamma^{-1})^n$ proportionality.

The difference $\Delta\tau_{\mathrm{rel}}$ is too small to show any clear behavior. The difference $\Delta\tau_{0,\,\mathrm{rel}}$ is linear with slope approximately 2, hence the approximation is better than the slope 1 expected from Theorem 38. In the same figure we compare our idealized network to random networks where all V_{kl} are chosen randomly between 0 and Θ and all E_k are chosen randomly between 0 and Γ, hence they have the same topology. The magnitudes of the errors are similar for the range considered, but the slopes are different. All the samples show an error slope of 1 for $\Delta\tau_{0,\,\mathrm{rel}}$, while the error slope for $\Delta\tau_{\mathrm{rel}}$ is varying, but in most parts steeper than the slope of $\Delta\tau_{0,\,\mathrm{rel}}$. This behavior is closer to the behavior expected from our bounds. Generally, the agreement is about six orders of magnitude better for the network N than the network N_0.

For the ideal highly connected network we derive the quantities used in Theorem 37 and 38 analytically in Appendix C.3. The resulting bounds are

$$\Delta\tau_{\mathrm{rel}} \leq c_1 n \Theta^2 \Gamma^{-2}$$
$$\Delta\tau_{1,\,\mathrm{rel}} \leq c_2 n \Theta \Gamma^{-1}$$

for dimension and scaling independent constants c_1 and c_2. The simulation of M has a relatively high

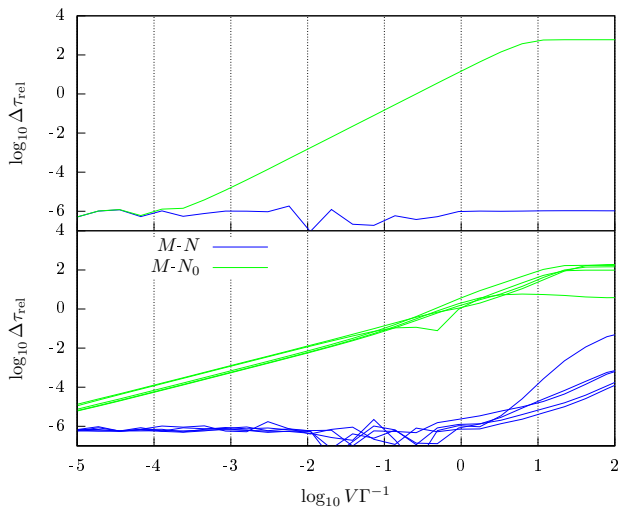


Figure 4.3: Relative error for the highly connected network

error and becomes slow very fast as n gets larger. Hence, we can only get meaningful results for $\Delta\tau_{1,\,\mathrm{rel}}$, the relaxation time difference of networks N and N_0. The result in Figure 4.4 actually shows that the difference increases with slope 2 or proportional to n^2. The reason is that in Theorem 38 we have the condition$\|\nu\| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$where the LHS is proportional to n and the RHS is constant (also discussed in the Appendix). If we increase the dimension at constant scaling, this condition and our bound break down. To still get a bound for large n we would need to readjust the scaling.

4.6.2 Linear network

Assume the sites are positioned on a circle and only nearest neighbors interact with strength Θ

$$V_{kl} = \begin{cases} \Theta & |k-l| = 1 \\ 0 & \text{else} \end{cases}$$

where we use the equivalence $n \equiv 0$. Further $\gamma_k = \Gamma$ and E_k such that $E_{kl} = \Gamma E$ when $|k-l| = 1$ which is possible for n even.

In Figure 4.5 we made a plot of the computed relative relaxation time differences $\Delta\tau_{\mathrm{rel}}$ and $\Delta\tau_{0,\,\mathrm{rel}}$ for different $\Theta\Gamma^{-1}$ with the initial state localized at site 1. Interestingly the quality of approximation by N_0 is improved over the highly connected model, while the quality of approximation by N has decreased. Also, both models show the same slope of about 2. We compare the ideal chain to random chains for which the V_{kl} that equal Θ in the idealized case are instead chosen randomly between 0 and Θ, and all E_k are chosen

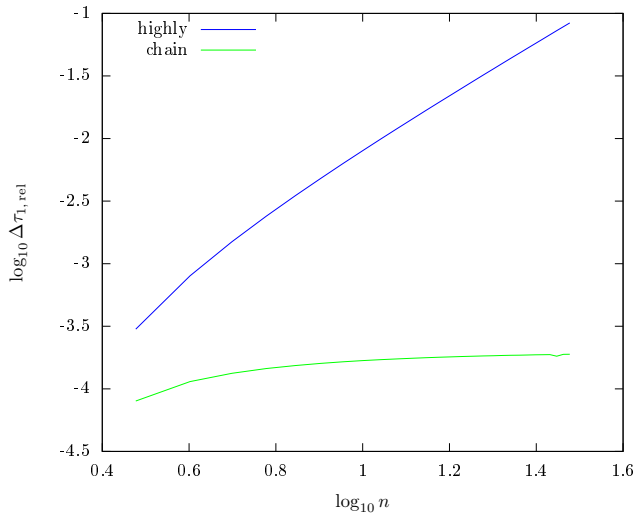


Figure 4.4: Relative errors between N and N_0 for the highly connected network and the cyclical chain with increasing dimension and $\Theta = 0.01$ and $\Gamma = 1$.

randomly between 0 and Γ. We get essentially the same behavior with all slopes being 2. That hints at a possible improvement of our bound in Theorem 38 in the case where the network is a chain, improving the proportionality from $\Theta\Gamma^{-1}$ to $\Theta^2\Gamma^{-2}$. Generally, the agreement is about five orders of magnitude better for the network N than the network N_0.

As in the last section, we can derive the necessary quantities for our bounds and get

$$\Delta\tau_{\mathrm{rel}} \leq c_3\Theta^2\Gamma^{-2}n^2$$

$$\Delta\tau_{1,\,\mathrm{rel}} \leq c_4\Theta\Gamma^{-1}n^2$$

for dimension and scaling independent constants c_3 and c_4. This time the condition$\|\nu\| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$ does not break down and the bounds hold for large dimensions as well. The n^2 terms are due to the lowest eigenvalue of N_0 being proportional to n^{-2}. This is a weakness of our strategy to use the operator norm for our bounds. Better bounds should be possible when only considering localized exciton as initial state. This initial state would the a superposition of all the eigenstates on N_0, and the average relaxation time would enter the bounds, instead of the longest relaxation time (the smallest eigenvalue of N_0).

As above we skip the simulation of M because the error is too large, and consider $\Delta\tau_{1,\,\mathrm{rel}}$ only. The result in Figure 4.4 shows that the difference seems to approximate a constant value for larger dimensions. So, both our bounds could be improved for large dimensions.

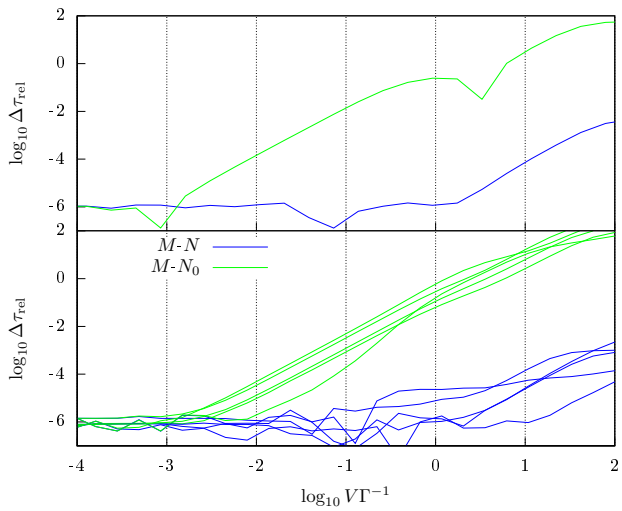


Figure 4.5: Relative error for the circular chain

4.6.3 The FMO-complex

The FMO-complex is pigment-protein with trimer structure. Each monomer contains seven bacteriochlorophyll a pigments that capture and transport light. The excitons start out at site 1 or 6 and the trapping occurs at site 3 [1], we set the initial state to be

$$\vec{p}_0 = (1/2, 0, 0, 0, 0, 1/2, 0)^\dagger \, .$$

We use the same numerical values as [58], with interactions and energies from [15]. The system Hamiltonian is

$$H + V = \begin{pmatrix} 280 & -106 & 8 & -5 & 6 & -8 & -4 \\ -106 & 420 & 28 & 6 & 2 & 13 & 1 \\ 8 & 28 & 0 & -62 & -1 & -9 & 17 \\ -5 & 6 & -62 & 175 & -70 & -19 & -57 \\ 6 & 2 & -1 & -70 & 320 & 40 & -2 \\ -8 & 13 & -9 & -19 & 40 & 360 & 32 \\ -4 & 1 & 17 & -57 & -2 & 32 & 260 \end{pmatrix}$$

with all the numbers in cm^{-1} (or $2.9978 \cdot 10^{10} s^{-1}$). Exciton recombination at rate $\kappa = 1ns^{-1}$ and reaction center trapping at rate $\kappa_3 = 1ps^{-1}$ enter the anti-hermitian operator

$$A = -\frac{i}{2}\left(\sum_k \kappa |k\rangle\langle k| + \kappa_3 |3\rangle\langle 3|\right) \, .$$

We use the same dephasing rate for every site $\gamma_k = \gamma$, and vary γ from 10^{-3} to $10^5 cm^{-1}$. Efficiency is calculated as

$$f = \kappa_3 \int_0^\infty \rho_{33}(t)\, dt$$

we calculated f for the three models in Figure 4.2. Peak efficiency is reached for $\gamma \approx 170 cm^{-1}$ close to the average energy gap along the chain which is $146 cm^{-1}$. The approximation N has less than 1% error, even for the lowest γ used, and the approximation N_0 gets below 1% error for $\gamma \approx 2 cm^{-1}$. Comparing this to our bounds we have

$$\|a\| = \left\|a^\dagger\right\| = 215 cm^{-1}$$

and for large γ

$$\left\|b^{-1}\right\|^{-1} = \gamma\,.$$

The numerical factor β is changing because of the changing ratio between energies and dephasing, for large γ however it is approximately equal to 100. Hence, our bound becomes

$$\Delta\tau_{\text{rel}} \lesssim 100 \left(215 cm^{-1}\gamma^{-1}\right)^2\,.$$

The 1% error margin is reached only when $\gamma = 21500 cm^{-1}$, so our numerical factors could certainly be much improved. But this is not unexpected, since our main goal was to find the leading behavior in $\Theta\Gamma^{-1}$.

We give N_0 for maximal transfer efficiency

$$N_0(\gamma = 170 cm^{-1}) = \begin{pmatrix} -80 & 79 & 0 & 0 & 0 & 1 & 0 \\ 79 & -82 & 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & -58 & 22 & 0 & 0 & 1 \\ 0 & 0 & 22 & -88 & 33 & 2 & 31 \\ 0 & 0 & 0 & 33 & -52 & 18 & 0 \\ 1 & 2 & 0 & 2 & 18 & -31 & 9 \\ 0 & 0 & 1 & 31 & 0 & 9 & -41 \end{pmatrix}.$$

It is interesting that the rate between sites 2 and 3 is actually smaller than the rate between sites 2 and 6 even though $|V_{23}| > |V_{26}|$. The reason is the large energy gap between sites 2 and 3 of $420 cm^{-1}$ while sites 2 and 6 have an energy gap of $60 cm^{-1}$. However, the values for site energies are still up to some debate [1, 15], and small changes can easily turn this behavior to the opposite again.

4.7 Resolvent difference bounds

The following three Lemmas are the main technical parts of our bounds. They all consist of bounding the operator norm of the resolvent difference

$$S(z) = \frac{1}{z - a^\dagger (b-z)^{-1} a} - \frac{1}{z - a^\dagger b^{-1} a}$$

for different values of z. Conceptually the bounding procedure is simple, we only employ the inverse bounds introduced in 4.3.3. Loosely speaking, if $|z| < \Gamma$ we can expand $(b-z)^{-1}$ and then the two terms in $S(z)$ only have a small difference in the denominator, so, using another inverse bound, they almost cancel. If $|z| > \Gamma$ then $|z| \gg \left\| a^\dagger b^{-1} a \right\|$ and we can directly use the second step from the case $|z| < \Gamma$.

Of course we also have to keep in mind where the poles of $S(z)$ are. According to Proposition 36 $(z-N)^{-1}$ has poles on the real axis below $-\mu$ which move according to the scaling $\Theta^2\Gamma^{-1}$. On the other hand $(z - a^\dagger (b-z)^{-1} a)^{-1}$ has poles close to the poles of $(z-N)^{-1}$ that approximately cancel each other, but it also has poles close to the eigenvalues of b which are approximately $\alpha_{ij} = -\gamma_{ij} + iE_{ij}$ and $\bar{\alpha}_{ij}$, scaling like Γ. Comparing the two sets of poles, the b-poles are much further to the left (negative real values) than the N-poles because $\Gamma \gg \Theta^2\Gamma^{-1}$. Our lemma steer clear of this poles by keeping $\operatorname{Re} z \geq -\mu/2$.

Lemma 43 contains bounds for $\operatorname{Re} z \geq 0$ which on the one hand ensures there are no poles on the right side of the complex plane, and on the other hand we use the bounds for $z = iy$ to bound the relaxation time. Lemma 44 contains bounds for the region $-\mu/2 \leq \operatorname{Re} z \leq 0$ the bounds are derived in a similar fashion as in Lemma 43, but there are some additional complications.

4.7.1 Bounds in the right half plane

Lemma 43. *If $\Theta\Gamma^{-1}$ is small enough and $\operatorname{Re} z \geq 0$ then $S(z)$ is bounded by*

1. $\|S(z)\| \leq 4\kappa\mu^{-2}|z|$ *if* $|z| \leq \alpha$, *where* $\alpha \propto \Gamma$ *depends on* a *and* b,
2. $\|S(z)\| \leq 4\beta\kappa|z|^{-1}$ *for any* z *with* $\operatorname{Re} z \geq 0$, *where* β *is a scaling independent constant depending on* a *and* b.

Proof. 1. Assume $\operatorname{Re} z \geq 0$ and $|z| \leq \alpha \propto \Gamma$, where

$$\alpha = \min\left\{ \frac{1}{2} \left\| b^{-1} \right\|^{-1}, \frac{1}{4}\kappa^{-1}\mu \right\}. \tag{4.23}$$

Set

$$X = (b-z)^{-1} - b^{-1}$$

and because $|z| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$ we can use (4.15) and have

$$\|X\| \leq 2\left\|b^{-1}\right\|^2 |z| .$$

Rewrite

$$S(z) = (z - a^\dagger b^{-1} a - a^\dagger X a)^{-1} - (z - a^\dagger b^{-1} a)^{-1} .$$

To use (4.15) on this expression notice that

$$|z| \leq \frac{1}{4}\kappa^{-1}\mu$$

and therefore

$$\begin{aligned}\left\|a^\dagger X a\right\| &\leq 2\kappa|z| \\ &\leq \frac{1}{2}\mu \\ &\leq \frac{1}{2}\left\|(z - a^\dagger b^{-1} a)^{-1}\right\|^{-1}\end{aligned}$$

where (4.16) was applied in the last step, using the fact that $a^\dagger b^{-1} a$ is self-adjoint from Proposition 36. This is just the condition for the bound

$$\begin{aligned}\|S(z)\| &\leq 2\left\|(z - a^\dagger b^{-1} a)^{-1}\right\|^2 \left\|a^\dagger X a\right\| \\ &\leq 4\kappa\left\|(z - a^\dagger b^{-1} a)^{-1}\right\|^2 |z|\end{aligned}$$

again using (4.16) and also (4.17) we get the bounds

$$\begin{aligned}\|S(z)\| &\leq 4\kappa\mu^{-2}|z| \\ \|S(z)\| &\leq 4\kappa|z|^{-1}\end{aligned} \tag{4.24}$$

for $|z| \leq \alpha$. The first bound is bound 1 of the Lemma, the second bound will be used below.

2. We now derive a bound when $|z| \geq \alpha$ and $\operatorname{Re} z \geq 0$, we will combine it with (4.24) to receive bound 2

for all $z \in \mathbb{R}$. If $\Theta\Gamma^{-1}$ is small enough then we have

$$\left\|a^\dagger b^{-1} a\right\| \leq \frac{\alpha}{2} \leq \frac{1}{2}|z|$$
$$\left\|a^\dagger (b-z)^{-1} a\right\| \leq \frac{\alpha}{2} \leq \frac{1}{2}|z|\,.$$

Where the latter inequality uses the fact that the spectrum of b approaches the spectrum of b_0 as $\Theta\Gamma^{-1}$ becomes small, and the spectrum of b_0, which is $-\gamma_{ij} \pm iE_{ij}$, has negative real part $-\gamma_{ij} < 0$. The last two inequalities are the conditions to use (4.15) and get the two bounds

$$\left\|(z - a^\dagger (b-z)^{-1} a)^{-1} - z^{-1}\right\| \leq 2|z|^{-2} \left\|a^\dagger (b-z)^{-1} a\right\|$$
$$\left\|(z - a^\dagger b^{-1} a)^{-1} - z^{-1}\right\| \leq 2|z|^{-2} \left\|a^\dagger b^{-1} a\right\|$$

set

$$b_{\min} = \min\left\{|\operatorname{Re}\lambda| | \lambda \in \operatorname{Spec} b\right\} \propto \Gamma \tag{4.25}$$

the closest any eigenvalue of b gets to the imaginary axis. Then $\left\|b^{-1}\right\| \leq b_{\min}^{-1}$ and $\left\|(b-z)^{-1}\right\| \leq b_{\min}^{-1}$ so

$$\|S(z)\| \leq 4|z|^{-2} \|a\|^2 d^{-1}\,.$$

Comparing to (4.24) with

$$\beta = \max\left\{1,\, 1/\left(\alpha b_{\min} \left\|b^{-1}\right\|^2\right)\right\} \propto 1$$

we have

$$4|z|^{-2} \|a\|^2 b_{\min}^{-1} \leq \beta \cdot 4\kappa |z|^{-1}$$

for $|z| \geq \alpha$ and therefore

$$\|S(z)\| \leq 4\beta\kappa|z|^{-1} \tag{4.26}$$

for all z with $\operatorname{Re} z \geq 0$. Which is bound 2 of the Lemma. □

4.7.2 Bounds parallel to the imaginary axis

The following Lemma establishes bounds along the imaginary axis $z = iy - \tilde{\mu}$. These bounds are used to prove the evolution bounds.

Lemma 44. *If we choose $\Theta\Gamma^{-1}$ small enough then for $0 \leq \tilde{\mu} \leq \mu/2$ the resolvent difference $S(iy - \tilde{\mu})$ is bounded by*

1. $\|S(iy-\tilde{\mu})\| \leq 16\kappa\mu^{-2}|iy-\tilde{\mu}|$ *and* $\|S(iy-\tilde{\mu})\| \leq 4\kappa|y|^{-2}|iy-\tilde{\mu}|$ *if* $|y| \leq \hat{\alpha}$, *where* $\hat{\alpha} \propto \Gamma$ *depends on* a *and* b,

2. $\|S(y)\| \leq 4|y|^{-2}\|a\|^2 (b_{\min}-\tilde{\mu})^{-1}$ *for* $|y| > \hat{\alpha}$ *with* $b_{\min} \propto \Gamma$.

Proof. We proceed almost identically as in the proof of Lemma 43 using the inverse bounds 4.15, (4.16) and (4.17) for the same parts of the resolvent terms.

1. We use the α from (4.23) to define

$$\hat{\alpha} = \min\left\{\frac{1}{2}\left\|b^{-1}\right\|^{-1}, \frac{1}{8}\kappa^{-1}\mu\right\} - \mu$$

notice that the scaling $\hat{\alpha} \propto \Gamma$ is only approximate and that $\Theta\Gamma^{-1}$ needs to be small enough such that $\hat{\alpha} > 0$. Now require $|y| \leq \hat{\alpha} \propto \Gamma$. Set

$$X = (b - iy + \tilde{\mu})^{-1} - b^{-1}$$

and because we have

$$\begin{aligned} |iy-\tilde{\mu}| &\leq |y| + \tilde{\mu} \\ &\leq \frac{1}{2}\left\|b^{-1}\right\|^{-1} - \mu + \tilde{\mu} \\ &\leq \frac{1}{2}\left\|b^{-1}\right\|^{-1} \end{aligned}$$

we can use (4.15) to get the bound

$$\|X\| \leq 2\left\|b^{-1}\right\|^2 |iy-\tilde{\mu}|.$$

Rewrite

$$S(iy-\tilde{\mu}) = (iy-\tilde{\mu}-a^\dagger b^{-1}a - a^\dagger X a)^{-1} - (iy-\tilde{\mu}-a^\dagger b^{-1}a)^{-1}.$$

To use (4.15) on this expression notice that we have

$$\begin{aligned} |iy-\tilde{\mu}| &\leq |y| + \tilde{\mu} \\ &\leq \left(\frac{1}{8}\kappa^{-1}\mu - \mu\right) + \tilde{\mu} \\ &\leq \frac{1}{8}\kappa^{-1}\mu \end{aligned}$$

and therefore

$$\begin{aligned}\left\|a^\dagger X a\right\| &\leq 2\kappa|iy-\tilde{\mu}| \\ &\leq \frac{1}{4}\mu \\ &\leq \frac{1}{2}\left\|(iy-\tilde{\mu}-a^\dagger b^{-1}a)^{-1}\right\|^{-1}\end{aligned}$$

where (4.16) was applied in the last step, using the fact that $\tilde{\mu}+a^\dagger b^{-1}a \leq -\mu/2$ from Proposition 36. This is just the condition for the bound

$$\begin{aligned}\|S(iy-\tilde{\mu})\| &\leq 2\left\|(iy-\tilde{\mu}-a^\dagger b^{-1}a)^{-1}\right\|^2\left\|a^\dagger X a\right\| \\ &\leq 4\kappa\left\|(iy-\tilde{\mu}-a^\dagger b^{-1}a)^{-1}\right\|^2|iy-\tilde{\mu}|\end{aligned}$$

again using (4.16) and also (4.17) we get the bounds

$$\begin{aligned}\|S(iy-\tilde{\mu})\| &\leq 16\kappa\mu^{-2}|iy-\tilde{\mu}| \\ \|S(iy-\tilde{\mu})\| &\leq 4\kappa|y|^{-2}|iy-\tilde{\mu}|\end{aligned}$$

for $|y| \leq \hat{\alpha}$. These are the bounds in part 1 of our Lemma.

2. We now derive a bound when $|y| \geq \hat{\alpha}$. If $\Theta\Gamma^{-1}$ is small enough then

$$\begin{aligned}\left\|a^\dagger b^{-1}a+\tilde{\mu}\right\| &\leq \frac{\hat{\alpha}}{2} \leq \frac{1}{2}|y| \\ \left\|a^\dagger(b-iy+\tilde{\mu})^{-1}a+\tilde{\mu}\right\| &\leq \frac{\hat{\alpha}}{2} \leq \frac{1}{2}|y|\,.\end{aligned}$$

The last two inequalities are the conditions to use (4.15) and get the two bounds

$$\begin{aligned}\left\|(iy-\tilde{\mu}-a^\dagger(b-iy+\tilde{\mu})^{-1}a)^{-1}-(iy-\tilde{\mu})^{-1}\right\| &\leq 2|y|^{-2}\left\|a^\dagger(b-iy+\tilde{\mu})^{-1}a\right\| \\ \left\|(iy-\tilde{\mu}-a^\dagger b^{-1}a)^{-1}-(iy-\tilde{\mu})^{-1}\right\| &\leq 2|y|^{-2}\left\|a^\dagger b^{-1}a\right\|\,.\end{aligned}$$

Use $b_{\min}$ from (4.25), giving

$$\begin{aligned}\left\|(b-iy+\tilde{\mu})^{-1}\right\| &\leq (b_{\min}-\tilde{\mu})^{-1} \\ \left\|(b-iy+\tilde{\mu})^{-1}\right\| &\leq (b_{\min}-\tilde{\mu})^{-1}\end{aligned}$$

and so

$$\|S(iy - \tilde{\mu})\| \leq 4|y|^{-2} \|a\|^2 (b_{\min} - \tilde{\mu})^{-1} .$$

for $|y| > \hat{\alpha}$, which is the bound in part 2 of our Lemma. □

4.8 Conclusion

We studied to kinetic networks that approximate the energy transfer in a quantum network subject to dephasing. The first network N_0 derives its rates only from nearest neighbor interactions, while the second N includes higher order corrections. We proved that the relaxation times are proportional to $\Theta\Gamma^{-1}$ and $\Theta^2\Gamma^{-2}$ respectively. Hence, the approximations are good if the interaction gets weak, or the dephasing and/or energy gaps get large. In the case of the FMO complex, both kinetic networks are good approximations in the regime of dephasing-assisted energy transfer. With simulations we found that the more complex kinetic network N provides approximations with a percentage error 5-6 magnitudes smaller than the simple kinetic network.

The study of these approximations could be extended in several ways. First, one could study the higher order corrections involved in N. Second, when the interactions V_{kl} are complex, N can be non-symmetric, meaning population exchange between sites is directed, this might relate to coherent cancellations along loops as mentioned in [13]. And finally, it would be interesting how our method of splitting population and coherence space to achieve kinetic network approximations could be generalized to other quantum networks and how it relates to existing models to approximate coherent evolution with incoherent statistical evolution.

4.9 Acknowledgments

I want to thank Chris King for his support, ideas and many useful discussions.

Appendix C

Appendix for Chapter 4

C.1 Three sites

In the following we write out parts of the master equation (4.3) for the case $n = 3$ and then derive the form of the matrix M. Then we explain how to generalize that form to higher n. For simplicity of notation we omit the scaling factors Γ and Θ, until we reach a block matrix expression. First note that with a standard calculation one finds $\mathcal{L}(\rho)$ to decrease the coherences in the manner

$$(\mathcal{L}(\rho))_{kl} = -\gamma_{kl}\rho_{kl}$$

where $k \neq l$ and $\gamma_{kl} = \frac{1}{2}(\gamma_k + \gamma_l)$ and $(\mathcal{L}(\rho))_{kk} = 0$. This gives a diagonal contribution $-\gamma_{kl}$ in the diagonal of the two rows corresponding to the real and imaginary part of ρ_{kl}.

Now, we evaluate the commutator

$$\left[\begin{pmatrix} E_1 & V_{12} & V_{13} \\ V_{21} & E_2 & V_{23} \\ V_{31} & V_{32} & E_3 \end{pmatrix}, \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix}\right].$$

From the 1x1 entry we get

$$
\begin{aligned}
\dot{\rho}_{11} &= -i(E_1\rho_{11} + V_{12}\rho_{21} + V_{13}\rho_{31} - E_1\rho_{11} - V_{21}\rho_{12} - V_{31}\rho_{13}) \\
&= -i(V_{12}\bar{\rho}_{12} + V_{13}\bar{\rho}_{13} - \overline{V}_{12}\rho_{12} - \overline{V}_{13}\rho_{13}) \\
&= 2\,\mathrm{Im}\,(V_{12}\bar{\rho}_{12} + V_{13}\bar{\rho}_{13}) \\
&= 2(-V_{12}^r\rho_{12}^i + V_{12}^i\rho_{12}^r - V_{13}^r\rho_{13}^i + V_{13}^i\rho_{13}^r)
\end{aligned}
$$

where superscripts r and i are shortcuts for real and imaginary parts, and from the 1x2 entry we get

$$
\begin{aligned}
\dot{\rho}_{12} &= -i(E_1\rho_{12} + V_{12}\rho_{22} + V_{13}\rho_{32} - V_{12}\rho_{11} - E_2\rho_{12} - V_{32}\rho_{13}) - \gamma_{12}\rho_{12} \\
&= -i((E_1 - E_2)\rho_{12} + V_{12}\rho_{22} - V_{12}\rho_{11} + V_{13}\bar{\rho}_{23} - \overline{V}_{23}\rho_{13}) - \gamma_{12}\rho_{12}
\end{aligned}
$$

with real and imaginary parts

$$
\begin{aligned}
\dot{\rho}_{12}^r &= -V_{12}^i\rho_{11} + V_{12}^i\rho_{22} - \gamma_{12}\rho_{12}^r + (E_1 - E_2)\rho_{12}^i + V_{23}^i\rho_{13}^r - V_{23}^r\rho_{13}^i + V_{13}^i\rho_{23}^r - V_{13}^r\rho_{23}^i \\
\dot{\rho}_{12}^i &= V_{12}^r\rho_{11} - V_{12}^r\rho_{22} - (E_1 - E_2)\rho_{12}^r - \gamma_{12}\rho_{12}^i + V_{23}^r\rho_{13}^r + V_{23}^i\rho_{13}^i - V_{13}^r\rho_{23}^r - V_{13}^i\rho_{23}^i\,.
\end{aligned}
$$

From these results we can read off lines 1, 4 and 5 of the following matrix and fill in the remaining lines in the same fashion

$$
M = \left(\begin{array}{ccc|cccccc}
 & & & \sqrt{2}V_{12}^i & -\sqrt{2}V_{12}^r & \sqrt{2}V_{13}^i & -\sqrt{2}V_{13}^r & & \\
 & 0 & & -\sqrt{2}V_{12}^i & \sqrt{2}V_{12}^r & & & \sqrt{2}V_{23}^i & -\sqrt{2}V_{23}^r \\
 & & & & & -\sqrt{2}V_{13}^i & \sqrt{2}V_{13}^r & -\sqrt{2}V_{23}^i & \sqrt{2}V_{23}^r \\
\hline
-\sqrt{2}V_{12}^i & \sqrt{2}V_{12}^i & & -\gamma_{12} & E_{12} & V_{23}^i & -V_{23}^r & V_{13}^i & -V_{13}^r \\
\sqrt{2}V_{12}^r & -\sqrt{2}V_{12}^r & & -E_{12} & -\gamma_{12} & V_{23}^r & V_{23}^i & -V_{13}^r & -V_{13}^i \\
-\sqrt{2}V_{13}^i & & \sqrt{2}V_{13}^i & -V_{23}^i & -V_{23}^r & -\gamma_{13} & E_{13} & V_{12}^i & V_{12}^r \\
\sqrt{2}V_{13}^r & & -\sqrt{2}V_{13}^r & V_{23}^r & -V_{23}^i & -E_{13} & -\gamma_{13} & -V_{12}^r & V_{12}^i \\
 & -\sqrt{2}V_{23}^i & \sqrt{2}V_{23}^i & -V_{13}^i & V_{13}^r & -V_{12}^i & V_{12}^r & -\gamma_{23} & E_{23} \\
 & \sqrt{2}V_{23}^r & -\sqrt{2}V_{23}^r & V_{13}^r & V_{13}^i & -V_{12}^r & -V_{12}^i & -E_{23} & -\gamma_{23}
\end{array}\right)
$$

where we define $E_{ij} = E_i - E_j$. Written as a block matrix

$$M = \begin{pmatrix} 0 & -a^\dagger \\ a & b \end{pmatrix}$$

one can see the explicit form of the matrices a, and b. Remember that we also separated b into two parts. We set the 2x2-block diagonal that scales like Γ (the E_{ij} and γ_{ij} entries) to be b_0 and we set the block-off-diagonal that scales like Θ (all the V_{ij} entries) to be ν. So $b = b_0 + \nu$.

In 4.2.5 in (4.8) we defined a transformation U to diagonalize b_0, if we extend this transformation to the entire space $P \oplus C$ as

$$\hat{U} = \mathbb{1}_n \oplus U$$

we can apply it to M directly and get

$$\tilde{M} = \hat{U}^\dagger M \hat{U} = \left(\begin{array}{ccc|cccccc}
 & & & -V_{12} & -\overline{V}_{12} & -V_{13} & -\overline{V}_{13} & & \\
 & 0 & & V_{12} & \overline{V}_{12} & & & -V_{23} & -\overline{V}_{23} \\
 & & & & & V_{13} & \overline{V}_{13} & V_{23} & \overline{V}_{23} \\
\hline
\overline{V}_{12} & -\overline{V}_{12} & & \alpha_{12} & & -iV_{23} & & & -i\overline{V}_{13} \\
V_{12} & -V_{12} & & & \overline{\alpha}_{12} & & i\overline{V}_{23} & iV_{13} & \\
\overline{V}_{13} & & -\overline{V}_{13} & -i\overline{V}_{23} & & \alpha_{13} & & i\overline{V}_{12} & \\
V_{13} & & -V_{13} & & iV_{23} & & \overline{\alpha}_{13} & & -iV_{12} \\
 & \overline{V}_{23} & -\overline{V}_{23} & & i\overline{V}_{13} & iV_{12} & & \alpha_{23} & \\
 & V_{23} & -V_{23} & -iV_{13} & & & -i\overline{V}_{12} & & \overline{\alpha}_{23}
\end{array}\right)$$

where $\alpha_{ij} = -\gamma_{ij} + iE_{ij}$. This new matrix consists of the matrices $\tilde{a}$ and $\tilde{b}_0$ also introduced in 4.2.5

$$\tilde{M} = \begin{pmatrix} 0 & -\tilde{a}^\dagger \\ \tilde{a} & \tilde{b} \end{pmatrix}$$

where

$$\begin{aligned} \tilde{b} &= U^\dagger b U \\ &= \tilde{b}_0 + \tilde{\nu} \end{aligned}$$

with $\tilde{\nu} = U^\dagger \nu U$. The two kinetic networks are

$$N_0 = \tilde{a}^\dagger \tilde{b}_0^{-1} \tilde{a}$$
$$N = \tilde{a}^\dagger \tilde{b}^{-1} \tilde{a}$$

which also holds with all the tildes removed.

It is straightforward to generalize the matrices $\tilde{a}$ and $\tilde{b}_0$ to $n > 3$. Matrix $\tilde{a}$ connects the population of site k to the coherences between site k and any other site l with strength V_{kl}, and matrix $\tilde{b}_0$ is a diagonal matrix with entries α_{ij} and $\overline{\alpha}_{ij}$. A bit more complicated is the matrix $\tilde{\nu}$ it is described in the next subsection.

C.2 General construction

Here we give a description of how to find $\tilde{a}$, $\tilde{b}_0$ and $\tilde{\nu}$ for general n. We number the n dimensions of population space P with k where $k = 1, 2, \ldots n$ and the $(n^2 - n)$ dimensions of coherence space C with kl and $\overline{kl}$ where $k < l$ are numbers from 1 to n. According to the order defined in 4.2.1 the first few dimensions of C are called 12, $\overline{12}$, 13, ..., 23, $\overline{23}$, 24, etc. .

C.2.1 Constructing $\tilde{a}$ and $\tilde{b}_0$

Matrix $\tilde{a}$ is an $n \times (n^2 - n)$ complex matrix, with the only nonzero entries

$$\tilde{a}_{k,kl} = \bar{V}_{kl} = -\tilde{a}_{k,lk}$$
$$\tilde{a}_{k,\overline{kl}} = V_{kl} = -\tilde{a}_{k,\overline{lk}}\,,$$

hence in every column there are only $(n-1)$ nonzero entries.

Matrix $\tilde{b}_0$ is diagonal with entries

$$\left(\tilde{b}_0\right)_{kl,kl} = -\gamma_{kl} + iE_{kl}$$
$$\left(\tilde{b}_0\right)_{\overline{kl},\overline{kl}} = -\gamma_{kl} - iE_{kl}\,.$$

C.2.2 Constructing $\tilde{\nu}$

The matrix $\tilde{\nu} = U^\dagger \nu U$ for any n is a somewhat complicated pattern of entries V_{kl}, signs and complex conjugates. It connects coherences between sites k and l with coherences between sites k and m with the

row	column	entry	row	column	entry
kl	km	$-iV_{lm}$	$\overline{kl}$	$\overline{km}$	$i\bar{V}_{lm}$
lk	mk	$i\bar{V}_{lm}$	$\overline{lk}$	$\overline{mk}$	$-iV_{lm}$
lk	$\overline{km}$	$-i\bar{V}_{lm}$	$\overline{lk}$	km	iV_{lm}

Table C.1: The non-zero entries of $\tilde{\nu}$, always $l \neq m$

strength V_{lm}. Entries of $\tilde{\nu}$ are only non-zero if one number of the two double indices match with further conditions on their conjugation. Table C.1 shows the rules for the nonzero entries.

C.3 Calculations for applications

C.3.1 Highly connected network

Assume all sites are equally interacting, and have the same energies and dephasing rates

$$V_{kl} = \Theta$$

$$E_k = 0$$

$$\gamma_k = \Gamma\,.$$

Then every column in a has $2(n-1)$ non-zero entries all equal to Θ. A simple calculation shows that

$$a^\dagger a = 2n\Theta^2 \left(\mathbb{1}_n - n\vec{e}\vec{e}^\dagger\right)$$

so for any $\vec{v} \in I$ we have $a^\dagger a\vec{v} = 2\Theta^2 n\vec{v}$, hence

$$\|a\| = \sqrt{2n}\Theta\,.$$

Obviously,$b = -\Gamma\mathbb{1}_C$ and $\left\|b_0^{-1}\right\| = \Gamma^{-1}$. This gives $\kappa = 2n\Theta^2\Gamma^{-2}$. Because

$$a^\dagger b_0^{-1} a = -\Gamma^{-1} a^\dagger a$$

we have $\mu_0 = 2n\Theta^2\Gamma^{-1}$. Using $\mu \approx \mu_0$ we find

$$\begin{aligned}
\alpha &= \min\left\{\frac{1}{2}\left\|b^{-1}\right\|^{-1},\ \frac{1}{4}\kappa^{-1}\mu\right\} \\
&= \min\left\{\frac{1}{2}\Gamma, \frac{1}{4}\frac{2n\Theta^2\Gamma^{-1}}{2n\Theta^2\Gamma^{-2}}\right\} \\
&= \Gamma/4
\end{aligned}$$

and so

$$\begin{aligned}
\beta &= \max\left\{1,\ \alpha^{-1}\left\|b^{-1}\right\|^{-1}\right\} \\
&= 4
\end{aligned}$$

and with Theorem 37 we get the bounds

$$\begin{aligned}
\Delta\tau &\leq \frac{20}{\pi}\Gamma^{-1} \\
\Delta\tau_{\mathrm{rel}} &\leq \frac{40}{\pi}n\Theta^2\Gamma^{-2}\,.
\end{aligned}$$

To get the bound on $\Delta\tau_{1,\,\mathrm{rel}}$ we also estimate $\|\nu\|$, we use the fact that each column and row of ν has $(n-2)$ nonzero entries and so

$$\|\nu\| \geq vn\Theta$$

for a scaling and dimension independent constant v. Then Theorem 38 gives the bound

$$\Delta\tau_{1,\,\mathrm{rel}} \leq 4vn\Theta\Gamma^{-1}\,.$$

The condition for this bound is

$$\|\nu\| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$$

the LHS is bounded from below by $vn\Theta$ and the RHS is constant, so the condition does not hold for large n.

C.3.2 Circular chain

Assume the sites are positioned on a circle and only nearest neighbors interact with strength Θ

$$V_{kl} = \begin{cases} \Theta & |k-l| = 1 \\ 0 & \text{else} \end{cases}$$

where we set equivalence $n \equiv 0$. Further $\gamma_k = \Gamma$ and E_k such that $E_{kl} = \Gamma E$ when $|k-l| = 1$ which is possible for n even.

Now, the column for site k in a has only 4 entries, two each for the coherences with $k-1$ and $k+1$. We calculate

$$\left(a^\dagger a\right)_{kl} = \begin{cases} 4\Theta^2 & k = l \\ -2\Theta^2 & |k-l| = 1 \\ 0 & \text{else} \end{cases}$$

So $\left\|a^\dagger a\right\| = 8\Theta^2$ and $\|a\| = \sqrt{8}\Theta$, in particular there is no n dependency. Also $\left\|b_0^{-1}\right\| = 1/\sqrt{\Gamma^2 + \Gamma^2 E^2}$ and so $\kappa = \frac{8}{1+E^2}\Theta^2\Gamma^{-2}$. We have

$$N_0 = \begin{cases} -\frac{4\Theta^2}{\Gamma(1+E^2)} & k = l \\ \frac{2\Theta^2}{\Gamma(1+E^2)} & |k-l| = 1 \\ 0 & \text{else} \end{cases}$$

which has the spectrum

$$\lambda_p = -\frac{4\Theta^2}{\Gamma(1+E^2)}\left(1 - \cos\left(\frac{2\pi p}{n}\right)\right) \tag{C.1}$$

with $p = 1 \ldots n$. The nonzero eigenvalue smallest in magnitude is μ_0, so for large n and small $\Theta\Gamma^{-1}$, approximately

$$\mu \approx \mu_0 \approx \frac{2\Theta^2}{\Gamma(1+E^2)}\left(\frac{2\pi}{n}\right)^2$$

so $\alpha = \frac{1}{16}\Gamma\left(\frac{2\pi}{n}\right)^2$ and

$$\begin{aligned} \beta &= \frac{16\Gamma\sqrt{1+E^2}}{\Gamma\left(\frac{2\pi}{n}\right)^2} \\ &= \left(\frac{2n}{\pi}\right)^2\sqrt{1+E^2}\,. \end{aligned}$$

Moving the numbers into constants k_1 and k_2, and dropping the 1 in $1+\beta$ (fine for large n), we have

$$\Delta\tau \leq k_1\sqrt{1+E^2}\Gamma^{-1}n^4$$
$$\Delta\tau_{\text{rel}} \leq \frac{k_2}{\sqrt{1+E^2}}\Theta^2\Gamma^{-2}n^2\,.$$

We again estimate $\|\nu\|$, now each column and row of ν has 2 or 4 nonzero entries and so

$$v_1\Theta \leq \|\nu\| \leq v_2\Theta$$

for some scaling and dimension independent constants v_1 and v_2. Then Theorem 38 gives the bound

$$\Delta\tau_{1,\,\text{rel}} \leq \frac{4}{\pi^2}v_2 n^2\Theta\Gamma^{-1}\,.$$

This time the condition

$$\|\nu\| \leq \frac{1}{2}\left\|b^{-1}\right\|^{-1}$$

does not break down for large dimensions, so the bound holds for all n when Θ and Γ are kept constant.

Bibliography

[1] J. Adolphs and T. Renger. How proteins trigger excitation energy transfer in the FMO complex of green sulfur bacteria. *Biophysical Journal*, 91(8):2778–2797, Oct. 2006. PMID: 16861264 PMCID: 1578489.

[2] G. G. Amosov. Remark on the additivity conjecture for a quantum depolarizing channel. *Problems of Information Transmission*, 42:69–76, June 2006.

[3] G. G. Amosov. Strong superadditivity conjecture holds for the quantum depolarizing channel in any dimension. *Physical Review A*, 75(6):060304, June 2007.

[4] G. G. Amosov, A. S. Holevo, and R. Werner. On some additivity problems in quantum information theory. *Probl. Inform. Transm.*, 36(N4):25–34, 2000.

[5] W. Arnold and H. I. Kohn. The chlorophyll unit in photosynthesis. *The Journal of General Physiology*, 18(1):109 –112, 1934.

[6] G. Aubrun, S. Szarek, and E. Werner. Hastings' additivity counterexample via Dvoretzky's theorem. *Communications in Mathematical Physics*, 305:85–97, Nov. 2010.

[7] T. Banachiewicz. Zur berechnung der determinanten, wie auch der inversen und zur darauf basierten auflosung der systeme linearer gleichungen. *Acta Astronom. Ser. C*, 3:41—67, 1937.

[8] R. Bhatia. *Matrix analysis*. Springer, 1997.

[9] R. E. Blankenship. *Molecular mechanisms of photosynthesis*. John Wiley & Sons, Feb. 2002.

[10] L. Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht*. kk Hof-und Staatsdruckerei, 1877.

[11] B. V. Bronk. Exponential ensemble for random matrices. *Journal of Mathematical Physics*, 6:228, 1965.

[12] D. Bruss, L. Faoro, C. Macchiavello, and G. M. Palma. Quantum entanglement and classical communication through a depolarizing channel. *Journal of Modern Optics*, 47:325–331, Feb. 2000.

[13] J. Cao and R. J. Silbey. Optimization of exciton trapping in energy transfer processes. *The Journal of Physical Chemistry A*, 113:13825–13838, Dec. 2009.

[14] S. Carnot. *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*. Gauthier-Villars, 1878.

[15] M. Cho, H. M. Vaswani, T. Brixner, J. Stenger, and G. R. Fleming. Exciton analysis in 2D electronic spectroscopy. *The Journal of Physical Chemistry B*, 109:10542–10556, June 2005.

[16] M. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, June 1975.

[17] R. Clausius. Ueber die wärmeleitung gasförmiger körper. *Annalen der Physik*, 191(1):1–56, Jan. 1862.

[18] R. M. Clegg, M. Sener, and Govindjee. From förster resonance energy transfer to coherent resonance energy transfer and back. pages 75610C–75610C–21, 2010.

[19] B. Collins and I. Nechita. Random quantum channels i: Graphical calculus and the bell state phenomenon. *Communications in Mathematical Physics*, 297:345–370, Feb. 2010.

[20] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, New York, NY, 2001.

[21] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, and A. Winter. Counterexamples to additivity of minimum output p-Rényi entropy for p close to 0. *Communications in Mathematical Physics*, 284:281–290, Sept. 2008.

[22] C. Curutchet, J. Kongsted, A. Munoz-Losa, H. Hossein-Nejad, G. Scholes, and B. Mennucci. Photosynthetic Light-Harvesting is tuned by the heterogeneous polarizable environment of the protein. *Journal of the American Chemical Society*, 2011.

[23] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57(2):830–839, Feb. 1998.

[24] G. S. Engel, T. R. Calhoun, E. L. Read, T. Ahn, T. Mancal, Y. Cheng, R. E. Blankenship, and G. R. Fleming. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, 446(7137):782–786, Apr. 2007.

[25] T. Förster. Transfer mechanisms of electronic excitation energy. *Radiation Research Supplement*, pages 326—339, 1960.

[26] A. Fujiwara and T. Hashizumé. Additivity of the capacity of depolarizing channels. *Physics Letters A*, 299(5-6):469–475, July 2002.

[27] M. Fukuda. Simplification of additivity conjecture in quantum information theory. *Quantum Information Processing*, 6:179–186, Mar. 2007.

[28] M. Fukuda and C. King. Entanglement of random subspaces via the hastings bound. *Journal of Mathematical Physics*, 51:042201, 2010.

[29] M. Fukuda and M. M. Wolf. Simplifying additivity problems using direct sum constructions. *Journal of Mathematical Physics*, 48:072101, 2007.

[30] H. Gaffron and K. Wohl. Zur theorie der assimilation. *Die Naturwissenschaften*, 24:81–90, Feb. 1936.

[31] M. A. Green, K. Emery, Y. Hishikawa, and W. Warta. Solar cell efficiency tables (version 37). *Progress in Photovoltaics: Research and Applications*, 19(1):84–92, Jan. 2011.

[32] M. B. Hastings. A counterexample to additivity of minimum output entropy. *arXiv:0809.3972*, Sept. 2008. Nature Physics 5, 255 (2009).

[33] M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nat Phys*, 5(4):255–257, Apr. 2009.

[34] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265:95–117, Mar. 2006.

[35] P. Hayden and A. Winter. Counterexamples to the maximal p-Norm multiplicativity conjecture for all $p > 1$. *Communications in Mathematical Physics*, 284:263–280, Sept. 2008.

[36] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20:37–39, Mar. 1953.

[37] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, Jan. 1998.

[38] A. S. Holevo. Quantum coding theorems. *Russian Mathematical Surveys*, 53:1295–1331, Dec. 1998.

[39] A. S. Holevo. Complementary channels and the additivity problem. *Theory of Probability and its Applications*, 51:92, 2007.

[40] M. Horodecki, P. W. Shor, and M. B. Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15:629–641, 2003.

[41] S. Hoyer, M. Sarovar, and K. Birgitta Whaley. Limits of quantum speedup in photosynthetic light harvesting. *New Journal of Physics*, 12:065041, June 2010.

[42] A. Ishizaki, T. R. Calhoun, G. S. Schlau-Cohen, and G. R. Fleming. Quantum coherence and its interplay with protein environments in photosynthetic electronic energy transfer. *Phys. Chem. Chem. Phys.*, 12(27):7319–7337, June 2010.

[43] A. Ishizaki and G. R. Fleming. Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: Reduced hierarchy equation approach. *The Journal of Chemical Physics*, 130:234111, 2009.

[44] C. King. Additivity for unital qubit channels. *Journal of Mathematical Physics*, 43:4641, 2002.

[45] C. King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49(1):221– 229, Jan. 2003.

[46] C. King, K. Matsumoto, M. Nathanson, and M. Ruskai. Properties of conjugate channels with applications to additivity and multiplicativity. *Markov Process. Relat. Fields*, 13(2):391–423, 2007.

[47] C. King, M. Nathanson, and M. B. Ruskai. Multiplicativity properties of entrywise positive maps. *Linear Algebra and its Applications*, 404(0):367–379, July 2005.

[48] C. King and M. B. Ruskai. Minimal entropy of states emerging from noisy quantum channels. *IEEE Transactions on Information Theory*, 47(1):192–209, Jan. 2001.

[49] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Soc., Aug. 2001.

[50] S. Lloyd and H. Pagels. Complexity as thermodynamic depth. *Annals of Physics*, 188(1):186–213, Nov. 1988.

[51] K. Matsumoto, T. Shimono, and A. Winter. Remarks on additivity of the holevo channel capacity and of the entanglement of formation. *Communications in Mathematical Physics*, 246:427–442, Apr. 2004.

[52] B. Matthews, R. Fenna, M. Bolognesi, M. Schmid, and J. Olson. Structure of a bacteriochlorophyll a-protein from the green photosynthetic bacterium prosthecochloris aestuarii. *Journal of Molecular Biology*, 131(2):259–285, June 1979.

[53] V. D. Milman and G. Schechtman. *Asymptotic theory of finite dimensional normed spaces*. Springer, 1986.

[54] J. v. Neumann. *The mathematical foundations of quantum mechanics*. Princeton University Press, 1932.

[55] G. Panitchayangkoon, D. Hayes, K. Fransted, J. Caram, E. Harel, J. Wen, R. Blankenship, and G. Engel. Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proceedings of the National Academy of Sciences*, 107(29):12766, 2010.

[56] M. B. Plenio and S. F. Huelga. Dephasing-assisted transport: quantum networks and biomolecules. *New Journal of Physics*, 10:113019, Nov. 2008.

[57] P. Rebentrost, R. Chakraborty, and A. Aspuru-Guzik. Non-Markovian quantum jumps in excitonic energy transfer. *The Journal of Chemical Physics*, 131:184102, 2009.

[58] P. Rebentrost, M. Mohseni, I. Kassal, S. Lloyd, and A. Aspuru-Guzik. Environment-assisted quantum transport. *New Journal of Physics*, 11:033003, 2009.

[59] A. Renyi. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, page 547–561, 1961.

[60] M. B. Ruskai. Qubit entanglement breaking channels. *Reviews in Mathematical Physics*, 15:643–662, 2003.

[61] M. B. Ruskai, S. Szarek, and E. Werner. An analysis of completely-positive trace-preserving maps on m2. *Linear Algebra and its Applications*, 347(1-3):159–187, May 2002.

[62] G. Scholes. Quantum-coherent electronic energy transfer: Did nature think of it first? *The Journal of Physical Chemistry Letters*, 1(1):2—8, 2010.

[63] B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, July 1997.

[64] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379—423, 1948.

[65] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246:453–472, Apr. 2004.

[66] G. Smith. Quantum channel capacities. In *2010 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, Sept. 2010.

[67] D. Weingarten. Asymptotic behavior of group integrals in the limit of infinite rank. *Journal of Mathematical Physics*, 19:999, 1978.

[68] J. Wen, H. Zhang, M. L. Gross, and R. E. Blankenship. Membrane orientation of the FMO antenna protein from chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *Proceedings of the National Academy of Sciences*, 106(15):6134 –6139, Apr. 2009.

[69] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43:4353, 2002.

[70] C. Wong, H. Hossein-Nejad, C. Curutchet, and G. Scholes. Quantum-Coherent energy transfer in marine algae at ambient temperature via ultrafast photon echo studies. In *International Conference on Ultrafast Phenomena*, 2010.

[71] F. Zhang. *The Schur complement and its applications*. Springer, 2005.

[72] X. Zhu, S. P. Long, and D. R. Ort. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology*, 19(2):153–159, Apr. 2008.

[73] K. Zyczkowski and H. Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A: Mathematical and General*, 34:7111–7125, Sept. 2001.